

Forecasting the Price of Oil

Ron Alquist*, Lutz Kilian†, and Robert J. Vigfusson‡

*Bank of Canada, Ottawa, ON, Canada

†University of Michigan, Ann Arbor, MI, USA

‡Federal Reserve Board, Washington, DC, USA

Contents

1. Introduction	428
2. Alternative Oil Price Measures	431
3. Alternative Oil Price Specifications	434
4. Granger Causality Tests	436
4.1. Nominal Oil Price Predictability	437
4.1.1. <i>The Pre-1973 Evidence</i>	437
4.1.2. <i>The Post-1973 Evidence</i>	437
4.1.3. <i>Reconciling the Pre- and Post-1973 Evidence on the Predictability of the Nominal Price of Oil</i>	440
4.2. Real Oil Price Predictability in the Post-1973 Period	441
5. Short-Horizon Forecasts of the Nominal Price of Oil	446
5.1. Forecasting Methods Based on Monthly Oil Futures Prices	447
5.2. Other Forecasting Methods	451
5.2.1. <i>Alternative Atheoretical Forecasting Models</i>	452
5.2.2. <i>Parsimonious Economic Forecasting Models</i>	452
5.3. Short-Horizon Forecasts Based on Daily Oil Futures Prices	454
6. Long-Horizon Forecasts of the Nominal Price of Oil Based on Oil Futures Prices	456
7. Survey Forecasts of the Nominal Price of Oil	459
8. What Have We Learned about Forecasting the Nominal Price of Oil?	463
9. Short-Horizon Forecasts of the Real Price of Oil	464
9.1. Real U.S. Refiners' Acquisition Cost for Imported Crude Oil	465
9.1.1. <i>Unrestricted AR, ARMA, and VAR Models</i>	465
9.1.2. <i>Real-Time Forecasts</i>	470
9.2. Real WTI Price	471
9.3. Restricted VAR Models	471
10. What Have We Learned about Forecasting the Real Price of Oil?	473
11. Structural VAR Forecasts of the Real Price of Oil	474
12. The Ability of Oil Prices to Forecast U.S. Real GDP	477
12.1. Linear Autoregressive Models	477
12.2. Non-Linear Dynamic Models	480
12.3. Does Using Pre-1974 Data Help in Forecasting Real GDP?	489
12.4. Non-Parametric Approaches	490
13. The Role of Oil Price Volatility	491
13.1. Nominal Oil Price Volatility	492

13.2. Real Oil Price Volatility	493
13.3. Quantifying Oil Price Risks	494
14. Avenues for Future Research	499
15. Conclusions	500
Acknowledgments	503
References	503

Abstract

We address some of the key questions that arise in forecasting the price of crude oil. What do applied forecasters need to know about the choice of sample period and about the tradeoffs between alternative oil price series and model specifications? Are real and nominal oil prices predictable based on macroeconomic aggregates? Does this predictability translate into gains in out-of-sample forecast accuracy compared with conventional no-change forecasts? How useful are oil futures prices in forecasting the spot price of oil? How useful are survey forecasts? How does one evaluate the sensitivity of a baseline oil price forecast to alternative assumptions about future oil demand and oil supply conditions? How does one quantify risks associated with oil price forecasts? Can joint forecasts of the price of oil and of U.S. real GDP growth be improved upon by allowing for asymmetries?

Keywords

Oil, No-change forecast, Futures price, Economic fundamentals, Asymmetries, Structural change, Structural oil market models, Predictability, Forecast scenarios, Risk

1. INTRODUCTION

There is widespread agreement that unexpected large and persistent fluctuations in the price of oil are detrimental to the welfare of both oil-importing and oil-producing economies, making reliable forecasts of the price of crude oil of interest for a wide range of applications. For example, central banks and private sector forecasters view the price of oil as one of the key variables in generating macroeconomic projections and in assessing macroeconomic risks. Of particular interest is the question of the extent to which the price of oil is helpful in predicting recessions. For example, [Hamilton \(2009\)](#), building on the analysis in [Edelstein and Kilian \(2009\)](#), provides evidence that the recession of late 2008 was amplified and preceded by an economic slowdown in the automobile industry and a deterioration in consumer sentiment.

Not only are more accurate forecasts of the price of oil likely to improve the accuracy of forecasts of macroeconomic outcomes, but, in addition, some sectors of the economy depend directly on forecasts of the price of oil for their business. For example, airlines rely on such forecasts in setting airfares, automobile companies decide their product menu and set product prices with oil price forecasts in mind, and utility companies use oil price forecasts in deciding whether to expand capacity or to build new plants. Likewise, homeowners rely on oil price forecasts in deciding the timing of their heating oil purchases or whether to invest in energy-saving home improvements.

Finally, forecasts of the price of oil (and the price of its derivatives such as gasoline or heating oil) are important in modeling purchases of energy-intensive durables goods such as automobiles or home heating systems.¹ They also play a role in generating projections of energy use, in modeling investment decisions in the energy sector, in predicting carbon emissions and climate change, and in designing regulatory policies such as the imposition of automotive fuel standards or gasoline taxes.²

This chapter provides a comprehensive analysis of the problem of forecasting the price of crude oil. In Section 2 we compare alternative measures of the price of oil. In Section 3 we discuss the rationales of alternative specifications of the oil price variable in empirical work. Section 4 studies the extent to which the nominal price of oil and the real price of oil are predictable based on macroeconomic aggregates. We document strong evidence of predictability in population. Predictability in population, however, need not translate into out-of-sample forecastability. The latter question is the main focus of Sections 5 through 10.

In Sections 5, 6 and 7, we compare a wide range of out-of-sample forecasting methods for the nominal price of oil. For example, it is common among policymakers to treat the price of oil futures contracts as the forecast of the nominal price of oil. We focus on the ability of daily and monthly oil futures prices to forecast the nominal price of oil in real time compared with a range of simple time series forecasting models. We find some evidence that the price of oil futures has additional predictive content compared with the current spot price at the 12-month horizon; the magnitude of the reduction in the mean-squared prediction error (MSPE) is modest even at the 12-month horizon, however, and there are indications that this result is sensitive to small changes in the sample period and in the forecast horizon. There is no evidence of significant forecast accuracy gains at shorter horizons, and at the long horizons of interest to policymakers, oil futures prices are clearly inferior to the no-change forecast.

Similarly, a forecasting model motivated by the analysis in [Hotelling \(1931\)](#), and a variety of simple time series regression models are not successful at significantly lowering the MSPE at short horizons. There is strong evidence, however, that recent percent changes in the nominal price of industrial raw materials (other than crude oil) can be used to substantially and significantly reduce the MSPE of the nominal price of oil at horizons of 1 and 3 months, building on insights in [Barsky and Kilian \(2002\)](#). The gains may be as large as 22% at the 3-month horizon. Similar short-run gains also are possible based on models that extrapolate the current nominal price of oil at the rate of recent percent changes in the dollar exchange rate of major broad-based commodity exporters such as Canada or Australia, consistent with insights provided in [Chen et al. \(2010\)](#).

By comparison, the predictive success of expert survey forecasts of the nominal price of oil proved disappointing. Only the one-quarter-ahead U.S. Energy Information

¹ See, e.g., [Kahn \(1986\)](#), [Davis and Kilian \(2011\)](#).

² See, e.g., [Goldberg \(1998\)](#), [Allcott and Wozny \(2011\)](#), [Busse et al. \(2011\)](#), [Kellogg \(2010\)](#).

Administration (EIA) forecast significantly improves on the no-change forecast and none of the expert forecasts we studied significantly improves on the MSPE of the no-change forecast at the 1-year horizon. Finally, forecasts obtained by adjusting the current price of oil for survey inflation expectation do little to improve accuracy at horizons up to 1 year, but outperform the no-change forecast by a wide margin at horizons of several years. Section 8 summarizes the lessons to be drawn from our analysis for forecasting the nominal price of oil.

Although the nominal price of crude oil receives much attention in the press, the variable most relevant for economic modeling is the real price of oil. Section 9 compares alternative forecasting models for the real price of oil. We provide evidence that reduced-form autoregressive and vector autoregressive models of the global oil market are more accurate than the random walk forecast of the real price of oil at short horizons. Even after taking account of the constraints on the real-time availability of these predictors, the MSPE reductions can be substantial in the short run. The accuracy gains tend to diminish at longer horizons, however, and, beyond 1 year, the no-change forecast of the real price of oil typically is the predictor with the lowest MSPE. Moreover, the extent of the MSPE reductions depends on the definition of the oil price series. The lessons to be drawn from this evidence are discussed in Section 10.

An important limitation of reduced-form forecasting models from a policy point of view is that they provide no insight into what is driving the forecast and do not allow the policymaker to explore alternative hypothetical forecast scenarios. In Section 11, we illustrate how recently developed structural vector autoregressive models of the global oil market may be used to generate conditional projections of how the oil price forecast would deviate from the unconditional forecast baseline, given alternative scenarios about future oil demand and oil supply conditions such as a surge in speculative demand triggered by Iran, a resurgence of the global business cycle, or an increase in U.S. oil production.

Section 12 focuses on the problem of jointly forecasting U.S. macroeconomic aggregates such as real GDP growth and the price of oil. Of particular interest is the forecasting ability of non-linear transformations of the price of oil such as the nominal net oil price increase or the real net oil price increase. The net oil price increase is a censored predictor that assigns zero weight to net oil price decreases. There is little evidence that this type of asymmetry is reflected in the responses of U.S. real GDP to innovations in the real price of oil, as documented in Kilian and Vigfusson (2011a,b), but Hamilton (2011) suggests that the net oil price increase specification is best thought of as a parsimonious forecasting device. We provide a comprehensive analysis of this conjecture.

Point forecasts of the price of oil are important, but they fail to convey the large uncertainty associated with oil price forecasts. This uncertainty can be captured by predictive densities. In Section 13 we discuss various approaches of conveying the information in the predictive density including measures of price volatility and of tail conditional expectations with particular emphasis on defining appropriate risk measures. Section 14

contains a discussion of directions for future research. The concluding remarks are in Section 15.

2. ALTERNATIVE OIL PRICE MEASURES

Figure 8.1 plots three alternative measures of the nominal price of oil. The longest available series is the West Texas Intermediate (WTI) price of crude oil. WTI is a particular grade of crude oil. The WTI price refers to the price of WTI oil to be delivered to Cushing, Oklahoma, and serves as a benchmark in pricing oil contracts. It is available in real time. An alternative measure of the oil price is the price paid by U.S. refiners purchasing crude oil. Data on the U.S. refiners' acquisition cost for domestically produced oil, for imported crude oil and for a composite of these series are available starting in 1974.1. These data become available only with a delay and are subject to revisions (see Baumeister and Kilian, 2012a).

Figure 3.1 highlights striking differences in the time series process for the price of oil prior to 1973 and after 1973. The WTI data until 1973 tend to exhibit a pattern resembling a step-function. The price remains constant for extended periods, followed by large adjustments. The U.S. wholesale price of oil for 1948–1972 used in Hamilton (1983) is very similar to the WTI series. As discussed in Hamilton (1983, 1985) the peculiar pattern of crude oil price changes during this period is explained by the specific

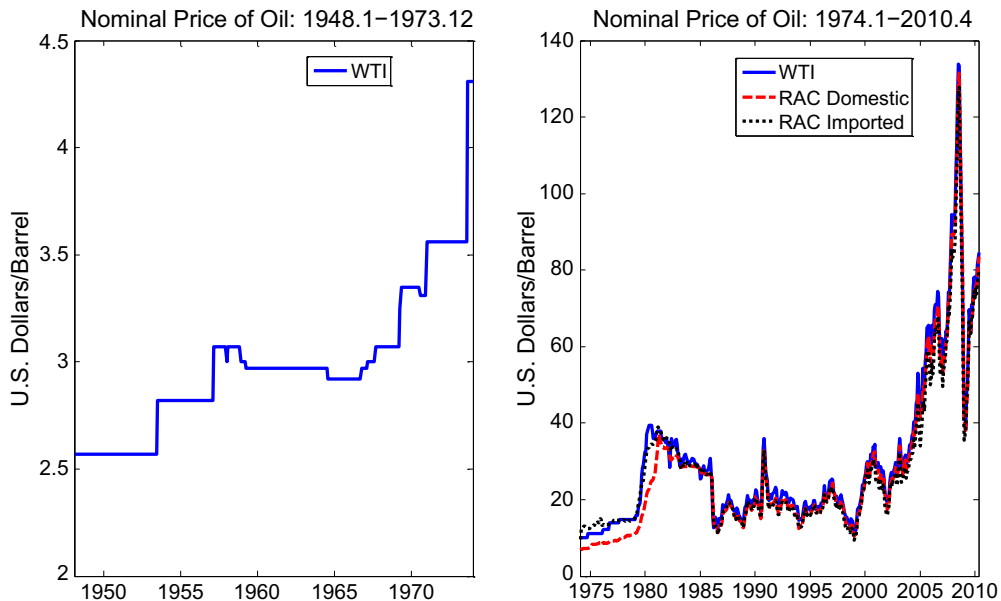


Figure 8.1 The nominal price of crude oil. *Notes:* WTI stands for the West Texas Intermediate price of crude oil and RAC for the U.S. refiners' acquisition cost.

regulatory structure of the oil industry during 1948–72. Each month the Texas Railroad Commission and other U.S. state regulatory agencies would forecast demand for oil for the subsequent month and would set the allowable production levels for wells in the state to meet demand. As a result, much of the cyclically endogenous component of oil demand was reflected in shifts in quantities rather than prices. The commission was generally unable or unwilling to accommodate sudden disruptions in oil production, preferring instead to exploit these events to implement sometimes dramatic price increases (Hamilton, 1983, p. 230).

Whereas the WTI price is a good proxy for the U.S. price for oil during 1948–72, when the U.S. was largely self-sufficient in oil, it becomes less representative after 1973, when the share of U.S. imports of oil rapidly expanded. The price discrepancy between unregulated foreign oil and regulated domestic oil created increasing pressure to deregulate the domestic market. As regulatory control weakened in the mid-1970s, adjustments to the WTI price became much more frequent and smaller in magnitude, as shown in the right panel of Figure 8.1. By the mid-1980s, the WTI had been deregulated to the point that there was strong comovement between all three oil price series most of the time.

Figure 8.2 shows the corresponding oil price data adjusted for U.S. CPI inflation. The left panel reveals that in real terms the price of oil had been falling considerably since the late 1950s. That decline was corrected only by the sharp rise in the real price of oil in 1973/74. There has been no pronounced trend in the real price of oil since 1974, but considerable volatility. The definition of the real price of oil is of lesser importance after 1986.

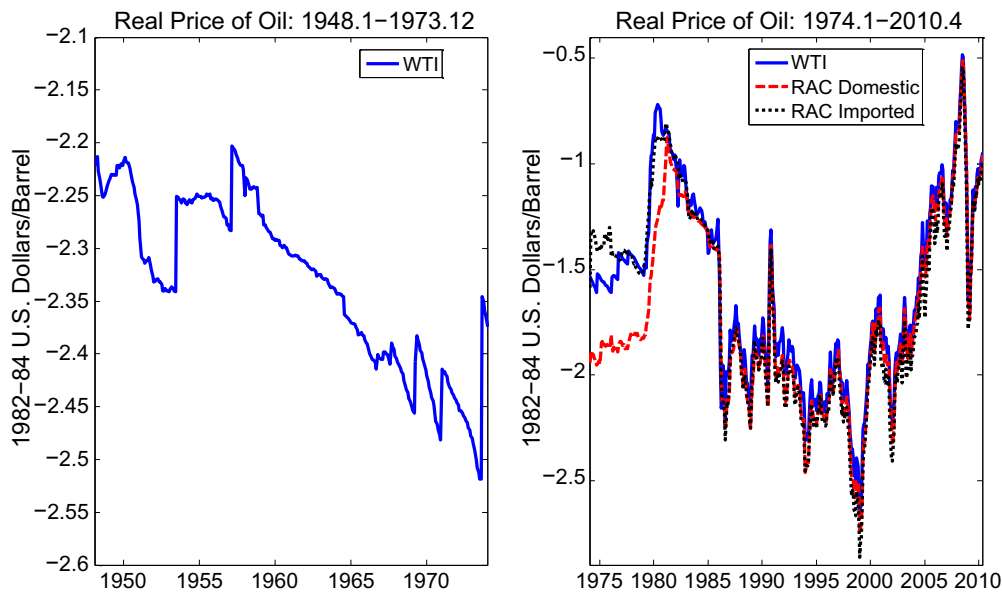


Figure 8.2 The real price of crude oil. *Notes:* Log scale. See Figure 8.1.

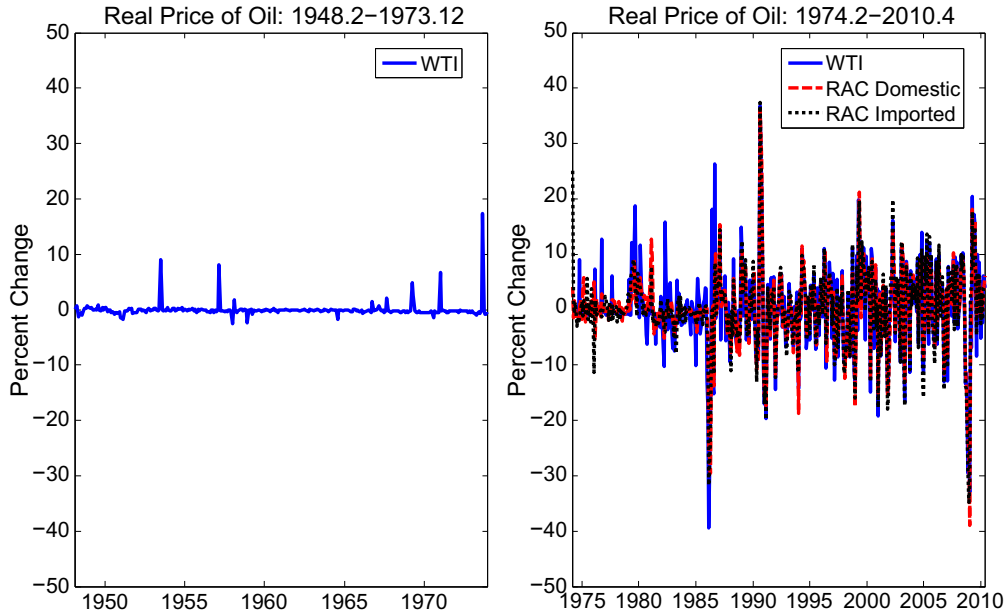


Figure 8.3 Percent changes in the real price of crude oil. *Notes:* See Figure 8.1. All price series deflated by the U.S. CPI for all urban consumers, seasonally adjusted.

Prior to 1986, one key difference is that the refiners' acquisition cost for imported crude oil fell in 1974–76, whereas the real WTI price rose. A second key difference is that the real WTI price spiked in 1980, whereas the real price of oil imports remained largely stable. This pattern was only reversed with the outbreak of the Iran–Iraq War in late 1980.

Figure 8.3 once more highlights the striking differences between the pre- and post-1973 period. It shows the percent growth rate of the real price of oil. A major structural change in the distribution of the price of oil in late 1973 is readily apparent.³ Whereas the pre-1973 period is characterized by long periods of low volatility interrupted by infrequent large positive price spikes, the post-1973 period is characterized by high month-to-month volatility. It has been suggested that perhaps this volatility has increased systematically after the collapse of OPEC in late 1985. The answer is somewhat sensitive to the exact choice of dates. If one were to date the OPEC period as 1973.10–1985.12, for example, there is no evidence of an increase in the variance of the percent change in the real WTI price of oil. The volatility in the OPEC period is virtually identical to that in the post-OPEC period of 1986.1–2010.6. Shifting the starting date of the OPEC period to 1974.1, in contrast, implies a considerable increase in volatility

³ In related work, *Dvir and Rogoff (2010)* present formal evidence of a structural break in the process driving the annual real price of oil in 1973. Given this evidence of instability, combining pre- and post-1973 real oil price data is not a valid option.

after 1985. Extending the ending date of the OPEC period to include the price collapse in 1986 induced by OPEC actions, on the other hand, renders the volatility much more similar across subperiods. Finally, combining the earlier starting date and the later ending date, there is evidence of a reduction in the real price volatility after the collapse of OPEC rather than an increase. Below we therefore treat the post-1973 data as homogeneous.

Which price series is more appropriate for the analysis of post-1973 data depends in part on the purpose of the study. The WTI price data (as well as other measures of the domestic U.S. price of oil) are questionable to the extent that these prices were regulated until the mid-1980s and do not reflect the true scarcity of oil or the price actually paid by U.S. refiners. The main advantage of the WTI price is that it is available in a timely manner and not subject to data revisions. The refiners' acquisition cost for imported crude oil provides a good proxy for oil price fluctuations in global oil markets, but may not be representative for the price that U.S. refineries paid for crude oil. The latter price may be captured better by a composite of the acquisition cost of domestic and imported crude oil, neither of which, however, is available before January 1974. The real price of oil imports also is the price relevant for theories interpreting oil price shocks as terms-of-trade shocks. Theories that interpret oil price shocks as allocative disturbances, on the other hand, require the use of retail energy prices, for which the composite refiners' acquisition cost may be a better proxy than the refiners' acquisition cost of oil imports. Below we will consider several alternative oil price series.⁴

3. ALTERNATIVE OIL PRICE SPECIFICATIONS

Although an increasing number of empirical studies of the post-1973 data focus on the real price of oil, many other studies have relied on the nominal price of oil. One argument for the use of nominal oil prices has been that the nominal price of oil – unlike the real price of oil – is considered exogenous with respect to U.S. macroeconomic conditions and hence linearly unpredictable on the basis of lagged U.S. macroeconomic conditions.⁵ This argument may have some merit for the pre-1973 period, but it is implausible for the post-1973 period. If the U.S. money supply unexpectedly doubles, for example, then, according to standard macroeconomic models, so will all nominal prices denominated in dollars (including the nominal price of oil), leaving the relative price or real price of crude oil unaffected (see Gillman and Nakov, 2009). Clearly, one would not want to interpret such an episode as an oil price shock involving a doubling of the nominal price of oil. Indeed, economic models of the impact of the price of oil on the U.S. economy correctly predict that such a nominal oil price shock should have no effect on the U.S.

⁴ For further discussion of the trade-offs between alternative oil price definitions from an economic point of view see Kilian and Vigfusson (2011b).

⁵ For a review of the relationship between the concepts of (strict) exogeneity and predictability in linear models see Cooley and LeRoy (1985).

economy because theoretical models inevitably are specified in terms of the real price of oil, which has not changed in this example.

Another argument in the literature has been that the nominal price of oil can be considered exogenous after 1973 because it is set by OPEC. This interpretation is without basis. First, there is little evidence to support the notion that OPEC has been successfully acting as a cartel in the 1970s and early 1980s, and the role of OPEC has diminished further since 1986 (see, e.g., Skeet, 1988; Smith, 2005; Almoguera et al., 2011). Second, even if we were to accept the notion that an OPEC cartel sets the nominal price of oil, economic theory predicts that this cartel price will endogenously respond to U.S. macroeconomic conditions. This theoretical prediction is consistent with anecdotal evidence of OPEC oil producers raising the price of oil (or equivalently lowering oil production) in response to unanticipated U.S. inflation, low U.S. interest rates, and the depreciation of the dollar. Moreover, as observed by Barsky and Kilian (2002), economic theory predicts that the strength of the oil cartel itself (measured by the extent to which individual cartel members choose to deviate from cartel guidelines) will be positively related to the state of the global business cycle (see Green and Porter, 1984). Thus, both nominal and real oil prices must be considered endogenous with respect to the global economy, unless proven otherwise.

A third and distinct argument has been that consumers of refined oil products choose to respond to changes in the nominal price of oil rather than the real price of oil, perhaps because the nominal price of oil is more visible. In other words, consumers suffer from money illusion. There is no direct empirical evidence in favor of this behavioral argument at the micro level. Rather the case for this specification, if there is one, has to be based on the predictive success of such models; a success that, however, has yet to be demonstrated empirically. We will address this question in Section 12.

Even proponents of using the nominal price in empirical models of the transmission of oil price shocks have concluded that there is no stable dynamic relationship between percent changes in the nominal price of oil and in U.S. macroeconomic aggregates. There is evidence from in-sample fitting exercises, however, of a predictive relationship between U.S. real GDP and suitable non-linear transformations of the nominal price of oil. The most successful of these transformations is the net oil price increase measure of Hamilton (1996, 2003). Let s_t denote the nominal price of oil in logs and Δ the difference operator. Then the net oil price increase is defined as:

$$\Delta s_t^{+,net} \equiv \max [0, s_t - s_t^*],$$

where s_t^* is the highest oil price in the preceding 12 months or, alternatively, the preceding 36 months. This transformation involves two distinct ideas. One is that consumers in oil-importing economies respond to increases in the price of oil only if the increase is large relative to the recent past. If correct, the same logic by construction should apply to decreases in the price of oil, suggesting a net change transformation that is symmetric in increases and decreases.

The second idea implicit in Hamilton's definition is that consumers do not respond to net decreases in the price of oil, allowing us to omit the net decreases from the model. In other words, consumers respond asymmetrically to net oil price increases and to net oil price decreases, and they do so in a very specific fashion. Although there are theoretical models that imply the existence of an asymmetry in the response of the economy to oil price increases and decreases, these models do not imply the specific non-linear structure embodied in the net increase measure nor do they imply that the net decrease measure should receive zero weight. Nevertheless, Hamilton's nominal net oil price increase variable has become one of the leading specifications in the literature on predictive relationships between the price of oil and the U.S. economy. Hamilton (2011), for example, interprets this specification as capturing non-linear changes in consumer sentiment in response to nominal oil price increases.⁶

As with other oil price specifications there is reason to expect lagged feedback from global macroeconomic aggregates to the net oil price increase. Whereas Hamilton (2003) made the case that net oil price increases in the 1970s, 1980s, and 1990s were capturing exogenous events in the Middle East, Hamilton (2009) concedes that the net oil price increase of 2003–08 was driven in large part by a surge in the demand for oil. Kilian (2009a,b, 2010), on the other hand, provides evidence based on structural vector autoregressive (VAR) models that in fact most net oil price increases have contained a large demand component driven by global macroeconomic conditions, even prior to 2003. This finding is also consistent with the empirical results in Kilian and Murphy (2013) and Baumeister and Peersman (forthcoming), among others.

For now we set aside all non-linear transformations of the price of oil and focus on linear forecasting models for the nominal price of oil and for the real price of oil. Non-linear joint forecasting models for U.S. real GDP and the price of oil based on net oil price increases are discussed in Section 12.

4. GRANGER CAUSALITY TESTS

Much of the existing work on predicting the price of oil has focused on testing for the existence of a predictive relationship from macroeconomic aggregates to the price of oil. In the absence of structural change, the existence of predictability in population is a necessary precondition for out-of-sample forecastability (see Inoue and Kilian, 2004a). Within the linear VAR framework the absence of predictability from one variable to another in population may be tested using Granger non-causality tests. We distinguish between predictability for the nominal price of oil and for the real price of oil. These are

⁶ The behavioral rationale for the net oil price increase measure applies equally to the nominal price of oil and the real price of oil. Although Hamilton (2003) applied this transformation to the nominal price of oil, several other studies have recently explored models that apply the same transformation to the real price of oil (see, e.g., Kilian and Vigfusson, 2011a; Herrera et al., 2011).

potentially quite different questions. For example, if the U.S. CPI is predictable and the nominal price of oil is not, then one could find that changes in the real price of oil are predictable simply because inflation is predictable. On the other hand, if the real price of oil is unpredictable, but the U.S. CPI is predictable, then one would expect to be able to forecast the nominal price simply because inflation is predictable. Finally, the real price of oil, the nominal price of oil, and the U.S. CPI may all be predictable to varying degrees.

4.1. Nominal Oil Price Predictability

4.1.1. *The Pre-1973 Evidence*

Granger causality from macroeconomic aggregates to the price of oil has received attention in part because Granger non-causality is one of the testable implications of strict exogeneity. The notion that the percent change in the nominal price of oil may be considered exogenous with respect to the U.S. economy was bolstered by evidence in [Hamilton \(1983\)](#), who observed that there is no apparent Granger causality from U.S. domestic macroeconomic aggregates to the percent change in the nominal price of oil during 1948–1972. Of course, the absence of Granger causality is merely a necessary condition for strict exogeneity. Moreover, a failure to reject the null of no Granger causality is at best suggestive; it does not establish the validity of the null hypothesis. Hamilton's case for the exogeneity of the nominal price of oil with respect to the U.S. economy therefore rested primarily on the unique institutional features of the oil market during this period, discussed in Section 2, and on historical evidence that unexpected supply disruptions under this institutional regime appear to be associated with exogenous political events in the Middle East, allowing us to treat the resulting price spikes as exogenous with respect to the U.S. economy. Even if we accept Hamilton's interpretation of the pre-1973 period, the institutional conditions that [Hamilton \(1983\)](#) appeals to ceased to exist in the early 1970s. The question that matters for our purposes is to what extent there is evidence that oil prices can be predicted from macroeconomic aggregates in the post-1973 period.

