NBER WORKING PAPER SERIES

# FORECASTING NATURAL GAS PRICES IN REAL TIME

Christiane Baumeister
Florian Huber
Thomas K. Lee
Francesco Ravazzolo

Working Paper 33156
http://www.nber.org/papers/w33156

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2024

We thank Joe Bozeman, Pedro Gete, Jim Hamilton, and participants at the EIA seminar for helpful suggestions. Ray Bae and Pascal Frank provided excellent research assistance. The work of Christiane Baumeister was supported by grant DE-EI0003268 from the U.S. Energy Information Administration. Francesco Ravazzolo acknowledges financial support from the European Union NextGenerationEU program (Piano Nazionale di Ripresa e Resilienza (PNRR) Missione 4 Componente 2, Investimento 1.5 D.D. 1058 23/06/2022, ECS-00000043). The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the U.S. Energy Information Administration or the National Bureau of Economic Research.

Forecasting Natural Gas Prices in Real Time
Christiane Baumeister, Florian Huber, Thomas K. Lee, and Francesco Ravazzolo
NBER Working Paper No. 33156
November 2024
JEL No. C11, C32, C52, Q41, Q47

**ABSTRACT**

This paper provides a comprehensive analysis of the forecastability of the real price of natural gas in the United States at the monthly frequency considering a universe of models that differ in their complexity and economic content. Our key finding is that considerable reductions in mean-squared prediction error relative to a random walk benchmark can be achieved in real time for forecast horizons of up to two years. A particularly promising model is a six-variable Bayesian vector autoregressive model that includes the fundamental determinants of the supply and demand for natural gas. To capture real-time data constraints of these and other predictor variables, we assemble a rich database of historical vintages from multiple sources. We also compare our model-based forecasts to readily available model-free forecasts provided by experts and futures markets. Given that no single forecasting method dominates all others, we explore the usefulness of pooling forecasts and find that combining forecasts from individual models selected in real time based on their most recent performance delivers the most accurate forecasts.

Christiane Baumeister
Department of Economics
University of Notre Dame
3028 Jenkins Nanovic Hall
Notre Dame, IN 46556
and CEPR
and also NBER
cbaumeis@nd.edu

Florian Huber
University of Salzburg
Mönchsberg 2A
5020 Salzburg
Austria
florian.huber@plus.ac.at

Thomas K. Lee
U.S. Energy Information Administration
1000 Independence Ave., SW
Washington, DC 20585
Thomas.Lee@eia.gov

Francesco Ravazzolo
BI Norwegian Business School
Nydalsveien 37
Oslo 0484
Norway
francesco.ravazzolo@unibz.it

A real-time database for the US natural gas market is available at
https://drive.google.com/uc?export=download&id=1a5orTP7zcJcj9cHYS3cmojpQNx0lsuwe

# 1 Introduction

Natural gas is an important primary source of energy that is used across all sectors in the economy in varying amounts: it is one of the most popular fuels for residential and commercial heating, it plays a major role in electricity generation, and it has a myriad of industrial uses such as feedstock for fertilizer, hydrogen, and other petrochemical products, and as heat source for steel, glass, and paper manufacturing. Natural gas is also considered the "cleanest" among the traditional fossil fuels given that it produces less carbon dioxide when it is burned compared to coal or petroleum and thus it takes center stage in reducing pollution and fostering the energy transition. Given that natural gas plays such a critical role in today's energy system, price developments in the natural gas market matter significantly for a range of energy and environmental policies since they directly impact the costs of power generation and input factors firms face and the electricity and gas bills households pay, which in turn will influence firms' and consumers' energy choices. They also drive policy decisions regarding the energy mix, the promotion of renewables, the supply of affordable energy, infrastructure development, tax subsidies, and energy security. Thus, to design and implement such policies it is key for federal and state governments as well as other stakeholders such as environnmental organizations, utility and clean power companies, regulatory agencies, and natural gas producers to have access to reliable forecasts of natural gas prices.

There exist few sources that can be tapped for obtaining regular forecasts of the price of natural gas. The U.S. Energy Information Administration releases monthly forecasts of nominal natural gas prices in its *Short-Term Energy Outlook* for horizons varying between one and two years ahead. Besides these institutional forecasts, market-based forecasts can be constructed from daily trading prices of futures contracts that deliver natural gas for several years into the future. Yet another strategy is to assume that the price over the forecast horizon is the same as the one observed today. However, little is know about the accuracy of these readily available forecasts.

What is even more surprising is that, despite the growing importance of natural gas as a transition fuel, there have been limited efforts by academic researchers to date to explore alternative, model-based foreasting approaches. One exception is Ferrari, Ravazzolo, and Vespignani (2021) who forecast a nominal natural gas price index, which is a weighted average of natural gas prices in the US, Europe, and Japan, at the quarterly frequency using a large dataset with around 200 variables for the 33 largest economies in the world whose information is summarized by a sparse dynamic factor model. They document some out-of-sample forecasting success relative to a random walk. Another recent contribution is Gao, Hou, and Nguyen (2020) who examine the forecasting ability of monthly autoregressive models with time-varying parameters and stochastic volatility, with Markov-switching dynamics, and a hybrid of the two for nominal natural gas prices across different regional markets; none of these models consistently outperforms their autoregressive benchmark.

In this paper, we provide a systematic evaluation of the forecastability of the real price of natural gas in the US against a common benchmark, considering a broad range of modeling approaches that exploit insights from economic theory, market structure, and time-series properties and that differ in their complexity and information content. We focus on the US given that competitive spot markets have been in place since the early 1990s, while in other countries natural gas prices are determined by long-term contracts that are linked to oil or other non-market-pricing mechanisms (see, for example, Halser, Paraschiv, and Russo, 2023; Hupka, Popova, Simkins, and Lee, 2023; Stock and Zaragoza-Watkins, 2024).[1] Our objective is to forecast the monthly average level of the real price of natural gas for horizons up to 24 months in a real-time setting.

We consider four classes of models. First, we investigate the predictive content of standard univariate models where forecasts depend only on past price dynamics. In particular, we study the accuracy of autoregressive, autoregressive-moving average, and exponential smoothing forecasts. Second, motivated by recently developed structural models for the U.S. natural gas market (see Arora and Lieskovsky, 2014; Wiggins and Etienne, 2017; Rubaszek, Szafranek, and Uddin, 2021; Winkler, 2023), we explore the predictive power of various supply-side and demand-side determinants of the real price of natural gas which we combine in the form of a vector autoregressive (VAR) model. Relevant fundamental drivers include natural gas production and consumption, the number of active drilling rigs for natural gas wells, the amount of gas in underground storage facilities, measures of meteorological conditions, and domestic economic activity indicators. Given that most of these predictor variables are subject to real-time data constraints in the form of publication lags and revisions to preliminary data releases, one of the contributions of this paper is to assemble an extensive real-time database of monthly vintages from several different sources. Accounting for these real-time aspects is crucial to avoid look-ahead bias when assessing the forecasting performance of our economic models. Mindful that the dynamic relationships in the domestic natural gas market might be subject to structural change and unusual episodes such as extreme weather events and infrastructure bottlenecks over the course of our evaluation period, we examine whether introducing time-varying volatility impacts the accuracy of natural gas price forecasts. Third, we consider several energy price models that are derived from the interaction between natural gas and oil markets. Because of fuel substitutability in the power sector, U.S. natural gas prices are closely tied to the evolution of oil prices, with both prices moving together in the long run (see, e.g., Brown and Yücel, 2008; Hartley et al., 2008). We quantify the predictive ability of price spread models as well as bivariate VAR models with and without cointegration restrictions imposed. Given that over time the oil-gas nexus has been affected by technological advances in power generation, regulatory changes, the shale gas revolution, and shortages of transportation facilities (see, e.g., Ramberg and Parsons, 2012; Stock and Zaragoza-Watkins, 2024), we also examine the benefit of allowing for regime switches in the relationship between the two fuel prices. Fourth, we propose a hybrid model that accommodates arbitrage between natural gas and oil prices as well as fundamental drivers of the natural gas market. This integrated-markets model is the most comprehensive model since it contains both own-market and cross-market determinants of natural gas prices.

[1] For example, in 2005, only 15% of the natural gas purchased in Europe on wholesale markets was determined by spot prices according to the International Gas Union. While this share increased to 64% by 2015 and to 81% by 2020, oil-indexed contracts are still a distinctive feature of the European gas market. By contrast, 99% of natural gas in the US is sold at spot prices.

Our analysis reveals that several forecasting methods perform quite well in real time compared to the no-change benchmark for horizons up to 24 months. At the nearest horizon, using the most recent daily observation of the natural gas price to forecast the average price next month delivers the most accurate forecast, but this approach is superseded at the three- and six-month horizons by more precise forecasts based on futures prices and various economic models of the natural gas market that differ in the number of predictor variables and how these are measured but have in common parsimonious dynamics of order one. Forecasting the real price of natural gas at intermediate horizons, ranging from nine to 15 months out, is most successful with futures prices, exponential smoothing, and an economic model that includes the full set of fundamental drivers. Exponential smoothing remains competitive at the longest horizons, but adding stochastic volatility to some of the economic models also achieves substantial improvements in forecasting performance. While these are the most promising models across horizons, the differences in average performance with the next best tier of models are often small.

These findings highlight that there is no clear winner in this forecasting horserace which begs the question of which model to rely on to produce accurate real-time out-of-sample forecasts of the real price of natural gas going forward. We also show that the predictive content of individual models changes considerably over time. Taken together, this evidence suggests that combining forecasts from a larger set of models might be beneficial. We conclude that a pooled forecast that assigns equal weight to the most successful individual forecasting models selected in real time yields impressive accuracy gains at all horizons dominating all other individual forecasting approaches.

The plan for the paper is as follows. Section 2 describes the forecasting environment and evaluates the performance of three model-free forecasting approaches. Section 3 proposes a diverse set of forecasting models and explores their relative merits. Section 4 conducts a joint assessment of the predictive accuracy of the entire model space and examines the promise of forecast combinations. Section 5 offers some concluding remarks.

# 2 The Forecasting Environment

## 2.1 Forecast Target and Evaluation Metric

Our goal is to forecast the average monthly value of the real Henry Hub spot price of natural gas which serves as the benchmark price for the entire North American natural gas market and parts of the global liquefied natural gas (LNG) market. Daily and monthly spot prices have been published by the U.S. Energy Information Administration (EIA) since January 1997. Monthly data for Henry Hub natural gas prices going back to November 1993 were released by the *Wall Street Journal* (WSJ) but their publication was discontinued in March 2014. This series is made available in the FRED database of the Federal Reserve Bank of St. Louis. Before that, the natural gas wellhead price, which is the average sales price across several trading hubs in the Southwestern United States from which Henry Hub emerged as the leading pricing point in the early 1990s when it became the official delivery location for natural gas futures contracts traded on the New York Mercantile Exchange (NYMEX), was used as the reference price. This monthly series is provided by the EIA starting in January 1976 but ceased to be reported at the end of 2012. We construct a historical natural gas price series based on the wellhead price for the period 1976M1 to 1993M10, the WSJ Henry Hub price for 1993M11 to 1996M12, and the EIA spot price from 1997M1 to 2024M2.[2]

Throughout the analysis, we mimic as closely as possible the situation of a real-life forecaster who can rely only on the information available at the point in time the forecast is generated. This means working with preliminary data that are subject to revisions later on and taking delays in data releases into account to accurately reflect real-time data constraints when assessing the out-of-sample performance of forecasting methods. Since our focus is on the real price of natural gas, we deflate the price by the U.S. consumer price index (CPI) in real time. For this purpose, we update the real-time vintages of the monthly seasonally adjusted U.S. consumer price index for all urban consumers originally compiled by Baumeister and Kilian (2012) up to February 2024. These real-time data are obtained from *Economic Indicators* published by the Council of Economic Advisers and made available in the FRASER database of the Federal Reserve Bank of St. Louis. While the nominal spot price is available in real time, CPI data are published with a one-month lag; we nowcast the missing observation for each vintage using the past average inflation rate.

We produce out-of-sample forecasts for monthly horizons $h$ of up to two years. Model-based forecasts are obtained by recursively re-estimating the model at each forecast origin $t$ based on data contained in the real-time vintage $t$. The evaluation period starts in February 1997 determined by the EIA's earliest reporting of the Henry Hub spot price. Thus, the initial recursive estimation window runs from 1976M1 to 1997M1 using data from the January 1997 vintage. After generating $h$-step-ahead forecasts, we move to the February 1997 data vintage which adds one observation to our sample. We repeat estimation and forecasting for all the vintages until February 2024 which is the last available vintage and thus the endpoint of our evaluation period.

Since our interest centers on point forecasts of the real price of natural gas, we use the mean-squared prediction error (MSPE) as our metric to assess the forecasting ability of alternative models. The real-time forecasts are evaluated against the final release of the real price of natural gas which we take to be the values in the May 2024 vintage of data. The MSPE results of all our candidate models are normalized relative to the monthly random walk without drift, which is the established benchmark in the energy price forecasting literature (see, e.g., Hamilton, 2009; Baumeister, Kilian, and Lee, 2017; Ferrari, Ravazzolo, and Vespignani, 2021; Baumeister, Korobilis, and Lee, 2022; Baumeister, Huber, and Marcellino, 2024). An MSPE ratio below one indicates an improvement in accuracy, while a value above one indicates a deterioration in accuracy.

[2]The wellhead price is quoted in dollars per thousand cubic feet, whereas the Henry Hub spot price is quoted in dollars per million British thermal units (Btu). We convert the wellhead price to dollars per million Btu by dividing it by 1.038.

## 2.2 Model-Free Forecasts

A question that has been raised recently is whether the monthly random walk is the appropriate benchmark for forecast comparisons of energy prices. In the context of oil price forecasting, Ellwanger and Snudden (2023) argue that the end-of-month price is a stricter benchmark than the average price over the month even if the goal is to predict the monthly average price. This issue was first examined by Baumeister and Kilian (2014) in relation to forecasting the quarterly price of oil where they show that the most recent monthly observation is indeed tougher to beat than the most recent price averaged over the quarter. Ellwanger and Snudden (2023) arrive at the same conclusion when comparing the last-day-of-the-month no-change forecast with the monthly average no-change forecast.

We start by addressing this question for natural gas price forecasts. The first column of Table 1 presents the average MSPE ratios for forecasts generated using the closing price on the last trading day of each month deflated by the nowcasted CPI relative to the monthly average price deflated by the same value for horizons $h = 1, 3, 6, 9, 12, 15, 18, 21, 24$ months ahead. Using real end-of-month prices yields a substantial gain in forecast accuracy at the one-month horizon with an MSPE reduction of 33%. This forecasting success is not entirely surprising given that the last observed daily price contains more recent information and thus is "closer" to the monthly average of the subsequent month which explains its superior performance one month ahead; however, this informational advantage quickly dissipates. While there is still a small improvement of 2% at the three-month horizon, from $h = 6$ onward the end-of-month no-change forecast is dominated by the monthly no-change benchmark. In sum, for the forecast horizons that we are interested in, the monthly random walk remains the relevant benchmark.

A potentially valuable and easily accessible source for forecasting natural gas prices is the futures market. Since futures contracts make it possible to lock in a price today for delivery of a fixed quantity of natural gas at a specified date in the future, the market price of futures contracts with different maturities can be viewed as the aggregated expected value of the spot price at expiry (see Baumeister, 2023). In other words, the prices of futures should have predictive content for future spot prices (see Chinn and Coibion, 2014). As in Baumeister and Kilian (2012, 2014), we produce forecasts using the following futures-spot spread relation:

$$R_{t+h|t}^{HH} = R_t^{HH} \left(1 + f_t^{HH,h} - s_t^{HH} - E_t(\pi_{t+h})\right)$$

where $R_t^{HH}$ denotes the average level of the real natural gas price in month $t$, $f_t^{HH,h}$ denotes the log of the current price for a natural gas futures contract deliverable at Henry Hub with maturity $h$, $s_t^{HH}$ denotes the log of the Henry Hub spot price, and $E_t(\pi_{t+h})$ denotes the expected inflation rate over the next $h$ months which we approximate with average inflation over a rolling window of ten years.[3] Column 3 of Table 1 shows that market-based forecasts beat the monthly no-change forecast

[3] We use the monthly average of daily prices for natural gas futures contracts with maturities of up to 24 months obtained from Bloomberg.

at all horizons. Futures perform best in the medium term for $h = 9, 12$, and 15 with accuracy gains between 19% and 24%, while MSPE reductions for short and long horizons range from 10% to 16%. For six out of the nine forecast horizons, improvements are statistically significant according to the Diebold and Mariano (1995) test for equal MSPE.

Another model-free approach is to rely on existing expert forecasts. The EIA regularly publishes monthly forecasts of several types of nominal natural gas prices in its *Short-Term Energy Outlook* (*STEO*). We compiled the real-time forecasts for the Henry Hub spot price from past issues of the *STEO* which were reported for the first time in August 2005. The forecast horizon varies across issues from a minimum of 12 months ahead to a maximum of 23 months ahead. These features restrict both the sample and the horizons over which we can evaluate these expert forecasts. We deflate the nominal price forecasts with the same measure of expected inflation as for the futures-spot spread model. The last column of Table 1 reveals that expert forecasts result in considerable losses compared to the no-change forecast one and three months ahead, but they greatly outperform the benchmark for subsequent horizons with statistically significant gains in accuracy of 35% at $h = 9$ and 38% at $h = 12$.

To put these results into perspective, we also report the MSPE ratios for the end-of-month random walk and the market-based forecasts for the same shorter evaluation period in columns 2 and 4, respectively. This not only enables a direct comparison with the EIA expert forecasts, but also speaks to the robustness of the forecasting performance over the more recent period and its stability over time. A similar pattern emerges for both forecasts. As before, accuracy gains of the last-day-of-the-month no-change forecast do not extend beyond three months. While the short-run MSPE reductions are even more impressive, these improvements vanish quickly as the forecast horizon lengthens. The futures market offers better forecasts at the longest horizons with additional gains of 4% at $h = 21$ and of 10% at $h = 24$ compared to the longer evaluation sample, but loses its edge at intermediate horizons with a relative deterioration of 5% on average.

The evidence so far suggests that the simple end-of-month random walk is only useful for very near-term forecasts and market-based forecasts are in the lead for long horizons, whereas the EIA experts have a clear advantage at medium-term forecast horizons (absent the possibility to evaluate their performance beyond one year). While all these model-free approaches yield readily available forecasts of the real price of natural gas that can be used in policy- and decision-making, it remains unclear what drives those forecasts. This begs the question of whether we can improve upon existing market-based and expert forecasts by designing alternative forecasting models for which we choose the determinants.

## 3 A Universe of Models to Forecast the Real Price of Natural Gas

We explore the relative forecasting performance of a diverse set of candidate models that differ in information set and economic content. In Section 3.1, we examine several simple time-series models that derive information solely based on past price dynamics. In Section 3.2, we turn to economic models of the natural gas market that feature the key determinants of the real price of natural gas on both the supply and the demand side which we measure drawing on our newly-assembled real-time database. In Section 3.3, we consider forecasting approaches based on the potentially time-varying link between natural gas and crude oil markets that finds its origin in the (imperfect) substitutability between these two energy sources. In Section 3.4, we jointly model own- and cross-market fundamentals. The regression models in this section include the real natural gas price either in logs or in growth rates for estimation. Given that our loss function is specified in levels, we exponentiate the log forecasts and accumulate and exponentiate the growth forecasts to get forecasts of the level of the real price of natural gas for evaluation purposes.

## 3.1 Models of Past Price Dynamics

It is well known that parsimonious time-series models that exploit the dynamic relationship of the current price with its own past often deliver quite accurate forecasts. A useful starting point is the ARMA$(1,1)$ model for the real price of natural gas in logs which we estimate numerically by Gaussian maximum likelihood methods. Conditional on the estimated coefficients and the most recent observations, we iterate the model forward to produce $h$-step-ahead forecasts for the log of the real price of natural gas and then convert it to levels. The first column of Table 2 shows that this simple ARMA$(1,1)$ model outperforms the no-change forecast at all horizons except for the shortest one, but yields lower MSPE reductions compared to the futures-spread-based forecast, in particular for intermediate horizons. Since we cannot rule out a priori that the log of the real price of natural gas is a unit root process, we consider an MA(1) model in percent changes, thus an IMA(1), as an alternative specification. Column 2 reveals that imposing a unit root on the ARMA$(1,1)$ process produces MSPE ratios well above one across all horizons, indicating that this model does much worse than the random walk. This result suggests that the process is not well described by a unit root process or a near unit root process. However, relaxing this constraint by estimating an ARIMA$(1,1)$ model does not alter the message (see column 3).

