



HAL
open science

Real-time demand in natural gas price forecasting: The role of temperature data

Zakaria Moussa, Arthur Thomas, Benoît Sévi

► **To cite this version:**

Zakaria Moussa, Arthur Thomas, Benoît Sévi. Real-time demand in natural gas price forecasting: The role of temperature data. 2023. hal-04333458

HAL Id: hal-04333458

<https://hal.science/hal-04333458v1>

Preprint submitted on 9 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Document de travail du LEMNA N° 2023-08

Décembre 2023



Real-time demand in natural gas price forecasting: The role of temperature data

Zakaria Moussa, Arthur Thomas & Benoît Sévi



Real-time demand in natural gas price forecasting: The role of temperature data*

Zakaria Moussa¹, Arthur Thomas², and Benoît Sévi^{†,1}

¹LEMNA, Nantes Université, 44322 Nantes Cedex 3, France

²LEDa-CGEMP, Université Paris-Dauphine, 75775 Paris Cedex 16, France

December 6, 2023

Abstract

This paper provides evidence of the pivotal role data on temperature conditions can play in forecasting the real natural gas prices at the Henry Hub *in real time*. We consider a recently proposed index to track deviations from normal temperature during winter and include it as an additional exogenous variable in various Bayesian vector autoregressive (BVAR) models. The inclusion of the index significantly increases forecast accuracy at horizons up to 12 months, which puts forward the long-lasting impact of temperature shocks and their associated consequences on the gas market. Our approach is novel to energy price forecasting as it considers both supply and demand simultaneously and further includes temperature data as a proxy of real-time demand for natural gas.

JEL classification: C11, C32, C53, Q41, Q47

Keywords: energy prices, natural gas, Bayesian VAR, price forecasting, real-time data, temperature.

1 Introduction

This paper presents a forecasting analysis of real natural gas prices at the Henry Hub in real time. We use *state-of-the-art* Bayesian VAR econometrics allowing for stochastic volatility and/or fat tails in an environment where variable selection can take place. We first show that considering a model *à la* [Baumeister and Kilian \(2012\)](#) for natural gas combined with the price of crude oil as an additional variable significantly improves forecast accuracy at horizons of up to 12 months. The originality of our approach lies in considering the

*This study is supported by the Economics of Natural Gas Chair. We would like to thank participants at the 12th International Conference on Computational and Financial Econometrics in Pisa (Italy), the Workshop in Financial Econometrics in Nantes (France), the 3rd Commodity Markets Winter Workshop in Hanover (Germany) and the CEMA annual meeting in Chicago (USA) for their valuable feedback. Christiane Baumeister and Derek Bunn are particularly acknowledged. Any remaining errors are, of course, ours. This work was made possible by Nantes University's computing labs (CCIPL), which allowed us to estimate our computationally-intensive models.

[†]Corresponding author: benoit.sevi@univ-nantes.fr

demand side directly - through the consumption of gas - as in [Baumeister et al. \(2022\)](#) in addition to the supply side already present in [Baumeister and Kilian \(2012\)](#).¹ We then provide evidence that considering an additional variable which is available in real time, temperature, increases forecast accuracy at all horizons. Using temperature data allows us to include real-time variations in demand in our models and is shown to be critical to the performance of our forecasts. To the best of our knowledge, this is the very first time such an approach has been applied to forecasting commodity prices in a VAR setting.²

Natural gas is a key industry in the U.S., with a total revenue for distributors amounting to \$101 billion in 2018. There are more than 75 million residential, commercial, and industrial natural gas customers in the U.S., and natural gas currently meets more than one-fourth of the United States' energy needs ([American Gas Association, 2019](#)). The U.S. natural gas price is closely watched by consumers, as it directly affects their budget as well as their home heating choices. It is also a key commodity price for the U.S. economy and has an impact on inflation expectations, consumption level, and consumer confidence in the economy. Hence, it is of vital importance for central bankers. Moreover, natural gas is playing an increasingly prominent role in generating power.³

Investments in the gas industry are significant and often non-reversible. Improved price visibility is important for all players in the gas industry as they are in other industries where good forecasts are of equal importance (see [Auffhammer \(2007\)](#)). Natural gas infrastructure operators (TSO) can establish the transportation cost of operating, balancing, and maintaining the pipeline networks in order to ensure security of supply. Power generators rely on forecasts to compete more efficiently in the natural gas market, especially in a world where the role of natural gas is evolving. Indeed, because of the rise of intermittent renewable energy sources of electricity, natural gas-based thermal generation is increasingly used as a back-up technology. For these reasons, providing better-quality long-term forecasts offers market participants improved visibility regarding the evolution of their investments, as well as being critical for all natural gas stakeholders.