4.1.2. *The Post-1973 Evidence*

There is widespread agreement among oil economists that, starting in 1973, nominal oil prices must be considered endogenous with respect to U.S. macroeconomic variables (see, e.g., [Kilian, 2008a](#)). This fact alone is not sufficient, however, for the nominal price of oil to be predictable on the basis of lagged U.S. macroeconomic aggregates. If the nominal price of oil instantaneously incorporates information about expected U.S. macroeconomic conditions, as suggested by some rational expectations models, this could render the nominal price of oil linearly unpredictable on the basis of lagged U.S. macroeconomic aggregates. This line of reasoning is familiar from the analysis of stock and bond prices as well as exchange rates.⁷

⁷ [Hamilton \(1994 p. 306\)](#) illustrates this point in the context of a model of stock prices and expected dividends.

A recent study by Kilian and Vega (2010) helps resolve this question. Kilian and Vega find no evidence of systematic feedback from news about a wide range of U.S. macroeconomic aggregates to the nominal price of oil at any horizon between 1 day and 1 month. This lack of evidence is in sharp contrast to the results obtained for other asset prices based on the same sample, so a lack of power of the procedure employed by Kilian and Vega cannot explain the absence of significant feedback from U.S. macroeconomic news to the nominal price of oil. These two results in conjunction allow us to rule out contemporaneous feedback from U.S. macroeconomic aggregates to the nominal price of oil and imply that lagged U.S. macroeconomic aggregates must have predictive power in population, if the nominal price of oil is indeed endogenous with respect to these macroeconomic aggregates.

Predictability in the context of linear vector autoregressions may be tested using Granger causality tests. Table 8.1 investigates the evidence of Granger causality from selected nominal U.S. and global macroeconomic predictors to the nominal price of oil. All results are based on pairwise vector autoregressions. The lag order is fixed at 12. We consider four alternative nominal oil price series. The evaluation period is 1975.2–2009.12. There are several reasons to expect the dollar-denominated nominal price of oil to respond to changes in nominal U.S. macroeconomic aggregates. One channel

Table 8.1 Marginal Significance Levels for Predictability from Selected Nominal Aggregates to Nominal Price of Oil Since 1973

Predictors	WTI	RAC Oil Imports	RAC Domestic Oil	RAC Composite
<i>Monthly Predictors:</i>				
<i>U.S. CPI</i>	0.072	0.024	0.234	0.097
<i>U.S. M1</i>	0.057	0.003	0.000	0.000
<i>U.S. M2</i>	0.101	0.184	0.014	0.062
<i>CRB Industrial Raw Materials Price Index</i>	0.000	0.000	0.001	0.002
<i>CRB Metals Price Index</i>	0.004	0.009	0.001	0.011
<i>U.S. 3-Month T-Bill Rate</i>	0.232	0.438	0.361	0.377
<i>Trade-Weighted U.S. Exchange Rate</i>	0.736	0.725	0.490	0.706
<i>Australian/U.S. Exchange Rate</i>	0.223	0.207	0.176	0.172
<i>Canadian/U.S. Exchange Rate</i>	0.022	0.003	0.155	0.019
<i>New Zealand/U.S. Exchange Rate</i>	0.548	0.667	0.218	0.438
<i>South African/U.S. Exchange Rate</i>	0.184	0.208	0.082	0.123

Notes: *p*-values for Wald tests of Granger non-causality causality based on heteroskedasticity-robust variance estimator. Boldface indicates significance at the 10% level. All test results are based on bivariate VAR(12) models. The evaluation period is 1975.2–2009.12. RAC stands for U.S. refiners' acquisition cost and CRB for the Commodity Research Bureau. All variables but the interest rate are expressed in percent changes. Interest rates are expressed in changes. In some cases, one needs to consider the possibility of cointegration in levels. In those cases, all rejections above remain significant if we follow Dolado and Lütkepohl (1996) in conducting a lag-augmented Granger non-causality test.

of transmission is purely monetary and operates through U.S. inflation. For example, Gillman and Nakov (2009) stress that changes in the nominal price of oil must occur in equilibrium just to offset persistent shifts in U.S. inflation, given that the price of oil is denominated in dollars. Indeed, the Granger causality tests in Table 8.1 indicate significant lagged feedback from U.S. headline CPI inflation to the percent change in the nominal price of oil except for the domestic refiners' acquisition cost, consistent with the findings in Gillman and Nakov (2009). The evidence of predictability is weaker for the domestic oil price series than for the price of oil imports.

Gillman and Nakov view changes in inflation in the post-1973 period as rooted in persistent changes in the growth rate of money.⁸ Thus, an alternative approach of testing the hypothesis of Gillman and Nakov (2009) is to focus on Granger causality from monetary aggregates to the nominal price of oil. Given the general instability in the link from changes in monetary aggregates to inflation, one would not necessarily expect changes in monetary aggregates to have much predictive power for the price of oil, except perhaps in the 1970s (see Barsky and Kilian, 2002). Table 8.1 nevertheless shows that there is statistically significant lagged feedback from narrow measures of money such as M1 to all four nominal oil price measures. The evidence for broader monetary aggregates such as M2 is weaker, with only two tests statistically significant.

A third approach to testing for a role for U.S. monetary conditions relies on the fact that rising dollar-denominated non-oil commodity prices are thought to presage rising U.S. inflation. To the extent that oil price adjustments are more sluggish than adjustments in other industrial commodity prices, one would expect changes in nominal Commodity Research Bureau (CRB) spot prices to Granger cause changes in the nominal price of oil. Indeed, Table 8.1 indicates highly statistically significant lagged feedback from CRB sub-indices for industrial raw materials and for metals. This evidence is also consistent with rising non-oil commodity prices being an indicator of shifts in the global demand for industrial commodities, in which case the predictability of the nominal price would arise because of the predictability of the real price of oil.

In contrast, neither short-term interest rates nor trade-weighted exchange rates have significant predictive power for the nominal price of oil. According to the Hotelling model, one would expect the nominal price of oil to grow at the nominal rate of interest if the marginal extraction cost is zero, providing yet another link from U.S. macroeconomic aggregates to the nominal price of oil.⁹ Table 8.1, however, shows no evidence

⁸ For an earlier exposition of the role of monetary factors in determining the price of oil see Barsky and Kilian (2002). Both Barsky and Kilian (2002) and Gillman and Nakov (2009) view the shifts in U.S. inflation in the early 1970s as caused by persistent changes in the growth rate of the money supply, but there are important differences in emphasis. Whereas Barsky and Kilian stress the effects of unanticipated monetary expansions on real domestic output, on the demand for oil and hence on the real price of oil, Gillman and Nakov stress that the relative price of oil must not decline in response to a monetary expansion, necessitating a higher nominal price of oil, consistent with anecdotal evidence on OPEC price decisions (see, e.g., Kilian, 2008b). These two explanations are complementary.

⁹ Specifically, we use the 3-month, 6-month, and 12-month constant-maturity Treasury bill rates from the Federal Reserve Board's website <http://federalreserve.gov/releases/H15/data.htm>

of statistically significant feedback from the 3-month T-Bill rate to the price of oil. This finding is not surprising as the price of oil was not even approximately growing at the rate of interest (see Figure 8.1). Nor is there evidence of significant feedback from lagged changes in the trade-weighted nominal U.S. exchange rate. This does not necessarily mean that all bilateral exchange rates lack predictive power. In related work, [Chen et al. \(2010\)](#) show that the floating exchange rates of small commodity exporters (including Australia, Canada, New Zealand, South Africa and Chile) with respect to the dollar in some cases have remarkably robust forecasting power for the global prices of their commodity exports. The explanation presumably is that these exchange rates are forward looking and embody information about future movements in commodity export markets that cannot easily be captured by other means.

Although Chen et al.'s analysis cannot be extended to oil exporters such as Saudi Arabia because Saudi Arabia's exchange rate has not been floating freely, the bilateral dollar exchange rates of Australia, Canada, New Zealand, and South Africa may serve as a proxy for expected broad-based movements in industrial commodity prices that may also be helpful in predicting changes in the nominal price of oil. According to Chen et al., the share of non-agricultural commodity exports is largest in South Africa, followed by Australia, Canada, and New Zealand. In general, the larger the share of non-agricultural exports, the higher one would expect the predictive power for industrial commodities to be. For the price of oil, the share of energy exports such as crude oil, coal, and natural gas may be an even better indicator of predictive power, suggesting that Canada should have the highest predictive power for the price of oil, followed by Australia, South Africa, and New Zealand. Table 8.1 shows strong evidence of predictability for the bilateral exchange rate of Canada, consistent with the intuition that the share of oil in commodity exports matters. There is no such evidence for the other commodity exporters. Moreover, when using the dollar exchange rate of the Japanese Yen and of the British Pound as a control group, there is no significant evidence of Granger causality from exchange rates to the price of oil.¹⁰

4.1.3. Reconciling the Pre- and Post-1973 Evidence on the Predictability of the Nominal Price of Oil

Table 8.1 suggests that indicators of U.S. inflation have significant predictive power for the nominal price of oil. This result is in striking contrast to the pre-1973 period. As shown in [Hamilton \(1983\)](#) using quarterly data and in [Gillman and Nakov \(2009\)](#) using monthly data, there is no significant Granger causality from U.S. inflation to the percent change in the nominal price of oil in the 1950s and 1960s. This difference in results is suggestive of a structural break in late 1973 in the predictive relationship between the price of oil and the U.S. economy. One reason that the pre-1973 predictive regressions

¹⁰ Although the U.K. has been exporting crude oil from the late 1970s until recently, its average share of petroleum exports is too low to consider the U.K. a commodity exporter (see [Kilian et al., 2009](#)).

differ from the post-1973 regressions is that prior to 1973 the nominal price of oil was adjusted only at discrete intervals (see Figure 8.1). Because the nominal oil price data were generated by a discrete-continuous choice model, conventional autoregressive or moving average time series processes are inappropriate for these data and tests of the predictability of the price of oil based on such models have to be viewed with caution.

This problem with the pre-1973 data may be ameliorated by deflating the nominal price of oil, which renders the oil price data continuous and more amenable to VAR analysis (see Figure 8.2). Additional problems arise, however, when combining oil price data generated by a discrete-continuous choice process with data from the post-Texas Railroad Commission era that are fully continuous. Concern over low power has prompted many applied researchers to combine oil price data for the pre-1973 and post-1973 period in the same model when studying the predictive relationship from macroeconomic aggregates to the price of oil. This approach is obviously inadvisable when dealing with nominal oil price data. Perhaps less obviously, this approach is equally unappealing when dealing with vector autoregressions involving the real price of oil. The problem is that the nature and speed of the feedback from U.S. macroeconomic aggregates to the real price of oil differs by construction, depending on whether the nominal price of oil is temporarily fixed or not. This instability manifests itself in a structural break in the predictive regressions commonly used to test for lagged potentially non-linear feedback from the real price of oil to real GDP growth (see, e.g., Balke et al., 2002). The p -value for the null hypothesis that there is no break in 1973:Q4 in the coefficients of this predictive regression is 0.001 (see Kilian and Vigfusson, 2011b).¹¹ For that reason, regression estimates of the relationship between the real price of oil and domestic macroeconomic aggregates obtained from the entire post-war period are not informative about the strength of these predictive relationships in post-1973 data.¹² In the analysis of the real price of oil below we therefore restrict the evaluation period to start no earlier than 1974.1.

4.2. Real Oil Price Predictability in the Post-1973 Period

It is well established in natural resource theory that the real price of oil increases in response to low expected real interest rates and in response to high real aggregate output. Any analysis of the role of expected real interest rates is complicated by the fact that inflation expectations are difficult to pin down, especially at longer horizons, and that the relevant horizon for resource extraction is not clear. We therefore focus on the predictive power of fluctuations in real aggregate output. Table 8.2 reports p -values for tests of the

¹¹ Even allowing for the possibility of data mining, this break remains statistically significant at the 5% level.

¹² This situation is analogous to that of combining real exchange rate data for the pre- and post-Bretton Woods periods in studying the speed of mean reversion toward purchasing power parity. Clearly, the speed of adjustment toward purchasing power parity will differ if one of the adjustment channels is shut down, as was the case under the fixed exchange rate system, than when both prices and exchange rates are free to adjust, as has been the case under the floating rate system. Thus, regressions on long time spans of real exchange rate data produce average estimates that by construction are not informative about the speed of adjustment after the end of the Bretton Woods system.

Table 8.2 Marginal Significance Levels for Predictability from Selected Real Aggregates to Real Price of Oil Since 1973

Predictors	WTI		RAC Oil Imports		RAC Domestic Oil		RAC Composite	
	p=4	p=8	p=4	p=8	p=4	p=8	p=4	p=8
Quarterly Predictors:								
<i>U.S. Real GDP</i>								
<i>LT</i>	0.615	0.765	0.419	0.269	0.268	0.269	0.323	0.268
<i>HP</i>	0.594	0.705	0.401	0.883	0.304	0.630	0.343	0.691
<i>DIF</i>	0.746	0.457	0.386	0.567	0.234	0.668	0.271	0.620
<i>World Industrial Production^a</i>								
<i>LT</i>	0.014	0.108	0.050	0.149	0.016	0.130	0.024	0.135
<i>HP</i>	0.578	0.909	0.821	0.909	0.499	0.912	0.534	0.900
<i>DIF</i>	0.518	0.846	0.734	0.869	0.564	0.823	0.595	0.851
Monthly Predictors:								
<i>Chicago Fed National Activity Index (CFNAI)</i>								
	0.489	0.781	0.696	0.960	0.409	0.864	0.393	0.923
<i>U.S. Industrial Production</i>								
<i>LT</i>	0.179	0.126	0.195	0.512	0.023	0.184	0.031	0.193
<i>HP</i>	0.199	0.205	0.243	0.746	0.045	0.234	0.052	0.298
<i>DIF</i>	0.061	0.148	0.169	0.266	0.024	0.193	0.027	0.757
<i>OECD+6 Industrial Production^b</i>								
<i>LT</i>	0.001	0.004	0.002	0.019	0.002	0.012	0.001	0.015
<i>HP</i>	0.093	0.196	0.069	0.486	0.074	0.249	0.040	0.404
<i>DIF</i>	0.061	0.148	0.169	0.266	0.024	0.193	0.027	0.757
<i>Global Real Activity Index^c</i>								
	0.034	0.002	0.082	0.007	0.006	0.002	0.016	0.005

Notes: *p*-values for Wald tests of Granger non-causality based on heteroskedasticity-robust variance estimator. Boldface indicates significance at the 10% level. The evaluation period is 1976:2-2009:12 and 1976:II-2009:IV, respectively, unless noted otherwise. LT denotes linear detrending, HP denotes HP filtering with smoothing parameter $\lambda = 1600$ and $\lambda = 129600$, respectively (see Ravn and Uhlig, 2002), and DIF denotes first differencing. The CFNAI and the global real activity index are constructed to be stationary. RAC stands for U.S. refiners' acquisition cost. All test results are based on bivariate VAR(*p*) models. In the baseline specification the real price of oil is expressed in log levels, while the output variable is suitably transformed to remove trends. Similar results are obtained when both variables are detrended by the same method.

^a Data source: U.N. Monthly Bulletin of Statistics. These data end in 2008:III because the U.N. has temporarily suspended updates of this series, resulting in a shorter evaluation period.

^b Data source: OECD Main Economic Indicators.

^c Data source: Updated version of the index developed in Kilian (2009a).

hypothesis of Granger non-causality from selected measures of real aggregate output to the real price of oil.

A natural starting point is U.S. real GDP. Economic theory implies that U.S. real GDP and the real price of oil are mutually endogenous and determined jointly. For example, one would expect an unexpected increase in U.S. real GDP, all else equal, to increase the flow demand for crude oil and hence the real price of oil. Unless the real price of oil is forward looking and already embodies all information about future U.S. real GDP, a reasonable conjecture therefore is that lagged U.S. real GDP should help predict the real price of oil. Recent research by Kilian and Murphy (2013) has shown that the real price of oil indeed contains an asset price component, but that this component most of the time explains only a small fraction of the historical variation in the real price of oil. Thus, we would expect fluctuations in U.S. real GDP to predict the real price of oil at least in population. Under the assumption that the joint process can be approximated by a linear vector autoregression, this implies the existence of Granger causality from U.S. real GDP to the real price of oil

Notwithstanding this presumption, Table 8.2 indicates no evidence of Granger causality from U.S. real GDP growth to the real price of oil. This finding is robust to alternative methods of detrending and alternative lag orders. In the absence of instantaneous feedback from U.S. real GDP to the real price of oil, a finding of Granger non-causality from U.S. real GDP to the real price of oil – in conjunction with evidence that the real price of oil Granger causes U.S. real GDP – would be consistent with the real price of oil being strictly exogenous with respect to U.S. real GDP. It can be shown, however, that the evidence of reverse Granger causality from the real price of oil to U.S. real GDP is not much stronger, suggesting that the test is simply not informative because of low power. In fact, this is precisely the argument that prompted some researchers to combine data from the pre-1973 and post-1973 period – a strategy that we do not recommend for the reasons discussed in Section 4.1.3.

Another likely explanation of the failure to reject the null of no predictability is model misspecification. It is well known that Granger causality in a bivariate model may be due to an omitted third variable, but equally relevant is the possibility of Granger non-causality in a bivariate model arising from omitted variables (see Lütkepohl, 1982). This possibility is more than a theoretical curiosity in our context. Recent models of the determination of the real price of oil after 1973 have stressed that this price is determined in global markets (see, e.g., Kilian, 2009a; Kilian and Murphy, 2013). In particular, the demand for oil depends not merely on U.S. demand, but on global demand. The bivariate model for the real price of oil and U.S. real GDP by construction omits fluctuations in real GDP in the rest of the world. The relevance of this point is that offsetting movements in real GDP abroad can easily obscure the effect of changes in U.S. real GDP and eclipse the dynamic relationship of interest, lowering the power of the Granger causality test. Only when real GDP fluctuations are highly correlated across countries would we expect U.S. real GDP

to be a good proxy for world real GDP.¹³ In addition, as the U.S. share in world GDP evolves, by construction so do the predictive correlations underlying Table 8.2. In this regard, Kilian and Hicks (forthcoming) have documented dramatic changes in the PPP-adjusted share in GDP of the major industrialized economies and of the main emerging economies in recent years that cast further doubt on the U.S. real GDP results in Table 8.2. For example, China, and India combined have almost as high a share in world GDP today as the United States.

A closely related third point is that fluctuations in real GDP are a poor proxy for business-cycle driven fluctuations in the demand for oil. It is well known, for example, that in recent decades the share of services in U.S. real GDP has greatly expanded at the cost of manufacturing and other sectors. Clearly, real GDP growth driven by the non-service sector will be associated with disproportionately higher demand for oil and other industrial commodities than real GDP growth in the service sector. This provides one more reason why one would not expect a strong or stable predictive relationship between U.S. real GDP and the real price of oil.

An alternative quarterly predictor that partially addresses these last two concerns is quarterly world industrial production from the *U.N. Monthly Bulletin of Statistics*. This series has recently been introduced by Baumeister and Peersman (forthcoming) in the context of modeling the demand for oil. Although there are serious methodological concerns regarding the construction of any such index, as discussed in Beyer et al. (2001), one would expect this series to be a better proxy for global fluctuations in the demand for crude oil than U.S. real GDP. Indeed, Table 8.2 shows evidence of Granger causality from world industrial production to all four real oil price series for the LT model specification with four lags. Likewise, for the LT specification with eight lags, the reduction in p -values compared with U.S. real GDP is dramatic. The fact that there is evidence of predictability only for the linearly detrended series is consistent with the view, expressed in Kilian (2009b), that the demand for industrial commodities such as crude oil is subject to long swings. Detrending methods such as HP filtering or first differencing eliminate much of this low frequency covariation in the data, making it more difficult to detect predictability.

Additional insights may be gained by focusing on monthly rather than quarterly predictors in the lower panel of Table 8.2. The first contender is the Chicago Fed National Activity Index (CFNAI). This is a broad measure of monthly real economic activity in the United States obtained from applying principal components analysis to a wide range of monthly indicators of real activity expressed in growth rates (see Stock and Watson, 1999).

¹³ For example, the conjunction of rising growth in emerging Asia with unchanged growth in the U.S. all else equal would cause world GDP growth and hence the real price of oil to increase, but would imply a zero correlation between U.S. real GDP growth and changes in the real price of oil. Alternatively, slowing growth in Japan and Europe may offset rising growth in the U.S., keeping the real price of oil stable and implying a zero correlation of U.S. growth with changes in the real price of oil. This does not mean that there is no feedback from lagged U.S. real GDP. Indeed, with lower U.S. growth the increase in the real price of oil would have slowed in the first example and without offsetting U.S. growth the real price of oil would have dropped in the second example.

As in the case of quarterly U.S. real GDP, there is no evidence of Granger causality. If we rely on U.S. industrial production as the predictor, the p -values drop across the board. There even is statistically significant evidence of feedback to the real price of oil for some specifications, but the evidence is weak, especially for the import price. There are no monthly data on world industrial production, but the OECD provides an industrial production index for OECD economies and six selected non-OECD countries. As expected, the rejections of Granger non-causality become much stronger when we focus on OECD+6 industrial production. Table 8.2 indicates strong and systematic Granger causality, especially for the LT specification. Even OECD+6 industrial production, however, is an imperfect proxy for business-cycle driven fluctuations in the global demand for industrial commodities such as crude oil.

An alternative is the index of global real activity recently proposed in Kilian (2009a). This index does not rely on any country weights and has global coverage. It has been constructed with the explicit purpose of capturing unexpected fluctuations in the broad-based demand for industrial commodities associated with the global business cycle in the context of structural oil market models.¹⁴ It also is a good predictor of the real price of oil. The last row of Table 8.2 indicates strong evidence of Granger causality from this index to the real price of oil, regardless of the definition of the oil price series. That evidence becomes even stronger, once we allow for 2 years' worth of lags rather than 1 year. This finding mirrors the point made in Hamilton and Herrera (2004) that it is essential to allow for a rich lag structure in studying the dynamic relationship between the economy and the price of oil.

Although none of the proxies for global fluctuations in demand is without limitations, we conclude that there is a robust pattern of Granger causality, once we correct for problems of model misspecification and of data measurement that undermine the power of the test. This conclusion is further strengthened by evidence in Kilian and Hicks (forthcoming) based on distributed lag models that revisions to professional real GDP growth forecasts have significant predictive power for the real price of oil during 2000.11–2008.12 after weighting each country's forecast revision by its share in PPP weighted world real GDP. Predictability in population, of course, does not necessarily imply out-of-sample

¹⁴ This index is constructed from ocean shipping freight rates. The idea of using fluctuations in shipping freight rates as indicators of shifts in the global real activity dates back to Isserlis (1938) and Tinbergen (1959). The panel of monthly freight-rate data underlying the global real activity index was collected manually from *Drewry's Shipping Monthly* using various issues since 1970. The data set is restricted to dry cargo rates. The earliest raw data are indices of iron ore, coal and grain shipping rates compiled by *Drewry's*. The remaining series are differentiated by cargo, route and ship size and may include in addition shipping rates for oilseeds, fertilizer and scrap metal. In the 1980s, there are about 15 different rates for each month; by 2000 that number rises to about 25; more recently that number has dropped to about 15. The index was constructed by extracting the common component in the nominal spot rates. The resulting nominal index is expressed in dollars per metric ton, deflated using the U.S. CPI and detrended to account for the secular decline in shipping rates. For this paper, this series has been extended based on the Baltic Exchange Dry Index, which is available from Bloomberg. The latter index, which is commonly discussed in the financial press, is essentially identical to the nominal data underlying the Kilian (2009a) index, but only available since 1985.

forecastability in practice (see [Inoue and Kilian, 2004a](#)). Not only is the objective of predictability testing fundamentally different from that of forecasting out of sample, but the smaller samples available for the construction of forecasting models necessitate simpler models and cause estimates to be less precise. In addition, structural changes in the data generating process could potentially undermine a model's out-of-sample accuracy. The next sections therefore examine alternative approaches to forecasting the nominal and the real price of oil out-of-sample.

5. SHORT-HORIZON FORECASTS OF THE NOMINAL PRICE OF OIL

The most common approach to forecasting the nominal price of oil is to treat the price of the oil futures contract of maturity h as the h -period forecast of the price of oil.¹⁵ In particular, many central banks and the International Monetary Fund (IMF) use the price of NYMEX oil futures as a proxy for the market's expectation of the spot price of crude oil. A widespread view is that prices of NYMEX futures contracts are not only good proxies for the expected spot price of oil, but also better predictors of oil prices than econometric forecasts. Forecasts of the spot price of oil are used as inputs in the macroeconomic forecasting exercises that these institutions produce. For example, the European Central Bank (ECB) employs oil futures prices in constructing the inflation and output-gap forecasts that guide monetary policy (see [Svensson, 2005](#)). Likewise the IMF relies on futures prices as a predictor of future spot prices (see, e.g., [International Monetary Fund 2005, p. 67; 2007, p. 42](#)). Futures-based forecasts of the price of oil also play a role in policy discussions at the Federal Reserve Board. This is not to say that forecasters do not recognize the potential limitations of futures-based forecasts of the price of oil. Nevertheless, the perception among many macroeconomists, financial analysts and policymakers is that oil futures prices, imperfect as they may be, are the best available forecasts of the spot price of oil. Such attitudes have persisted notwithstanding recent empirical evidence to the contrary and notwithstanding the development of theoretical models aimed at explaining the lack of predictive ability of oil futures prices and spreads (see, e.g., [Knetsch, 2007; Alquist and Kilian, 2010](#)).

Interestingly, the conventional wisdom in macroeconomics and finance is at odds with long-held views about storable commodities in agricultural economics. For example, [Peck \(1985\)](#) emphasized that "expectations are reflected nearly equally in current and in futures prices. In this sense cash prices will be nearly as good predictions of subsequent

¹⁵ Futures contracts are financial instruments that allow traders to lock in today a price at which to buy or sell a fixed quantity of the commodity at a predetermined date in the future. Futures contracts can be retraded between inception and maturity on a futures exchange such as the New York Mercantile Exchange (NYMEX, now CME Group). This exchange offers institutional features that allow traders to transact anonymously. These features reduce individual default risk and ensure homogeneity of the traded commodity, making the futures market a low-cost and liquid mechanism for hedging against and for speculating on oil price risks. The NYMEX light sweet crude contract traditionally has been the most liquid and largest volume market for crude oil trading.

cash prices as futures prices,” echoing in turn the discussion in Working (1942) who was critical of the “general opinion among economists that prices of commodity futures are . . . the market expression of consciously formed opinions on probable prices in the future” whereas “spot prices are not generally supposed to reflect anticipation of the future in the same degree as futures prices.” Working specifically criticized the error of “supposing that the prices of futures . . . tend to be more strongly influenced by these anticipations than are spot prices.” The next section investigates the empirical merits of these competing views in the context of oil markets.

5.1. Forecasting Methods Based on Monthly Oil Futures Prices

Alquist and Kilian (2010) recently provided a comprehensive evaluation of the forecast accuracy of models based on monthly oil futures prices using data ending in 2007.2, superseding earlier analysis by Wu and McCallum (2005), among others. Below we update their analysis until 2009.12 and expand the range of alternative forecasting models under consideration.¹⁶ In this subsection, attention is limited to forecast horizons of up to 1 year. Let $F_t^{(h)}$ denote the current nominal price of the futures contract that matures in h periods, S_t the current nominal spot price of oil, and $E_t[S_{t+h}]$ the expected future spot price at date $t+h$ conditional on information available at t .

A natural benchmark for forecasts based on the price of oil futures is provided by the random walk model without drift. This model implies that changes in the spot price are unpredictable, so the best forecast of the spot price of crude oil is simply the current spot price:

$$\hat{S}_{t+h|t} = S_t \quad h = 1, 3, 6, 9, 12 \quad (1)$$

This forecast is also known as the no-change forecast. In contrast, the common view that oil futures prices are the best available predictor of future oil prices implies the forecasting model:

$$\hat{S}_{t+h|t} = F_t^{(h)} \quad h = 1, 3, 6, 9, 12 \quad (2)$$

A closely related approach to forecasting the spot price of oil is to use the spread between the futures price and the spot price as an indicator of whether the price of oil is likely to go up or down. If the futures price equals the expected spot price, the spread should be an indicator of the expected change in spot prices. We explore the forecasting accuracy of the spread based on several alternative forecasting models. The simplest model is:

$$\hat{S}_{t+h|t} = S_t \left(1 + \ln (F_t^{(h)} / S_t) \right), \quad h = 1, 3, 6, 9, 12 \quad (3)$$

¹⁶ Because the Datastream data for the daily WTI spot price of oil used in Alquist and Kilian (2010) were discontinued, we rely instead on data from the Energy Information Administration and from Bloomberg. As a result the estimation window for the forecast comparison is somewhat shorter in some cases than in Alquist and Kilian (2010).