Another standard choice is to use purely autoregressive AR$(p)$ forecasting models which can alternatively be estimated by unrestricted least squares or by Bayesian methods that shrink unconstrained models toward a parsimonious benchmark. The latter have been shown to pay off in highly parameterized models where overfitting is a concern (see, e.g., Doan, Litterman, and Sims, 1984; Huber and Feldkircher, 2019). As in Baumeister and Kilian (2012), we rely on the data-driven approach of Giannone, Lenza, and Primiceri (2015) for selecting the optimal degree of shrinkage in real time based on the marginal data density. We set the lag length $p = 12$ as is common for monthly data. Column 4 shows that the AR(12) model produces hardly any reductions in MSPEs compared to the no-change forecast except at horizon 24 where we find a marginal improvement. Interestingly, while Bayesian shrinkage achieves some accuracy gains over the random walk benchmark from horizon $h = 6$ onward (see column 5), the BAR(12) model is not competitive with the ARMA$(1,1)$ at any horizon. An alternative strategy to determine the lag order is to rely on the

Akaike Information Criterion (see, e.g., Marcellino, Stock, and Watson, 2006), allowing the number of lags to be optimally chosen for each data vintage. To foster parsimony, we impose an upper limit of six autoregressive lags. Columns 6 and 7 of Table 2 indicate that this approach improves considerably on the (B)AR models with more lags and makes the forecasts competitive with the ARMA(1, 1) model for horizons up to one year and even better thereafter with MSPE reductions of 15-17% relative to the random walk. What is noteworthy is that once we use the AIC to select the lag length, there is little difference between the AR(AIC) and BAR(AIC) forecasts. In other words, Bayesian shrinkage does not play much of a role. Reducing the lag length to $p = 1$ helps outperform the no-change forecast at short horizons, but at the cost of losing some forecast accuracy at longer horizons relative to their AIC counterparts (see columns 8 and 9).

A very different forecasting method is recursive exponential smoothing which uses a weighted average of past realizations to predict future values with exponentially decaying weights for observations further in the past, allowing the forecast to adapt smoothly. This approach is suitable for data that are not trending over time. Given that the log level of the real price of natural gas has no discernible trend, it is an obvious candidate for the application of exponential smoothing. Following Baumeister, Kilian, and Lee (2017), we set the smoothing parameter to 0.8 which implies a moderate degree of smoothing in line with other economic applications (see, e.g., Faust and Wright, 2013). The last column of Table 2 shows that exponential smoothing performs very poorly at horizons 1 and 3, but yields substantial accuracy gains as the forecast horizon lengthens. Beyond the one-year horizon, it dominates market-based forecasts with statistically significant MSPE reductions of around 20%.

## 3.2 Economic Models of the Natural Gas Market

The price formation in the natural gas market results from the interaction of standard supply and demand forces. This suggests moving beyond modeling the dynamics of the equilibrium outcome since additional information can be derived from the fundamental determinants of the real price of natural gas suggested by economic theory and market structure. In Section 3.2.1, we develop a new real-time dataset that covers economically-motivated predictor variables that provide the basis for a rich economic model of the domestic natural gas market, which in Section 3.2.2 we estimate in the form of a vector autoregression and evaluate for the purpose of out-of-sample forecasting.

### 3.2.1 A Real-Time Dataset of Economic Determinants of U.S. Natural Gas Prices

We construct a comprehensive real-time database that contains variables that according to economic theory should have predictive power for the future path of natural gas prices and that allows us to account for publication delays as well as subsequent revisions in some of these data when assessing the accuracy of our forecasting models. The dataset is at the monthly frequency and has been assembled from a variety of sources. It consists of vintages from January 1991 to February 2024, each covering data going back to January 1973 where available.[4] Table 1A in the appendix provides an overview of the components of the real-time database, the data sources, the start date of vintages and data as well as lags in data releases and nowcasting rules to fill those gaps.

*Natural Gas Market Fundamentals.* The amount of natural gas produced domestically is one of the fundamental factors for price setting in the national natural gas market. The standard measure is dry natural gas production which is the quantity of natural gas extracted from natural gas and crude oil wells purified from other compounds and of high enough quality to be eligible for pipeline transportation. An alternative measure is marketed natural gas withdrawals which not only contain the consumer-grade dry natural gas but also other natural gas liquids. These "wet" components tend to be sold separately which is why we consider this second production concept less suitable to respresent the supply side of the market but the quantitative difference is relatively small. U.S. natural gas rig counts which measure the number of active drilling rigs employed in the search for natural gas are an indicator for the intensity of resource development and thus trends in future production. It stands to reason that movements in rig counts affect expectations about the future balance of supply and demand thereby inducing price swings. Another important price determinant is the volume of natural gas held in underground storage facilities since injections into and withdrawals from stockpiles allow to balance the more seasonal demand with the more steady supply. Natural gas inventories are measured as working natural gas in underground storage which is the flow component that adjusts to market conditions and seasonal factors. Price developments will also be influenced by the demand behavior of multiple end users. The quantity of natural gas delivered to households, manufacturing industries, electric power plants, commercial entities, and the transportation sector is measured by total natural gas consumption. We also consider residential consumption separately since households' demand patterns differ from those of other sectors given their limited substitution possibilities which can exacerbate changes in natural gas prices.

*Weather Conditions.* A more direct measure for consumption are variations in temperature and other weather phenomena since they are important drivers of heating demand and demand for air conditioning in residential homes and commercial buildings which account for roughly one third of natural gas use. For example, colder than usual winters typically lead to increased heating demand and low winds during the summer lead to the substitution from wind-generated to gas-generated energy, exerting upward pressure on natural gas prices. However, severe weather not only affects the demand side, but can also interfere with natural gas supply by disrupting production or transportation (e.g., hurricanes, winter storms, freezing temperatures). We consider several measures to capture the influence of meteorological conditions on price developments in the natural

[4] While we only use vintages from February 1997 onward, we make the entire dataset available here. The date of January 1991 for the first vintage is a natural starting point since the publication of the *Monthly Energy Review* (*MER*) was temporarily suspended between October and December 1990. The vintages up to December 2023 were hand-collected from electronic copies of historical issues of the *MER* available on the EIA's website. In January 2024, the EIA started to release real-time vintages in electronic format for all the data contained in the *MER*. Thus, our historical real-time database for variables related to the U.S. natural gas market can now be easily extended.

gas market. A widely used measure are heating and cooling degree days which are based on outdoor air temperature recorded daily at weather stations nationwide and converted into a quantitative index of energy requirements to heat or cool spaces weighted by population. A related indicator constructed by the National Oceanic and Atmospheric Administration (NOAA) is the Residential Energy Demand Temperature Index (REDTI) which uses the heating and cooling degree days as input to determine residential energy needs. A very different measure is NOAA's national index for average temperature anomalies which focuses on periods of extreme heat or cold snaps computed as deviations from long-term averages.

*Macroeconomic Environment.* Another key factor of demand for natural gas which has a direct impact on the price is the state of the domestic economy. Growth-driven increases in natural gas consumption can be particularly strong in the industrial sector which uses natural gas as a fuel and a feedstock in the production process to make goods such as fertilizer and pharmaceuticals. This suggests using U.S. industrial production or the related capacity utilization rates as cyclical indicators since both measures cover manufacturing, mining, and electric and gas utilities. A broader measure that is routinely used to gauge overall economic activity is the Chicago Fed National Activity Index (CFNAI). Among the three proxies for the U.S. business cycle, the degree of capacity utilization seems to be the one that is the least affected by the sharp COVID-19 contraction, which makes this our first choice.

*Backcasting.* Given that the historical data for the natural gas market fundamentals as well as the heating and cooling degree days are only available in print format in published issues of the *Monthly Energy Review* (*MER*), they cover at most a period of 54 months. We therefore assume that observations that are dropped from the historical editions are no longer revised and we use the last record in the construction of subsequent vintages. In fact, to backcast the data all the way to January 1973, we collect data from vintages going back to August 1976 where possible. For those series that were not reported in all the past issues of the *MER*, we approximate the missing observations in the early part of the sample using data from the most recent vintage. In this manner, all variables from the *MER* can be extended back in time to at least January 1976 except for natural gas drilling rigs in operation which started to be measured separately only in 1987M8. We extrapolate the data before that back to 1973M1 using the growth rate of crude oil and natural gas rig counts. This is facilitated by the fact that these two series are not subject to revisions. NOAA does not provide vintages for REDTI and the index of temperature anomalies, so we assume that these data are not revised. However, we account for the publication lag of one month in REDTI when compiling pseudo vintages. While we have a complete set of vintages for industrial production, the first vintage for capacity utilization is available in 1996M11 and for the CFNAI in 2001M1, both containing data back to 1973M1. Earlier vintages are based on data in these first actual vintages adjusted for the one-month delay in their release.

*Nowcasting.* To obtain a balanced dataset, we fill in missing observations at the end of each vintage by devising a set of nowcasting rules that are informed by the time-series properties of our predictor variables, such as seasonal patterns or trend behavior. As shown by Baumeister and

Kilian (2012) for oil prices and by Baumeister, Kilian, and Lee (2017) for gasoline prices, simple, variable-specific rules work particularly well in energy price forecasting models. For natural gas production which features between one and three missing values depending on the vintage, we use the annual log difference from the preceding year to take seasonality into account when filling the gaps. To accommodate the seasonal pattern of working gas underground inventories, we apply the month-on-month change one year prior to the last observation to nowcast the most recent values. For natural gas consumption, we assume that the same amount is consumed as in the same month the year before. Given the seasonal behavior of the three temperature series that are updated with a delay, a sensible nowcasting rule is to rely on the average value for the same month over the ten most recent years to account for slow-moving climatic changes. We extrapolate the one missing observation for industrial production based on its average monthly growth rate. Nowcasts for the current values of capacity utilization and the CFNAI are obtained with exponentional smoothing using a smoothing parameter of 0.95 which is the same weight that Faust and Wright (2013) applied to the unemployment rate.

### 3.2.2 Forecasting with Vector Autoregressive Models

We model the dynamic relationship between the real price of natural gas and its economic determinants as a reduced-form vector autoregression (VAR):

$$\boldsymbol{y}_t = \boldsymbol{c} + \boldsymbol{\Phi}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}) \tag{1}$$

where $\boldsymbol{y}_t$ is a $n \times 1$ vector of monthly data, $\boldsymbol{c}$ is a $n \times 1$ vector of intercepts, $\boldsymbol{\Phi}_i, i = 1, ..., p$, are $n \times n$ coefficient matrices with $p$ denoting the number of lags, and $\boldsymbol{\varepsilon}_t$ are normally-distributed disturbances with covariance matrix $\boldsymbol{\Sigma}$. In addition to the log of the real price of natural gas, $\boldsymbol{y}_t$ in our baseline model includes dry natural gas production, natural gas rig counts, working gas inventories, capacity utilization, and the REDTI as predictor variables with $n = 6$. For the three natural gas fundamentals, we consider transformations to log-levels and to monthly growth rates by taking the first difference of the natural logarithm. We also examine the sensitivity of our forecasting results with regard to alternative choices for the latter two predictor variables and we explore a few additional model specifications that have been proposed for the structural analysis of natural gas price fluctuations.

As in the case of autoregressive models, we estimate the VAR with unrestricted least squares as well as Bayesian shrinkage methods and we start with a fixed lag order $p = 12$. The first column of Table 3 shows that the VAR(12) is unsuccessful in improving upon the no-change forecast independent of the transformation of the gas market fundamentals, with the MSPE ratios exceeding 1 for all horizons. This does not mean that the economic variables have no predictive power but rather that the VAR(12) is overparameterized which hurts its out-of-sample forecasting performance. In fact, applying Bayesian shrinkage yields more precise forecasts; so much so that the BVAR(12) beats the random walk at all horizons except for $h = 1$ but only for the growth-rate specification (see column 2). Reducing the lag length to six or less by means of the AIC leads to more substantial accuracy gains whether the variables are included in log-levels or changes. The size of the MSPE reductions for both BVAR(AIC) models is comparable to the BAR(1) from horizon 3 onward. Columns 5 and 6 show that further restricting the number of lags to 1 delivers the best performance with little to choose between the VAR(1) and the BVAR(1). Both models do considerably better overall than any of the other model-based forecasts. The only model that produces larger MSPE reductions at horizons 21 and 24 is exponential smoothing but this method records big losses at horizons 1 and 3. The log-level specification also dominates the futures-based forecasts at most horizons. We conclude that the economic drivers of natural gas prices greatly enhance the forecast accuracy if the dynamics are kept parsimonious.

Table 4 compares the performance of our baseline six-variable BVAR(1) model with predictors in logs (column 1) to models where we replace one variable at a time with alternative indicators for U.S. meteorological and macroeconomic conditions. Columns 2 to 4 reveal that using an index for temperature anomalies or heating and cooling degree days in deviations from the historical average in the same month instead of the REDTI delivers equally accurate forecasts. While the CFNAI achieves similar reductions in MSPE as capacity utilization across all horizons, industrial production does slightly better in the near term but its accuracy decreases as the forecast horizon lengthens with losses as high as 7% two-years-ahead relative to the other two business cycle indicators (see columns 5 and 6).[5]

Table 5 reports results for a set of other models that have not been designed for forecasting but for studying the structural dynamics of the U.S. natural gas market. Given that any structural model is associated with a reduced form, it is useful to evaluate the success of these models in real-time out-of-sample forecasting. The first model is a four-variable VAR model developed by Arora and Lieskovsky (2014) that includes marketed natural gas production on the supply side and industrial production and residential natural gas consumption on the demand side with variables transformed to annual growth rates to remove seasonality and trends. While the authors set the lag length to four months based on the AIC, we stick with one lag since we have already shown that for forecasting natural gas prices parsimony is key.[6] Column 3 shows that while this smaller model beats the random walk across all horizons, it has on average 6% higher MSPE ratios than the more comprehensive baseline model (column 1). The second small-scale model is motivated by Wiggins and Etienne (2017) and Rubaszek, Szafranek, and Uddin (2021) who represent the U.S. natural gas market with seasonally-adjusted data for production, underground inventories, and industrial production in monthly growth rates. Column 5 reveals that this model also does not outperform the baseline BVAR, but their choices of variables lead to lower MSPE ratios at longer horizons compared to the other four-variable model. Model 3 is of the same size as the baseline model but features a different selection of variables proposed by Winkler (2023). Active gas drilling rigs and marketed production describe the supply side, while industrial production is the main demand-side

[5] Table 2A in the appendix shows that this finding is not due to the inclusion of the COVID-19 observations, but is a genuine pattern of industrial production. In fact, the MSPE ratios are remarkably robust to the pandemic period.

[6] Selecting the lag length recursively with the AIC does not yield any systematic improvements in forecast accuracy (see Table 3A in the appendix).

determinant of prices; we add heating and cooling degree days to the demand side which Winkler (2023) includes as exogenous variables. The MSPE ratios reported in column 7 indicate that this model does not provide more accurate forecasts than the baseline but is a serious competitor for horizons up to $h = 6$. Model 4 is a modified version of model 3 that does not distinguish between the sources of natural gas demand but lumps it all together via total natural gas consumption. This model yields MSPE reductions similar to the baseline up to nine-months-ahead; beyond that, there is a gap of about 3% but overall it performs best among the alternative economic models (see column 9).

### 3.2.3 The Role of Time-Varying Volatility

Over the decades, the U.S. natural gas market has experienced important changes in its regulatory and market structure which can impact the forecasting performance of our economic models. Hupka et al. (2023) document that structural shifts often happen quickly and for a wide variety of reasons. Prominent examples include abrupt production disruptions due to hurricanes and winter storms and consumption spikes due to other extreme weather events like hard freezes and heat waves. Technological advances in drilling and delivery methods or shifts in consumer preferences toward cleaner burning fuels can also trigger sudden changes in the energy mix. Another possible source of periodic instabilities are infrastructure bottlenecks in transportation or capacity constraints in underground storage. All these unexpected events might cause an increase in price volatility that interferes with the models' predictive ability.

It is by now well established that introducing nonlinearities in the form of stochastic volatility greatly enhances the accuracy of point forecasts of a broad range of macroeconomic variables and other energy prices compared to models that impose homoskedasticity (see, e.g., Clark, 2011; D'Agostino, Gambetti, and Giannone, 2013; Clark and Ravazzolo, 2015; Baumeister, Korobilis, and Lee, 2022). As noted by Primiceri (2005) and Carriero, Clark, and Marcellino (2019), the advantage of allowing for time variation in error variances is that it not only captures jumps in volatility but it also interacts with the model dynamics in a way that might further improve the precision of point forecasts. We now investigate the benefit of modeling time variation in the volatilities of shocks to $\boldsymbol{y}_t$ by postulating the following specification for the VAR residuals $\boldsymbol{\varepsilon}_t$ as in Kastner and Huber (2020):

$$\boldsymbol{\varepsilon}_t = \boldsymbol{\Lambda q}_t + \boldsymbol{\nu}_t, \quad \boldsymbol{q}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{H}_t), \quad \boldsymbol{\nu}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega}_t) \tag{2}$$

where $\boldsymbol{\Lambda}$ denotes an $n \times m$ matrix of factor loadings with $m$ referring to a small number of latent factors in $\boldsymbol{q}_t$. These factors are conditionally heteroskedastic with a time-varying diagonal covariance matrix $\boldsymbol{H}_t = diag(e^{h_{1t}}, ..., e^{h_{qt}})$ and $\boldsymbol{\nu}_t$ are measurement errors with a time-varying diagonal covariance matrix $\boldsymbol{\Omega}_t = diag(e^{h_{q+1,t}}, ..., e^{h_{q+n,t}})$. The law of motion for the $m + n$ logarithmic volatilities $h_{it}$ is given by independent AR(1) processes:

$$h_{it} = \mu_{hi} + \rho_{hi}(h_{it-1} - \mu_{hi}) + \sigma_{hi}\eta_{it}, \quad \eta_{it} \sim \mathcal{N}(0, 1) \quad \text{for } i = 1, ..., m + n, \tag{3}$$

with $\mu_{hi}$ the unconditional mean of the log volatilities, $\rho_{hi}$ their persistence parameter, and $\sigma_{hi}$ their standard deviation. The time-varying variance of $\boldsymbol{\varepsilon}_t$ can thus be written as $\boldsymbol{\Sigma}_t = \boldsymbol{\Lambda}\boldsymbol{H}_t\boldsymbol{\Lambda} + \boldsymbol{\Omega}_t$ which amounts to the factor stochastic volatility (FSV) model proposed by Pitt and Shephard (1999) and Aguilar and West (2000).[7] This volatility specification has several attractive features. First, in contrast to popular alternatives, the FSV model does not depend on the ordering of the variables in the VAR model. Second, the factor structure makes the FSV model more parsimonious since it reduces the number of free parameters if $m < n$.[8] Third, despite its parsimonious nature, this model offers rich dynamics by capturing both common and idiosyncratic sources of volatility. Details on the prior choices and the estimation are provided in Appendix A.1.

Table 5 compares the forecast accuracy of our economic models with and without stochastic volatility. While all the models with time variation in the error variances outperform the no-change forecast at all horizons, only models 1 and 2 improve upon their constant-variance counterparts beyond the 12-month forecast horizon, with the largest MSPE reductions of around 20% two years out (see columns 4 and 6). Given that these two models are smaller in size compared to the baseline and model 3, it might be the case that the FSV model favorably influences the model dynamics to produce better point forecasts at longer horizons, whereas adding more variables compensates for the lack of time-varying volatility. As can be seen in column 10, this is not borne out by model 4 though which contains the same number of variables as models 1 and 2 but displays a weaker performance than the same model without stochastic volatility. Another difference is that the predictor variables in models 1 and 2 are included in growth rates, while all other models are specified in log-levels, suggesting that FSV has more bite for models estimated in first differences. Thus, while yielding accuracy gains for some models, BVAR-FSV models only do better than the baseline BVAR at horizons $h = 21$ and $h = 24$.[9]

### 3.3 Energy Price Models: A Tale of Two Markets

So far we considered economic fundamentals within the domestic natural gas market as the main predictors of natural gas price dynamics, but natural gas prices also have close ties with other energy prices, in particular crude oil because of the possibility of fuel switching. In fact, historically, natural gas was often considered a by-product of crude oil production and thus production was not necessarily determined in response to market forces. Instead, the ability of large-volume fuel consumers such as power plants and iron, steel, and paper mills to switch between natural gas and petroleum as a function of their cost establishes a link between the two markets which suggests to jointly model natural gas and oil prices giving rise to bivariate energy price models; yet, the two

[7] While we cannot separately identify $\boldsymbol{\Lambda}$ and $\boldsymbol{q}_t$ absent further restrictions, this is not important since we are only interested in $\boldsymbol{\Sigma}_t$.