A substantial number of papers have tackled forecasting commodity prices. Some papers rely only on the price time-series itself ([Wang et al. \(2020\)](#), [Gao et al. \(2020\)](#)) while others

¹[Baumeister and Kilian \(2012\)](#) also consider the demand side through a quite general activity index which is likely to proxy for the state of the economy. Such an index is less likely to be relevant for natural gas, whose price is determined in a regional setting and where the majority of consumption is domestic. As such, our approach is more direct and relies on the consumption of natural gas itself to proxy the demand for gas.

²An exception is [Nick and Thoenes \(2014\)](#) which makes use of temperature data in a Structural VAR analysis for the European case. In their paper, the measure of temperature as well as the aim of the empirical analysis, not focused on forecasting, are notably different.

³Steam turbines (which can also be powered by oil or coal) combust fuel to generate steam, which then generates electricity. Combined-cycle units heat up fuel and use the fuel-air mixture to spin gas turbines and generate electricity. The waste heat from the gas turbine is used to generate steam for a steam turbine that generates additional electricity. Based on the U.S. Energy Information Administration's (EIA hereafter) December 2018 monthly electric generator inventory of utility-scale generation, 31.3 gigawatts (GW) of generating capacity were added in the United States in 2018 and 18.7 GW of capacity were retired. The 2018 annual capacity additions were the largest since the 48.8 GW added in 2003. More than 60% of electric-generating capacity installed in 2018 was fueled by natural gas and, since April 2018, the U.S. has produced more power using natural gas than using coal ([Energy Information Administration, 2019](#)). As of January 2019, U.S. generating capacity at NGCC power plants totaled 264 gigawatts (GW), compared with 243 GW at coal-fired power plants. See also [Doyle and Fell \(2018\)](#) and [Brehm \(2019\)](#) about the switch from gas to coal for power generation.

make use of information on a few fundamental variables, for example financial variables like equity and bond yields (Bessembinder and Chan (1992), Baumeister et al. (2015)), exchange rates (Chen et al. (2010)), futures prices (Alquist and Kilian (2010) and Baumeister et al. (2018)) or open interest in futures markets (Hong and Yogo (2012)). A quite different approach was developed for oil prices in Baumeister and Kilian (2012) and Alquist et al. (2013) and aims at modeling the oil market in a vector autoregression (VAR) framework where the variables at stake are commonly the oil price, oil production, oil inventories and an index of economic activity. This method delivered very promising results and was extended in various directions through both methodological and data improvements. In particular, the recent paper by Baumeister et al. (2022) focuses on enhancing the economic activity index used in Baumeister and Kilian (2012) and Alquist et al. (2013) as a proxy for oil demand. More relevant to our present framework, Baumeister et al. (2022) develop a consumption-based VAR model, replacing oil production with petroleum consumption on the grounds that “fluctuations in the demand for refined products will translate into changes in the demand for crude oil and thus have predictive power for the future path of the real price of crude oil.” (Baumeister et al. (2022), p. 11). This consumption-based model improves forecast accuracy in most cases, thereby highlighting the central role of demand in oil price dynamics.

Our approach is in the vein of the original contribution of Baumeister and Kilian (2012) and Alquist et al. (2013) but includes some major features specifically related to natural gas. We estimate a four-variable VAR model including the U.S. real price of natural gas (Henry Hub), U.S. natural gas production, U.S industrial production, and the real price of crude oil.⁴ In so doing, our model partly follows Nguyen and Okimoto (2019). While it is quite standard in the literature to include energy-commodity production in VAR models when considering the supply side, the use of U.S. industrial production is less common and is motivated, in our setting, by the regional nature of the gas market.⁵