To allow for the possibility that the spread may be a biased predictor, it is common to relax the assumption of a zero intercept:

$$\hat{S}_{t+h|t} = S_t \left(1 + \hat{\alpha} + \ln (F_t^{(h)} / S_t) \right), \quad h = 1, 3, 6, 9, 12 \quad (4)$$

Alternatively, one can relax the proportionality restriction:

$$\hat{S}_{t+h|t} = S_t \left(1 + \hat{\beta} \ln (F_t^{(h)} / S_t) \right), \quad h = 1, 3, 6, 9, 12 \quad (5)$$

Finally, one can relax both the unbiasedness and proportionality restrictions:

$$\hat{S}_{t+h|t} = S_t \left(1 + \hat{\alpha} + \hat{\beta} \ln (F_t^{(h)} / S_t) \right), \quad h = 1, 3, 6, 9, 12 \quad (6)$$

Here $\hat{\alpha}$ and $\hat{\beta}$ denote least-squares estimates obtained in real time from recursive regressions.

Our empirical analysis is based on daily prices of crude oil futures on the NYMEX from the commercial provider *Price-Data.com*. The time series begins in March 30, 1983, when crude oil futures were first traded on the NYMEX, and extends through December 31, 2009. Contracts are for delivery at Cushing, Oklahoma. Trading ends four days prior to the 25th calendar day preceding the delivery month. If the 25th is not a business day, trading ends on the third business day prior to the last business day before the 25th calendar day. A common problem in constructing monthly futures prices of a given maturity is that an h -month contract may not trade on a given day. We use as the end-of-month value the price of the h -month futures contract which trades closest to the last trading day of this month among all daily contracts in that month. Our approach is motivated by the objective of computing in a consistent manner end-of-month time series of oil futures prices for different maturities. This allows us to match up end-of-month spot prices and futures prices as closely as possible. The daily spot price data are obtained from the webpage of the Energy Information Administration and refer to the price of West Texas Intermediate crude oil available for delivery at Cushing, Oklahoma.

Table 8.3 assesses the predictive accuracy of various forecasting models against the benchmark of a random walk without drift for horizons of 1, 3, 6, 9, and 12 months. The forecast evaluation period is 1991.1–2009.12 with suitable adjustments, as the forecast horizon is varied. The assessment of which forecasting model is most accurate may depend on the loss function of the forecaster (see Elliott and Timmermann, 2008). We report results for the MSPE ratio relative to the no-change forecast and for the relative frequency with which a forecasting model correctly predicts the sign of the change in the spot price based on the success ratio statistic of Pesaran and Timmermann (2009). We also formally test the null hypothesis that a given candidate forecasting model is as accurate as the no-change forecast against the alternative that the candidate model is more accurate than the no-change forecast. It should be noted that commonly used tests of equal predictive accuracy for nested models (including the tests we rely on in this chapter) by

Table 8.3 1-Month, 3-Month, 6-Month, 9-Month, and 12-Month Ahead Forecast Error Diagnostics for Nominal WTI Price

$\hat{S}_{t+h t}$	$h=1$		$h=3$		$h=6$		$h=9$		$h=12$	
	MSPE Ratio	Success Ratio	MSPE Ratio	Success Ratio	MSPE Ratio	Success Ratio	MSPE Ratio	Success Ratio	MSPE Ratio	Success Ratio
S_t	20.325	N.A.	95.451	N.A.	222.28	N.A.	282.32	N.A.	302.54	N.A.
$F_t^{(h)}$	0.988	0.465	0.998	0.465	0.991	0.509	0.978	0.548	0.941	0.557
$S_t \left(1 + \hat{\alpha} + \hat{\beta} \ln \left(F_t^{(h)} / S_t \right) \right)$	1.001	0.539	1.044	0.531	1.051	0.535	1.042	0.583	1.240	0.537
$S_t \left(1 + \hat{\beta} \ln \left(F_t^{(h)} / S_t \right) \right)$	0.995	0.531	0.990	0.474	0.978	0.535	0.989	0.553	1.052	0.528
$S_t \left(1 + \hat{\alpha} + \ln \left(F_t^{(h)} / S_t \right) \right)$	1.002	0.513	1.026	0.518	1.024	0.544	1.019	0.561	1.281	0.528
$S_t \left(1 + \ln \left(F_t^{(h)} / S_t \right) \right)$	0.988	0.465	0.998	0.465	0.995	0.509	0.985	0.548	0.950	0.557
$S_t \left(1 + \Delta s_t \right)^h$	1.397	0.504	2.325	0.535	8.580	0.539	29.179	0.509	179.77	0.496
$S_t \left(1 + \hat{\alpha} \right)$	1.006	0.531	1.032	0.561	1.057	0.557	1.066	0.447	1.093	0.407
$S_t \left(1 + \Delta \bar{s}_t^{(h)} \right)$	1.397	0.504	1.678	0.539	2.225	0.504	2.816	0.487	3.746	0.439
$S_t \left(1 + i_{t,h} \right)^{h/12}$	N.A.	N.A.	1.000	0.575	1.002	0.575	N.A.	N.A.	0.998	0.566
$S_t \left(1 + \Delta e_{t,AUS} \right)^h$	0.865	0.513	0.811	0.553	1.071	0.561	1.352	0.583	1.678	0.583
$S_t \left(1 + \Delta e_{t,CAN} \right)^h$	0.930	0.478	0.918	0.496	0.966	0.526	0.990	0.539	1.144	0.504
$S_t \left(1 + \Delta \bar{e}_{t,h}^{AUS} \right)$	0.865	0.513	0.872	0.579	1.080	0.662	1.148	0.618	1.180	0.610
$S_t \left(1 + \Delta \bar{e}_{t,h}^{CAN} \right)$	0.931	0.478	0.926	0.526	0.917	0.601	1.022	0.592	1.185	0.544

(Continued)

Table 8.3 Continued

$\hat{S}_{t+h t}$	$h=1$		$h=3$		$h=6$		$h=9$		$h=12$	
	MSPE Ratio	Success Ratio	MSPE Ratio	Success Ratio	MSPE Ratio	Success Ratio	MSPE Ratio	Success Ratio	MSPE Ratio	Success Ratio
$S_t(1 + \Delta \bar{p}_t^{GRB,ind})_h$	0.913	0.583	0.802	0.605	0.976	0.614	1.402	0.570	1.846	0.566
$S_t(1 + \Delta \bar{p}_t^{GRB,met})_h$	1.031	0.579	0.942	0.636	1.574	0.592	3.374	0.561	7.170	0.548
$S_t(1 + \Delta \bar{p}_t^{GRB,ind,(h)})$	0.913	0.583	0.782	0.601	1.055	0.583	1.076	0.553	1.035	0.548
$S_t(1 + \Delta \bar{p}_t^{GRB,met,(h)})$	1.031	0.579	0.750	0.601	1.219	0.623	1.304	0.575	1.278	0.539

Notes: The forecast evaluation period is 1991.1–2009.12. The initial estimation window is 1986.1–1990.12. For regressions based on 6-month futures prices the estimation window begins in 1986.1; for the 9-month futures price in 1986.12; for the 12-month futures price in 1989.1. $F_t^{(h)}$ is the futures price that matures in h periods; $i_{t,m}$ is the m month interest rate; S_t is the percent change in S_t in the most recent month; and $\Delta \bar{p}_t^{(h)}$ is the percent change in the spot price over the most recent h months. All MSPE results are presented as ratios relative to the benchmark no-change forecast model, for which we report the level of the MSPE. The success ratio is defined as the fraction of forecasts that correctly predict the sign of the change in the price of oil. Results that are statistically significant at the 10% level are shown in boldface. All tests of statistical significance refer to pairwise tests of the null of equal predictive accuracy with the no-change forecast. Comparisons of non-nested models without estimated parameters are based on the DM-test of Diebold and Mariano (1995) using $N(0,1)$ critical values; the results for other non-nested comparisons are obtained by bootstrapping. Nested model comparisons with estimated parameters are obtained by bootstrapping the DM-test statistic as in Clark and McCracken (2005a) and Killian (1999). The sign test in the last column is based on Pesaran and Timmermann, 2009. This test cannot be applied when there is no variability in the predicted sign.

construction are tests of the null of no predictability in population rather than tests of equal out-of-sample MSPEs (see, e.g., Inoue and Kilian (2004a,b); Clark and McCracken, 2010). This means that these tests will reject the null of equal predictive accuracy more often than they should under the null, suggesting caution in interpreting test results that are only marginally statistically significant. We will discuss this point in more detail further below. This concern does not affect non-nested forecast accuracy comparisons.

Row (2) of the results in Table 8.3 shows that the oil futures price has lower MSPE than the no-change forecast at all horizons considered, but the differences are small and none of the differences is statistically significant. For all practical purposes, the forecasts are equally accurate. Nor do futures forecasts have important advantages when it comes to predicting the sign of the change in the nominal price of oil. Only at the 12-month horizon is the success ratio significant at the 10% level. The improvement in this case is 5.7%. At the 1-month and 3-month horizon, the success ratio of the futures price forecast actually is inferior to tossing a coin. Similarly, rows (3)–(6) in Table 8.3 show no systematic difference between the MSPE of the spread-based forecasts and that of the random walk forecast. In no case is there a statistically significant reduction in the MSPE from using the spread model. In the rare cases in which one of the spread models significantly helps predict the direction of change, the gains in accuracy are quite moderate. No spread model is uniformly superior to the others.

We conclude that there is no compelling evidence that, over this sample period, monthly oil futures prices were more accurate predictors of the nominal price of oil than simple no-change forecasts. Put differently, a forecaster using the most recent spot price would have done just as well in forecasting the nominal price of oil. This finding is broadly consistent with the empirical results in Alquist and Kilian (2010). To the extent that some earlier studies have reported evidence more favorable to oil futures prices, the difference in results can be traced to the use of shorter samples.¹⁷

5.2. Other Forecasting Methods

The preceding subsection demonstrated that simple no-change forecasts of the price of oil tend to be about as accurate in the MSPE sense as forecasts based on oil futures prices, but this does not rule out that there are alternative predictors with even lower MSPE. Next we broaden the range of forecasting methods to include some additional predictors that are of practical interest. One approach is the use of parsimonious regression-based forecasting models of the spot price of crude oil. Another approach is to adjust the no-change forecast on the basis of readily available macroeconomic indicators. For example, we may exploit the implication of the standard Hotelling (1931) model that the price of oil should grow at the rate of interest. Similarly parsimonious forecasting models may be

¹⁷ Although we have focused on the WTI price of oil, qualitatively similar results would also be obtained on the basis of Brent spot and Brent futures prices, which are available from the same data sources. The evaluation period for the Brent price series, however, is shorter, casting doubt on the reliability of the results, which is why we focus on the WTI data.

constructed by adjusting the no-change forecast for recent percent changes in bilateral exchange rates or in non-oil industrial commodity prices.

5.2.1. Alternative Atheoretical Forecasting Models

One alternative is the double-differenced forecasting model proposed in [Hendry \(2006\)](#). Hendry observed that, when time series are subject to infrequent trend changes, the no-change forecast may be improved upon by extrapolating today's oil price at the most recent growth rate:

$$\hat{S}_{t+h|t} = S_t (1 + \Delta s_t)^h \quad h = 1, 3, 6, 9, 12 \quad (7)$$

where Δs_t denotes the percent growth rate between $t - 1$ and t . In other words, we apply the no-change forecast to the growth rate rather than the level. Although there are no obvious indications of structural change in our sample period, it is worth exploring this alternative method, given the presence of occasional large fluctuations in the price of oil. Row (7) in [Table 8.3](#) shows that the double-differenced specification does not work well in this case. Especially at longer horizons, this forecasting method becomes erratic and suffers from very large MSPEs. Nor is this method particularly adept at predicting the sign of the change in the nominal price of oil.

Yet another strategy is to extrapolate from recent trends. Given that oil prices have been persistently trending upward (or downward) at times, it is natural to consider a random walk model with drift. One possibility is to estimate this drift recursively, resulting in the forecasting model:

$$\hat{S}_{t+h|t} = S_t (1 + \hat{\alpha}) \quad h = 1, 3, 6, 9, 12 \quad (8)$$

Alternatively, a local drift term may be estimated using rolling regressions:

$$\hat{S}_{t+h|t} = S_t (1 + \Delta \bar{s}_t^{(h)}) \quad h = 1, 3, 6, 9, 12, \quad (9)$$

where $\hat{S}_{t+h|t}$ is the forecast of the spot price at $t + h$; and $\Delta \bar{s}_t^{(h)}$ is the percent change in the spot price over the most recent h months. This local drift model postulates that traders extrapolate from the spot price's recent behavior when they form expectations about the future spot price. The local drift model is designed to capture "short-term forecastability" that arises from local trends in the oil price data. Rows (8)–(9) in [Table 8.3](#) document that allowing for a drift typically increases the MSPE and in no case significantly lowers the MSPE relative to the no-change forecast, whether the drift is estimated based on rolling regressions or is estimated recursively. Nor does allowing for a drift significantly improve the ability to predict the sign of the change in the nominal price of oil.

5.2.2. Parsimonious Economic Forecasting Models

An alternative approach to forecasting the nominal price of oil is to rely on economic predictors. One forecast rule is motivated by [Hotelling's \(1931\)](#) model, which predicts that the price of an exhaustible resource such as oil appreciates at the risk-free rate of

interest, provided the marginal extraction cost is zero:

$$\hat{S}_{t+h|t} = S_t(1 + i_{t,h})^{h/12} \quad h = 3, 6, 12, \quad (10)$$

where $i_{t,h}$ refers to the annualized interest rate at the relevant maturity h .¹⁸ Although the Hotelling model may seem too stylized to generate realistic predictions, we include it in this forecast accuracy comparison. We employ the Treasury bill rate as a proxy for the risk free rate. Row (10) in Table 8.3 shows no evidence that adjusting the no-change forecast for the interest rate significantly lowers the MSPE. The Hotelling model is better at predicting the sign of the change in the nominal price of oil than the no-change forecast, although we cannot assess the statistical significance of the improvement, given that there is no variability at all in the sign forecast.

An alternative approach builds on the insights of Chen et al. (2010) we discussed earlier and uses recent percent changes in the bilateral nominal dollar exchange rate of selected commodity exporters to extrapolate the current price of oil:

$$\hat{S}_{t+h|t} = S_t(1 + \Delta e_t^i)^h \quad h = 1, 3, 6, 9, 12, \quad (11)$$

where $i \in \{\text{Canada, Australia}\}$. To conserve space we do not include results for New Zealand and South Africa, but note that the latter results are typically statistically insignificant and indicate little evidence of forecast ability. Table 8.3 shows that model (11) based on the Canadian dollar and the Australian dollar exchange rate yield reductions in the MSPE as high as 19% at short horizons, but these gains are not statistically significant. There is evidence that the Australian exchange rate has significant predictive power for the sign of the change in the nominal price of oil at horizons 3, 6, 9, and 12. We also considered the alternative specification

$$\hat{S}_{t+h|t} = S_t(1 + \Delta \bar{e}_{t,h}^i) \quad h = 1, 3, 6, 9, 12, \quad (12)$$

based on the percent change in the exchange rate over the most recent h months, where $i \in \{\text{Canada, Australia}\}$ as before. This specification helps smooth the exchange rate changes. The use of model (12) improves the directional accuracy of both exchange rate models. It also implies statistically significant short-run MSPE reductions as high as 13% up for the Australian dollar and as high as 8% for the Canadian dollar. The MSPE reductions for the Canadian dollar remain significant at the 6-month horizon. We conclude that forecasting models for the nominal price of oil based on bilateral exchange-rates have some predictive power at horizons up to 1 year, provided we focus on broad-based commodity exporters such as Canada and Australia.

A third approach is to explore the forecasting value of recent percent changes in non-oil CRB commodity prices, building on insights about the link between oil and

¹⁸ Assuming perfect competition, no arbitrage, and no uncertainty, oil companies extract oil at a rate that equates: (1) the value today of selling the oil less the costs of extraction; (2) and the present value of owning the oil, which, given the model's assumptions, is discounted at the risk free rate. In competitive equilibrium, oil companies extract crude oil at the socially optimal rate.

non-oil commodity prices discussed in Barsky and Kilian (2002) and Kilian (2009b), among others. One such forecasting model is

$$\hat{S}_{t+h|t} = S_t(1 + \Delta p_t^{com})^h \quad h = 1, 3, 6, 9, 12, com \in \{ind, met\}. \quad (13)$$

It can be shown that model (13) does not produce statistically significant reductions in the MSPE, presumably because month-to-month changes in commodity prices tend to be noisy. In fact, model (13) tends to worsen the MSPE ratio at long horizons, although it significantly improves directional accuracy at horizons up to 9 months for metals prices and up to 12 months for prices of industrial raw materials. An alternative model specification is based on the percent change in the CRB price index over the most recent h months:

$$\hat{S}_{t+h|t} = S_t(1 + \Delta \bar{p}_{t,h}^{com}) \quad h = 1, 3, 6, 9, 12, \quad com \in \{ind, met\}. \quad (14)$$

Model (14) is designed to capture persistent changes in commodity prices in the recent past. This specification is less successful at predicting the direction of change at horizons beyond 6 months, but can yield significant reductions in the MSPE at short horizons. For example, the model using metals prices significantly lowers the MSPE at horizon 3 and the model using prices of industrial raw materials significantly reduces the MSPE at horizons 1 and 3. The MSPE reductions may be as large as 25% at horizon 3.

We conclude that forecasting models extrapolating persistent changes in CRB industrial commodity prices significantly reduce the MSPE of the nominal price of oil relative to the no-change forecast at short horizons. Beyond the 3-month horizon, based on the MSPE criterion, the no-change forecast for all practical purposes remains the most accurate model for forecasting the nominal price of oil in real time. Based on directional accuracy somewhat more favorable results are obtained. The model based on persistent changes in recent commodity prices model is significantly more accurate than the no-change forecast even at the 6-month horizon. These results are encouraging in that they demonstrate that even very simple models may outperform the random walk for the nominal price of oil.

5.3. Short-Horizon Forecasts Based on Daily Oil Futures Prices

Following the extant literature, our analysis so far has relied on monthly data for oil futures prices and spreads constructed from daily observations. The construction of monthly data allows one to compare the accuracy of these forecasts to that of alternative forecasts based on data only available at monthly frequency. A complementary approach is to utilize all daily oil futures prices and compare their forecasting accuracy to the no-change forecast only. This alternative approach makes use of all oil-futures price data and hence may have more accurate size and higher power. It is not without drawbacks, however. Ideally, one would like to compare the price of a futures contract for delivery in h months with the price of delivery exactly h months later, where one month corresponds to 21 business days. That price, however, is not observed. The spot price quoted on the day of delivery

Table 8.4 Short-Horizon Forecasts of the Nominal WTI Price of Oil from Daily Oil Futures Prices since January 1986

	Start of evaluation period: January 1986									
	<i>h</i> =1		<i>h</i> =3		<i>h</i> =6		<i>h</i> =9		<i>h</i> =12	
	MSPE	SR	MSPE	SR	MSPE	SR	MSPE	SR	MSPE	SR
$F_t^{(h)}$	0.963 (0.009)	0.522 (0.040)	0.972 (0.053)	0.516 (0.072)	0.973 (0.077)	0.535 (0.002)	0.964 (0.063)	0.534 (0.001)	0.929 (0.001)	0.562 (0.000)

Notes: There are 5968, 5926, 5861, 5744, and 5028 daily observations at horizons of 1 through 12 months, respectively. Following Leamer's (1978) rule for adjusting the threshold for statistical significance with changes in the sample size, *p*-values below about 0.0035 are considered statistically significant and are shown in boldface.

instead will be the price for delivery sometime in the month following the date on which the futures contract matures. In fact, the date of delivery associated with a given spot price can never be made exact. We therefore follow the convention of evaluating futures price forecasts against the spot price prevailing when the futures contract matures. A reasonable case can be made that this is what practitioners view as the relevant forecasting exercise.

Note that the daily data are sparse in that there are many days for which no price quotes exist. We eliminate these dates from the sample and stack the remaining observations similar to the approach taken in Kilian and Vega (2010) in the context of modeling the impact of U.S. macroeconomic news on the nominal price of oil. Table 8.4 summarizes our findings. The MSPE ratios in Table 8.4 indicate somewhat larger gains in forecasting accuracy from using oil futures prices than in Table 8.3. There are a number of caveats, however. First, the *h*-month oil futures forecasts are not forecasts for a horizon of *h* months as in Table 8.3, but rather for a horizon that may vary arbitrarily between *h* and *h* + 1 months. For example, an oil futures contract quoted on August 13 for delivery starting on October 1 would be considered a 1-month contract for the purpose of Table 8.4, but so would an oil-futures contract quoted on August 25 for delivery starting on October 1. This is an inherent limitation of working with daily oil futures price data. This concern suggests caution in interpreting short-horizon results, but obviously becomes less important as *h* increases. A second concern is that the sample period spanned by the daily data extends back to January 1986, whereas the data in Table 8.3 start in 1990. This difference is not driving the results in Table 8.4. It can be shown that making the sample period compatible with that in the earlier tables would yield substantively identical results.

The third and most important concern is the statistical significance of the results in Table 8.4. Given that the sample size in Table 8.4 is larger than in Table 8.3 by a factor of about 10, care must be exercised in interpreting the *p*-values. As is well known, for sufficiently large sample sizes, any null hypothesis is bound to be rejected at conventional significance levels, making it inappropriate to apply the same significance level as

in Table 8.3. In recognition of this problem, Leamer (1978 p. 108-120) proposes a rule for constructing sample-size dependent critical values. For example, for the F -statistic, the appropriate level of statistical significance is $\alpha = 1 - f_{cdf}((t-1) \times (t^{(1/t)} - 1), 1, t)$. For $t = 216$, as in Table 8.3, this rule of thumb implies a threshold for rejecting the null hypothesis of $\alpha = 0.0209$. In contrast, for $t = 5968$ the same rule implies a much higher threshold of $\alpha = 0.0032$. Applying this rule to the p -values in Table 8.4, none of the MSPE reductions is statistically significant except at the 12-month horizon. The MSPE ratio at the 12-month horizon of 0.93 is similar to the ratio of 0.94 reported in Table 8.3 based on monthly data. There also is evidence that at horizons 6, 9 and 12, the oil futures price has statistically significant directional accuracy, but the gains are quantitatively negligible except perhaps at horizon 12.

These results lead us to revise somewhat our earlier findings. We conclude that there is statistically significant evidence that oil futures prices improve on the accuracy of the no-change forecast of the nominal price of oil at the 1-year horizon, but not at shorter horizons. The magnitude of these gains in accuracy is modest – at least by the standards of the literature on forecasting macroeconomic aggregates such as inflation rates. Moreover, there are indications that this result is sensitive to changes in the sample period and may not be robust as more data accumulate. After eliminating the data beyond March 2008, for example, the MSPE ratio of the 12-month futures price exceeds 1 and only when extending the sample period beyond July 2008 is the MSPE reduction statistically significant. This result, together with the lack of evidence for slightly shorter or slightly longer futures contracts, suggests caution in interpreting the evidence for the 12-month contract in Table 8.4.

6. LONG-HORIZON FORECASTS OF THE NOMINAL PRICE OF OIL BASED ON OIL FUTURES PRICES

For oil industry managers facing investment decisions or for policymakers pondering the medium-term economic outlook a horizon of 1 year is too short. Crude oil futures may have maturities as long as 7 years. Notwithstanding the low liquidity of oil futures markets at such long horizons, documented in Alquist and Kilian (2010), it is precisely these long horizons that many policymakers focus on. For example, Greenspan (2004a) explicitly referred to the 6-year oil futures contract in assessing effective long-term supply prices. For similar statements also see Greenspan (2004b), Gramlich (2004), and Bernanke (2004). In this section we focus on forecasting the nominal price of oil at horizons up to 7 years.

It can be shown that the daily data are too sparse at horizons beyond 1 year to allow the construction of time series of end-of-month observations for oil futures prices. However, we can instead evaluate each daily futures price quote for contracts of any given maturity against the spot price that is realized on the day the contract expires. We already used

this approach in Table 8.4 for horizons up to 1 year. One drawback of extending this approach to longer horizons is that the evaluation period for long-horizon contracts may exclude many of the particularly informative observations at the end of our sample period. Another drawback is that long-horizon futures prices are sparsely quoted, greatly reducing the sample size as the horizon is lengthened. For that reason, one would expect the results to be far less reliable than the earlier short-horizon results. Nevertheless, they provide the only indication we have of the usefulness of oil futures prices at the horizons at which they are employed by many policymakers.

Table 8.5 shows the results for horizons of 2, 3, 4, 5, 6, and 7 years. In sharp contrast with Table 8.4 the MSPE ratios are consistently above 1, indicating that oil futures prices are less accurate than the no-change forecast. In no case is there evidence of significant reductions in the MSPE. The test for directional accuracy is statistically significant at the 2 year horizon, but not at longer horizons. In fact, in many cases the success ratios at longer horizons are distinctly worse than tossing a coin. Table 8.5 provides no evidence in support of the common practice at central banks of appealing to the price of long-horizon oil futures contracts as an indication of future spot prices. In particular, at a horizon of 6 years, which figures prominently in policy statements and speeches, central bankers would have been much better off relying on the no-change forecast than on oil futures prices.

An interesting question is whether the poor accuracy of forecasts from oil futures prices beyond 1 year simply reflects a sharp drop-off in the liquidity of oil futures markets at longer horizons. This does not appear to be the case. Figure 8.4 plots two measures of the liquidity of the oil futures market by horizon. Open interest is the total number of futures contracts, either long or short, that have been entered into for a given delivery

Table 8.5 Long-Horizon Forecasts of the Nominal WTI Price of Oil from Daily Oil Futures Prices

<i>h</i> (in years)	Starting Date	Sample Size	MSPE	SR
2	11/20/90	3283	1.159 (1.000)	0.515 (0.000)
3	05/29/91	515	1.168 (0.996)	0.518 (0.281)
4	11/01/95	194	1.212 (1.000)	0.294 (N.A.)
5	11/03/97	154	1.280 (1.000)	0.247 (N.A.)
6	11/03/97	134	1.158 (0.999)	0.276 (N.A.)
7	11/21/97	22	1.237 (0.957)	0.500 (N.A.)

Notes: Following Leamer's (1978) rule for adjusting the threshold for statistical significance with changes in the sample size, *p*-values below 0.0044 for a horizon of 2 years are considered statistically significant and are shown in boldface.

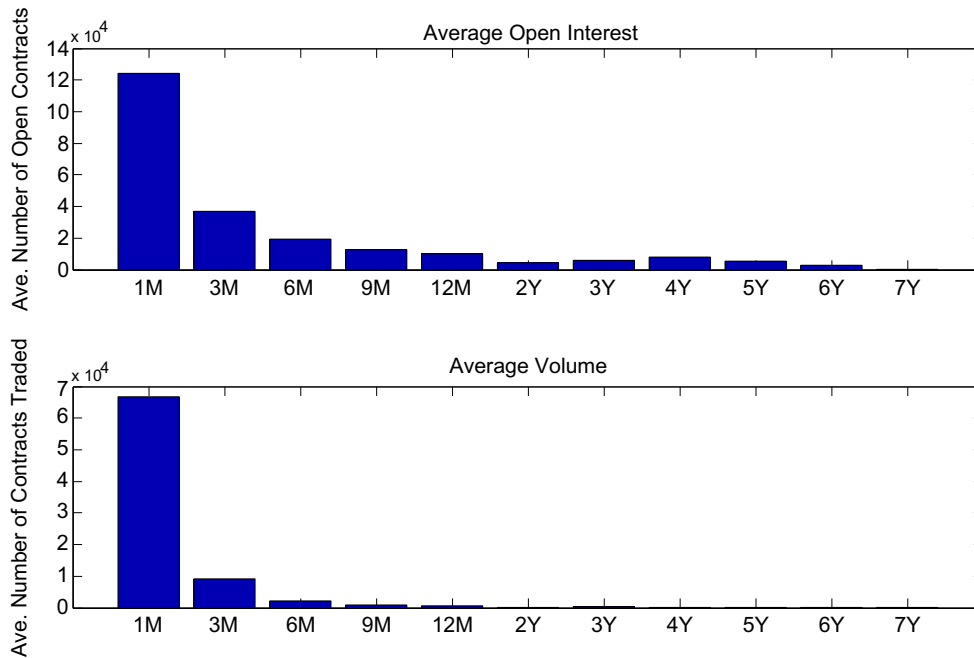


Figure 8.4 Measures of liquidity in the oil futures market (by Maturity). *Notes:* Computations by the authors based on CRB data.

month and have not yet been offset by another transaction or by physical delivery of the oil. It measures the total number of contracts outstanding for delivery in a specific month. Volume is the total number of contracts traded during a specific period of time. Contracts are denoted in units of 1,000 barrels of crude oil. Although both average open interest and average trading volume drop off quickly with increasing maturity, it is not the case that average liquidity at the daily frequency is discontinuously lower at horizons beyond 1 year than at the 12-month horizon. Rather the decline in average liquidity is smooth.