[8] We select $m$ according to the Ledermann bound which determines the largest number of factors that implies a unique decomposition of the covariance matrix $\boldsymbol{\Sigma}_t$ and is derived by solving $(n-m)^2 \geq n+m$.

[9] For completeness, Table 4A in the appendix also provides a comparison of the forecasting performance of the BVAR models with homoskedastic and heteroskedastic factor structure on the covariance matrix.

fuels are not perfect substitutes due to differences in heat content, transportation costs, and market structure. In contrast to other countries where oil indexation is the dominant pricing mechanism for natural gas, energy markets in the US are more competitive and relative price movements determine the marginal market (see Hupka et al., 2023). This means that there is no mechanical relationship between these two prices and thus the forecasting content of bivariate fuel-pricing models is an empirical question.

Figure 1 shows that there is pretty strong comovement between the Henry Hub natural gas and West Texas Intermediate (WTI) crude oil spot prices from 1993 to 2010 but no clear pattern thereafter.[10] For this first period, Brown and Yücel (2008) and Hartley, Medlock, and Rosthal (2008) find a cointegration relationship between natural gas and oil prices and use an error correction model to study the short-run and long-run interaction effects among both markets. Bachmeier and Griffin (2006) also provide empirical evidence for energy market integration. However, the stability of this long-run relationship has been questioned by Ramberg and Parsons (2012) who show that the price of oil has only weak explanatory power for short-term natural gas price fluctuations. The advent of the shale gas revolution in 2006 marks a major turning point that first weakened, then severed the linkages between the global oil market and the U.S. natural gas market. The steep increase in gas production from fracking led to the decoupling of natural gas prices from oil prices given that shale gas was landlocked (due to limited pipeline capabilities) until 2016 when LNG export terminals enabled shipments initiating the recoupling of U.S. natural gas prices to international energy price developments (see Stock and Zaragoza-Watkins, 2024).

Against this background, we investigate the promise of two types of models that explicitly or implicitly feature a cointegrating relationship. Specifically, we explore the predictive power of price spread models in Section 3.3.1 and of bivariate vector autoregressive models in Section 3.3.2.

### 3.3.1 Price Spread Models

This forecasting model is predicated on the assumption that the nominal spot prices of natural gas and crude oil are cointegrated. In that case, current deviations of the spot price of natural gas from the spot price of crude oil would be expected to have predictive power for cumulative changes in the nominal spot price of natural gas:

$$\Delta s_{t+h|t}^{h,HH} = \alpha + \beta(s_t^{HH} - s_t^{WTI}) + \varepsilon_{t+h} \tag{4}$$

where $s_t^{HH}$ is the log of the nominal spot price of natural gas, $s_t^{WTI}$ is the log of the nominal spot price of WTI crude oil, and $\Delta s_{t+h|t}^{h,HH}$ denotes the cumulative change in $s_t^{HH}$ over the next $h$ months.[11] Following the approach of Baumeister, Kilian, and Lee (2017) for gasoline prices, we map

[10] Monthly data for the WTI spot oil price are taken from FRED (WTISPLC). We convert the oil price which is quoted in dollars per barrel to dollars per million Btu by dividing it by 5.8 such that it has the same unit as the Henry Hub spot price.

[11] We estimate the spread model starting in 1993.1 because the Federal Energy Regulatory Commission Order 636 issued in 1992 marks the completion of the deregulation process of wellhead prices of natural gas (see Joskow, 2013).

recursive estimates of this relationship into a forecast for the real price of natural gas as follows:

$$R_{t+h|t}^{HH} = R_t^{HH} \exp\left[\widehat{\alpha} + \widehat{\beta}(s_t^{HH} - s_t^{WTI}) - E_t(\pi_{t+h})\right] \tag{5}$$

where $E_t(\pi_{t+h})$ denotes expected inflation over the next $h$ months extrapolated using the average inflation rate over a ten-year rolling window.

Columns 1 and 2 in Table 6 report the average MSPE ratios for two alternative specifications of the price spread models, an unrestricted variant and a restricted variant where, in the interest of parsimony, $\alpha$ is set to zero. Both models display a dismal forecasting performance with MSPE ratios as high as 1.5. Only at horizon 1 do price spread models marginally improve on the random walk. This result suggests that the maintained hypothesis of cointegration likely fails to hold for the entire evaluation period in line with evidence provided by Ramberg and Parsons (2012) and Stock and Zaragoza-Watkins (2024). To get a better sense of the evolution of the forecasting performance over time, Figure 2 presents recursively updated MSPE ratios for selected forecast horizons. The message that emerges from these plots is that price spread models delivered accurate forecasts until the breakdown of the cointegrating relationship dated January 2009 by statistical tests (see Table 1 in Stock and Zaragoza-Watkins, 2024). Thus, the shift to a regime without cointegration as a result of the U.S. shale gas revolution explains the sudden deterioration in the forecasting ability of these models.

This evidence raises the question of whether a model that explicitly allows for a switch in regime would have been able to detect this change in real time thereby preventing the massive forecast errors in the post-2009 period. To examine this question empirically, we extend the model in equation (4) with parameters driven by a Markov-switching (MS) process which takes the following form:

$$\Delta s_{t+h|t}^{h,HH} = \alpha_{s_{t+h}} + \beta_{s_{t+h}}(s_t^{HH} - s_t^{WTI}) + \varepsilon_{t+h}, \quad \varepsilon_{t+h} \sim \mathcal{N}(0, \sigma_{s_{t+h}}^2) \tag{6}$$

where $\alpha_{s_{t+h}}$ is the MS intercept, $\beta_{s_{t+h}}$ is the MS spread coefficient, $\sigma_{s_{t+h}}$ is the MS volatility, and $\{s_{t+h}\}$ follows an $m$-states ergodic and aperiodic Markov-chain process that determines the regime. This process is an unobservable variable which takes integer values $s_{t+h} \in \{1, \ldots, m\}$ and has transition probabilities $\mathbb{P}(s_{t+h} = j | s_{t+h-1} = i) = p_{ij}$, with $i, j \in \{1, \ldots, m\}$. The transition matrix $\boldsymbol{P}$ of the chain is

$$\boldsymbol{P} = \begin{pmatrix} p_{11} & \cdots & p_{1m} \\ \vdots & \ddots & \vdots \\ p_{m1} & \cdots & p_{mm} \end{pmatrix} \tag{7}$$

and features, as a special case, the one-forever-shift model that is widely used in structural break analysis. We set $m = 2$ postulating a switch between a cointegration regime and non-cointegration regime. We apply a Bayesian approach to estimation which offers a natural way to conduct inference for latent-variable models using simulation-based methods. Moreover, priors on parameters in different regimes can be imposed to resolve the label-switching problem that arises in MS models (see Appendix A.2 for details on the implementation).

The last column of Table 6 shows that accounting for the possibility of a shift in regime over the evaluation period improves upon the constant-coefficient counterparts at most horizons but only outperforms the random walk benchmark at $h = 3$ and $h = 6$.

### 3.3.2 Bivariate Vector Autoregressive Models

Given the economic interaction between natural gas and crude oil markets, an alternative modeling approach that only implicitly allows for a long-term relationship between natural gas and oil prices is a bivariate VAR($p$) model with both prices included in log-levels. Thus, in equation (1), $\boldsymbol{y}_t = [r_t^{HH}, r_t^{WTI}]'$ is a $2 \times 1$ vector that contains the log of the Henry Hub natural gas price and of the WTI crude oil price, both deflated in real time with the U.S. consumer price index. Panel A of Table 7 examines the role of different lag orders for the forecasting performance of this model when estimated by unrestricted least squares (LS) and the Bayesian method of Giannone et al. (2015). Column 1 shows that even in a simple bivariate model 12 lags prevent the LS estimator at most horizons to deliver forecasts that are more accurate than the no-change forecast, whereas Bayesian shrinkage yields MSPE ratios below 1 except at $h = 24$. The largest gains of around 10% are found for medium-term horizons of six to twelve months ahead. Determining the lag length recursively based on the AIC or setting $p = 1$ helps the VAR outperform the random walk more often, but hurts the performance of the BVAR both in terms of the size of MSPE reductions and the number of horizons for which it improves upon the no-change forecast. It is also interesting to note that the VAR(1) does better than the BVAR(1) for horizons up to 1 year (see columns 5 and 6). Comparing the results to the (B)AR($p$) models in Table 2, which are nested in the bivariate (B)VAR($p$) models, suggests that information about past oil price dynamics is not particularly useful for obtaining more accurate forecasts except maybe in the short run.

While the VAR model estimated in log-levels remains agnostic about the existence of a cointegrating relationship, it suffers an efficiency loss should the variables effectively be cointegrated. Despite the evidence derived for the price spread models and the narrative about the decoupling of natural gas and oil prices, we investigate the usefulness of imposing cointegration restrictions on the bivariate VARs for the purpose of forecasting. The vector error correction (VEC) representation is implied by the following VAR specification with $\boldsymbol{y}_t = [\Delta r_t^{HH}, r_t^{HH} - r_t^{WTI}]'$. Panel B of Table 7 confirms our earlier results that enforcing cointegration is detrimental to the forecasting performance. Except for some small accuracy gains at horizons 3 to 9, the VEC($p$) models perform poorly independent of the lag order and the estimation method and they are dominated by the log-level specification.

## 3.4 Integrating Own- and Cross-Market Fundamentals

A natural question is whether jointly modeling natural gas fundamentals and oil prices can further improve the accuracy of natural gas price forecasts. We augment our baseline economic model in (1) with the log of the real price of WTI which results in a model of size $n = 7$. The first three columns of Table 8 show that among the BVAR models with constant variance, the model where the AIC determines the lag length recursively performs best, beating the random walk at all forecast horizons. This suggests that the inclusion of oil prices leverages past dynamics to yield more precise forecasts despite the higher dimensionality of the model. Allowing for stochastic volatility leads to relative improvements for the model with $p = 12$, but worsens the performance of the model with $p = 1$ (see columns 4 and 6); neither of them outperforms the BVAR(AIC) model with an upper bound of $p = 6$ except for $h = 1$. Only the BVAR-FSV model with $p = 6$ delivers some small accuracy gains at long horizons relative to its constant-variance counterpart. However, none of the integrated models that account for arbitrage across fuels does better than the baseline economic model of the natural gas market alone.

# 4 Model Comparison and Forecast Pooling

## 4.1 Joint Assessment Based on the Model Confidence Set

The preceding analysis highlights that while there are quite a few models that beat the monthly no-change forecast, there is no single forecasting method that, based on pairwise comparisons, dominates all the others across all horizons and thus it is unclear which model to choose. With such a rich assortment of competitive candidate models, it is indeed difficult to get a good sense of whether some alternatives are more useful than others for out-of-sample forecasting. We shed light on this question with the help of the model confidence set (MCS) procedure proposed by Hansen, Lunde, and Nason (2011) that allows us to jointly assess the entire model space for a fixed horizon $h$ to decide which models should be included in the MCS based on their predictive power. Specifically, the approach selects the subset of models $\widehat{M}^*_{1-\alpha}$ that form the MCS such that it contains the best model with probability $1-\alpha$. The pruning of the initial model space is data-driven and the number of surviving models reflects the informativeness of the sample.

Table 5A in the appendix reports the MCS $p$-values for the entire collection of models evaluated previously, including the random walk benchmark, and summarizes the number of models that enter the MCS at each forecast horizon $h$ for confidence levels $\alpha = 0.1$ and $\alpha = 0.25$ based on their squared forecasting errors over the full out-of-sample period.[12] Starting with an initial set of 60 individual models, we find that for horizons up to nine months, the MCS basically covers the complete model space. The fact that no models are discarded suggests that the information in the data cannot discriminate between good and bad models; that is, the MCS procedure finds the forecasting ability of any one model not to be significantly worse than any other and thus keeps them all. As emphasized by Hansen et al. (2011), the MCS approach tends to err on the side of caution and does not shrink the initial set if the data provide insufficient information about the relative performance of the forecasting models. As the forecast horizon lengthens, some competitors

[12] Following Hansen et al. (2011), we compute the MCS $p$-values with 10,000 block bootstrap replications using a block size of 12.

are eliminated; for example, for $h = 18$ and $h = 24$, only 72% of the models pass the threshold to enter $\widehat{M}^*_{75\%}$. But even for longer horizons, the set of superior models remains large.

In addition to determining the size of the set for a given level of confidence, the MCS $p$-values imply a ranking where models with larger $p$-values are more likely to be among the best-performing models. The top-5 models are marked in red in Table 5A. What stands out is that the economic models of the natural gas market whose dynamics are summarized by vector autoregressions of order one always show up at the head of the pack. At intermediate horizons from $h = 6$ to $h = 18$, the VAR(1) with the baseline predictors entering the model in log-levels receives a $p$-value of one three times; at $h = 3$, the same model wins but with industrial production replacing capacity utilization as indicator of U.S. macroeconomic conditions. At long horizons, some models that feature stochastic volatility start appearing in the upper echelon with one of the alternative economic models being assigned a $p$-value of one at $h = 24$. Exponential smoothing and simple autoregressive models with moderate lag length are also among the most promising models further out, while most energy price models are removed from $\widehat{M}^*_{75\%}$ at longer horizons. Most other univariate models make a poor showing, especially at short horizons. The forecasting ability of models that combine own- and cross-market fundamentals is rated as comparable to economic models of the natural gas market with alternative predictors at the shortest horizons, irrespective of their lag order and volatility component, but is ranked considerably lower at horizons beyond one year. Not surprisingly, for $h = 1$, the end-of-month no-change forecast takes the lead with a $p$-value of one, but its relevance diminishes quickly, and, at horizons 12 and 18, it is even kicked out of $\widehat{M}^*_{75\%}$. The futures-market forecast is a strong contender across all horizons and even makes first place at $h = 9$. While the benchmark is always included in $\widehat{M}^*_{75\%}$, it never ends up among the top models.

### 4.2 Tracking Forecasting Performance over Time

The evidence presented so far focuses on the average forecast quality of the entire set of models over the full evaluation period. However, as already shown in Section 3.3.1 for the case of constant-coefficient price spread models, the forecasting performance is not necessarily stable over time and might depend on the prevailing market structure.

To get a better idea to what extent the predictive ability varies also for other models across the out-of-sample period, Figure 3 displays the evolution of the cumulative MSPE ratios relative to the no-change benchmark for a subset of representative models from each class and four different forecast horizons. The key takeaway from these plots is that there is substantial heterogeneity in the forecasting accuracy across models and horizons, and that the relative performance of some models changes considerably over time. For example, at the three-month horizon, the end-of-period random walk forecast starts out poorly, while the futures-based forecast is quite successful, but from 2010 onward they converge and end up with the same average performance with an MSPE ratio of 0.98. Similarly, for $h = 6$, the Markov-switching price spread model outperforms all other models in the first ten years of the evaluation period with an average MSPE ratio of 0.7, but is almost tied with the benchmark toward the end, whereas the BVEC(12) model starts out at a disadvantage with an MSPE ratio above one until 2006 but improves gradually thereafter yielding an average MSPE ratio of 0.91. In contrast, the baseline economic model augmented with the real oil price that features stochastic volatility (OilBVAR(6)-FSV) displays a rather stable performance over time. For 24-month-ahead forecasts, the simple AR(AIC) model is one of the most promising candidates early on but then experiences a dramatic deterioration in the interim only to reclaim its spot among the best models at the end of the evaluation sample. Instead, the accuracy gains from exponential smoothing at $h = 24$ are relatively constant throughout.

We complement this illustrative evidence with a more formal assessment of changes in the forecasting performance of the entire universe of models by splitting the evaluation period into four subsamples that correspond to the institutional regimes defined by Stock and Zaragoza-Watkins (2024). Changes in the structure of the natural gas market, the availability of transportation and storage infrastructure, the composition of end uses, as well as other industry developments are likely to affect the forecast accuracy of some of the models contained in our collection. We make use of the MCS ranking to determine which type of models prevails in different subperiods and investigate whether this pattern lines up with the historical narrative. Tables 6A and 7A in the appendix show the MCS $p$-values for our set of models for short ($h = 3$ and $h = 6$) and long ($h = 12$ and $h = 24$) forecast horizons, respectively.

The first subsample which extends until March 2006 is characterized by a tight link between natural gas and oil prices due to the substitutability of the two fossil fuels in the power sector and for industry uses. The leading models during this period for horizons up to 12 months belong to the class of price spread models which are built on the premise of a stable long-run relation between the two energy prices. Equally valuable are market-based forecasts, which even take first place at $h = 6$ and $h = 12$, possibly as a result of hedging demand from downstream sectors. Interestingly, at the longest horizon, economic models of the natural gas market that feature either stochastic volatility or many lags are ranked highest.

The next subsample running from April 2006 to December 2009 marks the initial phase of shale gas fracking. In this transitory regime, the model that yields the highest MCS $p$-value at the three-month horizon is the integrated-markets model, BVAR(AIC), that combines arbitrage between natural gas and oil prices with other economic determinants, but it is closely followed by models that only include supply-and-demand fundamentals which emerge as the dominant type at $h = 6$ and $h = 12$. While these two classes of models continue to figure prominently at the longest horizon, the atheoretical model of exponential smoothing now ranks first. But overall the shift to models that focus more on domestic market forces aligns well with rising shale gas production and rig count activity during this period.

Stock and Zaragoza-Watkins (2024) describe the third episode that spans January 2010 to May 2016 as the period of 'shut-in fracking' during which domestic gas prices decoupled from global energy prices due to the shale gas boom and limited export possibilities. At short horizons, the futures-based forecasts take the lead; for $h = 3$, the data are not very informative since models from all types show up in second place, while for $h = 6$, the second-ranked forecasts are obtained with domestic natural gas market models and univariate price models. The same holds for $h = 12$, while at $h = 24$, the MCS procedure only selects three models into $\widehat{M}^*_{75\%}$ and they all belong to the univariate class with AR(AIC) at the top. This suggests that own price dynamics sometimes coupled with additional economic drivers as well as price discovery in the futures market play an important role in those years.

The last subperiod starts in June 2016 with the opening of LNG export facilities which reconnect the domestic market to international energy trade. While the best models pertain to the category of economic models at short horizons, both energy price and integrated-markets models are in the same league, especially at $h = 3$, which suggests that these forecasting models are making a comeback as a result of the ongoing process of price recoupling. No matter which category these models belong to, what they have in common is a lag length of $p = 12$ which implies more complex dynamics in the most recent period. At long horizons, the futures-based forecast is back in the game, but at $h = 12$, the economic model with predictors in growth rates as well as the hybrid oil price-economic model, both with 12 lags, are still the next best. Instead, at $h = 24$, the MCS approach is again pointing to a mix of univariate and economic models without being able to clearly distinguish between them.

Taken together, the graphical and subsample analyses reveal some striking changes in the predictive power and relevance of different models in our set both over time and across horizons. Given this variation in performance and the difficulty of settling on a single model for the purpose of out-of-sample forecasting, it might be better to combine forecasts not only to possibly enhance the forecast accuracy further but also to guard against forecast failures due to structural change and model misspecification (see, e.g., Granger and Jeon, 2004; Timmermann, 2006), an approach that we investigate next.