Our motivation for including the price of oil comes from a long tradition in energy economics of investigating the strong link between these two energy sources (see Bachmeier and Griffin (2006), Brown and Yücel (2008) and more recently, Bello et al. (2017), Bunn et al. (2017) and Jadidzadeh and Serletis (2017)) although this relationship has been challenged in recent years following the shale gas revolution (Caporin and Fontini (2017)). The intuition behind our choice is that oil and gas are partly substitutable and are extracted from the same wells. In contrast to Hou and Nguyen (2018), who find oil price shocks to be of minor importance to modeling U.S. gas prices, our results show that oil prices are very helpful in the VAR framework. Indeed, models which do not include crude oil prices perform poorly

⁴Our sample periods covers the last two decades (1997-2018) and the major reforms introduced in 1989 (Natural Gas Wellhead Decontrol Act) are not covered by our sample period. These major reforms motivated the Markov switching approach in, for instance Hou and Nguyen (2018), who cover the 1980-2016 period, but they are not relevant in our framework.

⁵Attempts to include a world economic activity index in our empirical framework were uniformly unfruitful. Discussing the shale oil-gas revolution, Kilian (2016) writes: “Access to inexpensive natural gas benefits the petrochemical industry, for example. U.S. natural gas prices have been low because the natural gas market has never been a global market, but a regional market. Natural gas is transported by pipeline. Although natural gas may be cooled down and liquefied, allowing it to be shipped as liquefied natural gas (LNG) to any port in the world, the cost of LNG shipping is high and the infrastructure required to load and unload LNG is expensive. This fact has prevented the integration of regional natural gas markets and the emergence of a global price thus far. This means that, for the time being, the price of U.S. natural gas has been determined by domestic demand rather than global demand, allowing for a greater price response to increased domestic supply.”

with respect to the no-change and generally have low predictive power.

We go beyond this standard VAR model, whose primary aim is to consider endogenous supply and demand, to take into account temperature and its critical role in shaping natural gas consumption (Energy Information Administration, 2014). Well established since at least Berrisford (1965), this role has been studied in numerous papers (Timmer and Lamb (2007) offer an emblematic analysis of this strong relationship for the U.S.). As temperature is, by nature, an exogenous variable over a reasonable period of time and real-time data is available, it is a perfect natural candidate to forecast gas prices.⁶ Including temperature in our VAR allows us to reinforce the consumption-based aspect of our model in line with Baumeister et al. (2022) but goes one step further. Unlike the variables used to proxy demand in the VAR price-forecasting literature, temperature has the exceptional characteristic of being truly available in real time, which is an essential feature in our setting. Using temperature data in such a setting is new, and our convincing forecasting results show that it genuinely deserves further consideration.

Our general empirical approach is through a Bayesian VAR model allowing for fat tails and/or stochastic volatility as in Chiu et al. (2017) so as to deal efficiently with extreme events and/or low-frequency variations in volatility. Our temperature variable which, as explained below, is related to changes in natural gas inventories, is entered into the model as an exogenous predictor. Moreover, to reduce overfitting in making out-of-sample predictions, we implement a variable selection procedure for all specifications and show its relevancy to forecasting

Our contribution to the literature is threefold. First and foremost, this paper is the first to provide monthly real-time forecasts of the real price of natural gas for the U.S. following the methodology initiated in Baumeister and Kilian (2012) who exploit data freely available in the *Energy Monthly Review* on the Energy Information Administration (EIA) website to construct a real-time database.⁷ As such, our paper will be useful for market participants, policymakers, and regulators as well as for central bankers. Our results, the first of their kind for natural gas, should also prove a valuable contribution to economic research in the field of energy economics. Second, we provide evidence that our Bayesian VAR models produce promising forecasts of the real natural gas price at horizons of from 1 to 12 months, for instance improving on the no-change forecast by up to 47% at a 12-month horizon.⁸ These large improvements highlight the crucial importance of considering economic fundamentals in modeling the natural gas market dynamics. Third, we provide evidence of the central

⁶In contrast to Huurman et al. (2012), who focus on power prices, we do not forecast the weather to forecast gas prices, our predictions covering a longer term (1 to 12 months). Moreover, note that only a few papers in the field of energy economics have considered temperature data so far (see Panagiotidis and Rutledge (2007) among others).