One concern with the results in Table 8.5 is that the most traded oil futures contracts are the June and December contracts. This suggests focusing on the most liquid daily contracts rather than averaging results across all daily contracts, as we did in Table 8.5. Below we report sensitivity analysis for this subset of daily oil futures contracts. Because long-term futures contracts only became available in recent years and because their use greatly reduces the effective sample size, we focus on June and December contracts with maturities of 1, 2, and 3 years. Based on the evaluation period of 1998–2010, we find that 1-year contracts have an MSPE ratio of 0.91 compared with the no-change forecast, 2-year contracts an MSPE ratio of 1.01 and 3-year contracts an MSPE ratio of 1.27. These results are qualitatively similar to those in Table 8.5 for the same maturities, suggesting that there are no gains in forecast accuracy from restricting the sample.

Finally, we note that these results may not have been apparent in the years when longer-term oil futures contracts were first introduced. As recently as in the late 1990s, a forecaster employing the same methods that we used in this section, would have found that the monthly price of oil futures contracts with 1-year maturity is much more accurate than the no-change forecast, although the MSPE reductions declined steadily throughout the 1990s, as more information became available, and the ratio has oscillated about 1 since then. Even 2- and 3-year daily contracts, which were introduced much more recently, initially seemed to forecast more accurately than the no-change forecast, but these MSPE reductions have been reversed more recently. Given that the forecast errors become more highly serially correlated, the higher the data frequency, very long samples are required for reliable estimates of relative MSPEs. Clearly, an evaluation period of 15 years, for example, is insufficient to learn about the forecasting ability of oil futures prices, as illustrated by the repeated sharp reversals in forecast rankings over time. Even our results must be considered tentative and could be reversed as more data become available.

One possible explanation for the unexpectedly low out-of-sample accuracy of oil futures-based forecasts may be the presence of transaction costs impeding arbitrage. An alternative forecasting strategy in which one uses the futures price only if the futures spread exceeds 5% in absolute terms and uses the spot price otherwise, yields MSPE reductions between 0% and 6% at short horizons. Notably the MSPE reductions at horizons of 3 and 6 months are statistically significant in both the daily and the monthly data. At horizons beyond 1 year, this alternative method is much less accurate than the no-change forecast, however. We conclude that the evidence from daily futures prices is broadly consistent with the earlier evidence from monthly data.

7. SURVEY FORECASTS OF THE NOMINAL PRICE OF OIL

A comparison of alternative oil price forecasts would not be complete without evidence on survey expectations of the nominal price of oil. Given the significance of crude oil to the international economy, it is surprising that there are few organizations that produce monthly forecasts of spot prices. In the oil industry, where the spot price of oil is critical to investment decisions, producers tend to make annual forecasts of spot prices for horizons as long as 15–20 years, but these are not publicly available. The U.S. Department of Energy's EIA has published quarterly forecasts of the nominal price of oil since 1983. The *Economist Intelligence Unit* has produced annual forecasts since the 1990s for horizons of up to 5 years. None of these sources provides monthly forecasts.

One common source of monthly forecasts of the price of crude oil is *Consensus Economics Inc.*, a U.K.-based company that compiles private sector forecasts in a variety of countries. Initially, the sample consisted of more than 100 private firms; it now contains about 70 firms. Of interest to us are the survey expectations for the 3- and 12-month-ahead spot price of West Texas Intermediate crude oil, which corresponds to the type

Table 8.6 Accuracy of Survey and Expert Forecasts Relative to No-Change Forecast

	<i>h</i> =3		<i>h</i> =12		<i>h</i> =60	
	MSPE Ratio	Success Ratio	MSPE Ratio	Success Ratio	MSPE Ratio	Success Ratio
$\hat{S}_{t+h t} = S_{t,h}^{CE}$	1.519	0.447	0.944	0.539	–	–
$\hat{S}_{t+h t} = S_{t,h}^{EIA}$	0.918	0.417	0.973	0.562	–	–
$\hat{S}_{t+h t} = S_t(1 + \pi_{t,h}^{MSC})$	–	–	1.047	0.566 ^a	–	–
$\hat{S}_{t+h t} = S_t(1 + \pi_{t,h}^{SPF})$	–	–	1.016	0.579 ^a	0.855	0.811 ^a

Notes: Boldface indicates statistical significance at the 10% level.

^a No significance test possible due to lack of variation in success ratio. SPF denotes the Survey of Professional Forecasters, EIA the Energy Information Administration and CE denotes Consensus Economics Inc. $\pi_{t,h}$ stands for the expected inflation rate between t and $t+h$.

and grade delivered under the NYMEX futures contract. The survey provides the arithmetic average, the minimum, the maximum, and the standard deviation for each survey month beginning in October 1989 and ending in December 2009. We use the arithmetic mean at the relevant horizon:

$$\hat{S}_{t+h|t} = S_{t,h}^{CE} \quad h = 3, 12. \quad (15)$$

Table 8.6 reveals that this survey forecast does not significantly reduce the MSPE relative to the no-change forecast and may increase the MSPE substantially. The survey forecast is particularly poor at the 3-month horizon. At the 12-month horizon the survey forecast has a lower MSPE than the no-change forecast, but the gain in accuracy is not statistically significant. There also is a statistically significant but negligible gain in directional accuracy.¹⁹

Further analysis shows that until 2008.12 the consensus survey forecast had a much higher MSPE than the no-change forecast at both the 3-month and 12-month horizons. This pattern changes only toward the end of the sample. There is evidence that the accuracy of the consensus survey forecasts improves at the 12-month horizon, especially in 2009 as the oil market recovers from its collapse in the second half of 2008. It appears that professional forecasters correctly predicted a long-term price recovery in this instance, although they were not successful at predicting the timing of the 2009 recovery. Notwithstanding these caveats, there is no compelling evidence overall that survey forecasts outperform the no-change forecast. Figure 8.5 shows that these professional

¹⁹ The CE forecast is typically released in the second week of each month. We evaluate this forecast against the realization of the nominal price of oil h months after the end of the month in which the forecast is released. This corresponds to the way that *Consensus Economics Inc.* interprets its forecasts. For example, the January 10 release of the 3-month forecast is intended to be a forecast for the price at the end of April.

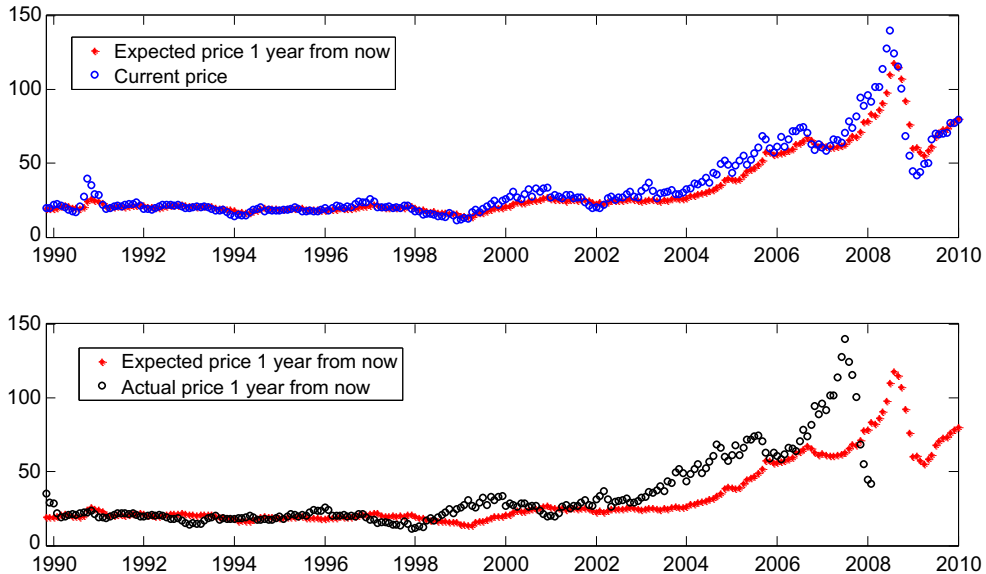


Figure 8.5 Consensus economics expectations of nominal price of oil (Dollars/Barrel): 1989.10–2009.12. *Notes:* Computations by the authors based on data from Consensus Economics Inc.

oil price forecasts most of the time are below the current price of oil. The upper panel of Figure 8.5 shows that professional forecasters tend to smooth the predicted path relative to the current price. This smoothing is especially apparent during large oil price fluctuations such as those in 1990/91, in 1999/2000, and in 2003–2009. This tendency contributes to the large and persistently negative forecast errors shown in the lower panel of Figure 8.5. It also helps explain why the consensus forecast typically fails to improve on the no-change forecast in Table 8.6.

One possible explanation of the less than satisfactory accuracy of these survey forecasts is that professional macroeconomic forecasters may not be experts on the oil market. Figure 8.6 focuses on an alternative time series of 1-quarter and 4-quarters-ahead forecasts of the U.S. nominal refiners' acquisition cost for imported crude oil:

$$\hat{S}_{t+h|t} = S_{t,h}^{EIA} \quad h = 3, 12. \quad (16)$$

These forecasts were collected from the U.S. Energy Information Administration's *Short-term Economic Outlook*, which is published by the U.S. Department of Energy. Given the difference in frequency and oil price definition the results are not strictly speaking comparable with our earlier analysis of the monthly WTI price. Nevertheless, these data are illuminating. Figure 8.6 illustrates that even these expert forecasts generally underpredicted the price of crude oil between 2004 and mid-2008, especially at longer horizons, while overpredicting it following the collapse of the price of oil in mid-2008 and underpredicting it again more recently. A natural question is how the EIA forecasts

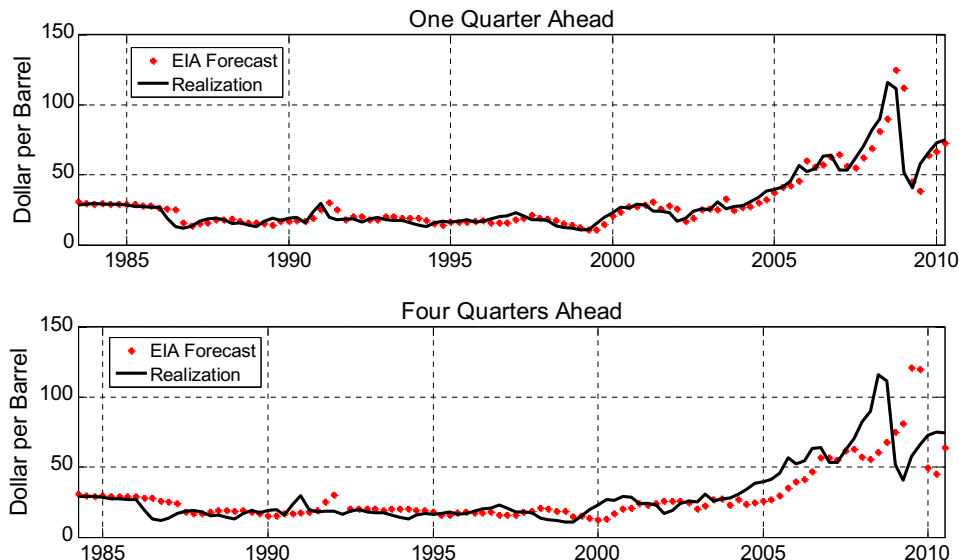


Figure 8.6 EIA forecasts of the U.S. refiners' acquisition cost for imported crude oil 1983.Q1-2009.Q4. *Notes:* The quarterly price forecasts were collected manually from the EIA's *Short-Term Economic Outlook* and compared with the ex-post realizations of the average quarterly nominal refiners' acquisition cost for imported crude oil. The plot shows the price realizations together with the EIA forecasts made for the same point in time one and four quarters earlier.

compare to the no-change forecast on the basis of the EIA's preliminary data releases for the current refiners' acquisition cost for imported crude oil. The latter data are provided by the same source. The *DM* test for equal predictive accuracy in Table 8.6 suggests that the MSPE ratio of 0.92 for the one-quarter-ahead forecast is statistically significant at the 10% level, but the MSPE ratio of 0.97 for the four-quarters-ahead forecast is not. We conclude that even the EIA has had at best modest success in forecasting the nominal price of oil in the short run and even less success at the 1-year horizon.

Professional surveys tend to focus on short forecast horizons. Although there are no long monthly or quarterly time series of long-horizon professional forecasts of the nominal price of oil, survey data on long-run inflation expectations may be used to construct surprisingly accurate real-time forecasts of the nominal price of oil at horizons of several years. The proposal is to forecast the nominal price of crude oil by extrapolating the current price at the rate of expected inflation. This amounts to treating changes in the real price of oil as unpredictable.²⁰ Measures of inflation expectations may be obtained from surveys of professional forecasts, even for very long horizons. Here we use the expected inflation rate at the 5-year horizon from the *Survey of Professional Forecasters*.

$$\hat{S}_{t+h|t} = S_t(1 + \pi_{t,h}^{SPF}) \quad h = 12, 60. \quad (17)$$

²⁰ This proposal builds on insights in Anderson et al. (2011) about how households form expectations about gasoline prices.

Table 8.6 shows that this forecast is quite accurate with an MSPE ratio of only 0.855. The p -value for the null hypothesis of equal predictive accuracy is 0.000. The success ratio of 0.811 is also extraordinarily high.²¹ The reason for these strong improvements in forecast accuracy is that at long horizons the inflation component of the nominal price of oil becomes very large. In other words, it is a fairly safe bet that the nominal price of oil must increase in nominal terms over a 5-year horizon. This simple forecasting rule is also much more accurate than the forecast implied by the 5-year oil futures price.

In contrast, at a horizon of 1 year, the effect of inflation on the nominal price is fairly small. As a result, forecasts based on inflation expectations show only modest gains in directional accuracy and have lower MSPEs than the no-change forecast. This is true whether we use the *Survey of Professional Forecasters* or the *Michigan Survey of Consumers* (see Table 8.6). The fact that these results are weaker than those obtained using inflation measures in Granger causality tests may simply reflect that there was not much variation in inflation expectations in our evaluation period, but considerable variation historically.

8. WHAT HAVE WE LEARNED ABOUT FORECASTING THE NOMINAL PRICE OF OIL?

Section 4 provided strong evidence that the nominal price of oil is predictable in population. An obvious concern is how useful this insight is when it comes to forecasting the nominal price of oil out of sample. There are many reasons why full-sample evidence of predictability need not translate into more accurate out-of-sample forecasts (see, e.g., Inoue and Kilian, 2004a). One reason is the inevitable bias-variance trade-off in out-of-sample forecasting. Another reason is that evidence of predictability on average over the full sample may obscure the disappearance of this predictability in the latter part of the sample (see Clark and McCracken, 2005b). For example, Barsky and Kilian (2002) observed that the 1970s were characterized by major shifts in monetary aggregates, whereas monetary aggregates remained much more stable following the Volcker disinflation. Hence, it would not be surprising to see the predictive content of monetary aggregates diminish after the mid-1980s.

In the presence of structural change, evidence of Granger causality in historical data would be of little relevance to applied forecasters interested in exploiting this predictability. This makes it important to assess the out-of-sample accuracy of models of the same set of predictors we considered in Section 4. To further complicate matters, some of the predictors studied in Section 4 cannot be evaluated out of sample without addressing the real-time data limitations faced by applied forecasters. A good example is U.S. M1. Among those predictors that lend themselves to real-time forecasts of the nominal price of oil, we found that inflation-based models are far less accurate out of sample than in the

²¹ The Pesaran–Timmermann test for directional accuracy cannot be applied because there is no variability in the predicted sign, making it impossible to judge the statistical significance of the success ratio.

full sample. It is all the more remarkable therefore that the forecast accuracy gains for the models based on industrial commodity prices (and to a lesser extent for models based on selected bilateral exchange rates) are large and as statistically significant out of sample as they were in the Granger causality analysis of Section 4. In fact, we found significant predictability out-of-sample for the model based on the Australian exchange rate even in the absence of Granger causality in the full sample. One possible explanation of this finding is that predictability increased in the second part of the sample rather than decreased.

One interpretation of the success of these forecasting models is that persistent increases in indices of non-oil industrial commodity prices reflect global demand pressures that affect all industrial commodities (see Barsky and Kilian, 2002; Kilian, 2009a), which in turn are reflected in the dollar exchange rates of broad-based commodity exporters such as Australia and Canada. This interpretation is consistent with additional evidence we will present in Section 9. It suggests some caution in using these forecasting models during times when the price of oil is primarily influenced by other demand or supply shocks. A case in point would be a major oil supply disruption causing an oil price spike. To the extent that the price of oil sometimes is driven by shocks not associated with the global business cycle, one would expect the accuracy gains from using model (14) to be smaller. On the other hand, events such as major oil supply disruptions or speculative frenzies in oil markets are rare, which explains why the forecasting model in question performs well over the full sample.

9. SHORT-HORIZON FORECASTS OF THE REAL PRICE OF OIL

Our analysis in Section 4 suggests that we stand a better chance of forecasting the real price of oil out-of-sample using monthly data, given the availability of more appropriate predictors at the monthly frequency. A natural benchmark for all forecasting models of the real price of oil is again the no-change forecast. At short horizons, inflation is expected to be at best moderate and ex ante there is every reason to expect the high forecast accuracy of the random walk model without drift relative to many less parsimonious regression models to carry over to the real price of oil.²² On the other hand, in forecasting the real price of oil we may rely on additional economic structure and on additional predictors that could potentially improve forecast accuracy. This section explores a number of such models. We consider two alternative measures of the real price of oil: The U.S. refiners' acquisition cost for imported crude oil, which may be thought of as a proxy for the price of oil in global oil markets, and the WTI price; in both cases the deflator is the U.S. CPI.

²² Such a finding would not necessarily imply that the real price of oil actually follows a random walk. It could merely reflect the fact that the bias-variance tradeoff favors parsimonious forecasting models in small samples. The local-to-zero asymptotic approximation of predictive models suggests that using the no-change forecast may lower the asymptotic MSPE even relative to the correctly specified non-random walk model, provided the local drift parameter governing the predictive relationship is close enough to zero (see, e.g., Inoue and Kilian, 2004b; Clark and McCracken, 2010).

9.1. Real U.S. Refiners' Acquisition Cost for Imported Crude Oil

9.1.1. Unrestricted AR, ARMA, and VAR Models

A useful starting point is a forecast accuracy comparison of selected monthly autoregressive (AR) and autoregressive-moving average (ARMA) models for the real price of oil in log levels and in log differences. Both classes of models are evaluated in terms of their ability to predict the log level of the real price of oil in recursive settings. Estimation starts in 1973.2, and the evaluation period is 1991.12–2009.8 to facilitate direct comparisons with VAR models of the global market for crude oil in this and the next section.²³ All MSPE results are expressed as fractions of the MSPE of the no-change forecast. Some models are based on fixed lag orders of 12 or 24, whereas others rely on the Schwarz Information Criterion (SIC) or the Akaike Information Criterion (AIC) for lag order selection (see Inoue and Kilian, 2006; Marcellino et al., 2006). We search over $p \in \{0, \dots, 12\}$. The forecast accuracy results are robust to allowing for a larger upper bound.

There are no theoretical results in the forecasting literature on how to assess the null of equal predictive accuracy when comparing iterated AR or ARMA forecasts to the no-change forecast. In particular, the standard tests discussed in Clark and McCracken (2001, 2005a) or Clark and West (2007) are only designed for direct forecasts. Below we assess the significance of the MSPE reductions based on bootstrap p -values for the MSPE ratio constructed under the null of a random walk model without drift.²⁴ The upper panel of Table 8.7 suggests that AR and ARMA models in log levels have lower recursive MSPE than the no-change forecast at short horizons. The accuracy gains may approach 17% in some cases and are highly statistically significant. Beyond the 6-month horizon, all gains in forecast accuracy evaporate. There also are statistically significant gains in directional accuracy at horizons 1 and 3, and in some cases at horizon 6. There is little to choose between the AR(12), ARMA(1,1), AR(SIC) and AR(AIC) specifications overall. The AR(24) model has slightly better directional accuracy at longer horizons, but at the cost of a higher MSPE ratio.

The next panel of Table 8.7 shows the corresponding forecasting models in log differences. Note that after imposing the unit root, the autoregressive lag order is reduced by one. For example, an ARMA(1,1) model in levels corresponds to an MA(1) model in differences. We find that models in log differences generally are about as accurate as models in log levels. There is robust evidence of statistically significant MSPE reductions at horizons 1 and 3 and there are statistically significant gains in directional accuracy at horizons of up to 6 months in some cases. There is little to choose between the five forecasting models in log differences. We conclude (1) that forecasting the real price of oil based on models in log levels is by no means inferior to forecasting based on models in log differences; (2) that simple AR or ARMA models with fixed lag orders perform quite

²³ The refiners' acquisition cost was extrapolated back to 1973.2 as in Barsky and Kilian (2002).

²⁴ Because there is no reason to expect the limiting distribution of the DM test statistic to be pivotal in this context, we bootstrap the average loss differential instead.

Table 8.7 Recursive Forecast Error Diagnostics for the Real Price of Oil from Unrestricted AR and ARMA Models

		Evaluation Period: 1991.12-2009.8									
		<i>h</i> =1		<i>h</i> =3		<i>h</i> =6		<i>h</i> =9		<i>h</i> =12	
		MSPE	SR	MSPE	SR	MSPE	SR	MSPE	SR	MSPE	SR
RAC	AR(12)	0.849	0.599	0.921	0.552	0.969	0.522	1.034	0.441	1.022	0.517
Imports	AR(24)	0.898	0.576	0.978	0.557	1.008	0.565	1.056	0.446	1.058	0.453
	AR(SIC)	0.826	0.613	0.936	0.557	1.015	0.488	1.039	0.515	1.007	0.532
	AR(AIC)	0.842	0.613	0.940	0.562	0.983	0.483	1.013	0.500	0.989	0.527
	ARMA(1,1)	0.837	0.580	0.932	0.514	0.982	0.493	1.006	0.510	0.992	0.527
RAC	ARI(11)	0.856	0.604	0.939	0.571	1.003	0.517	1.095	0.471	1.091	0.512
Imports	ARI(23)	0.898	0.561	0.978	0.538	1.009	0.546	1.068	0.500	1.068	0.508
	ARI(SIC)	0.833	0.594	0.951	0.605	1.041	0.546	1.053	0.505	1.016	0.527
	ARI(AIC)	0.849	0.604	0.958	0.605	1.008	0.556	1.042	0.500	1.015	0.527
	ARIMA(0,1)	0.841	0.599	0.945	0.581	1.009	0.546	1.032	0.515	1.017	0.512
WTI	AR(12)	0.972	0.500	0.974	0.533	1.011	0.459	1.037	0.441	1.034	0.478
	AR(24)	1.035	0.486	0.994	0.500	0.995	0.502	1.008	0.461	1.019	0.473
	AR(SIC)	0.947	0.505	0.979	0.491	1.022	0.464	1.052	0.471	1.058	0.508
	AR(AIC)	0.949	0.505	0.980	0.491	1.022	0.464	1.046	0.471	1.047	0.508
	ARMA(1,1)	0.956	0.500	0.982	0.491	1.010	0.473	1.036	0.476	1.040	0.508
WTI	ARI(11)	0.978	0.505	0.985	0.529	1.032	0.517	1.081	0.456	1.083	0.433
	ARI(23)	1.034	0.524	0.988	0.538	0.988	0.594	1.016	0.534	1.026	0.522
	ARI(SIC)	0.944	0.528	0.971	0.571	1.013	0.546	1.023	0.505	1.020	0.517
	ARI(AIC)	0.947	0.524	0.976	0.552	1.018	0.517	1.031	0.466	1.026	0.488
	ARIMA(0,1)	0.952	0.524	0.975	0.600	1.009	0.527	1.021	0.500	1.019	0.517

Notes: ARI and ARIMA, respectively, denote AR and ARMA models in log differences. The SIC and AIC are implemented with an upper bound of 12 lags. MSPE is expressed as a fraction of the MSPE of the no-change forecast. SR stands for success ratio. The *p*-values for the sign test are computed following Pesaran and Timmermann (2009); those for the test of equal MSPEs are computed by bootstrapping the VAR model under the null, adapting the bootstrap algorithm in Kilian (1999). Boldface indicates statistical significance at the 10% level.

well; and (3) that the no-change forecast of the real price of oil can be improved upon at horizons of 1 month and 3 months, but generally not at horizons beyond half a year.

All models in Table 8.7 have in common that the information set is restricted to past values of the real price of oil. The question we turn to next is whether suitably chosen macroeconomic predictors can be used to improve further on the no-change forecast. Recently, a number of structural vector autoregressive models of the global market for crude oil have been proposed (see, e.g., Kilian, 2009a). These models produce empirically plausible estimates of the impact of demand and supply shocks on the real price of oil. A natural conjecture is that such models may also have value for forecasting the real

price of oil. Here we focus on the reduced-form representation of the VAR(p) model in Kilian and Murphy (2013). The variables in this model include the percent change in global crude oil production ($\Delta prod_t$), the global real activity measure we already discussed in Section 4 (rea_t), the log of the real price of oil (r_t), and a proxy for the change in global above-ground crude oil inventories (ΔI). The model is $B(L)y_t = e_t$, where $B(L) = 1 - B_1L - \dots - B_pL^p$, e_t is white noise, and $y_t = (\Delta prod_t, rea_t, r_t, \Delta I_t)'$. The sample period is 1973.2-2009.8. For further discussion of the data see Kilian and Murphy (2013). The model may be consistently estimated without taking a stand on whether the real price of oil is I(0) or I(1) (see Sims et al., 1990). We focus on recursive rather than rolling regression forecasts throughout this section. This approach makes sense in the absence of structural change, given the greater efficiency of recursive regressions and the small sample size.²⁵

A natural starting point for the forecast accuracy comparison is the unrestricted VAR model. An obvious concern with forecasting from unrestricted vector autoregressions is that these highly parameterized models are subject to considerable estimation uncertainty which tends to inflate the out-of-sample MSPE. For that reason unrestricted VAR models are rarely used in applied forecasting. They nevertheless provide a useful point of departure. The upper panel of Table 8.8 shows results for unrestricted VAR models with 12 lags. Column (1) corresponds to the four-variable model used in Kilian and Murphy (2013). Table 8.8 shows that this unrestricted VAR forecast has lower recursive MSPE than the no-change forecast at all horizons but one and non-trivial directional accuracy.²⁶ Despite the lack of parsimony, the reductions in the MSPE are somewhat larger than for the AR and ARMA models in Table 8.7. Bootstrap p -values for the MSPE ratio constructed under the null of a random walk model without drift indicate statistically significant reductions in the MSPE at horizons 1, 3, and 6. At longer horizons it becomes harder to beat the no-change forecast benchmark and there are no statistically significant reductions in the MSPE. There also is evidence of statistically significant gains in directional accuracy at horizons 1 and 3.

The forecasting success of the VAR approach clearly depends on the choice of variables and of the lag length. The remaining columns of the upper panel of Table 8.8 show analogous results for five other unrestricted VAR(12) models obtained by dropping one or more of the variables included in model (1). None of these models performs as well as the original four-variable model with two exceptions. The bivariate model (4) which includes only the change in oil inventories and the real price of oil has slightly lower MSPE than the four-variable VAR(12) model and similar directional accuracy, as does

²⁵ Rolling regression forecasts would not protect us from structural change in any case. It has been shown that the presence of structural breaks at unknown points in the future invalidates the use of forecasting model rankings obtained in forecast accuracy comparisons whether one uses rolling or recursive regression forecasts (see Inoue and Kilian, 2006).

²⁶ It also outperforms the random walk model with drift in both of these dimensions, whether the drift is estimated recursively or as the average growth rate over the most recent h months. These results are not shown to conserve space.