### 4.3 Combining Forecasts

To explore the benefits of aggregating forecasts from different models, we consider a real-life forecaster who had good reasons for compiling the entire universe of models in January 1997 based on the insights discussed above but without any knowledge of their out-of-sample forecasting performance. During the first year, he pools the forecasts from all 60 models by taking the arithmetic average, but then employs the MCS approach to reduce the set of competing models to a smaller set that contains the best model with a specified level of confidence $\alpha$. An attractive feature of the MCS for real-time model selection is that it explicitly acknowledges the fact that more than one model can be the best (see also Granger and Jeon, 2004). Given that the procedure is tied to the informativeness in the data, the set of superior models evolves dynamically and thus differs at each point in time. We then combine the forecasts of the best-performing models where "best-performing" is defined in two different ways. First, we pool the forecasts from all the models above the threshold for inclusion in the MCS which we set at the $\alpha = 25\%$ significance level. Second, we exploit the ranking of models implicit in the MCS testing procedure and only select models that yield the highest MCS $p$-values within $\widehat{M}^*_{75\%}$ which essentially means that we pool forecasts from the models with the highest and second-highest $p$-values given that these are associated with the smallest sample loss of all forecasts.[13] The sample loss based on squared forecasting errors is determined either recursively or rolling, where, in the former case, the real-life forecaster uses the evidence on the usefulness of each model accumulated since the beginning, while, in the latter case, he only uses the model performance observed over the past year.

Table 9 presents the MSPE ratios for these alternative forecast combinations relative to the random walk. We start with including all 60 models in the forecast combination in line with the thick modeling approach of Granger and Jeon (2004) which serves as a useful benchmark. Column 1 shows that these pooled forecasts beat the no-change forecast across all horizons, but none of them is more accurate than the best individual model by horizon. If we restrict the forecast combination to models that end up in $\widehat{M}^*_{75\%}$ in real time, we find substantial MSPE reductions of around 24% from $h = 12$ onward which are either tied with or outperform individual models (see column 3). For horizons up to nine months, improvements are minor relative to pooling all models. This difference between short and long horizons can be traced back to the weaker sample information at short horizons which results in a large number of surviving models closer to the 'all-models' case.[14] Since it apparently pays to be selective, we now limit pooling to the models ranked first and second according to the MCS $p$-values. Column 5 reveals that using this stricter criterion results in fewer but more promising models yielding further accuracy gains with MSPE reductions ranging from 16% at the three-month horizon to 30% at the 12-month horizon. The combined forecasts now perform better than individual models at all horizons except for $h = 1$ where the end-of-month no-change forecast still dominates. Applying the same selection criterion but on a rolling basis leads to even more impressive improvements in MSPE ratios with the lowest value being a staggering 0.46 (see column 7). Thus, narrowing the evaluation of losses to a short window increases the power of the MCS procedure to make a precise selection of the relevant models whose forecasts when pooled achieve the largest MSPE reductions overall. These results are based on weighing all models equally when aggregating their forecasts. As can be seen from columns 2, 4, 6, and 8, using weights proportional to the recursive inverse MSPE instead makes no material difference for the accuracy of pooled forecasts; all that matters for the forecasting success is how the models for the forecast combination are chosen.

In sum, while this approach requires updating all the models each month to be able to track their latest forecasting performance in order to select the ones with the highest MCS $p$-values for the purpose of pooling their forecasts, it holds the most promise for out-of-sample forecasting of the real price of natural gas.

[13] The highest $p$-value is always one, whereas the second highest $p$-value can either be above or below the threshold for inclusion in the MCS. Thus, the minimum is one selected model.

[14] For example, at $h = 3$, the minimum number of models included is four, the maximum is 60, and the average is 57, as opposed to a minimum of one, a maximum of 46, and an average of 35 selected models at $h = 24$.

# 5 Conclusions

Accurate forecasts of the natural gas price inform the design of a range of policies on climate change, energy affordability, incentives for fuel shifting, and the use of public lands for mineral exploration and extraction. In this paper, we evaluated the usefulness of a variety of candidate models that find their origin in economic theory, cross-market dynamics, long-run relationships, and futures trading in terms of their out-of-sample forecasting performance for the monthly real price of natural gas for horizons up to 24 months. The assessment was conducted in a real-time setting that accounted for delays in the availability of predictor variables and revisions as preliminary data were updated. For this purpose, we compiled a rich database of fundamental determinants of the real price of natural gas from multiple sources. It consists of vintages from January 1991 to February 2024, each covering data going back to January 1973, that report only the information that a real-life forecaster would have had at his disposal at the time the forecasts were generated.

Our analysis offers several key takeaways. First, at short horizons, model-free forecasts based on futures prices and last-day-of-the-month prices alongside forecasts from economic models of the natural gas market with a selective choice of fundamental drivers perform best. Second, at longer horizons, forecasts derived from exponential smoothing, futures prices, and a set of economic models with and without time-varying volatility display the most success. Third, given that no single model wins across all horizons and that the relative forecasting performance of individual models varies across time, we show that forecast combinations are the best option to produce reliable forecasts. Pooled forecasts from individual models that are ranked highest by a selection criterion that adapts to the most recent performance in real time achieve the largest reductions in mean-squared prediction error and thus promise to deliver the most accurate forecasts going forward.

# A Prior Specifications and Bayesian Estimation Algorithms

## A.1 BVAR-FSV Models

To estimate the model in equations (1)-(3) using Bayesian methods, we specify the following priors for the various sets of parameters.

*Priors for VAR coefficients.* Let $\mathbf{B}_j = [c_j, \mathbf{\Phi}_{j,1}, ..., \mathbf{\Phi}_{j,p}]'$ for $j = 1, ...n$ contain the VAR coefficients pertaining to the $j^{th}$ equation. We postulate that the (conditional) prior on the rows of $\mathbf{B}$ is Gaussian, $p(\mathbf{B}_j|\mathbf{\Xi}_{B,j}) \sim \mathcal{N}(\underline{\mathbf{b}}_j, \underline{\mathbf{V}}_{B,j})$, where $\underline{\mathbf{b}}_j$ is a vector composed of elements $\underline{b}_{js}$ and $\mathbf{\Xi}_{B,j}$ is a vector of hyperparameters that determine the prior variances of $\underline{\mathbf{b}}_j$. The prior means incorporate the idea that series in log levels follow a random walk process by setting $\underline{b}_{js} = 1$ if $j = s$ and 0 if $j \neq s$. The prior covariance matrix $\underline{\mathbf{V}}_{B,j}$ is a diagonal matrix whose entries are determined by the product of two shrinkage parameters, a local one denoted $\gamma_{js}^2$ and an equation-specific one denoted $\lambda_j^2$. These two parameters follow a standard half-Cauchy prior distribution $\mathcal{C}^+(0, 1)$ which gives rise to the standard horseshoe prior (Carvalho, Polson, and Scott, 2010). The insight underlying this particular global-local shrinkage prior is that it exerts 'global' shrinking by assigning a high probability that the elements in $\mathbf{B}$ that are not associated with the first own lag of an equation are zero, while making sure that if this is not the case 'local' shrinkage kicks in and pulls prior mass away from zero (see also Huber and Feldkircher, 2019, for an empirical assessment of this type of prior in forecasting applications).

*Priors for factor loadings.* For the loadings $\mathbf{\Lambda} = (\boldsymbol{\lambda}_1, ..., \boldsymbol{\lambda}_m)$ in the FSV model, we assume a Gaussian prior with mean zero and an $n \times n$ diagonal covariance matrix $\underline{\mathbf{V}}_{\Lambda,j}$ for the $m$ columns of $\mathbf{\Lambda}$, $\boldsymbol{\lambda}_q|\mathbf{\Xi}_{\Lambda,q} \sim \mathcal{N}(\mathbf{0}, \underline{\mathbf{V}}_{\Lambda,j})$, for which we rely on a horseshoe prior as before except that the equation-specific parameters are replaced by column-specific parameters. Thus, the prior variance for the $i^{th}$ element in $\boldsymbol{\lambda}_q$, $\lambda_{iq}$, is $[\underline{\mathbf{V}}_{\Lambda,j}]_{ii} = Var(\lambda_{iq}) = \phi_{iq}^2 \xi_q^2$, where $\phi_{iq}$ is again a local shrinkage parameter and $\xi_q$ is a columwise shrinkage parameter, both of which have a standard half-Cauchy distribution. Columnwise shrinkage allows the number of factors to be selected in a data-driven manner since it implies that if a factor is irrelevant, it will be removed, thereby avoiding overfitting (see, e.g., Bhattacharya and Dunson, 2011; Kastner, 2019).

*Priors for stochastic volatilities.* For the priors on the parameters in the law of motion of the stochastic volatilities, we follow the setup of Kastner and Frühwirth-Schnatter (2014). In particular, for the unconditional mean parameter $\mu_{h_i}$, we postulate a zero-mean Gaussian prior with variance $10^2$; for the (transformed) persistence parameter $\rho_{h_i}$, we use a Beta-distributed prior $(\rho_{h_i} + 1)/2 \sim \mathcal{B}(5, 1.5)$ to push the log-volatilities toward a stationary process; and for the state innovation variances, we use a Gamma prior $\sigma_{h_i}^2 \sim \mathcal{G}(1/2, 1/(2 * c))$ where $c = 0.01$ which implies shrinkage toward homoskedasticity.

Combining these priors with the likelihood, we estimate the Bayesian VAR with factor stochastic volatility based on the Gibbs sampling algorithm detailed in Kastner and Huber (2020). After an initial burn-in period of 5,000 draws, we perform 5,000 iterations of the Gibbs sampler. For each draw from the posterior, we simulate the predictive distribution $h$ periods into the future via Monte Carlo integration. The point forecast at horizon $h$ is the median of that distribution.

## A.2 Markov-Switching Price Spread Models

For estimation of the model in equation (6), we apply a Bayesian approach given that inference for latent-variable models calls for simulation-based methods, which can be easily accommodated in a Bayesian setting. Moreover, priors on parameters in different regimes can be used for regime labeling.

The Bayesian inference framework relies on data augmentation (see Tanner and Wong, 1987) and on a Monte Carlo approximation of the posterior distributions. Following this approach, we introduce the allocation variable $\boldsymbol{\xi}_{t+h} = (\xi_{1,t+h}, \ldots, \xi_{m,t+h})$, where $\xi_{k,t+h} = \mathbb{I}_{\{k\}}(s_{t+h})$ indicates the regime to which the current observation $\Delta s_{t+h|t}^{h,HH}$ belongs, and $\mathbb{I}_A(x)$ is the indicator function that takes value 1 if $x \in A$ and 0 otherwise. The allocation variables cluster the observations in different groups. Each group corresponds to a regime and is characterized by regime-specific parameters in the regression equation. For expository convenience, we follow Frühwirth-Schnatter (2006) and define the vectors of time-variant regressors as follows: $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)'$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)'$, and $\boldsymbol{\sigma} = (\sigma_1^2, \ldots, \sigma_m^2)'$. With this notation in hand, the regression model in equation (6) can be rewritten as follows:

$$\Delta s_{t+h|t}^{h,HH} = \boldsymbol{\xi}_t'\boldsymbol{\alpha} + \boldsymbol{\xi}_t'\boldsymbol{\beta}(s_t^{HH} - s_t^{WTI}) + \varepsilon_{t+h}, \quad \varepsilon_{t+h} \sim \mathcal{N}\left(0, \gamma_{t+h}\right) \tag{8}$$

where $\gamma_{t+h} = \boldsymbol{\xi}_t'\boldsymbol{\sigma}$ is the MS stochastic volatility process. The data-augmentation procedure yields the completed likelihood function of model (6) (see also Frühwirth-Schnatter, 2006).

When using MS processes, one has to deal with the identification issue associated with label switching (see, for example, Frühwirth-Schnatter (2001) for a discussion about the effects that label switching and the lack of identification have on the results of Markov chain Monte Carlo-based Bayesian inference). One efficient approach is the permutation sampler (see Frühwirth-Schnatter, 2001), which can be applied under the assumption of exchangeability of the posterior density. This assumption is satisfied when one assumes symmetric priors on the transition probabilities of the switching process. As an alternative, one might impose identification constraints on the parameters. This practice is followed to a large extent in macroeconomics and is related to the natural interpretation of the different regimes as different phases of the business cycle (e.g. recession and expansion). We follow the latter approach and include inequality constraints on the parameters. In particular, we set $m = 2$ regimes and apply the following restrictions: $\alpha_1 < -0.005$; $\alpha_1 >= -0.005$; $\beta_1 <= -0.001$; $-0.001 < \beta_1 < 0.001$. The first regime corresponds to average estimates in the constant-coefficient price spread model. The second regime corresponds to a white noise process where the relationship between the cumulative change in the price of natural gas and the gas-oil price spread is non-existent and the intercept oscillates around 0.

We follow Diebolt and Robert (1994) and consider a conjugate partially improper prior. Conjugate improper priors are numerically close to Jeffreys' prior, provide similar inferences, but facilitate posterior simulation. We assume uniform prior distributions for the $\beta_{s_{t+h}}$ coefficient and the $\alpha_{s_{t+h}}$ intercept. For the error variance $\sigma^2_{s_{t+h}}$, we also apply a uniform prior distribution on the precision matrix: $\sigma^2_{s_{t+h}} \propto \frac{1}{\sigma^2_{s_{t+h}}} \mathbb{I}_{\mathbb{R}_+}(\sigma^2_{s_{t+h}})$. We assume standard conjugate prior distributions for the transition probabilities in equation (7), which are independent and identical Dirichlet distributions, one for each row of the transition matrix, $(p_{k1}, \ldots, p_{km})' \sim \mathcal{D}(\delta_1, \ldots, \delta_m)$ with $k = 1, \ldots, m$.

Samples from the joint posterior distribution of the parameters and the allocation variables are generated by iterating on the Gibbs sampling algorithm presented in Billio et al. (2012) where also the full conditional distributions can be found. We discard the first 1,000 Gibbs draws and keep 2,000 draws from which we use only every second draw to obtain uncorrelated draws for posterior inference. The $h$-step-ahead point forecast is the median of the predictive density produced with all 1,000 posterior draws using equation (5).

# References

Aguilar, Omar, and Mike West (2000). "Bayesian Dynamic Factor Models and Portfolio Allocation," *Journal of Business and Economic Statistics* 18: 338-357.

Arora, Vipin, and Jozef Lieskovsky (2014). "Natural Gas and U.S. Economic Activity," *The Energy Journal* 35(3): 167-182.

Bachmeier, Lance J., and James M. Griffin (2006). "Testing for Market Integration: Crude Oil, Coal, and Natural Gas," *The Energy Journal* 27(2): 55-71.

Baumeister, Christiane (2023). "Measuring Market Expectations," in: Rüdiger Bachmann, Giorgio Topa, and Wilbert van der Klaauw (eds.), *Handbook of Economic Expectations*, Academic Press, 413-442.

Baumeister, Christiane, Florian Huber, and Massimiliano Marcellino (2024). "Risky Oil: It's All in the Tails," NBER Working Paper 32524.

Baumeister, Christiane, and Lutz Kilian (2012). "Real-Time Forecasts of the Real Price of Oil," *Journal of Business and Economic Statistics* 30(2): 326-336.

Baumeister, Christiane, and Lutz Kilian (2014). "What Central Bankers Need to Know About Forecasting Oil Prices," *International Economic Review* 55(3): 869-889.

Baumeister, Christiane, and Lutz Kilian (2015). "Forecasting the Real Price of Oil in a Changing World: A Forecast Combination Approach," *Journal of Business and Economic Statistics* 33(3): 338-351.

Baumeister, Christiane, Lutz Kilian, and Thomas K. Lee (2017). "Inside the Crystal Ball: New Approaches to Predicting the Gasoline Price at the Pump," *Journal of Applied Econometrics* 32(2): 275-295.

Baumeister, Christiane, Dimitris Korobilis, and Thomas K. Lee (2022). "Energy Markets and Global Economic Conditions," *Review of Economics and Statistics* 104(4): 828-844.

Bhattacharya, Anirban, and David B. Dunson (2011). "Sparse Bayesian Infinite Factor Models," *Biometrika* 98: 291-306.

Billio, Monica, Roberto Casarin, Francesco Ravazzolo, and Herman K. van Dijk (2012). "Combination Schemes for Turning Point Predictions," *The Quarterly Review of Economics and Finance* 52(4): 402-412.

Brown, Stephen P.A., and Mine K. Yücel (2008). "What Drives Natural Gas Prices?" *The Energy Journal* 29(2): 45-60.

Carriero, Andrea, Todd E. Clark, and Massimiliano Marcellino (2019). "Large Vector Autoregressions with Stochastic Volatility and Non-Conjugate Priors," *Journal of Econometrics* 212(1): 137-154.

Carvalho, Carlos M., Nicholas G. Polson, and James G. Scott (2010). "The Horseshoe Estimator for Sparse Signals," *Biometrika* 97: 465-480.

Chinn, Menzie D., and Olivier Coibion (2014). "The Predictive Content of Commodity Futures," *Journal of Futures Markets* 34(7): 607–636.

Clark, Todd E. (2011). "Real-Time Density Forecasts From Bayesian Vector Autoregressions With Stochastic Volatility," *Journal of Business and Economic Statistics* 29(3): 327-341.

Clark, Todd E., and Francesco Ravazzolo (2015). "Macroeconomic Forecasting Performance Under Alternative Specifications of Time-Varying Volatility," *Journal of Applied Econometrics* 30(4): 551-575.

D'Agostino, Antonello, Luca Gambetti, and Domenico Giannone (2013). "Macroeconomic Forecasting and Structural Change," *Journal of Applied Econometrics* 28(1): 82-101.

Diebold, Francis X., and Roberto S. Mariano (1995). "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics* 13(3): 253-263.

Diebolt, Jean, and Christian P. Robert (1994). "Estimation of Finite Mixture Distributions through Bayesian Sampling," *Journal of the Royal Statistical Society B*, 56(2): 363-375.

Doan, Thomas, Robert B. Litterman, and Christopher A. Sims (1984). "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Reviews* 3(1): 1-100.

Ellwanger, Reinhard, and Stephen Snudden (2023). "Forecasts of the Real Price of Oil Revisited: Do They Beat the Random Walk?" *Journal of Banking and Finance* 154: 106962.

Faust, Jon, and Jonathan H. Wright (2013). "Forecasting Inflation," in: Graham Elliott and Allan Timmermann (eds.), *Handbook of Economic Forecasting*, Volume 2A, Amsterdam: North-Holland, 2-56.

Ferrari, Davide, Francesco Ravazzolo, and Joaquin Vespignani (2021). "Forecasting Energy Commodity Prices: A Large Global Dataset Sparse Approach," *Energy Economics* 98: 105268.

Frühwirth-Schnatter, Sylvia (2001). "Markov Chain Monte Carlo Estimation of Classical and Dynamic Switching and Mixture Models," *Journal of the American Statistical Association* 96 (453): 194-209.

Frühwirth-Schnatter, Sylvia (2006). *Mixture and Markov-swithing Models.* Springer, New York, 2006.

Gao, Shen, Chenghan Hou, and Bao H. Nguyen (2021). "Forecasting Natural Gas Prices Using Highly Flexible Time-Varying Parameter Models," *Economic Modelling* 105: 105652.

Giannone, Domenico, Michele Lenza, and Giorgio Primiceri (2015). "Prior Selection for Vector Autoregressions," *Review of Economics and Statistics* 97(2): 436-451.

Granger, Clive W.J., and Yongil Jeon (2004). "Thick Modeling," *Economic Modelling* 21: 323-343.

Halser, Christoph, Florentina Paraschiv, and Marianna Russo (2023). "Oil-Gas Price Relationships on Three Continents: Disruptions and Equilibria," *Journal of Commodity Markets* 31: 100347.

Hamilton, James D. (2009). "Understanding Crude Oil Prices," *The Energy Journal* 30(2): 179-206.

Hansen, Peter R., Asger Lunde, and James M. Nason (2011). "The Model Confidence Set," *Econometrica* 79(2): 453-497.

Hartley, Peter R., Kenneth B. Medlock, and Jennifer E. Rosthal (2008). "The Relationship of Natural Gas to Oil Prices," *The Energy Journal* 29(3): 47-66.

Huber, Florian, and Martin Feldkircher (2019). "Adaptive Shrinkage in Bayesian Vector Autoregressive Models," *Journal of Business and Economic Statistics* 37: 27-39.

Hupka, Yuri, Ivilina Popova, Betty Simkins, and Thomas Lee (2023). "A Review of the Literature on LNG: Hubs Development, Market Integration, and Price Discovery," *Journal of Commodity Markets* 31: 100349.

Joskow, Paul L. (2013). "Natural Gas: From Shortages to Abundance in the United States," *American Economic Review* 103(3): 338-343.