⁷On natural gas price forecasts, we are only aware of Ferrari et al. (2021) who provide energy price forecasts, including for natural gas, quarterly. Their empirical approach is very different from ours and relies on a global dataset including a large number of macroeconomic and financial variables. Gao et al. (2020) use various univariate models allowing for time-variation in the parameters and/or stochastic volatility, but AR models are known to perform poorly on energy price forecasting (see Baumeister and Kilian (2012) among others). Other important contributions focused on forecasting non-oil energy prices include Baumeister et al. (2017) who forecast the gasoline price in the U.S. To the best of our knowledge, no empirical study has to date attempted to provide natural gas price forecasts in an environment considering fundamental variables of the natural gas market.

⁸As will be clear below, such an improvement is expressed in terms of the mean squared prediction error (MSPE) ratio for a given model relative to the MSPE of the no-change forecast, as in Baumeister and Kilian (2012).

role of considering demand in real-time through temperature data, thereby avoiding the issue of using revised data for estimation. While such an approach is senseless in the case of the world oil market, it is particularly relevant in the case of U.S. natural gas which is a very large but regional market. As will be shown below, models including temperature as an exogenous variable unambiguously deliver superior forecasts, particularly over longer horizons.

The rest of the paper proceeds as follows. The next section outlines the data and how it is treated before estimation. Models are presented in Section 3 and estimation algorithms and forecasting procedure are discussed in Section 4. Section 5 details the empirical results and evaluates the accuracy of the various forecasts. Finally, Section 6 provides concluding remarks and possible avenues for future research.

2 Data

The various VAR models in our study rely on five endogenous variables where, in some cases, we include an additional exogenous variable. As main drivers of the real gas price, we consider both supply and demand, namely the log-difference in U.S. gas production and the log in U.S. gas consumption. As real crude oil price, we choose the West Texas Intermediate (WTI) spot oil price, which is available in real time and is very similar to other benchmarks, such as, for instance the U.S. Refiner Acquisition Cost (RAC) which is more often used in empirical analysis.⁹ As is standard in the energy price forecast literature, we also consider the level of storage through a measure for gas inventories, namely the U.S Natural Gas in Underground Storage taken in log difference. Finally, we include an indirect measure of temperature conditions as an exogenous variable. As explained above, this choice is motivated by the fact that gas demand, from both industrial and residential heating sectors, is highly sensitive to temperature (Müller et al. (2015)). We now return to the details of the construction of these variables.

2.1 Real-time data sources

As Baumeister and Kilian (2012) do establish for crude oil, using real-time data markedly improves forecast accuracy. Similarly, some of the data used in our work are available in real time and are not subject to revision over time, as are the WTI and the Henry Hub natural gas price. For the rest of the variables, however, we need to extract a real-time data set from various sources. Therefore, our first contribution in this study lies in constructing a real-time data set for the U.S. gas market, as Baumeister and Kilian (2012) originally did for crude oil. Gas production, consumption and inventories were hand-collected from the *Monthly Energy Review* published by the EIA.¹⁰ Monthly averages of the daily WTI spot price and Henry Hub gas prices were obtained from the FRED database.¹¹ The dataset is monthly and runs from January 1997 to August 2018, Henry Hub data being available from January 1997 onward. As our focus in this paper is on forecasting the real gas price, nominal

⁹The correlation between the WTI spot prices at the monthly frequency and the RAC is 0.99 on our sample. Using one or another thus makes little difference.

¹⁰As a control, the *Monthly Energy Review* is digitized and collected data are compared with the hand-collected series.

¹¹<https://fred.stlouisfed.org/>

oil and gas prices were therefore deflated by the real-time U.S. consumer price index for all urban consumers.¹²

Much of the existing literature on oil price forecasting via VAR models uses inventories as a proxy for precautionary demand (Baumeister and Kilian, 2012; Kilian and Murphy, 2014; Baumeister and Hamilton, 2019). However, storage operators understandably exploit the highly predictable seasonality in the gas market. Therefore, only unexpected shifts in gas demand due to extraordinary weather conditions are likely to be relevant for explaining gas market prices.¹³

For that reason, we pay particular attention to deviations from the normal seasonal meteorological pattern as a determinant of gas prices. Following Müller et al. (2015), we use the normalized Cumulated Heating Degree Days ($nCHDD$). More explicitly, we cumulate heating degree days over a winter, from which we subtract the average of cumulated heating degree days over previous winters. Our measure of the normalized cumulated heating degree days is :

$$nCHDD_{d,w} = CHDD_{d,w} - \frac{1}{w-1} \sum_{l=1}^{w-1} CHDD_{d,w} \quad \text{for } 1 \leq d \leq 182$$

where 182 is the number of days in a winter w (from the 1st of October to the 31st of March). Then, Λ returns to zero during next summer.¹⁴

We emphasize the fact that this measure allows to consider the effects of unexpected temperature conditions on gas prices. In relation to our results, cumulating HDD makes the measure highly relevant with respect to the consumption side of the market. This could explain the long-lasting impact of temperature anomalies during the winter on natural gas prices.