Table 8.8 Recursive Forecast Error Diagnostics for the Real Price of Oil from Unrestricted VAR Models

		Evaluation Period: 1991.12-2009.8													
Model:		(1)		(2)		(3)		(4)		(5)		(6)			
Price	p	h	MSPE	SR	MSPE	SR	MSPE	SR	MSPE	SR	MSPE	SR	MSPE	SR	
RAC Imports	12	1	0.814	0.561	0.876	0.594	0.863	0.613	0.801	0.613	0.863	0.580	0.798	0.585	
	3	1	0.834	0.567	0.960	0.562	0.947	0.576	0.833	0.614	0.944	0.524	0.833	0.586	
		6	0.940	0.546	1.011	0.507	0.991	0.536	0.920	0.551	0.996	0.527	0.922	0.511	
		9	1.047	0.564	1.085	0.534	1.060	0.539	0.999	0.544	1.063	0.471	1.000	0.569	
		12	0.985	0.632	1.055	0.562	1.036	0.567	0.948	0.617	1.045	0.503	0.931	0.647	
RAC Imports	24	1	0.961	0.561	0.954	0.552	0.912	0.580	0.892	0.571	0.912	0.561	0.895	0.561	
	3	1	1.081	0.614	1.151	0.591	1.048	0.619	0.924	0.591	1.005	0.548	0.978	0.605	
		6	1.298	0.604	1.271	0.585	1.078	0.594	1.052	0.546	1.073	0.522	1.129	0.585	
		9	1.476	0.583	1.441	0.593	1.153	0.632	1.150	0.431	1.158	0.422	1.255	0.593	
		12	1.415	0.647	1.407	0.612	1.137	0.642	1.137	0.468	1.169	0.458	1.208	0.617	
WTI	12	1	0.896	0.519	0.981	0.467	0.976	0.481	0.893	0.547	0.983	0.505	0.882	0.547	
	3	1	0.843	0.538	0.979	0.524	0.968	0.548	0.877	0.552	0.994	0.529	0.841	0.548	
		6	0.988	0.517	1.035	0.541	1.011	0.551	0.984	0.541	1.037	0.464	0.973	0.541	
		9	1.053	0.534	1.080	0.485	1.049	0.510	1.021	0.564	1.067	0.441	1.014	0.539	
		12	1.007	0.562	1.062	0.498	1.041	0.498	0.988	0.602	1.059	0.438	0.968	0.592	
WTI	24	1	1.109	0.509	1.118	0.491	1.053	0.538	1.011	0.552	1.063	0.500	1.013	0.509	
	3	1	1.112	0.581	1.185	0.552	1.017	0.562	0.970	0.562	1.049	0.481	0.962	0.619	
		6	1.369	0.570	1.312	0.541	1.030	0.594	1.107	0.483	1.075	0.488	1.127	0.589	
		9	1.455	0.564	1.340	0.520	1.060	0.583	1.160	0.446	1.106	0.490	1.153	0.583	
		12	1.369	0.562	1.378	0.503	1.054	0.592	1.167	0.478	1.119	0.478	1.086	0.602	

Notes: MSPE is expressed as a fraction of the MSPE of the no-change forecast. SR stands for success ratio. The p -values for the sign test are computed following Pesaran and Timmermann (2009); those for the test of equal MSPEs are computed by bootstrapping the VAR model under the null, adapting the bootstrap algorithm in Kilian (1999). Boldface indicates statistical significance at the 10% level. Model (1) includes all four variables used in the VAR model of Kilian and Murphy (2010); model (2) excludes oil inventories; model (3) excludes both oil inventories and oil production; model (4) excludes real activity and oil production; model (5) excludes real activity and oil inventories; and model (6) excludes oil production.

the trivariate model (6) specification that drops oil production from the baseline model. The differences are negligible, however.

The next panel of Table 8.8 suggests that including 24 lags in the unrestricted model tends to reduce the MSPE reductions. All VAR(24) models but model (2) still significantly improve on the MSPE of the no-change forecast at horizons 1 and 3, but their MSPE ratio tends to exceed unity at longer horizons. Likewise, all six VAR(24) models yield statistically significant gains in directional accuracy at short horizons. Only the four VAR(24) models that include the global real activity variable in the model, however, retain their superior directional accuracy at all horizons. Unlike in the corresponding

VAR(12) models, the gains in directional accuracy are statistically significant at all horizons.

We conclude that there is important predictive information in the change in oil inventories and in global real activity in particular, whereas the inclusion of oil production growth appears less important for forecasting. Moreover, based on the MSPE metric, suitably chosen VAR models systematically outperform the no-change forecast at short horizons. At horizons of 9 and 12 months, the no-change forecast remains unbeaten, except based on the sign metric. Beyond 1 year, none of these VAR forecasting models systematically outperforms the no-change forecast.

It is important to keep in mind, however, that Table 8.8 may overstate the statistical significance of the short-horizon MSPE reductions. One indication of this problem is that Table 8.8 sometimes indicates statistically significant rejections of the no-change forecast benchmark even when the MSPE ratio exceeds 1, indicating that the VAR has a strictly higher recursive MSPE. The reason for this counterintuitive result is that, as discussed earlier, standard tests of equal predictive accuracy do not test the null of equal out-of-sample MSPEs, but actually test the null of no predictability in population – much like the Granger causality tests we applied earlier – as pointed out by Inoue and Kilian (2004a). This point is readily apparent from the underlying proofs of asymptotic validity as well as the way in which critical values are simulated.

The distinction between population predictability and out-of-sample predictability does not matter asymptotically under fixed parameter asymptotics, but fixed parameter asymptotics typically provide a poor approximation to the finite-sample accuracy of forecasting models. Under more appropriate local asymptotics (designed to mimic the weak predictive power of many regressors) it can be shown that the null of no predictability in population is distinct from the null of equal out-of-sample MSPEs. It is always easier to reject the former than the latter. In other words, conventional tests of equal predictive accuracy test the wrong null hypothesis and may spuriously reject the no-change forecast in favor of the alternative. This is the deeper reason for the very low p -value obtained, for example, for model (1) with 24 lags at horizon 3. The intuition for this rejection is that under the null that the real price of oil is unpredictable one would expect much higher MSPE ratios than 1.047, so the fact that the MSPE of the VAR model is so close to 1 actually is evidence in favor of the VAR model being the population model.

Which model is the population model, of course, is irrelevant for the question of which model generates more accurate forecasts in finite samples, so we have to interpret this rejection with some caution. This type of insight recently has prompted the development of alternative tests of equal predictive accuracy based on local-to-zero asymptotic approximations to the predictive regression. Clark and McCracken (2010) for the first time proposed a correctly specified test of the null of equal out-of-sample MSPEs. Their analysis is limited to direct forecasts from much simpler forecasting models, however, and

cannot be applied in Table 8.8.²⁷ This caveat suggests that we discount only marginally statistically significant rejections of the no predictability null hypothesis in Table 8.8 and focus on the highly statistically significant test results. The tests for directional accuracy are not affected by this concern.

9.1.2. *Real-Time Forecasts*

The results so far are encouraging in that they suggest that VAR models (even more so than AR or ARMA models) may produce useful short-horizon forecasts of the real price of oil. An important caveat regarding the results in Tables 8.7 and 8.8 is that the forecast accuracy comparison is not conducted in real time. There are two rather distinct concerns. One is that not all useful predictors may be available to the forecaster in real time. The other concern is that many predictors and indeed some measures of the price of oil are subject to data revisions. This caveat applies even to the no-change forecast. For example, the refiners' acquisition cost data become available only with a delay of about 3 months and the CPI data used to deflate the refiners' acquisition cost become available only with a 1-month delay.

Additional caveats apply to the VAR evidence. Although the dry cargo shipping rate data underlying the real activity index are available in real time and not subject to revisions, the construction of the real activity index involves real-time CPI data as well real-time estimates of the trend in real shipping rates. Moreover, the data on global crude oil production only become available with a delay of 4 months and the data used to approximate global crude oil inventories with a delay of 5 months. This is less of a concern for the oil production data which tend to evolve rather smoothly than for the more volatile data on changes in crude oil inventories for which there is no good real-time proxy. How imposing these real-time data constraints alters the relative accuracy of no-change benchmark model compared with VAR models is not clear a priori because both the benchmark model and the alternative model are affected.

The first study to investigate this question is [Baumeister and Kilian \(2012a\)](#) who recently developed a real-time data set for the variables in question. They find (based on a data set extending until 2010.6) that VAR forecasting models of the type considered in this section can also generate substantial improvements in real-time forecast accuracy. The MSPE reduction for unrestricted VAR models may be as high as 25% at the 1-month horizon and as high as 9% at the 3-month horizon. At longer horizons the MSPE reductions diminish even for the most accurate VAR models. Beyond 1

²⁷ The size problem of conventional tests of equal predictive accuracy gets worse, when the number of extra predictors under the alternative grows large relative to the sample size. This point has also been discussed in a much simpler context by [Anatolyev \(2012\)](#) who shows that modifying conventional test statistics for equal predictive accuracy may remove these size distortions. Also see [Calhoun \(2011a\)](#). Related results can be found in [Calhoun \(2011b\)](#) who shows that standard tests of equal predictive accuracy for nested models such as [Clark and McCracken \(2001\)](#) or [Clark and West \(2007\)](#) will choose the larger model too often when the smaller model is more accurate in out-of-sample forecasts and also proposes alternative asymptotic approximations based on many predictors. None of the remedies is directly applicable in the context of Tables 8.7, 8.8 and 8.9, however.

year, the no-change forecast usually has lower MSPE than the VAR model. Baumeister and Kilian also show that VAR forecasting models based on Kilian and Murphy (2013) exhibit significantly improved directional accuracy. The improved directional accuracy persists even at horizons at which the MSPE gains have vanished. The success ratios range from 0.51 to 0.60, depending on the model specification and horizon. These results are broadly consistent with the evidence based on ex-post revised data in Table 8.8.

9.2. Real WTI Price

The lower panels of Tables 8.7 and 8.8 show the corresponding results based on the real WTI price of oil instead of the real U.S. refiners' acquisition cost for imported crude oil. These results are not intended to validate those in the upper panels of Tables 8.7 and 8.8, given the inherent differences in the definition of the oil price data; rather they are of independent and complementary interest. The estimation and evaluation period is unchanged to allow direct comparisons. The nominal WTI price is available without delay and is not subject to revisions, reducing concerns over the real-time availability of the oil price data.

Table 8.7 provides robust evidence that AR and ARMA models improve on the no-change forecast of the real WTI price of oil at horizons 1 and 3 with the exception of models with 24 lags. The largest MSPE reductions are only 5%, however, and all such accuracy gains vanish at longer horizons. The results for the unrestricted VAR(12) models in Table 8.8 paint a similar picture with MSPE reductions as high as 16% at horizon 3. Unrestricted VAR(24) models, on the other hand, are far less accurate. We conclude that the definition of the real price of oil matters for the degree of forecastability, but the overall patterns are similar.

Broadly similar results are obtained for a real-time data set extending to 2010.6 (see Baumeister and Kilian, 2012a). Unlike for the real refiners' acquisition cost, the differences between real-time forecasts of the real WTI price and forecasts based on ex-post revised data tend to be small. In both cases, there are substantial gains in out-of-sample forecast accuracy up to a horizon of about 1 year, corroborating and strengthening the qualitative results in Table 8.8.

9.3. Restricted VAR Models

Although the results for the unrestricted VAR models in Tables 8.7 and 8.8 are encouraging, there is reason to believe that alternative estimation methods may reduce the MSPE of the VAR forecast even further. One candidate is the use of Bayesian estimation methods. In the VAR model at hand a natural starting point would be to shrink all lagged parameters toward zero under the maintained assumption of stationarity. This leaves open the question of how to determine the weights of the prior relative to the information in the likelihood. Giannone et al. (2010) recently proposed a simple and theoretically founded data-based method for the selection of priors in recursively estimated Bayesian VARs (BVARs). Their recommendation is to select priors using the marginal

Table 8.9 Recursive MSPE Ratios for the Real Price of Oil from Bayesian VAR Models

Evaluation Period: 1991.12-2009.8													
Model:		(1)		(2)		(3)		(4)		(5)		(6)	
<i>p</i>	<i>h</i>	RAC	WTI	RAC	WTI	RAC	WTI	RAC	WTI	RAC	WTI	RAC	WTI
12	1	0.800	0.892	0.825	0.938	0.828	0.945	0.798	0.896	0.827	0.951	0.795	0.883
	3	0.876	0.886	0.929	0.954	0.930	0.957	0.855	0.890	0.921	0.972	0.867	0.870
	6	0.967	0.990	0.988	1.008	0.987	1.006	0.943	0.985	0.971	1.011	0.962	0.984
	9	1.052	1.036	1.053	1.036	1.054	1.037	1.033	1.037	1.031	1.029	1.050	1.036
	12	1.004	1.005	1.024	1.024	1.028	1.028	0.994	1.008	1.015	1.022	1.004	1.003
24	1	0.801	0.894	0.826	0.939	0.828	0.947	0.800	0.902	0.829	0.952	0.795	0.886
	3	0.883	0.875	0.939	0.945	0.944	0.948	0.860	0.877	0.924	0.958	0.876	0.859
	6	0.993	0.990	1.012	1.007	1.015	1.000	0.955	0.980	0.970	0.991	0.991	0.986
	9	1.095	1.038	1.093	1.034	1.096	1.032	1.044	1.028	1.028	1.005	1.097	1.037
	12	1.059	1.002	1.073	1.016	1.078	1.018	1.016	1.010	1.026	1.008	1.058	0.998

Notes: The Bayesian VAR forecast relies on the data-based procedure proposed in Giannone et al. (2010) for selecting the optimal degree of shrinkage in real time. MSPE is expressed as a fraction of the MSPE of the no-change forecast. Boldface indicates MSPE ratios lower than for the corresponding unrestricted VAR forecasting model in Tables 8.12 and 8.14. RAC refers to the U.S. refiners' acquisition cost for imported crude oil and WTI to the price of West Texas Intermediate crude oil. Model (1) includes all four variables used in the VAR model of Kilian and Murphy (2010); model (2) excludes oil inventories; model (3) excludes both oil inventories and oil production; model (4) excludes real activity and oil production; model (5) excludes real activity and oil inventories; and model (6) excludes oil production.

data density (i.e., the likelihood function integrated over the model parameters), which only depends on the hyperparameters that characterize the relative weight of the prior and the information in the data. They provide empirical examples in which the forecasting accuracy of that model in recursive settings is not only superior to unrestricted VAR models, but is comparable to that of single-equation dynamic factor models (see Stock and Watson, 1999).

Table 8.9 compares the forecasting accuracy of this approach with that of the unrestricted VAR models considered in Table 8.8. In all cases, we shrink the model parameters toward a white noise prior mean with the desired degree of shrinkage being determined by the data-based procedure in Giannone et al. (2010). For models with 12 lags, there is no strong evidence that shrinkage estimation reduces the MSPE. Although there are some cases in which imposing Bayesian priors reduces the MSPE slightly, in other cases it increases the MSPE slightly. For models with 24 lags, however, shrinkage estimation often greatly reduces the MSPE ratio and typically produces forecasts about as accurate as forecasts from the corresponding model with 12 lags. As in Table 8.8, there is evidence of MSPE reductions at horizons at short horizons. For example, model (1) with 12 lags yields MSPE reductions of 20% at horizon 1, 12% at horizon 3, and 3% at horizon 6 with no further gains at longer horizons. Model (1) with 24 lags yields gains of 20%, 12% and 1%, respectively. Again, it can be shown that similar gains in accuracy are feasible

even using real-time data at horizons as high as 12 months with corresponding gains in directional accuracy (see [Baumeister and Kilian, 2012a](#)).

10. WHAT HAVE WE LEARNED ABOUT FORECASTING THE REAL PRICE OF OIL?

As we discussed in the context of nominal oil prices, it is important to assess the out-of-sample accuracy of models that help predict the real price of oil based on full-sample regressions. Table 8.2 demonstrated overwhelming evidence of population predictability of the real price of oil based on the [Kilian \(2009a\)](#) measure of global real activity. The evidence in column (3) of Table 8.8 demonstrates that the same model also achieves large reductions in the out-of-sample MSPE. These reductions are statistically significant, according to standard tests of equal predictive accuracy. We conclude that measures of global real activity are useful for reducing the MSPE of in-sample as well as out-of-sample forecasts of the real price of oil.

One obvious concern is that the forecasting success of the models in Tables 8.7, 8.8, and 8.9 may merely reflect our ability to forecast inflation. For example, if the U.S. CPI is predictable and the nominal price of oil is not, one should be able to predict changes in the real price of oil to the extent that inflation is predictable. It is clear that this phenomenon cannot explain our results. Inflation does not move nearly enough to explain the large observed variation in the real price, so even if one were to generate perfect short-run inflation forecasts, one would obtain a poor forecast of the real price of oil. It can be shown that our VAR models do a much better job at forecasting the real price of oil than would forecasting models based on expected inflation. In fact, it is easy to see that a forecast of the real price of oil constructed by subtracting expected inflation from the current price of oil not only would have failed to predict the surge in the real price of oil between 2003 and mid-2008, but would have predicted a decline in the real price of oil instead. Another illustration of this point is provided in [Baumeister and Kilian \(2012a\)](#), who contrast the performance of alternative real-time forecasts of the real price of oil generated in 2010.12. On the contrary, our results suggest that the nominal price of oil is predictable because the real price of oil is predictable. In fact, it is possible to adjust the forecast of the real price of oil generated by the VAR models of Section 9 for expected inflation to produce a forecast of the nominal price of oil that is more accurate than the no-change forecast.

It can also be shown that extrapolating the current nominal price of oil based on recent percent changes in the price of industrial raw materials, adjusted for expected inflation, produces a forecast of the real price of oil that is as accurate in the short run as the forecasts of the real price of oil from VAR models – and more accurate than forecasts based on the oil futures spread adjusted for expected inflation (see [Baumeister and Kilian, 2012a](#)). The fact that these models perform similarly is no accident. The VAR models

that are successful at beating the no-change forecast in the short run are consistent with economic models in which the real price of oil responds to changes in global real activity and other real variables such as changes in crude oil inventories. Both of these variables may be viewed as leading indicators of the real price of oil. Likewise models based on recent percent changes in the price index for non-oil industrial raw materials will capture the effect of persistent changes in the global business cycle on the real price of oil, because shifts in the demand for all industrial raw materials are also associated with shifts in the demand for crude oil. The advantage of the VAR approach is that it allows for a broader set of determinants of the real price of oil, making it potentially more robust.

Finally, we note that the evidence in favor of VAR forecasts also is consistent with the result that simple autoregressive models outperform the no-change forecast of the real price of oil. The latter models may be viewed as approximations to the marginalized VAR processes. The chief advantage of VAR models over AR models is their superior directional accuracy.

11. STRUCTURAL VAR FORECASTS OF THE REAL PRICE OF OIL

So far we have considered reduced-form VAR forecasting models. VAR models can also be useful for studying how baseline forecasts of the real price of oil must be adjusted under hypothetical forecasting scenarios. This does require the structural shocks in the VAR model to be identified, however. Recent research has shown that historical fluctuations in the real price of oil can be decomposed into the effects of distinct oil demand and oil supply shocks including a forward-looking or speculative element in the real price of oil (see, e.g., Kilian and Murphy, 2013). Changes in the composition of these shocks over time help explain why conventional regressions of macroeconomic aggregates on the price of oil tend to be unstable. They also are potentially important in interpreting oil price forecasts.

In Section 9 we showed that recursive forecasts of the real price of oil based on the type of oil market VAR model originally proposed by Kilian and Murphy (2013) for the purpose of structural analysis are superior to simple no-change forecasts at least at short horizons. The case for the use of VAR models, however, does not rest on their predictive accuracy alone. Policymakers expect oil price forecasts to be interpretable in light of an economic model. They also expect forecasters to be able to generate projections conditional on a variety of hypothetical economic scenarios. Questions of interest include, for example, what effects an unexpected slowing of Asian growth would have on the forecast of the real price of oil; or what the effect would be of an unexpected decline in global oil production associated with peak oil. Answering questions of this type is impossible using reduced-form time series models. It requires a fully structural VAR model.

In this section we illustrate how to generate such projections from the structural moving average representation of the VAR model of Kilian and Murphy (2013) estimated

on data extending to 2009.8. The methodology resembles that discussed in [Baumeister and Kilian \(2012b\)](#).²⁸ The structural model allows the identification of three shocks of economic interest: (1) a shock to the flow of the production of crude oil (“flow supply shock”), (2) a shock to the flow demand for crude oil (“flow demand shock”) that reflects unexpected fluctuations in the global business cycle, and (3) a shock to the demand for oil inventories arising from forward-looking behavior not otherwise captured by the model (“speculative demand shock”). The structural demand and supply shocks in this model are identified by a combination of sign restrictions and bounds on impact price elasticities. This model is set-identified, but the admissible models can be shown to be quite similar, allowing us to focus on one such model with little loss of generality. We focus on the same model that Kilian and Murphy use as the basis for their historical decompositions.

There is a strict correspondence between standard reduced-form VAR forecasts and forecasts from the structural moving average representation. The reduced-form forecast corresponds to the expected real price of oil conditional on all future shocks being zero. Departures from this benchmark can be constructed by feeding pre-specified sequences of future structural shocks into the structural moving-average representation. A forecast scenario is defined as a sequence of future structural shocks. The implied movements in the real price of oil relative to the baseline forecast obtained by setting all future structural shocks to zero correspond to the revision of the reduced-form forecast implied by this scenario.

Here we consider three scenarios of economic interest. Additional scenarios are discussed in [Baumeister and Kilian \(2012b\)](#). The forecast horizon is 24 months for illustrative purposes. The first scenario involves a successful stimulus to U.S. oil production, as had been considered by the Obama administration prior to the 2010 oil spill in the Gulf of Mexico. Here we consider the likely effects of a 20% increase in U.S. crude oil output in 2009.9, after the estimation sample of [Kilian and Murphy \(2013\)](#) ends. This is not to say that such a dramatic and sudden increase would be feasible, but that it would be an optimistic scenario. Such a U.S. oil supply stimulus would translate to a 1.5% increase in world oil production, which is well within the variation of historical data. We simulate the effects of this production increase by calibrating a one-time structural oil supply shock such that the impact response of global oil production growth in 2009.9 is 1.5%. All other future structural shocks are set to zero. Figure 8.7 shows that the resulting reduction in the real price of oil expressed in percent changes relative to the baseline forecast is negligible. Even a much larger U.S. oil supply stimulus would do little to affect the forecast of the real price of oil, suggesting that policies aimed at creating such a stimulus will be ineffective at lowering the real price of oil.

The second scenario involves a recovery of global demand for oil and other industrial commodities. We ask how an unexpected surge in the demand for oil similar to that

²⁸ Related work in a different context also includes [Waggoner and Zha \(1999\)](#).

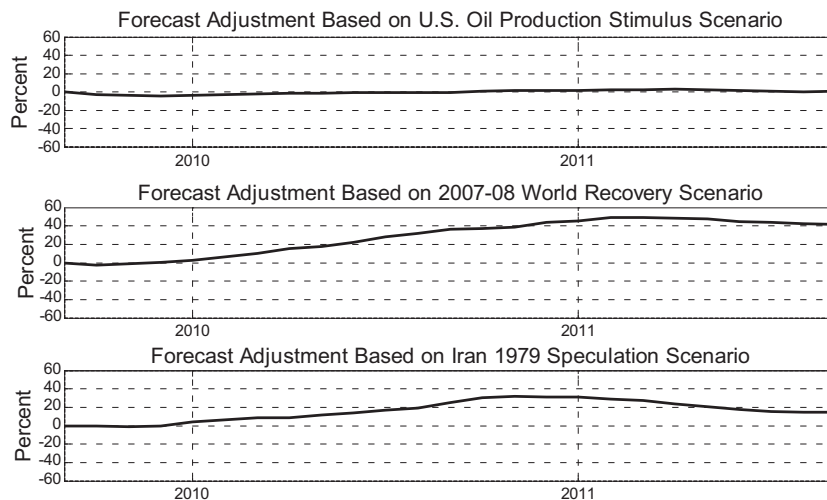


Figure 8.7 Forecasting scenarios for the real price of oil based on the structural VAR model of Kilian and Murphy (2013): Conditional projections expressed relative to baseline forecast. *Notes:* All results are based on the structural oil market model of Kilian and Murphy (2013). The U.S. oil production stimulus involves a 20% increase in U.S. oil production in 2009.9, which translates to a 1.5% increase in world oil production. For this purpose, a one-time structural oil supply shock is calibrated such that the impact response of global oil production is 1.5%. The 2007-08 world recovery scenario involves feeding in as future shocks the sequence of flow demand shocks that occurred in 2007.1–2008.6. The Iran 1979 speculation scenario involves feeding in as future shocks the speculative demand shocks that occurred between 1979.1 and 1980.2 and were a major contributor to the 1979/80 oil price shock episode.

occurring during 2007.1–2008.6, but starting in 2009.9, would affect the real price of oil. This scenario involves feeding into the structural moving average representation future flow demand shocks corresponding to the sequence of flow demand shocks that occurred in 2007.1–2008.6, while setting all other future structural shocks equal to their expected value of zero. Figure 8.7 shows a persistent increase in the real price of oil starting in early 2010 that peaks in early 2011 about 50% above the price of oil in 2009.8. Taking the no-change forecast as the benchmark, for example, this means that the peak occurs at a price of about 100 dollars.

Finally, we consider the possibility of a speculative frenzy such as occurred starting in mid-1979 after the Iranian Revolution (see Kilian and Murphy, 2013). This scenario involves feeding into the model future structural shocks corresponding to the sequence of speculative demand shocks that occurred between 1979.1 and 1980.2 and were a major contributor to the 1979/80 oil price shock episode. Figure 8.7 shows that this event would raise the baseline forecast temporarily by as much as 30%. Most of the effects would have dissipated by mid-2011.

These results, while necessarily tentative, illustrate how structural models of oil markets may be used to assess the sensitivity of reduced-form forecasts to specific economic

events, possibly in conjunction with the formal risk measures discussed in Section 13. Conditional projections, of course, are only as good as the underlying structural models. Our example highlights the importance of refining these models and of improving structural forecasting methods, perhaps in conjunction with Bayesian methods of estimating VAR forecasting models.

12. THE ABILITY OF OIL PRICES TO FORECAST U.S. REAL GDP

One of the main reasons why the price of oil is considered important by many macroeconomists is its perceived predictive power for U.S. real GDP. Assessing that predictive power requires a joint forecasting model for the price of oil and for domestic real activity. In this section we first examine the forecasting accuracy of linear models and then examine a variety of non-linear forecasting models. The baseline results are for the U.S. refiners' acquisition cost for imported crude oil. Toward the end of the section we discuss how these results are affected by other oil price choices. Our discussion draws on results in Kilian and Vigfusson (2013).

12.1. Linear Autoregressive Models

A natural starting point is a linear VAR(p) model for the real price of oil and for U.S. real GDP expressed in quarterly percent changes. The general structure of the model is $x_t = B(L)x_{t-1} + e_t$, where $x_t \equiv [\Delta r_t, \Delta y_t]'$, r_t denotes the log of real price of oil, y_t the log of real GDP, Δ is the difference operator, e_t the regression error, and $B(L) = B_1 + B_2L + B_3L^2 + \dots + B_pL^{p-1}$. The benchmark model for real GDP growth is the AR(p) model obtained with

$$B(L) = \begin{pmatrix} \times & \times \\ 0 & B_{22}(L) \end{pmatrix}.$$

The specification of the components of $B(L)$ marked as \times is irrelevant for this forecasting model. We determined the lag order of this benchmark model based on a forecast accuracy comparison involving all combinations of horizons $h \in \{1, \dots, 8\}$ and lag orders $p \in \{1, \dots, 24\}$. The AR(4) model for real GDP growth proved to have the lowest MSPE or about the same MSPE as the most accurate model at all horizons. The same AR(4) benchmark model has also been used by Hamilton (2003) and others, facilitating comparisons with existing results in the literature.

We compare the benchmark model with two alternative models. One model is the unrestricted VAR(p) model obtained with

$$B(L) = \begin{pmatrix} B_{11}(L) & B_{12}(L) \\ B_{21}(L) & B_{22}(L) \end{pmatrix}.$$

Table 8.10 MSPE Ratios of Linear Autoregressive Models Relative to the AR(4) Benchmark Model: Cumulative U.S. Real GDP Growth Rates

Horizon	Real RAC Price of Imports		Nominal RAC Price of Imports	
	Oil Price Endogenous	Oil Price Exogenous	Oil Price Endogenous	Oil Price Exogenous
1	1.09	1.09	1.10	1.10
2	1.03	1.03	1.04	1.04
3	0.99	0.98	1.00	0.99
4	0.97	0.96	0.98	0.97
5	0.96	0.95	0.97	0.95
6	0.95	0.94	0.95	0.94
7	0.92	0.92	0.92	0.92
8	0.92	0.92	0.92	0.92

Notes: The benchmark model is an AR(4) for U.S. real GDP growth. The first alternative is a VAR(4) model for real GDP growth and the percent change in the price of oil that allows for unrestricted feedback from U.S. real GDP growth to the price of oil. The second alternative is a restricted VAR(4) model that treats the price of oil as exogenous. Boldface indicates gains in accuracy relative to the benchmark model. No tests of statistical significance have been conducted, given that these models are economically indistinguishable.