Kastner, Gregor (2019). "Sparse Bayesian Time-Varying Covariance Estimation in Many Dimensions," *Journal of Econometrics* 210: 98-115.

Kastner, Gregor, and Sylvia Frühwirth-Schnatter (2014). "Ancillarity-Sufficiency Interweaving Strategy (ASIS) for Boosting MCMC Estimation of Stochastic Volatility Models," *Computational Statistics and Data Analysis* 76: 408-423.

Kastner, Gregor, and Florian Huber (2020). "Sparse Bayesian Vector Autoregressions in Huge Dimensions," *Journal of Forecasting* 39(7): 1142-1165.

Pitt, Michael K., and Neil Shephard (1999). "Time-Varying Covariances: A Factor Stochastic Volatility Approach," in: Bernardo, Jose M., James O. Berger, A. Philip Dawid, and Adrian F.M. Smith (eds.), *Bayesian Statistics–Proceedings of the Sixth Valencia International Meeting*, Oxford University Press, 547-570.

Primiceri, Giorgio E. (2005). "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies* 72(3): 821-852.

Ramberg, David J., and John E. Parsons (2012). "The Weak Tie Between Natural Gas and Oil Prices," *The Energy Journal*, 33(2): 13-35.

Rubaszek, Michał, Karol Szafranek, and Gazi Salah Uddin (2021). "The Dynamics and Elasticities on the U.S. Natural Gas Market. A Bayesian Structural VAR Analysis," *Energy Economics* 103: 105526.

Stock, James H., and Matthew Zaragoza-Watkins (2024). "The Market and Climate Implications of U.S. LNG Exports," NBER Working Paper 32228.

Tanner, Martin A., and Wing Hung Wong (1987). "The Calculation of Posterior Distributions by Data Augmentation," *Journal of the American Statistical Association* 82: 528-550.

Timmermann, Allan (2006). "Forecast Combinations," in: Graham Elliott, Clive W.J. Granger, and Allan Timmermann (eds.), *Handbook of Economic Forecasting*, Volume 1, Amsterdam: North-Holland, 135-196.

Wiggins, Seth, and Xiaoli L. Etienne (2017). "Turbulent Times: Uncovering the Origins of US Natural Gas Price Fluctuations Since Deregulation," *Energy Economics* 64: 196-205.

Winkler, Valentin (2023). "What Drives Drilling Up and Prices Down?" IMK Working Paper 217.

**Figure 1. Monthly Spot Prices for Henry Hub Natural Gas and West Texas Intermediate Crude Oil, 1993.11-2024.2**

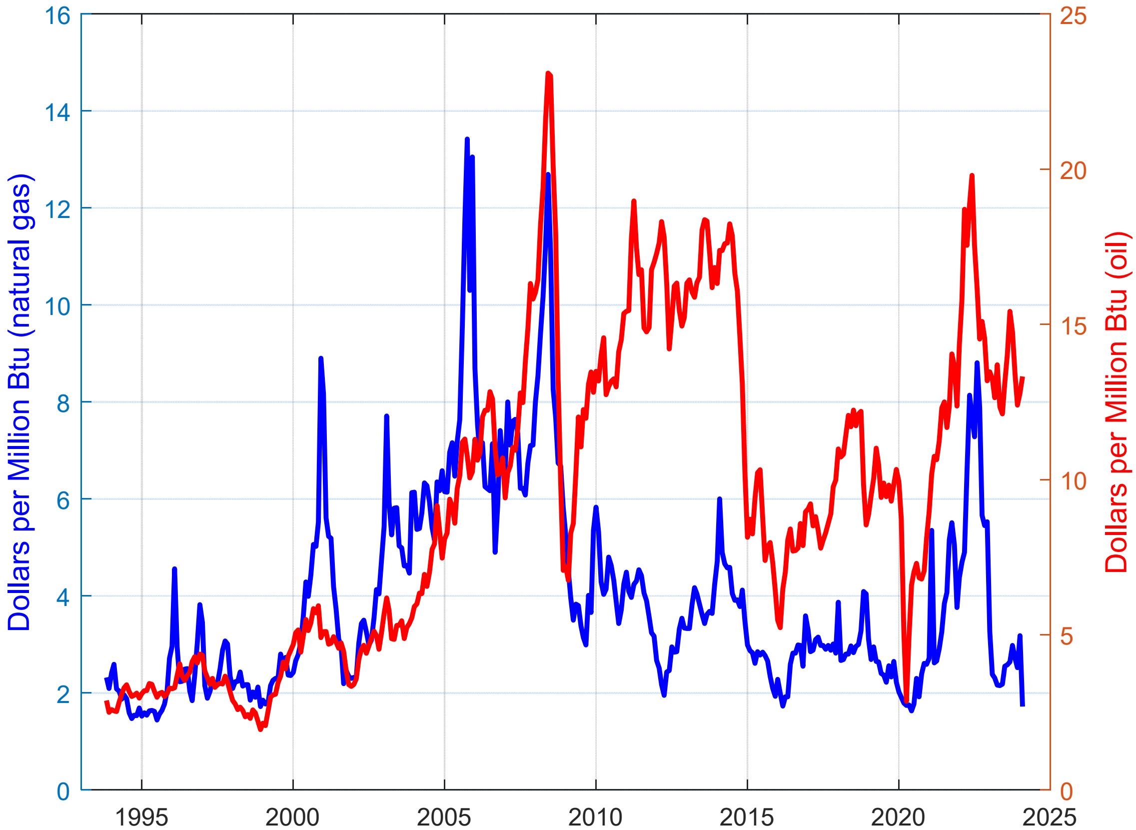
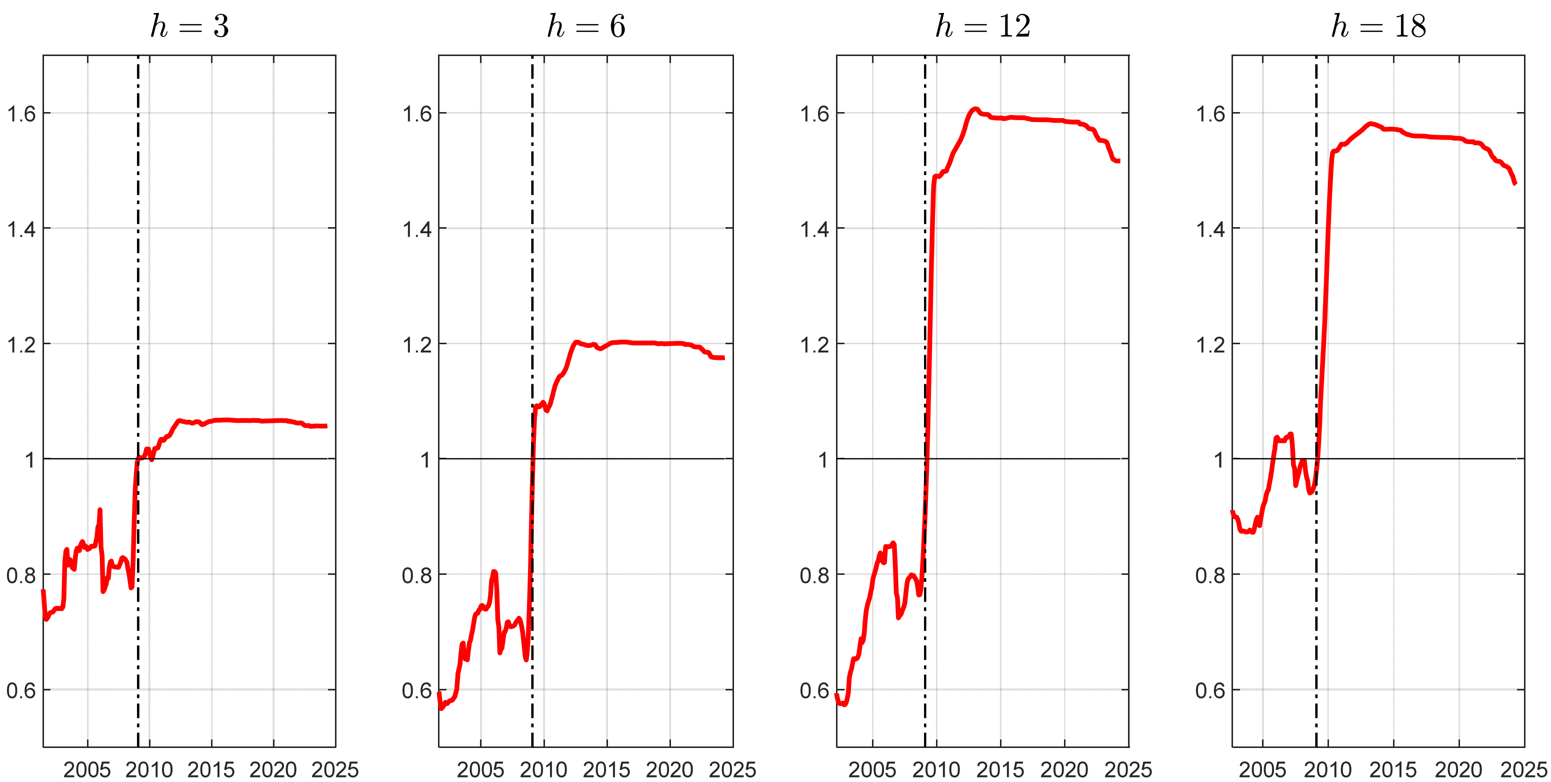
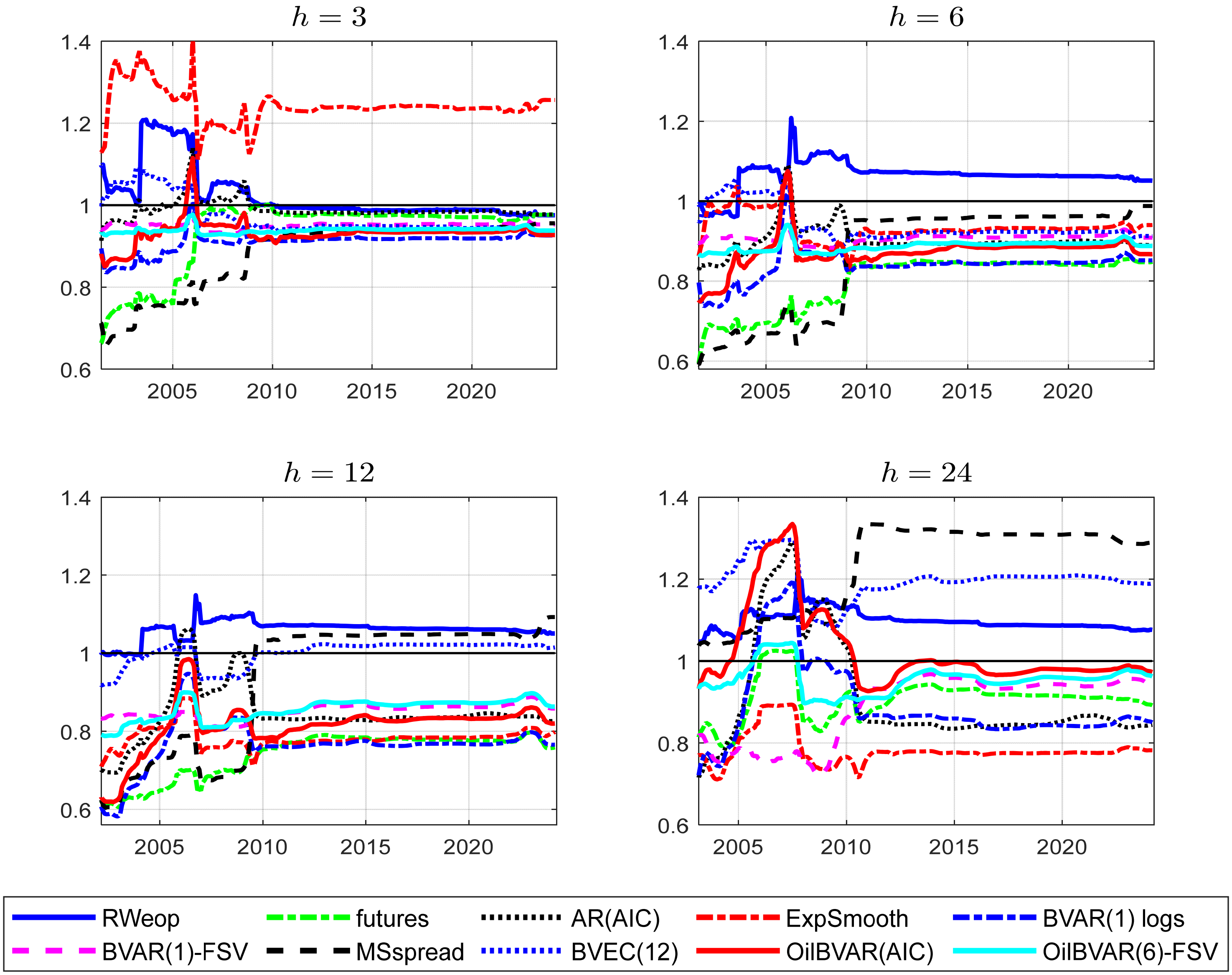


**Figure 2. Evolution of Real-Time Cumulative MSPE Ratios for Price Spread Model Forecast Relative to No-Change Forecast at Selected Horizons**




NOTES: A ratio below 1 indicates that the price spread model outperforms the no-change forecast. To allow the MSPE ratio to stabilize, we skip the first 50 forecast periods. The vertical line indicates January 2009 which is the estimated date of the breakdown of the cointegration relationship between the Henry Hub natural gas price and the WTI oil price reported in Table 1 of Stock and Zaragoza-Watkins (2024).

**Figure 3. Real-Time Cumulative MSPE Ratios Relative to Monthly No-Change Forecast for a Subset of Models at Selected Horizons**




NOTES: A ratio below 1 indicates that the model-based forecast outperforms the no-change forecast. To allow the MSPE ratio to stabilize, we skip the first 50 forecast periods.

**Table 1. Average MSPE Ratios Relative to the No-Change Forecast of the Real Natural Gas Spot Price for Model-Free Forecasting Approaches**

| | End-of-month No-Change Forecasts | | Futures-Market-Based Forecasts | | US EIA Expert Forecasts |
|---|---|---|---|---|---|
| Monthly horizon | (1) 1997.2-2024.2 | (2) 2005.9-2024.2 | (3) 1997.2-2024.2 | (4) 2005.9-2024.2 | (5) 2005.9-2024.2 |
| 1 | **0.671*** | **0.439**** | **0.897**** | **0.891**** | 1.631 |
| 3 | **0.976** | **0.852** | **0.978** | 1.068 | 1.136 |
| 6 | 1.051 | 1.044 | **0.848*** | **0.941** | **0.796** |
| 9 | 1.046 | 1.040 | **0.792**** | **0.849** | **0.645**** |
| 12 | 1.051 | 1.065 | **0.758**** | **0.802*** | **0.615**** |
| 15 | 1.054 | 1.070 | **0.811**** | **0.849** | – |
| 18 | 1.072 | 1.073 | **0.838*** | **0.851** | – |
| 21 | 1.065 | 1.050 | **0.859** | **0.813** | – |
| 24 | 1.076 | 1.050 | **0.892** | **0.799** | – |

NOTES: Boldface indicates improvements relative to the monthly no-change forecast. Significant at **5% and *10% based on the Diebold-Mariano (1995) test. Expert forecasts for the nominal Henry Hub spot price are reported in the US EIA's *Short-Term Energy Outlook* from August 2005 onward and consistently cover a maximum horizon of 12 months ahead.

**Table 2. Average MSPE Ratios Relative to No-Change Forecast of Real Natural Gas Spot Price for a Set of Univariate Models**
Evaluation Period: 1997.2-2024.2

| Monthly horizon | ARMA(1,1) (1) | IMA(1) (2) | ARIMA(1,1) (3) | AR(12) (4) | BAR(12) (5) | AR(AIC) (6) | BAR(AIC) (7) | AR(1) (8) | BAR(1) (9) | Exponential Smoothing (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.019 | 1.049 | 1.049 | 1.076 | 1.059 | 1.043 | 1.033 | **0.987** | **0.987** | 2.121 |
| 3 | **0.962** | 1.042 | 1.038 | 1.071 | 1.038 | **0.978** | **0.976** | **0.956** | **0.956** | 1.257 |
| 6 | **0.895** | 1.055 | 1.040 | 1.031 | **0.987** | **0.891** | **0.891** | **0.900** | **0.899** | **0.941** |
| 9 | **0.854** | 1.083 | 1.060 | **0.999** | **0.946** | **0.849** | **0.849** | **0.855** | **0.854** | **0.819***[*] |
| 12 | **0.843** | 1.099 | 1.073 | 1.001 | **0.937** | **0.827** | **0.828** | **0.837** | **0.837** | **0.798**[**] |
| 15 | **0.857** | 1.122 | 1.088 | **0.980** | **0.926** | **0.825** | **0.828** | **0.836** | **0.836** | **0.802**[**] |
| 18 | **0.874** | 1.141 | 1.106 | **0.987** | **0.934** | **0.830** | **0.834** | **0.843** | **0.843** | **0.808**[**] |
| 21 | **0.888** | 1.163 | 1.141 | **0.995** | **0.940** | **0.836** | **0.841** | **0.853** | **0.854** | **0.792**[**] |
| 24 | **0.915** | 1.190 | 1.175 | **0.966** | **0.923** | **0.844** | **0.850** | **0.871** | **0.871** | **0.782**[**] |

NOTES: Boldface indicates improvements relative to the no-change forecast. BAR refers to AR models estimated using the Bayesian method of Giannone et al. (2015). The AIC lag order estimates are based on an upper bound of 6 lags for parsimony. The exponential smoothing forecasts are based on a weight of 0.8. The statistical significance of MSPE reductions is only assessed for the exponential smoothing forecasts based on the Diebold-Mariano test with significance at **5% and *10%. All forecasts are generated recursively from data subject to real-time data constraints.

**Table 3. Average MSPE Ratios Relative to the No-Change Forecast of the Real Natural Gas Spot Price Based on a Comprehensive Economic Model of the U.S. Natural Gas Market: The Roles of Lag Length and Variable Transformation**

Evaluation Period: 1997.2-2024.2

| Monthly horizon | VAR(12) | BVAR(12) | VAR(AIC) | BVAR(AIC) | VAR(1) | BVAR(1) |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | PANEL A: Predictor Variables in Monthly Growth Rates | | | | | |
| 1 | 1.179 | 1.028 | 1.126 | 1.015 | **0.968** | **0.969** |
| 3 | 1.054 | **0.950** | **0.995** | **0.954** | **0.935** | **0.935** |
| 6 | 1.089 | **0.923** | **0.943** | **0.900** | **0.863** | **0.865** |
| 9 | 1.082 | **0.873** | **0.872** | **0.836** | **0.822** | **0.824** |
| 12 | 1.213 | **0.863** | **0.864** | **0.825** | **0.799** | **0.801** |
| 15 | 1.175 | **0.858** | **0.859** | **0.824** | **0.798** | **0.798** |
| 18 | 1.144 | **0.871** | **0.865** | **0.834** | **0.798** | **0.797** |
| 21 | 1.187 | **0.871** | **0.876** | **0.847** | **0.811** | **0.810** |
| 24 | 1.106 | **0.874** | **0.893** | **0.867** | **0.836** | **0.833** |
| | PANEL B: Predictor Variables in Log-Levels | | | | | |
| 1 | 1.169 | 1.035 | 1.110 | 1.006 | **0.970** | **0.969** |
| 3 | 1.078 | 1.029 | **1.000** | **0.972** | **0.926** | **0.926** |
| 6 | 1.305 | 1.122 | **0.952** | **0.898** | **0.845** | **0.852** |
| 9 | 1.525 | 1.198 | **0.881** | **0.827** | **0.794** | **0.804** |
| 12 | 1.857 | 1.228 | **0.870** | **0.810** | **0.758** | **0.766** |
| 15 | 1.777 | 1.255 | **0.887** | **0.827** | **0.762** | **0.767** |
| 18 | 1.677 | 1.282 | **0.906** | **0.852** | **0.783** | **0.785** |
| 21 | 1.450 | 1.293 | **0.894** | **0.856** | **0.817** | **0.819** |
| 24 | 1.432 | 1.343 | **0.918** | **0.883** | **0.851** | **0.850** |

NOTES: Boldface indicates improvements relative to the no-change forecast. BVAR refers to VAR models estimated using the Bayesian method of Giannone et al. (2015). The AIC lag order estimates are based on an upper bound of 6 lags. All forecasts are generated recursively from data subject to real-time data constraints.