2.2 Nowcasting

One complication from using real-time data is the usual delay in data availability, from one to three months depending on the variables. Real-time gas production becomes available with a three-month lag for all samples between January 1997 and August 2018, while the consumer price index only has a one-month lag. Gas consumption and gas stock share the same publication lags, namely one month between January 1997 and June 2002 and three months from July 2002 to the end of our sample.

To obtain nowcasts for vintages of gas production and the consumer price index, we follow Baumeister and Kilian (2012) and extrapolate data based on the average rate of

¹²Real-time data for the monthly seasonally-adjusted U.S. consumer price index for all urban consumers are obtained from the Economic Indicators published by the Council of Economic Advisers. These data are made available by the FRASER database of the Federal Reserve Bank of St. Louis (<https://fraser.stlouisfed.org/theme/economic-data>). Additional real-time consumer price index (CPI) data were obtained from the macroeconomic real-time database of the Federal Reserve Bank of Philadelphia (<https://www.philadelphiafed.org/surveys-and-data/real-time-data-research/real-time-data-set-for-macroeconomists>).

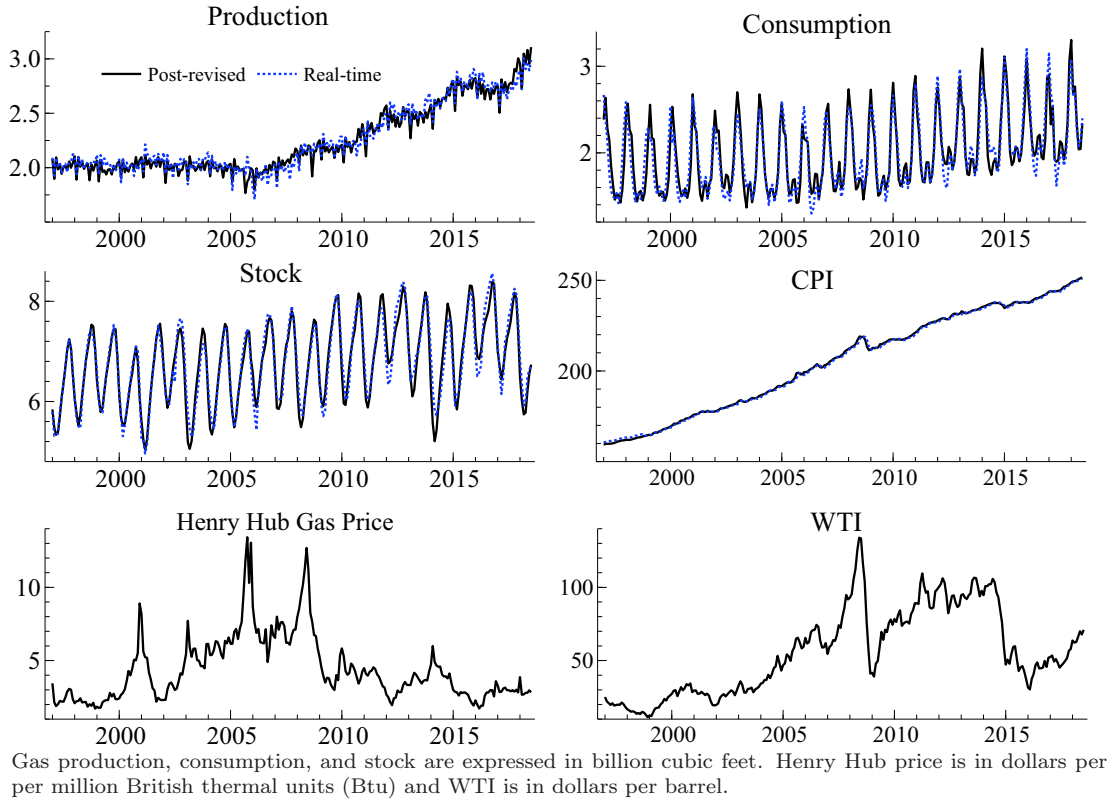
¹³In preliminary experiments, we estimated our VAR models using the variation of the real-time U.S. gas stock and, in all cases, forecasting performance was significantly poorer. Results are available upon request.