The other is a restricted VAR model of the form

$$B(L) = \begin{pmatrix} B_{11}(L) & 0 \\ B_{21}(L) & B_{22}(L) \end{pmatrix}.$$

The restriction $B_{12}(L) = 0$ is implied by the hypothesis of exogenous oil prices. Although that restriction is not literally true, in Section 4 we observed that in linear models the predictive content of U.S. real GDP for the real price of oil, while not zero, appears to be weak. Thus, a natural conjecture is that the added parsimony from imposing zero feedback from lagged real GDP to the real price of oil may help reduce the out-of-sample MSPE of multi-step-ahead real GDP forecasts.

The real price of oil is obtained by deflating the refiners' acquisition cost for imported crude oil by the U.S. CPI. All three models are estimated recursively on data starting in 1974.Q1. The initial estimation period ends in 1990.Q1, right before the invasion of Kuwait in August of 1990. The forecast evaluation ends in 2010.Q1. The maximum length of the recursive sample is restricted by the end of the data and the forecast horizon. We evaluate the MSPE of each model for the cumulative growth rates at horizons $h \in \{1, \dots, 8\}$, corresponding to the horizons of interest to policymakers.

The first column of Table 8.10 shows that, at horizons of three quarters and beyond, including the real price of oil in the autoregressive models may reduce the MSPE for real GDP growth by up to 8% relative to the AR(4) model for real GDP growth. The unrestricted VAR(4) model for the real price of oil is about as accurate as the restricted VAR(4) model in the second column. Imposing exogeneity marginally reduces the MSPE at some horizons, but the differences are all negligible. This fact is remarkable given the

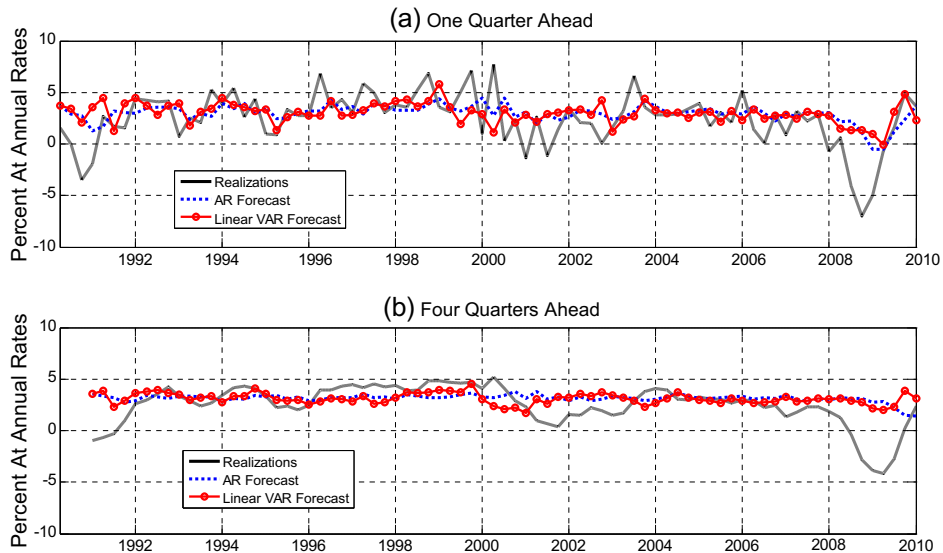


Figure 8.8 VAR forecasts of cumulative real GDP growth based on the real price of oil: U.S. refiners' acquisition cost for imports. *Notes:* The benchmark model is an AR(4) for real GDP growth. The alternative is an unrestricted linear VAR(4) model for real GDP growth and the percent change in the real price of oil. The price of oil is defined as the U.S. refiners' acquisition cost for imports.

greater parsimony of the model with exogenous oil prices. We conclude that there are no significant gains from imposing exogeneity in forecasting from linear models. Next consider a similar analysis for the nominal price of oil. Although the use of the nominal price of oil in predicting real GDP is not supported by standard economic models, it is useful to explore this alternative approach in light of the discussion in Section 3. Table 8.10 shows that the unrestricted VAR(4) model based on the real price of oil is consistently at least as accurate as the same model based on the nominal price of oil. We conclude that in linear models there are no gains in forecast accuracy from replacing the real price of oil by the nominal price. Imposing exogeneity, as shown in the last column, again makes little difference.

MSPE ratios are informative about relative forecasting accuracy, but are not informative about how accurate these models are in practice. Figure 8.8 focuses on the ability of recursively estimated AR(4) and VAR(4) models based on the real price of oil imports to predict the recessions of 1991, 2001, and 2007/8. The upper panel plots the one-quarter-ahead forecasts against the forecast realizations. AR and VAR forecasts are generally quite similar. Neither model is able to forecast the large economic declines in 1990/91, 2001, and 2008/09. The forecast accuracy deteriorates further at the one-year horizon, as shown in the lower panel.

One possible explanation is that this forecast failure simply reflects our inability to forecast more accurately the real price of oil. Put differently, the explanation could be that

the real GDP forecasts would be more accurate if only we had more accurate forecasts of the real price of oil. Conditioning on realized values of the future price of oil, however, does not greatly improve the forecast accuracy of the linear VAR model for cumulative real GDP growth, so this explanation can be ruled out. An alternative explanation could be that the predictive relationship between the price of oil and domestic macroeconomic aggregates is time-varying. One source of time variation is that the share of energy in domestic expenditures has varied considerably over time. This suggests that we replace the percent change in the real price of oil in the linear VAR model by the percent change in the real price of oil weighted by the time-varying share of oil in domestic expenditures, building on the analysis in [Edelstein and Kilian \(2009\)](#). [Hamilton \(2009\)](#) reported some success in employing a similar strategy.²⁹ Another source of time variation may be changes in the composition of the underlying oil demand and oil supply shocks, as discussed in [Kilian \(2009a\)](#). Finally, yet another potential explanation investigated below is that the linear forecasting model may be inherently misspecified. Of particular concern is the possibility that non-linear dynamic regression models may generate more accurate out-of-sample forecasts of cumulative real GDP growth.

12.2. Non-Linear Dynamic Models

In this regard, [Hamilton \(2003\)](#) suggested that the predictive relationship between oil prices and U.S. real GDP is non-linear in that (1) oil price increases matter only to the extent that they exceed the maximum oil price in recent years and that (2) oil price decreases do not matter at all. This view was based on the in-sample fit of a single-equation predictive model of the form:

$$\Delta y_t = \alpha + \sum_{i=1}^4 \beta_i \Delta y_{t-i} + \sum_{i=1}^4 \delta_i \Delta s_{t-i}^{net,+ ,3yr} + u_t, \quad (18)$$

where s_t denotes the log of the nominal price of oil and $\Delta s_t^{net,+ ,3yr}$ the corresponding 3-year net increase in the nominal price of oil.

Hamilton's line of reasoning has prompted many researchers to construct asymmetric responses to positive and negative oil price innovations from censored oil price VAR models. Censored oil price VAR models refer to linear VAR models for $[\Delta s_t^{net,+ ,3yr}, \Delta y_t]'$, possibly augmented by other variables. Recently, [Kilian and Vigfusson \(2011a\)](#) have shown that impulse response estimates from VAR models involving censored oil price variables are inconsistent. Specifically, that paper demonstrated, first, that asymmetric

²⁹ In related work, [Ramey and Vine \(2010\)](#) propose an alternative adjustment to the price of gasoline that reflects the time cost of queuing in gasoline markets during the 1970s. That adjustment as well serves to remove a non-linearity in the transmission process. Both the non-linearity postulated in [Edelstein and Kilian \(2009\)](#) and that postulated in [Ramey and Vine \(2010\)](#) is incompatible with the specific non-linearity embodied in the models of [Mork \(1989\)](#) and [Hamilton \(1996,2003\)](#). In fact, the aforementioned papers rely on linear regressions after adjusting the energy price data.

models of the transmission of oil price shocks cannot be represented as censored oil price VAR models and are fundamentally misspecified whether the data generating process is symmetric or asymmetric. Second, standard approaches to the construction of structural impulse responses in this literature are invalid, even when applied to correctly specified models. Instead, Kilian and Vigfusson proposed a modification of the procedure discussed in Koop et al. (1996). Third, standard tests for asymmetry based on the slope coefficients of single-equation predictive models are neither necessary nor sufficient for judging the degree of asymmetry in the structural response functions, which is the question of ultimate interest to users of these models. Kilian and Vigfusson proposed a direct test of the latter hypothesis and showed empirically that there is no statistically significant evidence of asymmetry in the response functions for U.S. real GDP.

Hamilton (2011) agrees with Kilian and Vigfusson on the lack of validity of impulse response analysis from censored oil price VAR models, but suggests that non-linear predictive models such as model (18) may still be useful for out-of-sample forecasting. Indeed, the predictive model (18) can be consistently estimated by OLS and used to form one-quarter-ahead forecasts. We explore the forecasting accuracy of this strategy below. We consider both one-quarter-ahead forecasts of real GDP growth and forecasts of the cumulative real GDP growth rate several quarters ahead. The latter forecasts require a generalization of the single-equation forecasting approach proposed by Hamilton (2011).

In implementing this approach, there are several potentially important modeling choices to be made. First, even granting the presence of asymmetries in the predictive model, one question is whether the predictive model should be specified as

$$\Delta y_t = \alpha + \sum_{i=1}^4 \beta_i \Delta y_{t-i} + \sum_{i=1}^4 \delta_i \Delta s_{t-i}^{net,+ ,3yr} + u_t, \quad (18)$$

as in Hamilton (2003), or rather as

$$\Delta y_t = \alpha + \sum_{i=1}^4 \beta_i \Delta y_{t-i} + \sum_{i=1}^4 \gamma_i \Delta s_{t-i} + \sum_{i=1}^4 \delta_i \Delta s_{t-i}^{net,+ ,3yr} + u_t, \quad (19)$$

as in Balke et al. (2002) or Herrera et al. (2011), for example. The latter specification encompasses the linear reduced-form model as a special case. If $\gamma_i = 0 \forall i$, or, alternatively, if the slopes γ_i are close enough to zero, the more parsimonious non-linear predictive model (18) will have lower out-of-sample MSPE in finite samples than the unrestricted encompassing model (19). Below we explore the merits of imposing $\gamma_i = 0 \forall i$ not only in the context of single-equation models designed for one-step-ahead forecasting, but for multivariate non-linear models as well.

A second point of contention is whether non-linear forecasting models should be specified in terms of the nominal price of oil or the real price of oil. For linear models, a strong economic case can be made for using the real price of oil. For non-linear models,

the situation is less clear, as noted by Hamilton (2011). Because the argument for using net oil price increases is behavioral, one specification appears as reasonable as the other. Below we therefore will consider models specified in real as well as in nominal oil prices.

A third issue that arises only in constructing iterated forecasts for higher horizons is how to specify the process governing the price of oil. The case can be made that treating this process as exogenous with respect to real GDP might help reduce the out-of-sample MSPE, even if that restriction is incorrect. Below we therefore consider specifications with and without imposing exogeneity.

In Table 8.11, we investigate whether there are MSPE reductions associated with the use of censored oil price variables at horizons $h \in \{1, \dots, 8\}$, drawing on the analysis in Kilian and Vigfusson (2011b) and Kilian and Vigfusson (2013). For completeness, we also include results for the percent increase specification proposed in Mork (1989), the forecasting performance of which has not been investigated to date. We consider non-linear models based on the real price of oil as in Kilian and Vigfusson and non-linear models based on the nominal price of oil as in Hamilton (2003). The unrestricted multivariate non-linear forecasting model takes the form

$$\begin{aligned} \Delta r_t &= \alpha_1 + \sum_{i=1}^4 B_{11,i} \Delta r_{t-i} + \sum_{i=1}^4 B_{12,i} \Delta y_{t-i} + e_{1,t} \\ \Delta y_t &= \alpha_2 + \sum_{i=1}^4 B_{21,i} \Delta r_{t-i} + \sum_{i=1}^4 B_{22,i} \Delta y_{t-i} + \sum_{i=1}^4 \delta_i \tilde{r}_{t-i} + e_{2,t}, \end{aligned} \tag{20}$$

where $\tilde{r}_t \in \left\{ \Delta r_t^{net,+ ,3yr}, \Delta r_t^{net,+ ,1yr}, \Delta r_t^+ \right\}$, $\Delta r_t^+ \equiv \Delta r_t I(\Delta r_t > 0)$ as in Mork (1989), and $I(\cdot)$ denotes the indicator function. Analogous non-linear forecasting models may be constructed based on the nominal price of oil, denoted in logs as s_t :

$$\begin{aligned} \Delta s_t &= \alpha_1 + \sum_{i=1}^4 B_{11,i} \Delta s_{t-i} + \sum_{i=1}^4 B_{12,i} \Delta y_{t-i} + e_{1,t} \\ \Delta y_t &= \alpha_2 + \sum_{i=1}^4 B_{21,i} \Delta s_{t-i} + \sum_{i=1}^4 B_{22,i} \Delta y_{t-i} + \sum_{i=1}^4 \delta_i \tilde{s}_{t-i} + e_{2,t}, \end{aligned} \tag{20'}$$

where $\tilde{s}_t \in \left\{ \Delta s_t^{net,+ ,3yr}, \Delta s_t^{net,+ ,1yr}, \Delta s_t^+ \right\}$.

In addition, we consider a restricted version of models (20) and (20') which imposes the hypothesis that the price of oil is exogenous such that:

$$\begin{aligned} \Delta r_t &= \alpha_1 + \sum_{i=1}^4 B_{11,i} \Delta r_{t-i} + e_{1,t} \\ \Delta y_t &= \alpha_2 + \sum_{i=1}^4 B_{21,i} \Delta r_{t-i} + \sum_{i=1}^4 B_{22,i} \Delta y_{t-i} + \sum_{i=1}^4 \delta_i \tilde{r}_{t-i} + e_{2,t} \end{aligned} \tag{21}$$

and

$$\Delta s_t = \alpha_1 + \sum_{i=1}^4 B_{11,i} \Delta s_{t-i} + e_{1,t} \quad (21')$$

$$\Delta y_t = \alpha_2 + \sum_{i=1}^4 B_{21,i} \Delta s_{t-i} + \sum_{i=1}^4 B_{22,i} \Delta y_{t-i} + \sum_{i=1}^4 \delta_i \tilde{s}_{t-i} + e_{2,t} .$$

Alternatively, we may restrict the feedback from lagged percent changes in the price of oil, as suggested by Hamilton (2003). After imposing $B_{21,i} = 0 \forall i$, the baseline non-linear

Table 11a MSPE Ratios of Nonlinear Dynamic Models Relative to the AR(4) Benchmark Model Cumulative U.S. Real GDP Growth Rates

Real Refiners' Acquisition Cost for Imported Crude Oil						
Horizon	Unrestricted Model (20)			Exogenous Model (21)		
	Mork Increase	Hamilton Net Increase		Mork Increase	Hamilton Net Increase	
		1 Year	3 Year		1 Year	3 Year
1	1.14	1.21	0.98	1.15	1.21	0.98
2	1.13	1.10	0.83	1.13	1.10	0.83
3	1.07	1.05	0.79	1.07	1.04	0.79
4	1.08	0.99	0.76	1.07	0.99	0.76
5	1.09	0.94	0.75	1.09	0.93	0.74
6	1.10	0.92	0.77	1.09	0.91	0.77
7	1.10	0.92	0.79	1.09	0.91	0.78
8	1.10	0.92	0.81	1.09	0.91	0.80

Nominal Refiners' Acquisition Cost for Imported Crude Oil						
Horizon	Unrestricted Model (20')			Exogenous Model (21')		
	Mork Increase	Hamilton Net Increase		Mork Increase	Hamilton Net Increase	
		1 Year	3 Year		1 Year	3 Year
1	1.12	1.20	1.08	1.12	1.20	1.08
2	1.10	1.07	0.87	1.09	1.07	0.87
3	1.04	1.01	0.81	1.03	1.01	0.81
4	1.04	0.96	0.77	1.04	0.96	0.77
5	1.05	0.92	0.77	1.04	0.92	0.77
6	1.05	0.91	0.79	1.04	0.91	0.79
7	1.05	0.91	0.82	1.04	0.91	0.81
8	1.06	0.92	0.83	1.05	0.91	0.83

Notes: The benchmark model is an AR(4) for U.S. real GDP growth. The nonlinear dynamic models are described in the text. Boldface indicates gains in accuracy relative to benchmark model. The exogenous model suppresses feedback from lagged real GDP growth to the current price of oil.

Table 11b MSPE Ratios of Nonlinear Dynamic Models Relative to the AR(4) Benchmark Model Cumulative U.S. Real GDP Growth Rates

Real Refiners' Acquisition Cost for Imported Crude Oil						
Horizon	Restricted Model (22)			Restricted Exogenous Model (23)		
	Mork Increase	Hamilton Net Increase		Mork Increase	Hamilton Net Increase	
		1 Year	3 Year		1 Year	3 Year
1	1.14	1.12	0.91	1.14	1.12	0.91
2	1.12	1.05	0.80	1.11	1.05	0.80
3	1.07	1.01	0.77	1.07	1.01	0.78
4	1.04	0.97	0.76	1.03	0.96	0.75
5	1.03	0.93	0.75	1.02	0.93	0.75
6	1.02	0.91	0.77	1.00	0.91	0.77
7	1.01	0.91	0.79	1.00	0.90	0.78
8	1.01	0.91	0.80	1.01	0.90	0.79

Nominal Refiners' Acquisition Cost for Imported Crude Oil						
Horizon	Restricted Model (22')			Restricted Exogenous Model (23')		
	Mork Increase	Hamilton Net Increase		Mork Increase	Hamilton Net Increase	
		1 Year	3 Year		1 Year	3 Year
1	1.12	1.12	1.01	1.12	1.12	1.01
2	1.09	0.02	0.84	1.08	1.02	0.83
3	1.04	0.98	0.79	1.03	0.97	0.79
4	1.01	0.94	0.76	1.00	0.94	0.76
5	0.99	0.92	0.77	0.98	0.91	0.77
6	0.99	0.91	0.80	0.97	0.90	0.79
7	0.98	0.90	0.81	0.97	0.90	0.80
8	0.98	0.91	0.83	0.97	0.90	0.82

Notes: The benchmark model is an AR(4) for U.S. real GDP growth. The nonlinear dynamic models are described in the text. Boldface indicates gains in accuracy relative to benchmark model. The restricted model suppresses feedback from lagged percent changes in the price of oil to current real GDP growth, as proposed by Hamilton (2003, 2011). The restricted exogenous model combines this restriction with that of exogenous oil prices, further increasing the parsimony of the model.

forecasting model reduces to:

$$\begin{aligned}
 \Delta r_t &= \alpha_1 + \sum_{i=1}^4 B_{11,i} \Delta r_{t-i} + \sum_{i=1}^4 B_{12,i} \Delta y_{t-i} + e_{1,t} \\
 \Delta y_t &= \alpha_2 + \sum_{i=1}^4 B_{22,i} \Delta y_{t-i} + \sum_{i=1}^4 \delta_i \tilde{r}_{t-i} + e_{2,t} .
 \end{aligned} \tag{22}$$

and

$$\begin{aligned}\Delta s_t &= \alpha_1 + \sum_{i=1}^4 B_{11,i} \Delta s_{t-i} + \sum_{i=1}^4 B_{12,i} \Delta y_{t-i} + e_{1,t} \\ \Delta y_t &= \alpha_2 + \sum_{i=1}^4 B_{22,i} \Delta y_{t-i} + \sum_{i=1}^4 \delta_i \tilde{s}_{t-i} + e_{2,t}.\end{aligned}\tag{22'}$$

Finally, we can combine the restrictions $B_{12,i} = 0 \forall i$ and $B_{21,i} = 0 \forall i$, resulting in forecasting models (23) and (23'):

$$\begin{aligned}\Delta r_t &= \alpha_1 + \sum_{i=1}^4 B_{11,i} \Delta r_{t-i} + e_{1,t} \\ \Delta y_t &= \alpha_2 + \sum_{i=1}^4 B_{22,i} \Delta y_{t-i} + \sum_{i=1}^4 \delta_i \tilde{r}_{t-i} + e_{2,t}\end{aligned}\tag{23}$$

and

$$\begin{aligned}\Delta s_t &= \alpha_1 + \sum_{i=1}^4 B_{11,i} \Delta s_{t-i} + e_{1,t} \\ \Delta y_t &= \alpha_2 + \sum_{i=1}^4 B_{22,i} \Delta y_{t-i} + \sum_{i=1}^4 \delta_i \tilde{s}_{t-i} + e_{2,t}.\end{aligned}\tag{23'}$$

At the one-quarter horizon, real GDP growth forecasts from model (22') and (23') only depend on the second equation, which is equivalent to using Hamilton's model (1). All models are estimated by least squares, as is standard in the literature. The forecasts are constructed by Monte Carlo integration based on 10,000 draws. The estimation and evaluation periods are the same as in Table 8.10.

Table 8.11 displays the MSPE ratios for all eight models by horizon. All results are normalized relative to the AR(4) model for real GDP growth. No tests of statistical significance have been conducted, given the computational cost of such tests. The first result is that no non-linear model is substantially more accurate than the AR(4) benchmark model at the one-quarter horizon except for models (22) and (23) with the 3-year net oil price increase. For the latter models, the reduction in MSPE is 9%. At the one-year horizon, the reduction in MSPE reaches 24% and 25% respectively. The accuracy gains at the two-year horizon are about 20%. Second, models that combine Hamilton's assumptions with that of exogenous oil prices and embed all these assumptions in a multivariate dynamic framework, help improve forecast accuracy at short horizons. Third, neither the percent increase model based on Mork (1989) nor the 1-year net increase model motivated by Hamilton (1996) is more accurate than the AR(4) benchmark at the one-quarter horizon. This is true regardless of whether the price of oil is specified in nominal or real terms and regardless of what additional restrictions we impose. At longer horizons, there

is weak evidence that some of these specifications reduce the MSPE at some horizons, but in no case as much as for the three-year net oil price increase. Fourth, there is a clear ranking between 3-year net increase models based on the real price of oil and the corresponding models based on the nominal price of oil. Forecasts from models 22 and 23 based on the real price of oil are more accurate, especially at short horizons, than forecasts from models (22') and (23') based on the nominal price.

An obvious question of interest is to what extent allowing for nonlinearities improves our ability to forecast major economic downturns in the U.S. The one-quarter ahead results in the upper panel of Figure 8.9 indicate that the 3-year net increase models are quite successful at forecasting the 2008 recession, about half of which is forecast by models 23 and (23'), whereas their performance during other episodes is less impressive. The corresponding lower panel in Figure 8.9 shows that models 23 and (23') are equally successful at forecasting the downturn of 2008 and the subsequent recovery four quarters ahead. If anything, these nonlinear models appear too successful in that their forecasting success in 2008 implies a comparatively minor role for the financial crisis. This success, however, comes at the price that these models forecast a number of economic declines in the pre-crisis period that did not materialize. For example, the net increase model

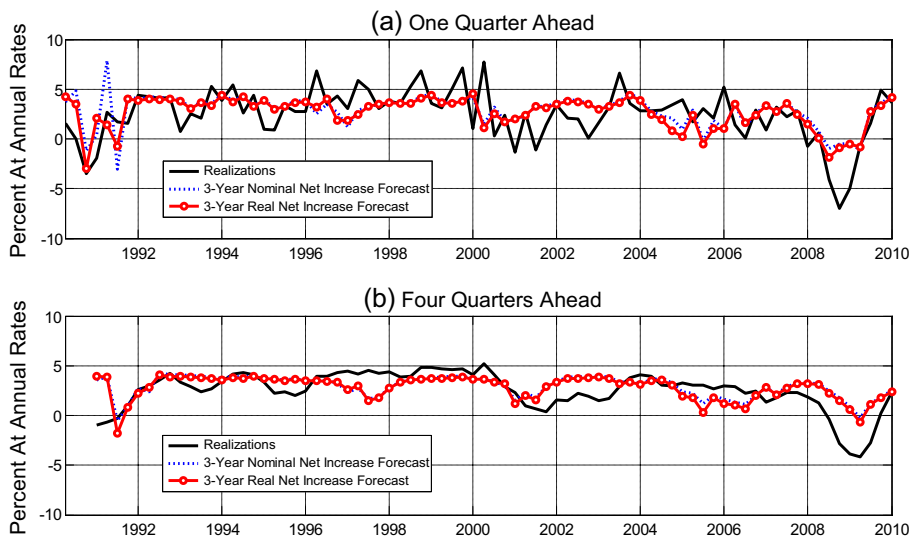


Figure 8.9 Nonlinear forecasts of cumulative real GDP growth from models (23) and (23') U.S. refiners' acquisition cost for imports. Notes: One forecasting model is a suitably restricted VAR(4) model for real GDP growth and the percent change in the real price of oil augmented by four lags of the 3-year real net oil price increase. The other model is a similarly restricted VAR(4) model for real GDP growth and the percent change in the nominal price of oil augmented by four lags of the 3-year nominal net oil price increase.

incorrectly forecast much lower growth than materialized in 1997 and in 2005/06, and it missed the beginning of the economic decline of 1990/91.

Plots of the recursive MSPE of these nonlinear models show that much of the forecasting success of nonlinear models is driven by one episode, namely the economic collapse in 2008/09 following the financial crisis. This point is illustrated in Figure 8.10. The left panel of Figure 8.10 is based on the nominal PPI used in Hamilton's original analysis; the right panel shows the corresponding results for the nominal refiners' acquisition cost for crude oil imports. The plot of the cumulative recursive MSPE for the PPI model (23') reveals that the overall gain in accuracy in this example is almost driven entirely by the 2008/09 recession. Excluding this episode, model (23') has higher MSPE than the linear AR model nearly throughout the evaluation period. Given this evidence a strong case can be made that few forecasters would have had the courage to stick with the predictions of this nonlinear model given the sustained failure of the model in the years leading up to the financial crisis. The corresponding results for the refiners' acquisition cost for imported crude oil in the right panel are somewhat more favorable, but also reveal a tendency of the net oil price increase model to have a higher recursive MSPE prior to the financial crisis than the AR(4) benchmark model for real GDP growth.

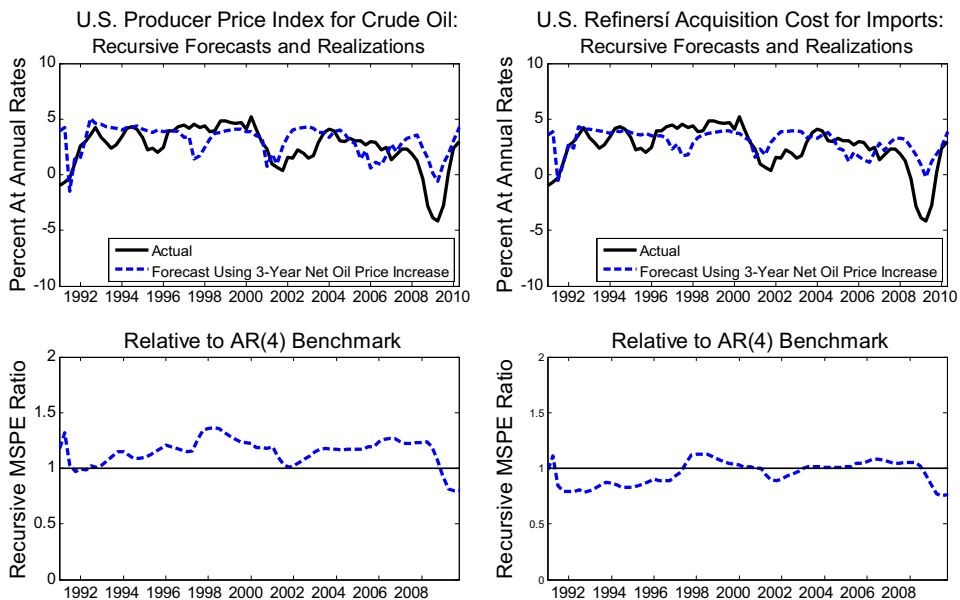


Figure 8.10 Nonlinear forecasts of cumulative real GDP growth from model (23'). Notes: The nonlinear forecasting model is a suitably restricted VAR(4) model for real GDP growth and the percent change in the nominal price of crude oil augmented by four lags of the corresponding 3-year nominal net oil price increase.

The unusually good fit of the 3-year net increase model in 2008/08 combined with the tendency of this model to incorrectly predict recessions prior to the crisis, is an indication of potential out-of-sample overfitting. It is usually thought that out-of-sample forecasts protect against overfitting. This intuition may fail in small samples because, under quadratic loss, the ability of the nonlinear model to predict correctly one extreme event (such as the sharp economic decline associated with the financial crisis) may more than offset repeated smaller forecast errors of the same nonlinear model during earlier episodes. Given that the financial crisis occurred immediately after a major surge in the price of oil, but itself was presumably not caused by that oil price surge, the obvious concern is that net increase models may have forecast the 2008 recession for the wrong reasons.³⁰ Only additional data will ultimately resolve the question of how much the low MSPE of the net increase model in Table 8.11 owes to overfitting. If the near-simultaneous occurrence of the financial crisis and the oil price surge is coincidental, then the forecasting accuracy of the nonlinear model can be expected to worsen, as the sample is extended. If the forecasting success of the nonlinear model were to persist even after the financial crisis is over, this would add credibility to the nonlinear real GDP growth forecasts.