**Table 4. Average MSPE Ratios Relative to the No-Change Forecast of the Real Natural Gas Spot Price Based on Six-Variable BVAR(1) Models of the U.S. Natural Gas Market:**
**The Role of Different Proxies for Weather and the Business Cycle**

Evaluation Period: 1997.2-2024.2

| Monthly horizon | Economic Model | US Meteorological Conditions | | | US Macroeconomic Conditions | |
|---|---|---|---|---|---|---|
| | Baseline predictors in logs | Temperature anomalies | HDD deviations | CDD deviations | Industrial production | CFNAI |
| | (1) | (2) | (3) | (4) | (6) | (7) |
| 1 | **0.969** | **0.973** | **0.966** | **0.969** | **0.958** | **0.968** |
| 3 | **0.926** | **0.932** | **0.926** | **0.927** | **0.908** | **0.925** |
| 6 | **0.852** | **0.857** | **0.851** | **0.854** | **0.852** | **0.854** |
| 9 | **0.804** | **0.807** | **0.801** | **0.804** | **0.824** | **0.808** |
| 12 | **0.766** | **0.768** | **0.764** | **0.766** | **0.801** | **0.772** |
| 15 | **0.767** | **0.768** | **0.765** | **0.767** | **0.810** | **0.774** |
| 18 | **0.785** | **0.787** | **0.784** | **0.786** | **0.838** | **0.792** |
| 21 | **0.819** | **0.821** | **0.817** | **0.819** | **0.880** | **0.824** |
| 24 | **0.850** | **0.852** | **0.848** | **0.850** | **0.924** | **0.853** |

NOTES: Boldface indicates improvements relative to the no-change forecast. HDD refers to nationwide, population-weighted heating degree days and CDD to nationwide, population-weighted cooling degree days, both in deviations of the average of the same month for all available past years. The alternative indicators of meteorological conditions replace the Residential Energy Demand Temperature Index (REDTI) and the alternative business cycle indicators replace the capacity utilization rate in the six-variable BVAR(1) model in column (1) with industrial production entering in log-levels. All forecasts are generated recursively from data subject to real-time data constraints.

**Table 5. Average MSPE Ratios Relative to the No-Change Forecast of the Real Natural Gas Spot Price for Baseline and Other BVAR(1) Models of the U.S. Natural Gas Market With and Without Stochastic Volatility**

Evaluation Period: 1997.2-2024.2

| Monthly horizon | Baseline | | Model 1 | | Model 2 | | Model 3 | | Model 4 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| | BVAR | BVAR-FSV | BVAR | BVAR-FSV | BVAR | BVAR-FSV | BVAR | BVAR-FSV | BVAR | BVAR-FSV |
| 1 | **0.969** | **0.978** | **0.981** | **0.981** | 1.016 | **0.985** | **0.960** | **0.989** | **0.967** | **0.984** |
| 3 | **0.926** | **0.953** | **0.950** | **0.950** | **0.961** | **0.954** | **0.909** | **0.966** | **0.920** | **0.950** |
| 6 | **0.852** | **0.909** | **0.903** | **0.913** | **0.904** | **0.920** | **0.861** | **0.936** | **0.853** | **0.914** |
| 9 | **0.804** | **0.876** | **0.866** | **0.874** | **0.858** | **0.887** | **0.827** | **0.914** | **0.813** | **0.882** |
| 12 | **0.766** | **0.858** | **0.854** | **0.851** | **0.844** | **0.865** | **0.820** | **0.903** | **0.798** | **0.864** |
| 15 | **0.767** | **0.857** | **0.862** | **0.832** | **0.842** | **0.841** | **0.828** | **0.906** | **0.800** | **0.861** |
| 18 | **0.785** | **0.870** | **0.875** | **0.816** | **0.849** | **0.830** | **0.850** | **0.913** | **0.811** | **0.866** |
| 21 | **0.819** | **0.891** | **0.887** | **0.796** | **0.854** | **0.807** | **0.872** | **0.937** | **0.839** | **0.897** |
| 24 | **0.850** | **0.944** | **0.906** | **0.777** | **0.870** | **0.792** | **0.906** | **0.979** | **0.880** | **0.951** |

NOTES: Boldface indicates improvements relative to the no-change forecast. All models contain the real price of natural gas in logs. The additional predictor variables are as follows: (a) Baseline model: log of dry natural gas production, log of working gas inventories, log of natural gas rig counts, capacity utilization, and REDTI; (b) Model 1: annual growth rates of marketed natural gas production, of residential natural gas consumption, and of industrial production; (c) Model 2: monthly growth rates of dry natural gas production, of working gas inventories, and of industrial production, all three deseasonalized; (d) Model 3: log of marketed natural gas production, log of natural gas rig counts, log of industrial production, heating and cooling degree days in deviations from their historical average; (e) Model 4: log of marketed natural gas production, log of natural gas rig counts, and log of total natural gas consumption. All forecasts are generated recursively from data subject to real-time data constraints.

**Table 6. Average MSPE Ratios Relative to No-Change Forecast for Price Spread Models**

Evaluation Period: 1997.2-2024.2

| Monthly horizon | Constant-Coefficient Models | | Markov-Switching Model |
|---|---|---|---|
| | $\hat{\alpha}, \hat{\beta}$ | $\alpha = 0, \hat{\beta}$ | $\hat{\alpha}^{MS}, \hat{\beta}^{MS}$ |
| 1 | **0.987** | **0.992** | 1.270 |
| 3 | 1.057 | 1.032 | **0.956** |
| 6 | 1.175 | 1.117 | **0.988** |
| 9 | 1.369 | 1.233 | 1.042 |
| 12 | 1.516 | 1.337 | 1.093 |
| 15 | 1.547 | 1.382 | 1.176 |
| 18 | 1.483 | 1.371 | 1.290 |
| 21 | 1.363 | 1.302 | 1.357 |
| 24 | 1.332 | 1.281 | 1.290 |

NOTES: Boldface indicates improvements relative to the no-change forecast. All forecasts are generated recursively from data subject to real-time data constraints.

**Table 7. Average MSPE Ratios Relative to No-Change Forecast of Bivariate VAR Models for Real Price of Natural Gas and Real WTI Oil Price**

Evaluation Period: 1997.2-2024.2

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | PANEL A: The Role of the Lag Length | | | | | |
| Monthly Horizon | VAR(12) | BVAR(12) | VAR(AIC) | BVAR(AIC) | VAR(1) | BVAR(1) |
| 1 | 1.033 | **0.991** | 1.031 | **0.986** | **0.985** | **0.998** |
| 3 | 1.006 | **0.947** | **0.964** | **0.951** | **0.962** | **0.994** |
| 6 | **0.971** | **0.891** | **0.923** | **0.914** | **0.925** | **0.957** |
| 9 | **0.968** | **0.871** | **0.921** | **0.908** | **0.915** | **0.944** |
| 12 | 1.010 | **0.899** | **0.957** | **0.944** | **0.949** | **0.960** |
| 15 | 1.033 | **0.941** | 1.010 | **0.998** | 1.003 | **0.993** |
| 18 | 1.062 | **0.978** | 1.048 | 1.042 | 1.048 | 1.031 |
| 21 | 1.074 | **0.998** | 1.065 | 1.066 | 1.082 | 1.059 |
| 24 | 1.081 | 1.040 | 1.107 | 1.116 | 1.141 | 1.108 |
| | PANEL B: The Role of Cointegration | | | | | |
| | VEC(12) | BVEC(12) | VEC(AIC) | BVEC(AIC) | VEC(1) | BVEC(1) |
| 1 | 1.052 | 1.023 | 1.026 | 1.021 | 1.014 | 1.013 |
| 3 | 1.027 | **0.937** | **0.974** | **0.969** | **0.976** | **0.976** |
| 6 | 1.043 | **0.913** | **0.945** | **0.946** | **0.968** | **0.968** |
| 9 | 1.098 | **0.939** | **0.968** | **0.971** | 1.015 | 1.014 |
| 12 | 1.205 | 1.015 | 1.029 | 1.034 | 1.109 | 1.108 |
| 15 | 1.267 | 1.078 | 1.107 | 1.114 | 1.229 | 1.228 |
| 18 | 1.327 | 1.133 | 1.181 | 1.188 | 1.337 | 1.334 |
| 21 | 1.381 | 1.152 | 1.219 | 1.228 | 1.410 | 1.406 |
| 24 | 1.372 | 1.189 | 1.283 | 1.297 | 1.530 | 1.525 |

NOTES: Boldface indicates improvements relative to the no-change forecast. BVAR refers to VAR models estimated using the Bayesian method of Giannone et al. (2015). The AIC lag order estimates are based on an upper bound of 6 lags. All forecasts are generated recursively from data subject to real-time data constraints.

**Table 8. Average MSPE Ratios Relative to No-Change Forecast for Economic Model Augmented with WTI Oil Price with and without Factor Stochastic Volatility**

Evaluation Period: 1997.2-2024.2

| Monthly horizon | BVAR(12) (1) | BVAR(AIC) (2) | BVAR(1) (3) | BVAR(12)-FSV (4) | BVAR(6)-FSV (5) | BVAR(1)-FSV (6) |
|---|---|---|---|---|---|---|
| 1 | 1.002 | **0.992** | **0.971** | **0.980** | **0.984** | **0.988** |
| 3 | **0.948** | **0.927** | **0.936** | **0.938** | **0.938** | **0.967** |
| 6 | **0.970** | **0.868** | **0.873** | **0.904** | **0.888** | **0.936** |
| 9 | **0.950** | **0.823** | **0.844** | **0.893** | **0.866** | **0.917** |
| 12 | **0.935** | **0.820** | **0.839** | **0.891** | **0.863** | **0.913** |
| 15 | **0.933** | **0.853** | **0.880** | **0.909** | **0.876** | **0.927** |
| 18 | **0.978** | **0.899** | **0.931** | **0.933** | **0.888** | **0.955** |
| 21 | 1.010 | **0.921** | **0.987** | **0.968** | **0.909** | **0.992** |
| 24 | 1.078 | **0.973** | 1.055 | 1.012 | **0.963** | 1.075 |

NOTES: Boldface indicates improvements relative to the no-change forecast. The AIC lag order is determined recursively based on an upper bound of 6 lags. All forecasts are generated recursively from data subject to real-time data constraints.

**Table 9. Real-Time Accuracy of Forecast Combinations**

Evaluation Period: 1997.2-2024.2

| Monthly horizon | All Models | | Models in $\widehat{M}^{*}_{75\%}$ set | | Models with highest $p$-value: recursive | | Models with highest $p$-value: rolling | |
|---|---|---|---|---|---|---|---|---|
| | Equal weights | Inverse MSPE weights | Equal weights | Inverse MSPE weights | Equal weights | Inverse MSPE weights | Equal weights | Inverse MSPE weights |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| 1 | **0.956** | **0.949** | **0.944** | **0.936** | **0.831** | **0.814** | **0.491** | **0.488** |
| 3 | **0.913** | **0.908** | **0.910** | **0.905** | **0.839** | **0.838** | **0.655** | **0.654** |
| 6 | **0.867** | **0.856** | **0.857** | **0.847** | **0.760** | **0.759** | **0.583** | **0.583** |
| 9 | **0.837** | **0.822** | **0.813** | **0.801** | **0.711** | **0.711** | **0.531** | **0.531** |
| 12 | **0.833** | **0.814** | **0.766** | **0.758** | **0.699** | **0.698** | **0.522** | **0.522** |
| 15 | **0.843** | **0.822** | **0.753** | **0.746** | **0.724** | **0.724** | **0.498** | **0.498** |
| 18 | **0.856** | **0.833** | **0.749** | **0.742** | **0.713** | **0.713** | **0.476** | **0.476** |
| 21 | **0.866** | **0.840** | **0.770** | **0.764** | **0.756** | **0.755** | **0.473** | **0.473** |
| 24 | **0.881** | **0.852** | **0.773** | **0.766** | **0.718** | **0.717** | **0.459** | **0.459** |

NOTES: Average MSPE ratios of pooled forecasts relative to the no-change forecast of the real natural gas spot price. Boldface indicates improvements relative to the no-change forecast. Models that enter the forecast combinations are selected in real time by the MCS procedure with 10,000 block bootstrap replications using a block size of 12 from the entire universe of 60 individual models.

**Table 1A. Components of Real-Time Dataset for the Analysis of the U.S. Natural Gas Market**

| Data category | Variable | First vintage | Start date | Unit | Data source | Data delay | Nowcast rule | Data revisions |
|---|---|---|---|---|---|---|---|---|
| Natural gas market fundamentals | Henry Hub spot price | n/a | 1976.1 | dollars/million Btu | EIA website, FRED | 0 | n/a | N |
| | Dry natural gas production | 1978.12 | 1973.1 | billion cubic feet | MER (Table 4.1) | 1,2,3 | ALD | Y |
| | Marketed natural gas production | 1976.8 | 1973.1 | billion cubic feet | MER (Table 4.1) | 1,2,3 | ALD | Y |
| | Natural gas working storage | 1979.2 | 1975.9 | billion cubic feet | MER (Table 4.4) | 1,2,3 | MC | Y |
| | Total natural gas consumption | 1976.8 | 1973.1 | billion cubic feet | MER (Table 4.3) | 1,3 | RWSM | Y |
| | Residential natural gas consumption | 1983.12 | 1973.1 | billion cubic feet | MER (Table 4.3) | 1,3 | RWSM | Y |
| | Natural gas drilling rigs in operation | 1993.3 | 1987.8 | number | MER (Table 5.1), BH | 0 | n/a | N |
| Weather conditions | Residential Energy Demand Temperature Index (REDTI) | n/a | 1973.1 | index | NOAA website | 1 | ASM | N |
| | Heating degree days (HDD) | 2015.9 | 1973.1 | degree days | MER (Table 1.10) | 3 | ASM | Y |
| | Cooling degree days (CDD) | 2015.9 | 1973.1 | degree days | MER (Table 1.11) | 3 | ASM | Y |
| | National temperature anomalies | n/a | 1973.1 | index | NOAA website | 0 | n/a | N |
| Macroeconomic environment | Consumer price index for all urban consumers | 1991.1 | 1973.1 | index | FRASER, BK (2012) database | 1 | AMG | Y |
| | Industrial production | 1927.1 | 1973.1 | index | ALFRED | 1 | AMG | Y |
| | Capacity utilization rate | 1996.11 | 1973.1 | percent | ALFRED | 1 | ES | Y |
| | Chicago Fed National Activity Index (CFNAI), 3-month moving average | 2001.1 | 1973.1 | index | Federal Reserve Bank of Chicago website | 1 | ES | Y |

NOTES: First vintage refers to when the actual vintages started (either used for backcasting data from the MER or effective vintages available). The start date indicates the first available observation; for series available before 1973.1, the start date is capped at 1973.1. Degree days are defined as the difference between the mean daily temperature and the threshold of 65°F. The delay in data release is measured in months. More than one number for data delay indicates that the number of missing observations has changed across vintages. Nowcasts are based on the annual log difference from the preceding year (ALD), the average monthly growth rate (AMG), the average value in the same month over the past 10 years (ASM), exponential smoothing (ES) with smoothing parameter $\alpha = 0.95$, the month-on-month change in the preceding year (MC), and value in the same month the year before (RWSM). The abbreviations for the data sources are as follows: BH – Baker Hughes (https://rigcount.bakerhughes.com/na-rig-count); EIA – Energy Information Administration; MER – Monthly Energy Review; NOAA – National Oceanic and Atmospheric Administration. The table numbers in the MER change over time; the numbers reported here refer to the tables in the real-time vintages released in electronic form by the EIA since January 2024.

**Table 2A. Average MSPE Ratios Relative to the No-Change Forecast of the Real Natural Gas Spot Price Based on Six-Variable BVAR(1) Models of the US Natural Gas Market: The Role of Excluding COVID-19 Observations**

Evaluation Period: 1997.2-2019.12

| Monthly horizon | Economic Model | US Macroeconomic Conditions | |
|---|---|---|---|
| | Baseline predictors in logs | Industrial production | CFNAI |
| | (1) | (2) | (3) |
| 1 | **0.966** | **0.954** | **0.965** |
| 3 | **0.919** | **0.899** | **0.917** |
| 6 | **0.846** | **0.847** | **0.848** |
| 9 | **0.802** | **0.824** | **0.808** |
| 12 | **0.767** | **0.806** | **0.775** |
| 15 | **0.771** | **0.820** | **0.780** |
| 18 | **0.791** | **0.850** | **0.798** |
| 21 | **0.816** | **0.883** | **0.821** |
| 24 | **0.843** | **0.924** | **0.845** |

NOTES: Boldface indicates improvements relative to the no-change forecast. The alternative business cycle indicators replace the capacity utilization rate in the six-variable BVAR(1) model in column (1). Industrial production enters in log-levels. All forecasts are generated recursively from data subject to real-time data constraints.

**Table 3A. Average MSPE Ratios Relative to the No-Change Forecast of the Real Natural Gas Spot Price for Other BVAR(1) Models of the U.S. Natural Gas Market**

Evaluation Period: 1997.2-2024.2

| Monthly horizon | Model 1 | | Model 2 | | Model 3 | | Model 4 | |
|---|---|---|---|---|---|---|---|---|
| | $p = 1$ | AIC (6) | $p = 1$ | AIC (6) | $p = 1$ | AIC (6) | $p = 1$ | AIC (6) |
| 1 | **0.981** | **0.998** | 1.016 | 1.024 | **0.960** | **0.959** | **0.967** | 1.042 |
| 3 | **0.950** | **0.965** | **0.961** | **0.968** | **0.909** | **0.884** | **0.920** | **0.967** |
| 6 | **0.903** | **0.905** | **0.904** | **0.903** | **0.861** | **0.831** | **0.853** | **0.895** |
| 9 | **0.866** | **0.870** | **0.858** | **0.863** | **0.827** | **0.806** | **0.813** | **0.844** |
| 12 | **0.854** | **0.856** | **0.844** | **0.852** | **0.820** | **0.807** | **0.798** | **0.826** |
| 15 | **0.862** | **0.871** | **0.842** | **0.861** | **0.828** | **0.824** | **0.800** | **0.834** |
| 18 | **0.875** | **0.883** | **0.849** | **0.871** | **0.850** | **0.851** | **0.811** | **0.847** |
| 21 | **0.887** | **0.894** | **0.854** | **0.879** | **0.872** | **0.879** | **0.839** | **0.848** |
| 24 | **0.906** | **0.918** | **0.870** | **0.898** | **0.906** | **0.926** | **0.880** | **0.878** |

NOTES: Boldface indicates improvements relative to the no-change forecast. All models contain the real price of natural gas in logs. The additional predictor variables are as follows: (a) Baseline model: log of dry natural gas production, log of working gas inventories, log of natural gas rig counts, capacity utilization, and REDTI; (b) Model 1: annual growth rates of marketed natural gas production, of residential natural gas consumption, and of industrial production; (c) Model 2: monthly growth rates of dry natural gas production, of working gas inventories, and of industrial production, all three deseasonalized; (d) Model 3: log of wet natural gas production, log of natural gas rig counts, log of industrial production, heating and cooling degree days in deviations from their historical average; (e) Model 4: log of wet natural gas production, log of natural gas rig counts, and log of total natural gas consumption. AIC(6) indicates that the lag length has been recursively determined based on the Akaike Information Criterion with an upper bound of 6. All forecasts are generated recursively from data subject to real-time data constraints.