The same concern regarding the financial crisis episode arises to varying degrees with other oil price series. Table 8.12 provides a systematic comparison of the performance of nonlinear forecasting models relative to the AR(4) benchmark model for real GDP growth for different oil price series and evaluation periods. To conserve space, we focus on models 23 and which tend to be the most accurate nonlinear forecasting models. Table 8.12 shows that the relative MSPE of nonlinear forecasting models can be highly sensitive to the choice of oil price series. The first two columns of Table 8.12 focus on the evaluation period 1990.Q1–2010.Q1. Column (1) shows that, for eight of ten model specifications, the one-quarter ahead nonlinear forecasting model proposed by Hamilton (2011) fails to outperform the AR(4) benchmark model for real GDP. Only for the real refiners' acquisition cost for imported crude oil and for the nominal WTI specification are there any gains in forecast accuracy. In particular, the nominal PPI specification favored by Hamilton (2011) on the basis of in-sample diagnostics is less accurate than the AR benchmark model. Much more favorable results are obtained at the one-year horizon in column (2) of Table 8.12. In the latter case, all nonlinear forecasting models yield sizable reductions in the MSPE. However, almost all evidence of forecast accuracy gains vanishes once the financial crisis episode is excluded, as shown in columns (3) and (4) of Table 8.12, suggesting once again an element of overfitting. Some nonlinear forecasting models have more than 60% higher MSPE than the AR benchmark model in that case. We conclude

³⁰ While the case can be made that higher oil prices contributed to the U.S. housing crisis, as discussed in Hamilton (2009, p. 255–257), for example, few observers would view higher oil prices as the primary cause of the subsequent financial market meltdown.

Table 8.12 MSPE Ratios for Cumulative U.S. Real GDP Growth Rate Relative to AR(4) Benchmark Model: Models (23) and (23') for Alternative Oil Price Specifications and Evaluation Periods

		1990.Q1-2010.Q1		1990.Q1-2007.Q4	
		Horizon		Horizon	
	Oil Price Series	<i>h</i> =1	<i>h</i> =4	<i>h</i> =1	<i>h</i> =4
Real	RAC imports	0.91	0.75	1.12	1.14
	RAC composite	1.16	0.81	1.50	1.33
	RAC domestic	1.22	0.82	1.56	1.23
	WTI	1.04	0.77	1.24	0.97
	PPI	1.23	1.86	1.64	1.50
Nominal	RAC imports	1.01	0.76	1.23	1.05
	RAC composite	1.25	0.80	1.59	1.16
	RAC domestic	1.22	0.83	1.50	1.12
	WTI	0.91	0.80	1.01	0.94
	PPI	1.23	0.80	1.59	1.23

Notes: To conserve space, we focus on the most accurate nonlinear forecasting models. The models are described in the text. Boldface indicates gains in accuracy relative to AR(4) benchmark model for real GDP growth. The out-of-sample evaluation period for $h = 4$ starts three quarters later by construction than for $h = 1$.

that the evidence that nonlinear oil price transformation help forecast cumulative U.S. real GDP growth is mixed at best.³¹

12.3. Does Using Pre-1974 Data Help in Forecasting Real GDP?

In Section 4.2.3 we made the case that using pre-1974 data in studying the predictive relationship between the price of oil and U.S. real GDP is inadvisable. This argument has not prevented some researchers from fitting forecasting models on samples dating back to 1949. Indeed, one can make the case that in forecasting the reduction in the variance of the model estimates from using these additional data may in practice outweigh any biases caused by ignoring structural breaks. For example, [Hamilton \(2011\)](#) reports the one-quarter-ahead MSPE ratio of his preferred forecasting model (18) with four lags of the 3-year net increase in the nominal PPI with the coefficients estimated on data for 1949.Q2 to 2001.Q3 relative to a model motivated by the analysis in [Kilian and Vigfusson \(2011a\)](#) with six lags of real GDP growth, six lags of the percent change in the oil price, and six lags of the net oil price increase. The latter model is estimated on data for 1974.Q1-2007.Q4. The out-of-sample MSPE of both models is evaluated over 2008.Q1 to 2010.Q1. The resulting MSPE ratio is 0.24.

³¹ The results in Tables 8.11 and 8.12

This result is not directly comparable to our analysis for three reasons. First, it is based on split-sample analysis rather than recursive analysis for model (18). Second, the relevant benchmark for our out-of-sample comparisons is the AR(4) model for real GDP growth, not the much less parsimonious model discussed in Kilian and Vigfusson (2011a). Third, Hamilton's out-of-sample evaluation window includes only two years of data, whereas we evaluate the model over a 20-year period. Moreover, the specific 2 years of data that Hamilton focuses on when evaluating the model correspond to the financial crisis episode, which we showed to be to the time period most favorable to Hamilton's model.

It is straightforward to conduct an exercise in the same spirit as Hamilton's such that the results can be compared directly to those in Table 8.12. When we conducted this additional analysis, we found that the one-quarter-ahead MSPE ratio of Hamilton's 3-year net increase model (18) recursively estimated on data back to 1949.Q2 and evaluated on the 1990.Q1–2010.Q1 period is 1.02 compared with an AR(4) model for real GDP estimated on post-break data. This compares with a ratio of 1.23 for the same model estimated on post-break data, as reported in Table 8.12, indicating that using the longer sample does improve the one-quarter-ahead MSPE, but does not overturn our conclusion that the AR(4) for real GDP growth is more accurate out of sample than Hamilton's model.³² We also found that the MSPE ratio of the corresponding 4-quarter ahead iterated forecast increases from 0.80 in Table 8.12 to 0.85, if we include the pre-1973 data in the recursive estimation of the model, indicating that the longer estimation sample lowers the model's forecast accuracy. Finally, when the nonlinear model in question is estimated based on data back to 1949.Q2, but we exclude the financial crisis period from the evaluation sample, the recursive MSPE ratios increase to 1.36 one quarter ahead and 1.53 four quarters ahead. In short, our results are robust to extending the estimation sample back to 1949.Q2.

12.4. Non-Parametric Approaches

Our approach in this section has been parametric. Alternatively, one could have used non-parametric econometric models to investigate the forecasting ability of the price of oil for real GDP. In related work, Bachmeier et al. (2008) used the integrated conditional moment test of Corradi and Swanson (2002, 2007) to investigate whether oil prices help forecast real GDP growth one-quarter-ahead. The advantage of this approach is that – while imposing linearity under the null – it allows for general non-linear models under the alternative; the disadvantage is that the test is less powerful than the parametric approach if the parametric structure is known. Bachmeier et al. report a p -value of 0.20 for the null that nominal net increases in the WTI price of oil do not help forecast U.S. real GDP. The p -value for percent changes in the WTI price of crude oil is 0.77. Similar results

³² The corresponding one-quarter-ahead MSPE ratio during the 1990.Q1–2007.Q4 evaluation period is 1.36 and the ratio during 2008.Q1–2010.Q1 is 0.32, consistent with our earlier remarks on the special nature of this 2-year period.

are obtained for real net increases and for percent changes in the real WTI price. These findings are broadly consistent with ours. Bachmeier et al. (2008) also report qualitatively similar results using a number of fully non-parametric approaches. An obvious caveat is that their analysis is based on data since 1949, which is not appropriate for the reasons discussed earlier, and ends before the 2008/09 recession. Using their non-parametric techniques on our much shorter sample period does not seem advisable, because there is no way of controlling the size of the test.

13. THE ROLE OF OIL PRICE VOLATILITY

Point forecasts of the price of oil are important, but they fail to convey the uncertainty associated with oil price forecasts. That uncertainty is captured by the predictive density. Figure 8.11 plots the 12-month-ahead predictive density for the real price of oil as of 2009.12, generated from the no-change forecasting model. Although it is obvious that there is tremendous uncertainty about the future real price of oil, even when using the best available forecasting methods, it is less obvious how to convey and interpret that information. For example, standard questions in the financial press about whether the price of oil *could* increase to \$200 a barrel, at the risk of being misunderstood, inevitably and always must be answered in the affirmative because the predictive distribution has unbounded support. That answer, however, is vacuous because it does not convey how likely such an event is or by how much the price of oil is expected to exceed the \$200 threshold in that event.

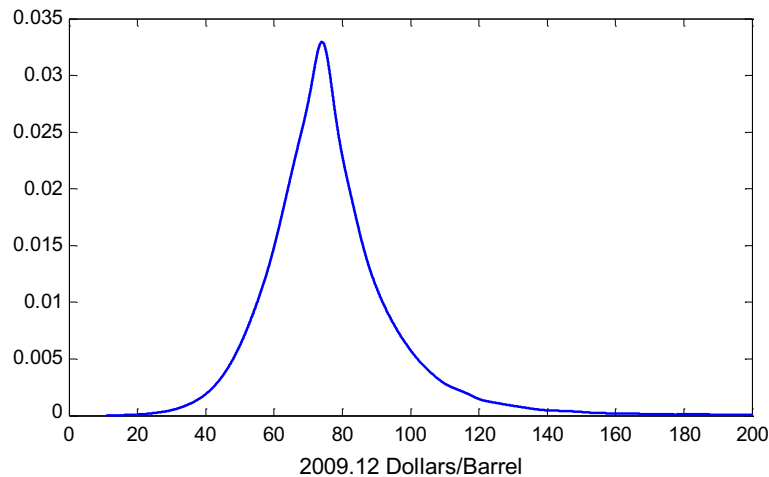


Figure 8.11 12-Month-ahead predictive density of the real WTI price of oil as of 2009.12 based on no-change forecast.

13.1. Nominal Oil Price Volatility

One seemingly natural way of summarizing the information in the predictive distribution is to report the variability of the forecasts. Interest in oil price volatility measures arises, for example, from financial analysts interested in pricing options and from portfolio managers interested in diversifying risks. Given that at short horizons CPI inflation is negligible, it is customary in financial applications to focus on nominal oil price volatility. One approach to measuring oil price volatility is to rely on the implied volatilities of put and call options, which are available from January 1989 on. Implied volatility measures are computed as the arithmetic average of the daily implied volatilities from the put and call options associated with a futures contract of a given maturity. The upper panel of Figure 8.12 shows the 1-month implied volatility time series for 2001.1–2009.12, computed from daily CRB data, following the same procedure as for the spot and futures prices in Section 5. Alternatively, we may use daily percent changes in the nominal WTI price of oil to construct measures of realized volatility, as shown in the second panel of Figure 8.12 (see, e.g., Bachmeier et al., 2008). Finally, yet another measure of volatility can be constructed from parametric GARCH or stochastic volatility models. The bottom panel of Figure 8.12 shows the 1-month-ahead conditional variance obtained from recursively

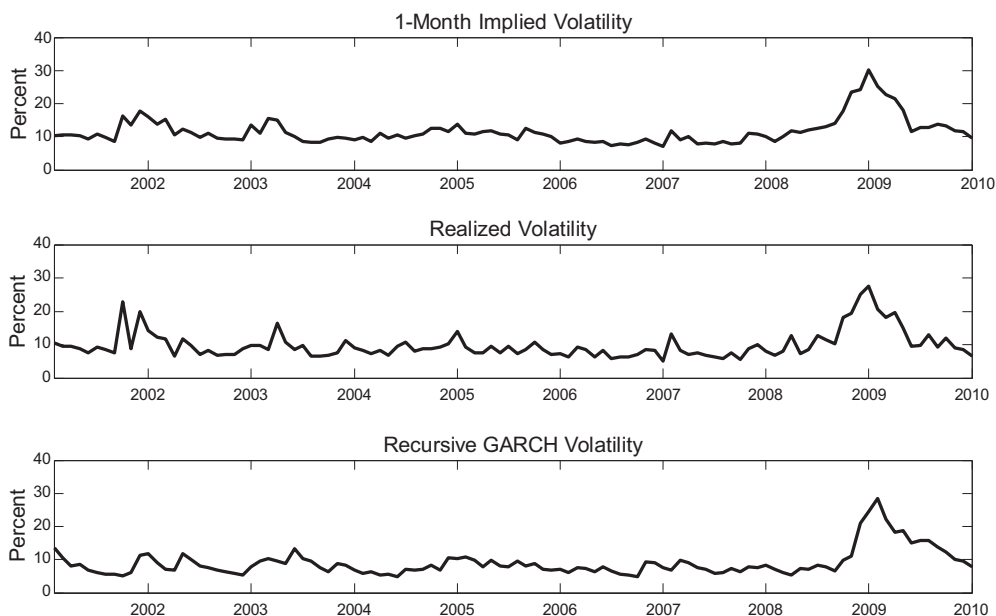


Figure 8.12 Alternative measures of nominal oil price volatility. *Notes:* The GARCH volatility estimate is for the percent change in the nominal WTI price. The realized volatility was obtained from daily WTI prices. The implied volatility measure refers to the arithmetic average of the daily implied volatilities from at-the-money put and call options associated with 1-month oil futures contracts and was constructed by the authors from CRB data. All volatility estimates are monthly and expressed as standard deviations, following the convention in the literature.

estimated Gaussian GARCH(1,1) models.³³ The initial estimation period is 1974.1–2000.12. The estimates are based on the percent change in the nominal WTI price; the corresponding results for the real WTI price are almost indistinguishable at the 1-month horizon.³⁴

Figure 8.12 plots all three volatility measures on the same scale. Although the volatility estimates agree that by far the largest peak in volatility occurred near the end of 2008, there are important differences. For example, the implied volatility measure increases steadily starting in early 2008 and peaks in December 2008. Realized volatility also peaks in December 2008, but does not increase substantially the second half of 2008. Finally, GARCH volatility is even slower to increase in 2008 and only peaks in January 2009. This ranking is consistent with the view that implied volatility is the most forward-looking volatility measure and GARCH volatility the most backward-looking volatility estimate (and hence the least representative measure of real time volatility). Similarly, the implied volatility and realized volatility measures indicate substantial secondary spikes in volatility in 2001/02 and 2003, whereas the spikes in the GARCH volatility estimate are much smaller and occur only with a delay.

It may seem that fluctuations in oil price volatility, defined in this manner, would be a good indicator of changes in oil price risks. It is important not to equate risk and uncertainty, however. Whereas the latter may be captured by the volatility of oil price forecasts, the former cannot. The standard risk that financial markets in oil-importing economies are concerned with is the risk of excessively high oil prices. That risk in general will be at best weakly correlated with the volatility of oil price forecasts because any reduction in risk, as the price of oil falls, all else equal, will be associated with increased oil price volatility. This is why in 1986, for example, oil price volatility increased, as OPEC collapsed and the price of oil dropped sharply, whereas by all accounts consumers were pleased with lower oil prices and the diminished risk of an OPEC-induced supply disruption. Hence, standard volatility measures are of limited use as summary statistics for the predictive distribution of oil price forecasts. We defer to Section 13.3 a more detailed exposition of how appropriate risk measures may be computed from the predictive distribution of the price of oil.

13.2. Real Oil Price Volatility

Interest in the volatility of oil prices also has been prompted by research aimed at establishing a direct link from oil price volatility to business cycle fluctuations in the real economy.

³³ The standard GARCH model is used for illustrative purposes. An alternative would be a GARCH-in-Mean model. Given that oil is only one of many assets handled by portfolio managers, however, it is not clear that the GARCH-in-Mean model for single-asset markets is appropriate in this context, while more general multivariate GARCH models are all but impossible to estimate reliably on the small samples available for our purposes (see, e.g., Bollerslev et al., 1992).

³⁴ We deliberately focus on oil price volatility at the 1-month horizon. Although from an economic point of view volatility forecasting at longer horizons would be of great interest, the sparsity of options price data makes it difficult to extend the implied volatility approach to longer horizons. Likewise, GARCH volatility estimates quickly converge to the unconditional variance at longer horizons.

For example, [Bernanke \(1983\)](#) and [Pindyck \(1991\)](#) showed that the uncertainty of the price of oil (measured by the volatility of the price of oil) matters for investment decisions if firms contemplate an irreversible investment, the cash flow of which depends on the price of oil. An analogous argument holds for consumers considering the purchase of energy-intensive durables such as cars. Real options theory implies that, all else equal, an increase in expected volatility will cause marginal investment decisions to be postponed, causing a reduction in investment expenditures. [Kellogg \(2010\)](#) provides evidence that such mechanisms are at work in the Texas oil industry, for example.

Unlike in empirical finance, the relevant volatility measure in these models is the volatility of the real price of oil at horizons relevant to purchase and investment decisions, which is typically measured in years or even decades rather than days or months, making standard measures of short-term nominal price volatility inappropriate. Measuring the volatility of the real price of oil at such long forecast horizons is inherently difficult given how short the available time series are, and indeed researchers in practice have typically asserted rather than measured these shifts in real price volatility or they have treated short-horizon volatility as a proxy for longer-horizon volatility (see, e.g., [Elder and Serletis, 2010](#)).³⁵ This approach is unlikely to work. Standard monthly or quarterly GARCH model cannot be used to quantify changes in the longer-run expected volatility of the real price of oil because GARCH forecasts of the conditional variance quickly revert to their time invariant unconditional expectation, as the forecasting horizon increases. If volatility at the economically relevant horizon is constant by construction, it cannot explain variation in real activity over time, suggesting that survey data may be better suited for characterizing changes in forecast uncertainty over time. Some progress in this direction may be expected from ongoing work conducted by [Anderson et al. \(2011\)](#) based on the distribution of Michigan consumer expectations of 5-year-ahead gasoline prices. For further discussion of this point also see [Kilian and Vigfusson \(2011b\)](#).

13.3. Quantifying Oil Price Risks

Although oil price volatility shifts play a prominent role in discussions of the impact of oil price shocks, it is important to keep in mind that volatility measures are not in general useful measures of the price risks faced by either producers or consumers of crude oil (or of refined products). Consider an oil producer capable of producing crude oil from existing wells as long as the price of oil exceeds his marginal cost of \$25 a barrel. One risk faced by that oil producer is that he will go out of business if the price of oil falls below that threshold. Excessively high oil prices, in contrast, are of no concern until they

³⁵ In rare cases, the relevant forecast horizon may be short enough for empirical analysis. For example, [Kellogg \(2010\)](#) makes the case that for the purpose of drilling oil wells in Texas, as opposed to Saudi Arabia, a forecast horizon of only 18 months is adequate. Even at that horizon, however, there are no oil-futures options price data that would allow the construction of implied volatility measures. [Kellogg \(2010\)](#) therefore converts the 1-month volatility to 18-month volatilities based on the term structure of oil futures. That approach relies on the assumption that oil futures prices are reliable predictors of future oil prices.

reach the point of making replacement technologies economically viable. That might be the case at a threshold of \$120 a barrel, for example, at which price major oil producers risk inducing the large-scale use of alternative technologies with adverse consequences for the long-run price of crude oil.³⁶ Thus, the oil producer will care about the risk of the price of oil not being contained in the range between \$25 and \$120, and the extent to which he is concerned with violations of that range depends on his risk aversion, which need not be symmetric in either direction.³⁷ There is no reason why oil producers should necessarily be concerned with a measure of the variability of the real price of oil. In fact, it can be shown that risk measures are not only quantitatively different from volatility measures, but in practice may move in the opposite direction.

Likewise, a consumer of retail motor gasoline (and hence indirectly of crude oil) is likely to be concerned with the price of gasoline exceeding what he can afford to spend each month (see [Edelstein and Kilian, 2009](#)). The threshold at which consumers might trade in their SUV more energy-efficient car is near \$3 a gallon perhaps. The threshold at which commuters may decide to relocate closer to their place of work might be at a price near \$5 a gallon. The possibility that the price of gasoline could fall below \$2, in contrast, is of comparatively little consequence to consumers' economic choices, making the volatility of oil prices and related statistics such as the value at risk irrelevant to the typical consumer.

In both examples above, the appropriate specification of these agents' decision problem is in terms of upside and downside price risks. The literature on risk management postulates that risk measures must satisfy two basic requirements. One requirement is that the measure of risk must be related to the probability distribution $F(\cdot)$ of the random variable of interest; the other requirement is that it must be linked to the preferences of the user, typically parameterized by a loss function (see [Machina and Rothschild, 1987](#)). Except in special cases these requirements rule out commonly used measures of risk based on the predictive distribution alone such as the sample moments, sample quantiles or the value at risk. In deriving appropriate risk measures that characterize the predictive distribution for the real price of oil, it is useful to start with the loss function. A reasonably general class of loss functions $l(\cdot)$ that encompasses the two empirical examples above is:

$$l(R_{t+h}) = \begin{cases} a(\underline{R} - R_{t+h})^\alpha & \text{if } R_{t+h} < \underline{R} \\ 0 & \text{if } \underline{R} \leq R_{t+h} \leq \bar{R} \\ (1 - a)(R_{t+h} - \bar{R})^\beta & \text{if } R_{t+h} > \bar{R}, \end{cases}$$

where R_{t+h} denotes the real price of oil in dollars h periods from date t , $0 \leq a \leq 1$ is the weight attached to downside risks, and $\alpha \geq 0$ and $\beta \geq 0$ are determined by the user's degree of risk aversion. Risks are associated with the event of R_{t+h} exceeding an upper

³⁶ A similar irreversible shift in OECD demand occurred after the oil price shocks of the 1970s when fuel oil was increasingly replaced by natural gas. The fuel oil market never recovered, even as the price of this fuel fell dramatically in the 1980s and 1990s (see [Dargay and Gately, 2010](#)).

³⁷ The threshold of \$120 in this example follows from adjusting the cost estimates for shale oil production in [Farrell and Brandt \(2006\)](#) for the cumulative inflation rate since 2000.

threshold of \bar{R} or falling below the lower threshold of R . It can be shown that under this loss function, the expected loss is a weighted average of upside and downside risks of the form

$$E(l) = -aDR_\alpha + (1 - a)UR_\beta,$$

where

$$DR_\alpha \equiv - \int_{-\infty}^{\underline{R}} (\underline{R} - R_{t+h})^\alpha dF(R_{t+h}), \quad \alpha \geq 0$$

$$UR_\beta \equiv \int_{\bar{R}}^{\infty} (R_{t+h} - \bar{R})^\beta dF(R_{t+h}), \quad \beta \geq 0$$

are the downside risk and upside risk, respectively. This definition encompasses a variety of risk definitions familiar from the finance literature. For example, for the special case of $\alpha = \beta = 0$ these expressions reduce to the (target) probabilities $DR_0 = -\Pr(R_{t+h} < \underline{R})$ and $UR_0 = \Pr(R_{t+h} > \bar{R})$ and for the special case of $\alpha = \beta = 1$ they reduce to the probability weighted expected shortfall $DR_1 = E(R_{t+h} - \underline{R} | R_{t+h} < \underline{R}) \Pr(R_{t+h} < \underline{R})$ and the probability weighted expected excess $UR_1 = E(R_{t+h} - \bar{R} | R_{t+h} > \bar{R}) \Pr(R_{t+h} > \bar{R})$. Note that the latter definition not only is concerned with the likelihood of a tail event, but also with how far the real price of oil is expected to be in the tail. The latter term is also known as the expected shortfall (or expected excess). The expectations and probabilities in question in practice can be estimated by their sample equivalent.³⁸

This digression highlights that the volatility of the real price of oil in general is not the relevant statistic for the analysis of risks. In particular, if and only if the loss function is quadratic and symmetric about zero, the variance of the price of oil about zero provides an adequate summary statistic for the risk in oil price forecasts. Even that target variance, however, is distinct from conventionally used measures of oil price volatility, defined as the variance about the sample mean of the predictive distribution. The latter measure under no circumstances can be interpreted as a risk measure because it depends entirely on the predictive distribution of the price of oil and not at all on the user's preferences.

Risk measures can be computed for any predictive distribution. Estimates of predictive distributions for iterated forecasts may be constructed by applying standard bootstrap methods (possibly adapted to allow for conditional heteroskedasticity) to the serially uncorrelated errors of the estimated regression model and iteratively constructing the bootstrap realizations of the multi-step-ahead forecast. When evaluating the distribution of direct forecasts rather than iterated forecasts, additional complications may arise. Applying standard bootstrap methods to the sequence of one-step-ahead forecast errors

³⁸ Measures of risk of this type were first introduced by Fishburn (1977), Holthausen (1981), Artzner et al. (1999), and Basak and Shapiro (2001) in the context of portfolio risk management and have become a standard tool in recent years (see, e.g., Engle and Brownlees, 2010). For a general exposition of risk measures and risk management in a different context see Kilian and Manganello (2007, 2008).

obtained from fitting the direct forecasting model to historical data requires the forecast errors to be serially uncorrelated. This will typically be the case when constructing forecasts at horizon $h = 1$. For example, when fitting a random walk model of the form $s_{t+1} = s_t + \varepsilon_{t+1}$, the forecast errors at horizon 1 may be resampled using standard bootstrap methods as in the case of iterated forecasts.

At longer horizons, however, the construction of direct multi-horizon forecasts from the random walk model is more involved. One option is to fit the forecasting model on non-overlapping observations and proceed as for $h = 1$. This approach is simple, but tends to involve a considerable reduction in estimation precision. For example, in constructing the predictive distribution of 1-year-ahead no-change forecasts from monthly data, one would construct for the current month the sequence of year-on-year percent changes relative to the same month in the preceding year and approximate the predictive distribution by resampling this sequence of year-on-year forecast errors. The other option is to construct forecast errors from overlapping observations and to recover the underlying white noise errors by fitting an $MA(h-1)$ process to the sequence of h -step-ahead forecast errors. This allows the construction of bootstrap approximations of the predictive density by first resampling the serially uncorrelated white noise residuals using suitable bootstrap methods and then constructing bootstrap replicates of the h -month-ahead forecast errors from the implied moving averages. Below we implement this approach in the context of a 12-month-ahead no-change forecast of the real WTI price of oil. The risk measures are constructed directly from the bootstrap estimate of the predictive distribution, as discussed above.

Figure 8.13 plots the risk that the price of oil (expressed in 2009.12 dollars) exceeds \$80 1 year later ($\bar{R} = 80$) and the risk that it drops below \$45 1 year later ($\underline{R} = 45$). These thresholds have been chosen for illustrative purposes. The upper panel of Figure 8.13 plots the upside and downside risks for $\alpha = \beta = 0$, whereas the lower panel plots the corresponding results for $\alpha = \beta = 1$. Note that by convention the downside risks have been defined as a negative number to improve the readability of the plots. Although the upside risks and downside risks respond to sustained changes in the conditional mean forecast by construction, the relationship is not one-for-one. Figure 8.13 shows that the ex ante probability of the real price of oil exceeding \$80 1 year later was small except during 2005-08 and after mid-2009; high probabilities of the real price of oil falling below \$45 occurred only in 2001-04 and 2009. The lower panel shows the corresponding probability-weighted expected excess and expected shortfall measures. For $\alpha = \beta = 1$, the upside risks in 2007-08 become disproportionately larger relative to earlier upside risks and relative to the downside risks. Regardless of the choice of α and β , the balance of risks since mid-2009 has been tilted in the upside direction. Recent upside risks are comparable to those in 2006.

It is immediately evident that the three standard volatility measures in Figure 8.12 are not good proxies for either of the two risks shown in Figure 8.13. For example, in

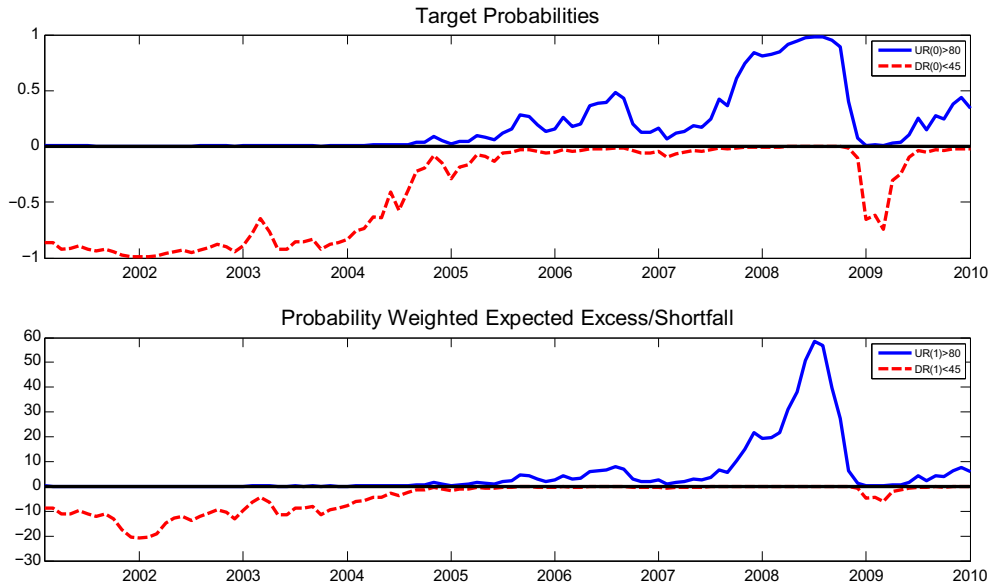


Figure 8.13 12-Month-ahead upside and downside risks in the real WTI price based on no-change forecast. *Notes:* Risks are defined in terms of the event that the price of oil (in 2009.12 dollars) exceeds 80 dollars or falls below 45 dollars. For further discussion of these risk measures see Kilian and Manganello (2007).

the second half of 2008 volatility skyrockets while the upside risk plummets. The upside risk peaks in mid-2008, when the real price of oil peaked, but volatility only peaks in December 2008 or January 2009, when the real price of oil had reached a trough, much to the relief of oil consumers. Moreover, the spikes in volatility in 2001/02 and 2003 are not mirrored by increases in upside risk, while the sustained increase in upside risk after 2004 is not mirrored by a sustained increase in volatility. Nor is volatility systematically related to downside oil price risks. Although both downside risks and volatility peak in 2001/02, the sustained increase in volatility in early and mid-2008 is not mirrored by an increase in downside risk. Furthermore, the decline in downside risks during 2004 and 2005 is not reflected in systematic changes in volatility.

It is worth emphasizing that none of these 12-month-ahead risk forecasts provided any warning of the collapse of the real price of oil in late 2008. To the extent that this collapse was unpredictable based on past data, this is not surprising. The problem is not with the risk measures but rather with the underlying predictive distribution that these risk measures have been applied to. In fact, none of the available forecasting methods provided an advance warning of the collapse of the real price of oil in late 2008, although some of the more accurate forecasting models predicted a decline in the real price.

14. AVENUES FOR FUTURE RESEARCH

There are a number of directions for future research on forecasting oil prices. One relates to the use of additional industry-level predictors not commonly considered by economists. Although crude oil is one of the more homogeneous commodities traded in global markets, not all refineries may process all grades of crude oil. Moreover, different grades of crude oil yield different mixes of refined products. Hence, shifts in the demand for one type of refined product, say, diesel fuel, have implications for the product mix of refined products (diesel, gasoline, kerosene, heating oil, etc.) and hence for the demand for different grades of crude oil, depending on the capacity utilization rates of different refineries. Situations can arise in which excess demand for one grade of crude oil may result in rising prices, while excess supply of another grade of crude oil is associated with falling prices.³⁹ Models that incorporate information about such spreads or about the underlying determinants of demand have the potential of improving forecasts of the price of a given grade of crude oil (see, e.g., Verleger Jr., 1982, 2011).

A second issue of interest is the role played by heterogeneous oil price and gasoline price expectations in modeling the demand for energy-intensive durables (see Anderson et al., 2011). There is strong evidence that not all households share the same expectations, casting doubt on standard rational expectations models with homogeneous agents. This also calls into question the use of a single price forecast in modeling purchasing decisions in the aggregate. This problem is compounded to the extent that different market participants (households, refiners, oil producers) in the same model may have very different risk assessments based on the same predictive oil price distribution. Both of these effects may undermine the predictive power of the price of oil for macroeconomic aggregates as well as the explanatory power of theoretical models based on oil price forecasts.

Third, we have deliberately refrained from exploring the use of factor models for forecasting the price of oil. In related work, Zagaglia (2010) reports some success in using a factor model in forecasting the nominal price of oil at short horizons, although his evaluation period only covers early 2003 to early 2008 given his data limitations, and it is unclear how sensitive the results would be to extending the evaluation period. An obvious concern is that there are no price reversals over the evaluation period, so any predictor experiencing sustained growth is likely to have some forecasting power. Moreover, we have shown in Section 5 that much simpler forecasting models appear capable of generating equally substantial reductions in the MSPE of the nominal price of oil at short horizons and do so for extended periods. The more important problem from an economic point of view, in any case, is forecasting the real price of oil. It seems

³⁹ A case in point is the persistent divergence of the Brent price and WTI price of crude oil in 2011. This unusual divergence can be explained by a combination of factors including (1) an excess supply of Canadian crude oil in the United States, given the lack of infrastructure for selling Canadian crude oil elsewhere, (2) a shortage of light sweet crude oil in the EU due to the recent Libyan production shortfall, and (3) increased demand for diesel fuel (caused in part by environmental policies) driving up the demand for light sweet crude oil in Europe.

unlikely that approximate factor models could be used to forecast the real price of oil. The variables that matter most for the determination of the real price of oil are global. Short of developing a comprehensive worldwide data set of real aggregates at monthly frequency, it is not clear whether there are enough predictors available for reliable real-time estimation of the factors. For example, drawing excessively on U.S. real aggregates as in Zagaglia (2010) is unlikely to be useful for forecasting the global price of oil for the reasons discussed in Section 4. Using a cross-section of data on energy prices, quantities, and other oil-market related indicators may be more promising, but almost half of the series used by Zagaglia are specific to the United States and unlikely to be representative of global markets.

15. CONCLUSIONS

Although there are a fair number of papers dealing with the problem of predicting the price of oil, it is difficult to reconcile the seemingly conflicting results in this literature. The problem is not only the precise definition of the oil price variable, but whether the price of oil is expressed in nominal or in real terms, what estimation and evaluation periods are chosen, how the forecast accuracy is evaluated, whether the conditional mean, conditional variance or conditional density is being forecast, whether the analysis is conducted in-sample or out-of-sample, whether the methods are parametric or non-parametric, and whether tests of statistical significance are provided or not. The most common problem in the literature is that results are sensitive to the choice of the sample period and vanish when the sample period is extended.

In this chapter, our objective has been to provide a benchmark based on data that include the recent collapse of the price of oil in late 2008 and its subsequent recovery. We started by discussing problems with combining data from the pre-1973 and post-1973 period, highlighting the need to discard the pre-1973 data because these data cannot be represented by standard time series models. We documented a structural break in the time series process of both the nominal and the real price of oil in late 1973. We also noted the presence of a structural break in the dynamic correlations between changes in the real price of oil and U.S. real GDP growth. This structural break invalidates predictive regressions based on data extending back further than 1973.

A natural starting point for our analysis was the question of whether changes in the price of oil are inherently unpredictable, as is sometimes claimed. We provided strong evidence that after 1973 changes in the nominal price of oil are predictable in population. The most successful predictors are recent percent changes in U.S. consumer prices and monetary aggregates as well as global non-oil industrial commodity prices. We also found strong evidence that after 1973 the real price of oil is predictable in population based on fluctuations in global real output. We illustrated how problems of omitted variables and of measurement can obscure this predictive relationship. We emphasized the importance

of accounting for structural changes in the composition of real output, of using measures with broad geographic coverage, and of using methods of detrending that can capture long swings in the demand for industrial commodities.

These results demonstrate that neither the nominal nor the real price of oil follows a random walk in population. Predictability in population, however, need not translate into out-of-sample forecast accuracy. One concern is that in small samples simple parsimonious forecasting models such as the no-change forecast often have lower MSPE than forecasts from larger-dimensional models suggested by economic theory. This may occur even if the large-dimensional model is correctly specified, provided the increase in the forecast variance from estimating the unknown parameters of the correctly specified model exceeds the reduction in the (squared) forecast bias from eliminating the model misspecification.

We provided evidence that this concern is unwarranted. We showed that suitably designed unrestricted vector autoregressive models estimated recursively on ex-post revised data tend to have lower out-of-sample MSPE than the no-change forecast of the real price of oil at horizons up to 12 months in some cases. The largest MSPE reductions occur at horizons 1 and 3. The MSPE reductions may be as high as 20%. The same models also have statistically significant directional accuracy at horizons up to 1 year. In addition, there is strong evidence that recursively estimated AR and ARMA models have lower MSPE than the no-change forecast, especially at horizons of 1 and 3 months. At horizons beyond 1 year, the no-change forecast of the real price of oil typically is the predictor with the lowest MSPE. These results are robust to the use of real time data.

Likewise, it is possible to forecast the nominal price of oil out of sample. There is strong evidence of statistically significant MSPE reductions in forecasting the nominal price of oil at horizons of 1 and 3 months based on recent percent changes in the price of non-oil industrial raw materials, for example. The reductions in the out-of-sample MSPE at the 3-month horizon are 22%. Similarly, information in recent changes in the dollar exchange rate of Canada and Australia helps reduce the MSPE of the forecast of the nominal price of oil at horizons up to 6 months. Both models also exhibit statistically significant directional accuracy. Finally, there is evidence that simply adjusting the no-change forecast for the real price of oil for expected inflation yields much more accurate forecasts of the nominal price of oil than the no-change forecast at horizons of several years. The same inflation-based model provides few if any accuracy gains at short or intermediate horizons, however.

More commonly used methods of forecasting the nominal price of oil based on the price of oil futures or the spread of the oil futures price relative to the spot price cannot be recommended. There is no reliable evidence that oil futures prices significantly lower the MSPE relative to the no-change forecast at short horizons, and long-term futures prices often cited by policymakers are distinctly less accurate than the no-change forecast. One possible explanation for the unexpectedly low out-of-sample accuracy of oil futures-based forecasts may be the presence of transaction costs impeding arbitrage. An alternative

forecasting strategy in which one uses the futures price only if the futures spread exceeds 5% in absolute terms and uses the spot price otherwise, yields MSPE reductions between 0% and 6% at short horizons (some of which are statistically significant), but performs much worse than the no-change forecast at longer horizons. Likewise professional and government forecasts of the nominal price of oil do not significantly improve on the no-change forecast, except in some cases in the short run, and can be much less accurate.

One of the main reasons for the importance that many macroeconomists attach to the price of oil is its perceived predictive power for U.S. real GDP. We assessed this predictive power based on a joint forecasting model for the price of oil and for domestic real activity. We showed that there are only small gains in using the price of oil in forecasting cumulative real GDP growth from VAR models. This finding is robust to whether the price of oil is specified in nominal or in real terms and whether it is treated as exogenous or endogenous. More importantly, bivariate linear autoregressive models fail to predict major economic downturns. One possible explanation of this forecast failure is that the predictive relationship is non-linear. We therefore evaluated and compared a wide range of non-linear joint forecasting models for the price of oil and real GDP growth. Except for the 3-year net oil price increase specification, we found no evidence at all of substantially improved forecast accuracy for real GDP growth. Even for the 3-year net increase model, the evidence was mixed at best. For example, we found no evidence that the nominal PPI 3-year net increase model is more accurate than linear models for real GDP growth at the one-quarter horizon. A multivariate generalization of the model proposed by [Hamilton \(2003, 2011\)](#) tended to provide MSPE gains of up to 20% relative to the AR(4) benchmark model at longer horizons. Even more accurate results were obtained with some alternative oil price series. All these forecasting successes, however, were driven mainly by the 2008/09 recession. Excluding that episode from the evaluation period, all nonlinear forecasting models were less accurate than the benchmark AR(4) model for real GDP growth at the one-quarter horizon, and only two of ten nonlinear models showed any improvement relative to the AR(4) model at the four-quarter horizon.

This evidence suggests that there is reason to be skeptical of the seeming forecasting success of many nonlinear models during the recent financial crisis. If these forecasts are to be believed, the financial crisis played a small role in the economic decline of 2008/09, which does not seem economically plausible. An alternative explanation is that the evaluation sample is too short for reliable inference. We observed that the net oil price increase model has a tendency to predict economic declines anytime the price of oil has increased substantially. Although such predictions repeatedly proved incorrect, most notably in 2005/06, the ability of some 3-year net increase models to forecast the extreme decline in U.S. real GDP in 2008/09 under quadratic loss more than compensates for earlier forecasting errors and hence accounts for their higher average out-of-sample forecast accuracy when the evaluation period includes the financial crisis.

We also discussed the use of structural forecasting models for the real price of oil. An important limitation of reduced-form forecasting models of the real price of oil from a policy point of view is that they do not allow the user to explore alternative hypothetical forecasting scenarios. We illustrated how recently developed structural vector autoregressive models of the global oil market not only generate quite accurate out-of-sample forecasts, but may be used to generate projections of how the oil price forecast would deviate from the unconditional baseline forecast, conditional on alternative economic scenarios such as a surge in speculative demand similar to previous historical episodes, a resurgence of the global business cycle, or increased U.S. oil production. The proposed method allows users to assess the risks associated with reduced-form oil price forecasts.

Finally, we showed that oil price volatility measures commonly used to characterize predictive densities for the price of oil are not adequate measures of the risks faced by market participants. We demonstrated how appropriate risk measures can be constructed. Those risk measures, however, are only as good as the underlying forecasting models. Although the best forecasting models are substantially more accurate at short horizons than the no-change forecast of the price of oil, it is fair to say that even the best available forecasting methods are far from accurate and substantial uncertainty remains about the evolution of the price of oil in nominal and in real terms.

ACKNOWLEDGMENTS

We thank Christiane Baumeister for providing access to the global industrial production data. We thank Domenico Giannone for providing the code generating the Bayesian VAR forecasts. We have benefited from discussions with Christiane Baumeister, Mike McCracken, James Hamilton, Ana María Herrera, Ryan Kellogg, Simone Manganello, and Keith Sill, as well as comments from two anonymous referees and the editors. We thank David Finer and William Wu for assisting us in collecting some of the data. The views in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Board of Governors of the Federal Reserve System or of the Bank of Canada or of any other person associated with the Federal Reserve System or with the Bank of Canada.

REFERENCES

- Allcott, H., Wozny N., 2011. Gasoline Prices, Fuel Economy, and the Energy Paradox. Mimeo, MIT.
- Almoguera, P.A., Douglas, C., Herrera, A.M., 2011. Testing for the cartel in OPEC: noncooperative collusion or just noncooperative? *Oxford Review of Economic Policy* 27, 144–168.
- Alquist, R., Kilian, L., 2010. What do we learn from the price of crude oil futures? *Journal of Applied Econometrics* 25, 539–573.
- Anatolyev, S., 2012. Inference in regression models with many regressors. *Journal of Econometrics* 170, 368–382.
- Anderson, S., Kellogg, R., Sallee, J., 2011. What Do Consumers Believe About the Price of Gasoline? Mimeo, Department of Economics, University of Michigan.
- Artzner, P., Delbaen, F., Eber, J.-M., Heath, D., 1999. Coherent measures of risk. *Mathematical Finance* 9, 203–228.

- 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 US economy. *Economic Inquiry* 46, 528–539.
- Balke, N.S., Brown, S.P.A., Yücel, M.K., 2002. Oil price shocks and the US economy: where does the asymmetry originate? *Energy Journal* 23, 27–52.
- Barsky, R.B., Kilian, L., 2002. Do we really know that oil caused the great stagflation? A monetary alternative. In: Bernanke, B.S., Rogoff, K. (Eds.), *NBER Macroeconomics Annual 2001*. MIT Press, Cambridge, MA, pp. 137–183.
- Basak, S., Shapiro, A., 2001. Value-at-risk based management: optimal policies and asset prices. *Review of Financial Studies* 14, 371–405.
- Baumeister, C., Kilian, L., 2012a. Real-time forecasts of the real price of oil. *Journal of Business and Economic Statistics* 30, 326–336.
- Baumeister, C., Kilian, L., 2012b. Real-Time Analysis of Oil Price Risks using Forecast Scenarios. Mimeo, Department of Economics, University of Michigan.
- Baumeister, C., Peersman, G., 2012. The role of time-varying price elasticities in accounting for volatility changes in the crude oil market. *Journal of Applied Econometrics*.
- Bernanke, B.S., 1983. Irreversibility, uncertainty, and cyclical investment. *Quarterly Journal of Economics* 98, 85–106.
- Bernanke, B.S., 2004. Oil and the Economy. Speech presented at Darton College, Albany, GA. <<http://www.federalreserve.gov/boarddocs/speeches/2004/20041021/default.htm>>.
- Beyer, A., Doornik, J.A., Hendry, D.F., 2001. Constructing historical euro-zone data. *Economic Journal* 111, 308–327.
- Bollerslev, T., Chou, R.Y., Kroner, K.F., 1992. ARCH modeling in finance. *Journal of Econometrics* 52, 5–59.
- Busse, M., Knittel, C., Zettelmeyer, F., 2011. Pain at the Pump: How Gasoline Prices Affect Automobile Purchasing. Mimeo, Northwestern University.
- Calhoun, G., 2011a. Hypothesis testing in linear regression when k/n is large. *Journal of Econometrics* 165, 163–174.
- Calhoun, G., 2011b. Out of Sample Comparisons of Overfit Models. Mimeo, Department of Economics, Iowa State University.
- Carlton, A.B., 2010. Oil Prices and Real-Time Output Growth. Mimeo, Department of Economics, University of Houston.
- Chen, Y.-C., Rogoff, K., Rossi, B., 2010. Can exchange rates forecast commodity prices? *Quarterly Journal of Economics* 125, 1145–1194.
- Clark, T.E., McCracken, M., 2001. Tests of equal predictive accuracy and encompassing for nested models. *Journal of Econometrics* 105, 85–101.
- Clark, T.E., McCracken, M., 2005a. Evaluating direct multistep forecasts. *Econometric Reviews* 24, 369–404.
- Clark, T.E., McCracken, M., 2005b. The power of tests of predictive ability in the presence of structural breaks. *Journal of Econometrics* 124, 1–31.
- Clark, T.E., McCracken, M., 2010. Nested Forecast Model Comparisons: A New Approach to Testing Equal Accuracy. Mimeo, Federal Reserve Bank of St. Louis.
- Clark, T.E., West, K.D., 2007. Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics* 138, 291–311.
- Cooley, T.F., LeRoy, S., 1985. Atheoretical macroeconomics: a critique. *Journal of Monetary Economics* 16, 283–308.
- Corradi, V., Swanson, N.R., 2002. A consistent test for nonlinear out of sample predictive accuracy. *Journal of Econometrics* 110, 353–381.
- Corradi, V., Swanson, N.R., 2007. Nonparametric bootstrap procedures for predictive inference based on recursive estimation schemes. *International Economic Review* 48, 67–109.
- Dargay, J.M., Gately, D., 2010. World oil demand's shift toward faster growing and less price-responsive products and regions. *Energy Policy* 38, 6261–6277.
- Davis, L.W., Kilian, L., 2011. The allocative cost of price ceilings in the US residential market for natural gas. *Journal of Political Economy* 119, 212–241.
- Diebold, F.X., Mariano, R.S., 1995. Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13, 253–263.

- Dolado, J.J., Lütkepohl, H., 1996. Making wald tests work for cointegrated VAR systems. *Econometric Reviews* 15, 369–386.
- Dvir, E., Rogoff, K., 2010. Three Epochs of Oil. Mimeo, Harvard University.
- Edelstein, P., Kilian, L., 2009. How sensitive are consumer expenditures to retail energy prices? *Journal of Monetary Economics* 56, 766–779.
- Elder, J., Serletis, A., 2010. Oil price uncertainty. *Journal of Money, Credit and Banking* 42, 1138–1159.
- Elliott, G., Timmermann, A., 2008. Economic forecasting. *Journal of Economic Literature* 46, 3–56.
- Engle, R.F., Brownlees, C.T., 2010. Volatility, Correlation and Tails for Systemic Risk Measurement. Stern School of Business. Mimeo, New York University.
- Farrell, A.E., Brandt, A.R., 2006. Risks of the oil transition. *Environmental Research Letters* 1, 1–6.
- Fishburn, P.C., 1977. Mean-risk analysis with risk associated with below-target returns. *American Economic Review* 67, 116–26.
- Giannone, D., Lenza, M., Primiceri, G., 2010. Prior Selection for Vector Autoregressions. Department of Economics. Mimeo, Free University of Brussels.
- Gillman, M., Nakov, A., 2009. Monetary effects on nominal oil prices. *North American Journal of Economics and Finance* 20, 239–254.
- Goldberg, P., 1998. The Effects of the corporate average fuel economy standards in the US. *Journal of Industrial Economics* 46, 1–33.
- Gramlich, E.M., 2004. Oil Shocks and Monetary Policy. Annual Economic Luncheon, Federal Reserve Bank of Kansas City, Kansas City, Missouri.
- Green, E.J., Porter, R.H., 1984. Noncooperative collusion under imperfect price information. *Econometrica* 52, 87–100.
- Greenspan, A., 2004a. Energy Remarks by Chairman Alan Greenspan Before the Center for Strategic and International Studies. Washington, DC. <<http://www.federalreserve.gov/boarddocs/speeches/2004/20040427/default.htm>>.
- Greenspan, A., 2004b. Oil, Speech presented at the National Italian American Foundation. Washington, DC. <<http://www.federalreserve.gov/boarddocs/speeches/2004/200410152/default.htm>>.
- Hamilton, J.D., 1983. Oil and the macroeconomy since world war II. *Journal of Political Economy* 91, 228–248.
- Hamilton, J.D., 1985. Historical causes of postwar oil shocks and recessions. *Energy Journal* 6, 97–116.
- Hamilton, J.D., 1994. *Time Series Analysis*. Princeton University Press, Princeton, NJ.
- 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., 2009. Causes and consequences of the oil shock of 2007–08. *Brookings Papers on Economic Activity*, vol. 1. Springer, pp. 215–261.
- Hamilton, J.D., 2011. Nonlinearities and the macroeconomic effects of oil prices. *Macroeconomic Dynamics* 15, 364–378.
- Hamilton, J.D., Herrera, A.M., 2004. Oil shocks and aggregate economic behavior: the role of monetary policy. *Journal of Money, Credit and Banking* 36, 265–286.
- Hendry, D., 2006. Robustifying forecasts from equilibrium-correction systems. *Journal of Econometrics* 135, 399–426.
- Herrera, A.M., Lagalo, L.G., Wada, T., 2011. Oil price shocks and industrial production: is the relationship linear? *Macroeconomic Dynamics* 15, 472–497.
- Holthausen, D.M., 1981. A risk–return model with risk and return measured in deviations from target return. *American Economic Review* 71, 182–88.
- Hotelling, H., 1931. The economics of exhaustible resources. *Journal of Political Economy* 39, 137–175.
- Inoue, A., Kilian, L., 2004a. In-sample or out-of-sample tests of predictability: which one should we use? *Econometric Reviews* 23, 371–402.
- Inoue, A., Kilian, L., 2004b. Bagging Time Series Models. CEPR Discussion Paper No. 4333.
- Inoue, A., Kilian, L., 2006. On the selection of forecasting models. *Journal of Econometrics* 130, 273–306.
- International Monetary Fund, 2005. *World Economic Outlook*. Washington, DC.
- International Monetary Fund, 2007. *World Economic Outlook*. Washington, DC.
- Isserlis, L., 1938. Tramp shipping cargoes and freights. *Journal of the Royal Statistical Society* 101 (1), 53–134.

- Kahn, J.A., 1986. Gasoline prices and the used automobile market: a rational expectations asset price approach. *Quarterly Journal of Economics* 101, 323–340.
- Kellogg, R., 2010. The Effect of Uncertainty on Investment: Evidence from Texas Oil Drilling. Mimeo, Department of Economics, University of Michigan.
- Kilian, L., 1999. Exchange rates and monetary fundamentals: what do we learn from long-horizon regressions? *Journal of Applied Econometrics* 14, 491–510.
- Kilian, L., 2008a. The economic effects of energy price shocks. *Journal of Economic Literature* 46 (4), 871–909.
- Kilian, L., 2008b. Exogenous oil supply shocks: how big are they and how much do they matter for the US economy? *Review of Economics and Statistics* 90, 216–240.
- Kilian, L., 2009a. 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., 2009b. In: James, D., Hamilton (Eds.), *Comment on Causes and Consequences of the Oil Shock of 2007–08*. *Brookings Papers on Economic Activity*, vol. 1. Springer, pp. 267–278.
- Kilian, L., 2010. Explaining fluctuations in US gasoline prices: a joint model of the global crude oil market and the US retail gasoline market. *Energy Journal* 31, 87–104.
- Knetsch, T.A., 2007. Forecasting the price of oil via convenience yield predictions. *Journal of Forecasting* 26, 527–549.
- Kilian, L., Hicks, B., forthcoming. Did unexpectedly strong economic growth cause the oil price shock of 2003–2008? *Journal of Forecasting*.
- Kilian, L., Manganelli, S., 2007. Quantifying the risk of deflation. *Journal of Money, Credit and Banking* 39, 561–590.
- 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., Murphy, D., 2013. The Role of Inventories and Speculative Trading in the Global Market for Crude Oil. Mimeo, University of Michigan.
- Kilian, L., Rebucci, A., Spatafora, N., 2009. Oil shocks and external balances. *Journal of International Economics* 77, 181–194.
- Kilian, L., Vega, C., 2010. Do energy prices respond to US macroeconomic news? A test of the hypothesis of predetermined energy prices. *Review of Economics and Statistics* 93, 660–671.
- Kilian, L., Vigfusson, R.J., 2011a. Are the responses of the US economy asymmetric in energy price increases and decreases? *Quantitative Economics* 2, 419–453.
- Kilian, L., Vigfusson, R.J., 2011b. Nonlinearities in the oil price-output relationship. *Macroeconomic Dynamics* 15, 337–363.
- Kilian, L., Vigfusson, R.J., 2013. Do oil price increases help forecast US real GDP? The role of nonlinearities and asymmetries. *Journal of Business and Economic Statistics* 31, 78–93.
- Koop, G., Pesaran, M.H., Potter, S.M., 1996. Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics* 74, 119–147.
- Leamer, E.E., 1978. *Specification Searches: Ad hoc Inference with Nonexperimental Data*. Wiley-Interscience, New York.
- Lütkepohl, H., 1982. Non-causality due to omitted variables. *Journal of Econometrics* 19, 367–378.
- Machina, M.J., Rothschild, M., 1987. Risk. In: Eatwell, J., Millgate, M., Newman, P. (Eds.), *The New Palgrave Dictionary of Economics*. MacMillan, London, pp. 203–205.
- 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.
- Peck, A.E., 1985. Economic role of traditional commodity futures markets. In: Peck, A.E. (Ed.), *Futures Markets: Their Economic Role*. American Enterprise Institute for Public Policy Research, Washington, DC, pp. 1–81.
- Pesaran, M.H., Timmermann, A., 2009. Testing dependence among serially correlated multicategory variables. *Journal of the American Statistical Association* 104, 325–337.
- Pindyck, R.S., 1991. Irreversibility, uncertainty and investment. *Journal of Economic Literature* 29, 1110–1148.

- Ramey, V.A., Vine, D.J., 2010. Oil, Automobiles, and the US economy: how much have things really changed. *NBER Macroeconomics Annual* 25, 333–368.
- Ravazzolo, F., Rothman, P., 2013. Oil and US GDP: a real time out-of-sample examination. *Journal of Money, Credit and Banking* 45, 449–463.
- Ravn, M.O., Uhlig, H., 2002. On adjusting the Hodrick-Prescott filter for the frequency of observations. *Review of Economics and Statistics* 84, 371–380.
- Sims, C.A., Stock, J.H., Watson, M.W., 1990. Inference in linear time series models with some unit roots. *Econometrica* 58, 113–144.
- Skeet, I., 1988. *OPEC: Twenty-Five Years of Prices and Politics*. Cambridge University Press, Cambridge.
- Smith, J.L., 2005. Inscrutable OPEC? Behavioral tests of the cartel hypothesis. *Energy Journal* 26, 51–82.
- Stock, J.H., Watson, M.W., 1999. Forecasting inflation. *Journal of Monetary Economics* 44, 293–335.
- Svensson, L.E.O., 2005. *Oil Prices and ECB Monetary Policy*. Briefing Paper for the Committee on Economic and Monetary Affairs of the European Parliament.
- Tinbergen, J., 1959. Tonnage and freight. In: *Jan Tinbergen Selected Papers*. North Holland, Amsterdam, pp. 93–111.
- Verleger, Jr., P.K., 1982. The determinants of official OPEC crude prices. *Review of Economics and Statistics* 64, 177–183.
- Verleger, Jr., P.K., 2011. The margin, currency, and the price of oil. *Business Economics* 46, 71–82.
- Waggoner, D.F., Zha, T., 1999. Conditional forecasts in dynamic multivariate models. *Review of Economics and Statistics* 81, 639–651.
- Working, H., 1942. Quotations on commodity futures as price forecasts. *Econometrica* 16, 39–52.
- Wu, T., McCallum, A., 2005. Do Oil Futures Prices Help Predict Future Oil Prices? *Federal Reserve Bank of San Francisco Economic Letter*, pp. 2005–38.
- Zagaglia, P., 2010. Macroeconomic factors and oil futures prices: a data-rich model. *Energy Economics* 32, 409–417.