**Table 4A. Average MSPE Ratios Relative to the No-Change Forecast of the Real Natural Gas Spot Price for Baseline and Other BVAR(1) Models of the U.S. Natural Gas Market With and Without Stochastic Volatility**

Evaluation Period: 1997.2-2024.2

| | Baseline | | Model 1 | | Model 2 | | Model 3 | | Model 4 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Monthly horizon | BVAR-F | BVAR-FSV | BVAR-F | BVAR-FSV | BVAR-F | BVAR-FSV | BVAR-F | BVAR-FSV | BVAR-F | BVAR-FSV |
| 1 | **0.983** | **0.978** | **0.987** | **0.981** | 1.003 | **0.985** | **0.998** | **0.989** | **0.987** | **0.984** |
| 3 | **0.965** | **0.953** | **0.955** | **0.950** | **0.957** | **0.954** | **0.980** | **0.966** | **0.951** | **0.950** |
| 6 | **0.932** | **0.909** | **0.903** | **0.913** | **0.897** | **0.920** | **0.963** | **0.936** | **0.893** | **0.914** |
| 9 | **0.906** | **0.876** | **0.860** | **0.874** | **0.851** | **0.887** | **0.938** | **0.914** | **0.851** | **0.882** |
| 12 | **0.871** | **0.858** | **0.853** | **0.851** | **0.843** | **0.865** | **0.921** | **0.903** | **0.834** | **0.864** |
| 15 | **0.860** | **0.857** | **0.863** | **0.832** | **0.851** | **0.841** | **0.908** | **0.906** | **0.823** | **0.861** |
| 18 | **0.865** | **0.870** | **0.883** | **0.816** | **0.864** | **0.830** | **0.901** | **0.913** | **0.829** | **0.866** |
| 21 | **0.883** | **0.891** | **0.897** | **0.796** | **0.876** | **0.807** | **0.902** | **0.937** | **0.844** | **0.897** |
| 24 | **0.892** | **0.944** | **0.921** | **0.777** | **0.896** | **0.792** | **0.900** | **0.979** | **0.872** | **0.951** |

NOTES: Boldface indicates improvements relative to the no-change forecast. BVAR-F indicates the homoskedastic BVAR with factor structure on the covariance matrix for parsimony. All models contain the real price of natural gas in logs. The additional predictor variables are as follows: (a) Baseline model: log of dry natural gas production, log of working gas inventories, log of natural gas rig counts, capacity utilization, and REDTI; (b) Model 1: annual growth rates of marketed natural gas production, of residential natural gas consumption, and of industrial production; (c) Model 2: monthly growth rates of dry natural gas production, of working gas inventories, and of industrial production, all three deseasonalized; (d) Model 3: log of marketed natural gas production, log of natural gas rig counts, log of industrial production, heating and cooling degree days in deviations from their historical average; (e) Model 4: log of marketed natural gas production, log of natural gas rig counts, and log of total natural gas consumption. All forecasts are generated recursively from data subject to real-time data constraints.

**Table 5A. Model Confidence Sets for Real Natural Gas Price Forecasts**
**Evaluation Period: 1997.2-2024.2**

| No. | Models | Monthly horizon | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 3 | 6 | 9 | 12 | 18 | 24 |
| (1) | Benchmark Model: random walk | 0.532** | 0.776** | 0.560** | 0.491** | 0.365** | 0.415** | 0.437** |
| | **Model-free Forecasts** | | | | | | | |
| (2) | End-of month no-change forecast | **1.000**** | 0.963** | 0.469** | 0.326** | 0.230* | 0.238* | 0.303** |
| (3) | Futures-market-based forecast | 0.632** | 0.965** | 0.979** | **1.000**** | 0.997** | 0.950** | 0.667** |
| | **Univariate Models** | | | | | | | |
| (4) | ARMA(1,1) | 0.500** | 0.948** | 0.860** | 0.915** | 0.648** | 0.738** | 0.492** |
| (5) | IMA(1) | 0.335** | 0.677** | 0.451** | 0.312** | 0.223* | 0.248* | 0.220* |
| (6) | ARIMA(1,1) | 0.359** | 0.707** | 0.510** | 0.395** | 0.262** | 0.307** | 0.264** |
| (7) | AR(12) | 0.312** | 0.613** | 0.489** | 0.441** | 0.250** | 0.320** | 0.354** |
| (8) | BAR(12) | 0.346** | 0.649** | 0.612** | 0.523** | 0.349** | 0.450** | 0.438** |
| (9) | AR(AIC) | 0.327** | 0.835** | 0.860** | 0.940** | 0.676** | 0.897** | 0.739** |
| (10) | BAR(AIC) | 0.370** | 0.852** | 0.860** | 0.932** | 0.674** | 0.867** | 0.739** |
| (11) | AR(1) | 0.613** | 0.944** | 0.833** | 0.810** | 0.591** | 0.807** | 0.667** |
| (12) | BAR(1) | 0.613** | 0.960** | 0.833** | 0.814** | 0.609** | 0.826** | 0.667** |
| (13) | Exponential Smoothing | 0.153* | 0.472** | 0.821** | 0.950** | 0.769** | 0.950** | 0.956** |
| | **Natural Gas Market Models** | | | | | | | |
| | *Baseline predictors in growth rates*: | | | | | | | |
| (14) | VAR(12) | 0.402** | 0.817** | 0.636** | 0.647** | 0.438** | 0.300** | 0.351** |
| (15) | BVAR(12) | 0.617** | 0.969** | 0.833** | 0.810** | 0.520** | 0.587** | 0.667** |
| (16) | VAR(AIC) | 0.458** | 0.880** | 0.708** | 0.770** | 0.511** | 0.685** | 0.600** |
| (17) | BVAR(AIC) | 0.632** | 0.969** | 0.860** | 0.950** | 0.676** | 0.847** | 0.667** |
| (18) | VAR(1) | 0.632** | 0.988** | 0.932** | 0.950** | 0.769** | 0.950** | 0.739** |
| (19) | BVAR(1) | 0.632** | 0.988** | 0.932** | 0.950** | 0.769** | 0.950** | 0.739** |
| | *Baseline predictors in logs*: | | | | | | | |
| (20) | VAR(12) | 0.378** | 0.765** | 0.417** | 0.274** | 0.184* | 0.112* | 0.215* |
| (21) | BVAR(12) | 0.515** | 0.755** | 0.439** | 0.364** | 0.272** | 0.231* | 0.133* |
| (22) | VAR(AIC) | 0.571** | 0.864** | 0.682** | 0.763** | 0.511** | 0.541** | 0.492** |
| (23) | BVAR(AIC) | 0.632** | 0.902** | 0.845** | 0.950** | 0.676** | 0.685** | 0.667** |
| (24) | VAR(1) | 0.632** | 0.988** | **1.000**** | 0.991** | **1.000**** | **1.000**** | 0.739** |
| (25) | BVAR(1) | 0.632** | 0.988** | 0.979** | 0.950** | 0.769** | 0.950** | 0.739** |
| (26) | BVAR(1)-FSV | 0.632** | 0.963** | 0.855** | 0.902** | 0.662** | 0.826** | 0.667** |
| | *Alternative predictors*: | | | | | | | |
| (27) | Temperature anomalies | 0.632** | 0.970** | 0.932** | 0.950** | 0.769** | 0.950** | 0.739** |
| (28) | HDD deviations | 0.632** | 0.988** | 0.979** | 0.950** | 0.769** | 0.950** | 0.739** |
| (29) | CDD deviations | 0.632** | 0.988** | 0.979** | 0.950** | 0.769** | 0.950** | 0.739** |
| (30) | Industrial production | 0.632** | **1.000**** | 0.979** | 0.950** | 0.769** | 0.897** | 0.595** |
| (31) | CFNAI | 0.632** | 0.988** | 0.979** | 0.950** | 0.769** | 0.950** | 0.739** |
| | *Alternative economic models*: | | | | | | | |
| (32) | Model 1: BVAR(1) | 0.632** | 0.964** | 0.834** | 0.784** | 0.550** | 0.611** | 0.492** |
| (33) | BVAR(1)-FSV | 0.632** | 0.965** | 0.833** | 0.806** | 0.581** | 0.950** | **1.000**** |
| (34) | Model 2: BVAR(1) | 0.554** | 0.956** | 0.833** | 0.810** | 0.581** | 0.779** | 0.667** |
| (35) | BVAR(1)-FSV | 0.617** | 0.963** | 0.829** | 0.762** | 0.520** | 0.942** | 0.739** |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| (36) | Model 3: BVAR(1) | $0.632^{**}$ | $0.988^{**}$ | $0.979^{**}$ | $0.950^{**}$ | $0.769^{**}$ | $0.847^{**}$ | $0.667^{**}$ |
| (37) | BVAR(1)-FSV | $0.617^{**}$ | $0.909^{**}$ | $0.792^{**}$ | $0.731^{**}$ | $0.501^{**}$ | $0.580^{**}$ | $0.492^{**}$ |
| (38) | Model 4: BVAR(1) | $0.632^{**}$ | $0.988^{**}$ | $0.979^{**}$ | $0.950^{**}$ | $0.676^{**}$ | $0.913^{**}$ | $0.588^{**}$ |
| (39) | BVAR(1)-FSV | $0.632^{**}$ | $0.965^{**}$ | $0.833^{**}$ | $0.806^{**}$ | $0.599^{**}$ | $0.779^{**}$ | $0.492^{**}$ |
| | **Price Spread Models** | | | | | | | |
| (40) | Constant coefficients: $\hat{\alpha}, \hat{\beta}$ | $0.632^{**}$ | $0.848^{**}$ | $0.645^{**}$ | $0.426^{**}$ | $0.202^{*}$ | $0.162^{*}$ | $0.161^{*}$ |
| (41) | Constant coefficients: $\alpha = 0, \hat{\beta}$ | $0.632^{**}$ | $0.884^{**}$ | $0.666^{**}$ | $0.547^{**}$ | $0.297^{**}$ | $0.259^{**}$ | $0.309^{**}$ |
| (42) | Markov switching: $\hat{\alpha}^{MS}, \hat{\beta}^{MS}$ | $0.235^{**}$ | $0.970^{**}$ | $0.833^{**}$ | $0.762^{**}$ | $0.482^{**}$ | $0.110^{*}$ | $0.132^{*}$ |
| | **Energy Price VAR Models** | | | | | | | |
| (43) | VAR(12) | $0.607^{**}$ | $0.796^{**}$ | $0.655^{**}$ | $0.339^{**}$ | $0.172^{*}$ | $0.141^{*}$ | $0.227^{*}$ |
| (44) | BVAR(12) | $0.632^{**}$ | $0.969^{**}$ | $0.860^{**}$ | $0.762^{**}$ | $0.453^{**}$ | $0.271^{**}$ | $0.272^{**}$ |
| (45) | VAR(AIC) | $0.433^{**}$ | $0.905^{**}$ | $0.756^{**}$ | $0.598^{**}$ | $0.304^{**}$ | $0.196^{*}$ | $0.176^{*}$ |
| (46) | BVAR(AIC) | $0.617^{**}$ | $0.964^{**}$ | $0.821^{**}$ | $0.672^{**}$ | $0.390^{**}$ | $0.169^{*}$ | $0.136^{*}$ |
| (47) | VAR(1) | $0.613^{**}$ | $0.891^{**}$ | $0.733^{**}$ | $0.569^{**}$ | $0.273^{**}$ | $0.127^{*}$ | 0.099 |
| (48) | BVAR(1) | $0.483^{**}$ | $0.735^{**}$ | $0.531^{**}$ | $0.375^{**}$ | $0.205^{*}$ | $0.115^{*}$ | $0.117^{*}$ |
| (49) | VEC(12) | $0.488^{**}$ | $0.807^{**}$ | $0.595^{**}$ | $0.349^{**}$ | $0.166^{*}$ | 0.096 | 0.092 |
| (50) | BVEC(12) | $0.613^{**}$ | $0.988^{**}$ | $0.860^{**}$ | $0.655^{**}$ | $0.237^{*}$ | $0.103^{*}$ | $0.114^{*}$ |
| (51) | VEC(AIC) | $0.448^{**}$ | $0.875^{**}$ | $0.728^{**}$ | $0.471^{**}$ | $0.195^{*}$ | 0.088 | 0.085 |
| (52) | BVEC(AIC) | $0.510^{**}$ | $0.899^{**}$ | $0.701^{**}$ | $0.411^{**}$ | $0.187^{*}$ | 0.082 | 0.078 |
| (53) | VEC(1) | $0.428^{**}$ | $0.821^{**}$ | $0.551^{**}$ | $0.285^{**}$ | $0.142^{*}$ | 0.058 | 0.049 |
| (54) | BVEC(1) | $0.442^{**}$ | $0.829^{**}$ | $0.582^{**}$ | $0.298^{**}$ | $0.153^{*}$ | 0.069 | 0.063 |
| | **Integrated-Markets Models** | | | | | | | |
| (55) | BVAR(12) | $0.632^{**}$ | $0.970^{**}$ | $0.754^{**}$ | $0.708^{**}$ | $0.501^{**}$ | $0.445^{**}$ | $0.258^{**}$ |
| (56) | BVAR(AIC) | $0.632^{**}$ | $0.988^{**}$ | $0.932^{**}$ | $0.950^{**}$ | $0.627^{**}$ | $0.426^{**}$ | $0.321^{**}$ |
| (57) | BVAR(1) | $0.632^{**}$ | $0.969^{**}$ | $0.860^{**}$ | $0.822^{**}$ | $0.520^{**}$ | $0.370^{**}$ | $0.163^{*}$ |
| (58) | BVAR(12)-FSV | $0.632^{**}$ | $0.988^{**}$ | $0.860^{**}$ | $0.765^{**}$ | $0.520^{**}$ | $0.570^{**}$ | $0.354^{**}$ |
| (59) | BVAR(6)-FSV | $0.632^{**}$ | $0.970^{**}$ | $0.927^{**}$ | $0.891^{**}$ | $0.581^{**}$ | $0.685^{**}$ | $0.437^{**}$ |
| (60) | BVAR(1)-FSV | $0.617^{**}$ | $0.929^{**}$ | $0.772^{**}$ | $0.708^{**}$ | $0.486^{**}$ | $0.541^{**}$ | $0.335^{**}$ |
| | Number of models in set $\widehat{M}^*_{90\%}$ | 60 | 60 | 60 | 60 | 60 | 55 | 54 |
| | Number of models in set $\widehat{M}^*_{75\%}$ | 59 | 60 | 60 | 60 | 48 | 43 | 43 |

NOTES: MCS *p*-values computed with 10,000 block bootstrap replications using a block size of 12. The forecasts in $\widehat{M}^*_{90\%}$ and $\widehat{M}^*_{75\%}$ are identified by one and two asterisks, respectively. The loss function is specified in terms of squared forecast errors. Red indicates the top 5 and blue the bottom 5 models according to the MCS ranking.

**Table 6A. Model Confidence Sets for Real Natural Gas Price Forecasts: Subsample Analysis Based on Institutional Changes for Short-Term Horizons**

| No. | Models | **1997.2-2006.3** | | **2006.4-2009.12** | | **2010.1-2016.5** | | **2016.6-2024.2** | |
|---|---|---|---|---|---|---|---|---|---|
| | | $h = 3$ | $h = 6$ | $h = 3$ | $h = 6$ | $h = 3$ | $h = 6$ | $h = 3$ | $h = 6$ |
| (1) | Benchmark Model: random walk | 0.408$^{**}$ | 0.260$^{**}$ | 0.347$^{**}$ | 0.225$^{*}$ | 0.604$^{**}$ | 0.902$^{**}$ | 0.569$^{**}$ | 0.379$^{**}$ |
| | **Model-free Forecasts** | | | | | | | | |
| (2) | End-of month no-change forecast | 0.323$^{**}$ | 0.240$^{*}$ | 0.946$^{**}$ | 0.209$^{*}$ | 0.638$^{**}$ | 0.973$^{**}$ | 0.995$^{**}$ | 0.406$^{**}$ |
| (3) | Futures-market-based forecast | 0.408$^{**}$ | **1.000**$^{**}$ | 0.252$^{**}$ | 0.183$^{*}$ | **1.000**$^{**}$ | **1.000**$^{**}$ | 0.878$^{**}$ | 0.506$^{**}$ |
| | **Univariate Models** | | | | | | | | |
| (4) | ARMA(1,1) | 0.350$^{**}$ | 0.232$^{*}$ | 0.711$^{**}$ | 0.731$^{**}$ | 0.638$^{**}$ | 0.993$^{**}$ | 0.795$^{**}$ | 0.506$^{**}$ |
| (5) | IMA(1) | 0.376$^{**}$ | 0.260$^{**}$ | 0.384$^{**}$ | 0.190$^{*}$ | 0.562$^{**}$ | 0.868$^{**}$ | 0.538$^{**}$ | 0.379$^{**}$ |
| (6) | ARIMA(1,1) | 0.386$^{**}$ | 0.260$^{**}$ | 0.305$^{**}$ | 0.175$^{*}$ | 0.548$^{**}$ | 0.839$^{**}$ | 0.452$^{**}$ | 0.379$^{**}$ |
| (7) | AR(12) | 0.316$^{**}$ | 0.175$^{*}$ | 0.320$^{**}$ | 0.442$^{**}$ | 0.482$^{**}$ | 0.920$^{**}$ | 0.651$^{**}$ | 0.506$^{**}$ |
| (8) | BAR(12) | 0.309$^{**}$ | 0.164$^{*}$ | 0.445$^{**}$ | 0.631$^{**}$ | 0.592$^{**}$ | 0.973$^{**}$ | 0.651$^{**}$ | 0.503$^{**}$ |
| (9) | AR(AIC) | 0.315$^{**}$ | 0.218$^{*}$ | 0.629$^{**}$ | 0.731$^{**}$ | 0.633$^{**}$ | 0.976$^{**}$ | 0.820$^{**}$ | 0.506$^{**}$ |
| (10) | BAR(AIC) | 0.314$^{**}$ | 0.218$^{*}$ | 0.640$^{**}$ | 0.731$^{**}$ | 0.638$^{**}$ | 0.976$^{**}$ | 0.851$^{**}$ | 0.506$^{**}$ |
| (11) | AR(1) | 0.350$^{**}$ | 0.218$^{*}$ | 0.650$^{**}$ | 0.679$^{**}$ | 0.638$^{**}$ | 0.976$^{**}$ | 0.874$^{**}$ | 0.506$^{**}$ |
| (12) | BAR(1) | 0.350$^{**}$ | 0.218$^{*}$ | 0.669$^{**}$ | 0.697$^{**}$ | 0.638$^{**}$ | 0.993$^{**}$ | 0.876$^{**}$ | 0.506$^{**}$ |
| (13) | Exponential Smoothing | 0.275$^{**}$ | 0.260$^{**}$ | 0.207$^{*}$ | 0.632$^{**}$ | 0.638$^{**}$ | 0.976$^{**}$ | 0.459$^{**}$ | 0.372$^{**}$ |
| | **Natural Gas Market Models** | | | | | | | | |
| | *Baseline predictors in growth rates*: | | | | | | | | |
| (14) | VAR(12) | 0.331$^{**}$ | 0.210$^{*}$ | 0.946$^{**}$ | 0.658$^{**}$ | 0.457$^{**}$ | 0.901$^{**}$ | 0.998$^{**}$ | **1.000**$^{**}$ |
| (15) | BVAR(12) | 0.354$^{**}$ | 0.190$^{*}$ | 0.946$^{**}$ | 0.731$^{**}$ | 0.512$^{**}$ | 0.774$^{**}$ | 0.958$^{**}$ | 0.592$^{**}$ |
| (16) | VAR(AIC) | 0.350$^{**}$ | 0.230$^{*}$ | 0.861$^{**}$ | 0.462$^{**}$ | 0.579$^{**}$ | 0.761$^{**}$ | 0.815$^{**}$ | 0.517$^{**}$ |
| (17) | BVAR(AIC) | 0.408$^{**}$ | 0.260$^{**}$ | 0.813$^{**}$ | 0.661$^{**}$ | 0.638$^{**}$ | 0.902$^{**}$ | 0.862$^{**}$ | 0.506$^{**}$ |
| (18) | VAR(1) | 0.408$^{**}$ | 0.260$^{**}$ | 0.711$^{**}$ | 0.731$^{**}$ | 0.633$^{**}$ | 0.972$^{**}$ | 0.878$^{**}$ | 0.506$^{**}$ |
| (19) | BVAR(1) | 0.408$^{**}$ | 0.260$^{**}$ | 0.694$^{**}$ | 0.712$^{**}$ | 0.633$^{**}$ | 0.967$^{**}$ | 0.877$^{**}$ | 0.506$^{**}$ |
| | *Baseline predictors in logs*: | | | | | | | | |
| (20) | VAR(12) | 0.408$^{**}$ | 0.260$^{**}$ | 0.761$^{**}$ | 0.331$^{**}$ | 0.424$^{**}$ | 0.491$^{**}$ | **1.000**$^{**}$ | 0.643$^{**}$ |
| (21) | BVAR(12) | 0.365$^{**}$ | 0.260$^{**}$ | 0.694$^{**}$ | 0.411$^{**}$ | 0.383$^{**}$ | 0.331$^{**}$ | 0.980$^{**}$ | 0.506$^{**}$ |
| (22) | VAR(AIC) | 0.339$^{**}$ | 0.259$^{**}$ | 0.946$^{**}$ | 0.534$^{**}$ | 0.442$^{**}$ | 0.516$^{**}$ | 0.888$^{**}$ | 0.506$^{**}$ |
| (23) | BVAR(AIC) | 0.361$^{**}$ | 0.260$^{**}$ | 0.919$^{**}$ | 0.731$^{**}$ | 0.525$^{**}$ | 0.669$^{**}$ | 0.911$^{**}$ | 0.506$^{**}$ |
| (24) | VAR(1) | 0.408$^{**}$ | 0.260$^{**}$ | 0.919$^{**}$ | **1.000**$^{**}$ | 0.638$^{**}$ | 0.973$^{**}$ | 0.632$^{**}$ | 0.425$^{**}$ |
| (25) | BVAR(1) | 0.408$^{**}$ | 0.260$^{**}$ | 0.761$^{**}$ | 0.731$^{**}$ | 0.638$^{**}$ | 0.993$^{**}$ | 0.599$^{**}$ | 0.401$^{**}$ |
| (26) | BVAR(1)-FSV | 0.408$^{**}$ | 0.260$^{**}$ | 0.432$^{**}$ | 0.305$^{**}$ | 0.588$^{**}$ | 0.902$^{**}$ | 0.602$^{**}$ | 0.391$^{**}$ |
| | *Alternative predictors*: | | | | | | | | |
| (27) | Temperature anomalies | 0.408$^{**}$ | 0.260$^{**}$ | 0.761$^{**}$ | 0.731$^{**}$ | 0.638$^{**}$ | 0.978$^{**}$ | 0.585$^{**}$ | 0.397$^{**}$ |
| (28) | HDD deviations | 0.408$^{**}$ | 0.260$^{**}$ | 0.761$^{**}$ | 0.731$^{**}$ | 0.638$^{**}$ | 0.993$^{**}$ | 0.645$^{**}$ | 0.418$^{**}$ |
| (29) | CDD deviations | 0.408$^{**}$ | 0.260$^{**}$ | 0.761$^{**}$ | 0.731$^{**}$ | 0.638$^{**}$ | 0.976$^{**}$ | 0.602$^{**}$ | 0.406$^{**}$ |
| (30) | Industrial production | 0.408$^{**}$ | 0.260$^{**}$ | 0.727$^{**}$ | 0.731$^{**}$ | 0.638$^{**}$ | 0.912$^{**}$ | 0.651$^{**}$ | 0.453$^{**}$ |
| (31) | CFNAI | 0.408$^{**}$ | 0.260$^{**}$ | 0.811$^{**}$ | 0.731$^{**}$ | 0.638$^{**}$ | 0.993$^{**}$ | 0.594$^{**}$ | 0.403$^{**}$ |
| | *Alternative economic models*: | | | | | | | | |
| (32) | Model 1: BVAR(1) | 0.354$^{**}$ | 0.218$^{*}$ | 0.792$^{**}$ | 0.731$^{**}$ | 0.616$^{**}$ | 0.816$^{**}$ | 0.942$^{**}$ | 0.506$^{**}$ |
| (33) | BVAR(1)-FSV | 0.396$^{**}$ | 0.260$^{**}$ | 0.567$^{**}$ | 0.386$^{**}$ | 0.638$^{**}$ | 0.993$^{**}$ | 0.651$^{**}$ | 0.440$^{**}$ |
| (34) | Model 2: BVAR(1) | 0.309$^{**}$ | 0.186$^{*}$ | 0.730$^{**}$ | 0.712$^{**}$ | 0.638$^{**}$ | 0.993$^{**}$ | 0.933$^{**}$ | 0.506$^{**}$ |
| (35) | BVAR(1)-FSV | 0.376$^{**}$ | 0.260$^{**}$ | 0.610$^{**}$ | 0.420$^{**}$ | 0.638$^{**}$ | 0.973$^{**}$ | 0.646$^{**}$ | 0.422$^{**}$ |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| (36) | Model 3: BVAR(1) | 0.408** | 0.260** | 0.891** | 0.712** | 0.638** | 0.952** | 0.780** | 0.479** |
| (37) | BVAR(1)-FSV | 0.408** | 0.260** | 0.411** | 0.256** | 0.633** | 0.972** | 0.529** | 0.365** |
| (38) | Model 4: BVAR(1) | 0.408** | 0.260** | 0.813** | 0.731** | 0.638** | 0.976** | 0.694** | 0.386** |
| (39) | BVAR(1)-FSV | 0.408** | 0.260** | 0.491** | 0.289** | 0.638** | 0.973** | 0.643** | 0.386** |
| | **Price Spread Models** | | | | | | | | |
| (40) | Constant coefficients: $\hat{\alpha}, \hat{\beta}$ | 0.408** | 0.752** | 0.538** | 0.378** | 0.248* | 0.228* | 0.742** | 0.410** |
| (41) | Constant coefficients: $\alpha = 0, \hat{\beta}$ | 0.433** | 0.982** | 0.265** | 0.201* | 0.217* | 0.267** | 0.606** | 0.391** |
| (42) | Markov switching: $\hat{\alpha}^{MS}, \hat{\beta}^{MS}$ | **1.000**** | 0.982** | 0.231* | 0.305** | 0.361** | 0.767** | 0.492** | 0.382** |
| | **Energy Price VAR Models** | | | | | | | | |
| (43) | VAR(12) | 0.353** | 0.153* | 0.626** | 0.697** | 0.525** | 0.902** | 0.958** | 0.506** |
| (44) | BVAR(12) | 0.354** | 0.186* | 0.813** | 0.731** | 0.638** | 0.993** | 0.998** | 0.506** |
| (45) | VAR(AIC) | 0.343** | 0.191* | 0.711** | 0.731** | 0.527** | 0.641** | 0.958** | 0.506** |
| (46) | BVAR(AIC) | 0.365** | 0.218* | 0.761** | 0.731** | 0.533** | 0.681** | 0.958** | 0.506** |
| (47) | VAR(1) | 0.354** | 0.242* | 0.640** | 0.643** | 0.503** | 0.615** | 0.882** | 0.506** |
| (48) | BVAR(1) | 0.301** | 0.186* | 0.479** | 0.335** | 0.522** | 0.735** | 0.878** | 0.506** |
| (49) | VEC(12) | 0.408** | 0.259** | 0.512** | 0.352** | 0.473** | 0.713** | 0.952** | 0.506** |
| (50) | BVEC(12) | 0.408** | 0.260** | 0.690** | 0.658** | 0.638** | 0.965** | 0.998** | 0.506** |
| (51) | VEC(AIC) | 0.408** | 0.260** | 0.292** | 0.284** | 0.603** | 0.582** | 0.925** | 0.506** |
| (52) | BVEC(AIC) | 0.408** | 0.260** | 0.364** | 0.324** | 0.540** | 0.545** | 0.900** | 0.503** |
| (53) | VEC(1) | 0.408** | 0.260** | 0.615** | 0.519** | 0.284** | 0.348** | 0.652** | 0.441** |
| (54) | BVEC(1) | 0.408** | 0.260** | 0.601** | 0.451** | 0.321** | 0.413** | 0.697** | 0.465** |
| | **Integrated-Markets Models** | | | | | | | | |
| (55) | BVAR(12) | 0.376** | 0.260** | 0.946** | 0.712** | 0.402** | 0.308** | 0.998** | 0.592** |
| (56) | BVAR(AIC) | 0.365** | 0.260** | **1.000**** | 0.733** | 0.633** | 0.793** | 0.952** | 0.506** |
| (57) | BVAR(1) | 0.408** | 0.260** | 0.690** | 0.731** | 0.638** | 0.973** | 0.646** | 0.421** |
| (58) | BVAR(12)-FSV | 0.408** | 0.260** | 0.576** | 0.275** | 0.518** | 0.796** | 0.952** | 0.506** |
| (59) | BVAR(6)-FSV | 0.408** | 0.260** | 0.592** | 0.360** | 0.616** | 0.961** | 0.864** | 0.503** |
| (60) | BVAR(1)-FSV | 0.408** | 0.260** | 0.279** | 0.242* | 0.509** | 0.777** | 0.646** | 0.379** |
| | Number of models in set $\widehat{M}^*_{90\%}$ | 60 | 60 | 60 | 60 | 60 | 60 | 60 | 60 |
| | Number of models in set $\widehat{M}^*_{75\%}$ | 60 | 41 | 58 | 54 | 58 | 59 | 60 | 60 |

NOTES: MCS *p*-values computed with 10,000 block bootstrap replications using a block size of 12. The forecasts in $\widehat{M}^*_{90\%}$ and $\widehat{M}^*_{75\%}$ are identified by one and two asterisks, respectively. The loss function is specified in terms of squared forecast errors. Red indicates the best model (in bold) and the second-best model, and blue the worst model according to the MCS ranking.

**Table 7A. Model Confidence Sets for Real Natural Gas Price Forecasts: Subsample Analysis Based on Institutional Changes for Long-Term Horizons**

| No. | Models | **1997.2-2006.3** | | **2006.4-2009.12** | | **2010.1-2016.5** | | **2016.6-2024.2** | |
|---|---|---|---|---|---|---|---|---|---|
| | | $h$ = 12 | $h$ = 24 | $h$ = 12 | $h$ = 24 | $h$ = 12 | $h$ = 24 | $h$ = 12 | $h$ = 24 |
| (1) | Benchmark Model: random walk | 0.365** | 0.324** | 0.237* | 0.316** | 0.521** | 0.093 | 0.360** | 0.380** |
| | **Model-free Forecasts** | | | | | | | | |
| (2) | End-of month no-change forecast | 0.351** | 0.177* | 0.207* | 0.261** | 0.490** | 0.142* | 0.366** | 0.428** |
| (3) | Futures-market-based forecast | **1.000**** | 0.324** | 0.378** | 0.789** | 0.354** | 0.064 | **1.000**** | **1.000**** |
| | **Univariate Models** | | | | | | | | |
| (4) | ARMA(1,1) | 0.266** | 0.153* | 0.748** | 0.316** | 0.831** | 0.902** | 0.479** | 0.779** |
| (5) | IMA(1) | 0.384** | 0.449** | 0.184* | 0.253** | 0.354** | 0.080 | 0.340** | 0.350** |
| (6) | ARIMA(1,1) | 0.384** | 0.474** | 0.190* | 0.257** | 0.385** | 0.080 | 0.364** | 0.493** |
| (7) | AR(12) | 0.121* | 0.130* | 0.331** | 0.261** | 0.702** | 0.169* | 0.460** | 0.562** |
| (8) | BAR(12) | 0.117* | 0.144* | 0.391** | 0.358** | 0.802** | 0.169* | 0.466** | 0.705** |
| (9) | AR(AIC) | 0.203* | 0.182* | 0.748** | 0.660** | 0.831** | **1.000**** | 0.479** | 0.779** |
| (10) | BAR(AIC) | 0.210* | 0.157* | 0.748** | 0.634** | 0.831** | 0.550** | 0.479** | 0.779** |
| (11) | AR(1) | 0.237* | 0.157* | 0.626** | 0.538** | 0.831** | 0.169* | 0.479** | 0.779** |
| (12) | BAR(1) | 0.250** | 0.157* | 0.626** | 0.470** | 0.831** | 0.169* | 0.479** | 0.779** |
| (13) | Exponential Smoothing | 0.384** | 0.523** | 0.748** | **1.000**** | 0.547** | 0.163* | 0.364** | 0.779** |
| | **Natural Gas Market Models** | | | | | | | | |
| | *Baseline predictors in growth rates*: | | | | | | | | |
| (14) | VAR(12) | 0.146* | 0.110* | 0.546** | 0.687** | 0.430** | 0.167* | 0.966** | 0.779** |
| (15) | BVAR(12) | 0.171* | 0.157* | 0.748** | 0.806** | 0.343** | 0.167* | 0.673** | 0.779** |
| (16) | VAR(AIC) | 0.184* | 0.154* | 0.626** | 0.772** | 0.328** | 0.167* | 0.580** | 0.531** |
| (17) | BVAR(AIC) | 0.317** | 0.157* | 0.748** | 0.789** | 0.338** | 0.167* | 0.570** | 0.562** |
| (18) | VAR(1) | 0.384** | 0.157* | 0.748** | 0.750** | 0.589** | 0.169* | 0.479** | 0.779** |
| (19) | BVAR(1) | 0.384** | 0.177* | 0.748** | 0.783** | 0.642** | 0.169* | 0.479** | 0.779** |
| | *Baseline predictors in logs*: | | | | | | | | |
| (20) | VAR(12) | 0.384** | 0.633** | 0.293** | 0.876** | 0.125* | 0.060 | 0.479** | 0.350** |
| (21) | BVAR(12) | 0.399** | 0.523** | 0.512** | 0.789** | 0.061 | 0.049 | 0.479** | 0.380** |
| (22) | VAR(AIC) | 0.384** | 0.370** | 0.612** | 0.789** | 0.120* | 0.119* | 0.479** | 0.562** |
| (23) | BVAR(AIC) | 0.384** | 0.324** | 0.748** | 0.876** | 0.221* | 0.130* | 0.479** | 0.562** |
| (24) | VAR(1) | 0.384** | 0.218* | **1.000**** | 0.783** | 0.831** | 0.169* | 0.466** | 0.562** |
| (25) | BVAR(1) | 0.384** | 0.224* | 0.748** | 0.768** | 0.831** | 0.169* | 0.479** | 0.779** |
| (26) | BVAR(1)-FSV | 0.384** | **1.000**** | 0.355** | 0.316** | 0.490** | 0.079 | 0.366** | 0.518** |
| | *Alternative predictors*: | | | | | | | | |
| (27) | Temperature anomalies | 0.384** | 0.207* | 0.748** | 0.720** | 0.831** | 0.169* | 0.479** | 0.779** |
| (28) | HDD deviations | 0.384** | 0.186* | 0.748** | 0.772** | 0.831** | 0.169* | 0.479** | 0.779** |
| (29) | CDD deviations | 0.384** | 0.263** | 0.748** | 0.783** | 0.831** | 0.169* | 0.466** | 0.674** |
| (30) | Industrial production | 0.384** | 0.452** | 0.612** | 0.626** | 0.557** | 0.071 | 0.479** | 0.674** |
| (31) | CFNAI | 0.384** | 0.288** | 0.748** | 0.626** | **1.000**** | 0.169* | 0.479** | 0.779** |
| | *Alternative economic models*: | | | | | | | | |
| (32) | Model 1: BVAR(1) | 0.196* | 0.157 | 0.748** | 0.394** | 0.452** | 0.169* | 0.479** | 0.428** |
| (33) | BVAR(1)-FSV | 0.223* | 0.257** | 0.475** | 0.876** | 0.831** | 0.167* | 0.430** | 0.779** |
| (34) | Model 2: BVAR(1) | 0.171* | 0.157* | 0.748** | 0.423** | 0.831** | 0.249* | 0.479** | 0.779** |
| (35) | BVAR(1)-FSV | 0.231* | 0.200* | 0.445** | 0.876** | 0.675** | 0.167* | 0.430** | 0.779** |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| (36) | Model 3: BVAR(1) | 0.384** | 0.523** | 0.626** | 0.694** | 0.571** | 0.066 | 0.430** | 0.527** |
| (37) | BVAR(1)-FSV | 0.384** | 0.523** | 0.313** | 0.357** | 0.600** | 0.127* | 0.346** | 0.358** |
| (38) | Model 4: BVAR(1) | 0.384** | 0.156* | 0.748** | 0.687** | 0.831** | 0.167* | 0.420** | 0.562** |
| (39) | BVAR(1)-FSV | 0.384** | 0.523** | 0.403** | 0.421** | 0.760** | 0.093 | 0.366** | 0.369** |
| | **Price Spread Models** | | | | | | | | |
| (40) | Constant coefficients: $\hat{\alpha}, \hat{\beta}$ | 0.384** | 0.483** | 0.261** | 0.238* | 0.046 | 0.081 | 0.342** | 0.380** |
| (41) | Constant coefficients: $\alpha = 0, \hat{\beta}$ | 0.762** | 0.343** | 0.198* | 0.261** | 0.062 | 0.078 | 0.336** | 0.380** |
| (42) | Markov switching: $\hat{\alpha}^{MS}, \hat{\beta}^{MS}$ | 0.399** | 0.157* | 0.330** | 0.255** | 0.186* | 0.154* | 0.332** | 0.316** |
| | **Energy Price VAR Models** | | | | | | | | |
| (43) | VAR(12) | 0.106* | 0.036 | 0.218* | 0.347** | 0.430** | 0.167* | 0.479** | 0.562** |
| (44) | BVAR(12) | 0.146* | 0.043 | 0.462** | 0.626** | 0.452** | 0.149* | 0.479** | 0.674** |
| (45) | VAR(AIC) | 0.137* | 0.068 | 0.500** | 0.316** | 0.204* | 0.097 | 0.469** | 0.499** |
| (46) | BVAR(AIC) | 0.177* | 0.050 | 0.530** | 0.345** | 0.231* | 0.081 | 0.466** | 0.483** |
| (47) | VAR(1) | 0.186* | 0.030 | 0.369** | 0.626** | 0.173* | 0.067 | 0.448** | 0.380** |
| (48) | BVAR(1) | 0.125* | 0.026 | 0.244* | 0.626** | 0.305** | 0.074 | 0.440** | 0.406** |
| (49) | VEC(12) | 0.265** | 0.101* | 0.177* | 0.394** | 0.163* | 0.080 | 0.345** | 0.347** |
| (50) | BVEC(12) | 0.278** | 0.116* | 0.271** | 0.626** | 0.254** | 0.064 | 0.363** | 0.350** |
| (51) | VEC(AIC) | 0.226* | 0.082 | 0.296** | 0.668** | 0.146* | 0.054 | 0.364** | 0.341** |
| (52) | BVEC(AIC) | 0.217* | 0.061 | 0.300** | 0.694** | 0.129* | 0.052 | 0.364** | 0.337** |
| (53) | VEC(1) | 0.161* | 0.018 | 0.327** | 0.789** | 0.076 | 0.035 | 0.364** | 0.303** |
| (54) | BVEC(1) | 0.165* | 0.020 | 0.325** | 0.789** | 0.095 | 0.040 | 0.364** | 0.320** |
| | **Integrated-Markets Models** | | | | | | | | |
| (55) | BVAR(12) | 0.384** | 0.324** | 0.626** | 0.789** | 0.058 | 0.057 | 0.936** | 0.562** |
| (56) | BVAR(AIC) | 0.380** | 0.094 | 0.748** | 0.876** | 0.285** | 0.093 | 0.479** | 0.562** |
| (57) | BVAR(1) | 0.330** | 0.023 | 0.612** | 0.772** | 0.589** | 0.080 | 0.466** | 0.518** |
| (58) | BVAR(12)-FSV | 0.384** | 0.483** | 0.305** | 0.789** | 0.273** | 0.048 | 0.479** | 0.463** |
| (59) | BVAR(6)-FSV | 0.384** | 0.235* | 0.423** | 0.806** | 0.433** | 0.063 | 0.421** | 0.380** |
| (60) | BVAR(1)-FSV | 0.384** | 0.382** | 0.278** | 0.316** | 0.350** | 0.063 | 0.364** | 0.483** |
| | Number of models in set $\widehat{M}^*_{90\%}$ | 60 | 47 | 60 | 60 | 54 | 32 | 60 | 60 |
| | Number of models in set $\widehat{M}^*_{75\%}$ | 38 | 22 | 52 | 59 | 44 | 3 | 60 | 60 |

NOTES: MCS *p*-values computed with 10,000 block bootstrap replications using a block size of 12. The forecasts in $\widehat{M}^*_{90\%}$ and $\widehat{M}^*_{75\%}$ are identified by one and two asterisks, respectively. The loss function is specified in terms of squared forecast errors. Red indicates the best model (in bold) and the second-best model, and blue the worst model according to the MCS ranking.