¹⁴Figure A.1 in Appendix A shows the original data on heating degree days and the resulting $nCHDD$ measure of temperature.

change in variables up to the required date. Certainly, such an approach is not applicable to gas consumption and inventory, which require specific processing to properly take into account their marked seasonality.¹⁵

The real-time series data we end with are plotted in Figure 1 along with the post-revised series (sub-figures production, Consumption, Stock and CPI). The gap between each pair of series is small, on average, but is more pronounced for gas production because of the three-month gap between the first release and the last data revision. Overall, Figure 1 visually confirms the goodness-of-fit of our approach.

Figure 1 – Real-time and post-revised data variables



3 Models under consideration

This section introduces the specification of the Bayesian VAR (BVAR) models and an overview of the estimation methods. Our framework includes several BVARs with possible stochastic volatility (SV) and/or shocks having Student’s t -distribution along with some exogenous regressors in a context where variable selection can take place.

Consider a VAR model and define n and n_e , the number of endogenous and exogenous variables, respectively:

$$y_t = c + B_1 y_{t-1} + \dots + B_p y_{t-p} + G_1 x_t + \dots + G_q x_{t-q} + u_t \quad \forall t \in [1, \dots, T] \quad (1)$$

¹⁵Based on the estimated (partial) autocorrelation functions (ACF, PACF) as well as on the Hannan–Quinn (HQ) information criterion, we select a SARIMA(0, 1, 1) × (0, 1, 1)₁₂ specification.

where y_t is an $n \times 1$ vector of observed endogenous variables, x_t is an $n_e \times 1$ vector of observed exogenous variables, and c is an $n \times 1$ vector of constants; $B_i, i = 1, \dots, p$ are $n \times n$ matrices of coefficients for the endogenous variables and $G_i, i = 1, \dots, q$ are $n_e \times n_e$ matrices of coefficients for the exogenous variables. Finally, u_t are heteroscedastic shocks associated with the VAR equations.

We assume that the covariance matrix of u_t is defined as:

$$\text{Cov}(u_t) = \Sigma_t^{-1} = A^{-1}H_tA^{-1'} \quad (2)$$

where $H_t = \text{diag}\left(\sigma_{1,t}^2 \frac{1}{\lambda_{1,t}}, \sigma_{2,t}^2 \frac{1}{\lambda_{2,t}}, \dots, \sigma_{n,t}^2 \frac{1}{\lambda_{n,t}}\right)$ with $\sigma_{k,t} = \ln \sigma_{k,t-1} + s_{k,t}$ for $k = 1, \dots, n$, where $s_{k,t}$ is the error term associated with the volatility, and $\text{var}(s_{k,t}) = h_k$, where h_k is the shocks to the volatility transition from [Cogley and Sargent \(2005\)](#).

As shown by [Geweke \(1993\)](#), assuming a Gamma prior for $\lambda_{k,t}$ of the form $p(\lambda_k) = \prod_{t=1}^T p(\lambda_{k,t}) = \prod_{t=1}^T \tilde{\Gamma}(1, \nu_{\lambda,k})$ leads to a scale mixture of normals for the orthogonal residuals $\tilde{\epsilon}_t = Au_t$ where $\tilde{\epsilon}_t = \{\tilde{\epsilon}_{1,t}, \tilde{\epsilon}_{2,t}, \dots, \tilde{\epsilon}_{n,t}\}$ and $\text{cov}(\tilde{\epsilon}_t) = H_t$. Note that $\tilde{\Gamma}(a, b)$ denotes a gamma density with mean a and b degrees of freedom b .

[Primiceri \(2005\)](#) and [Chiu et al. \(2017\)](#) provide evidence that stochastic volatility permits to capture the possible heteroscedasticity of the shocks along with potential nonlinearities in the dynamic relationships of the variables under consideration in the VAR specification, which are related to *low-frequency* variations in volatility. [Baumeister et al. \(2022\)](#) also rely on such extensions of their genuine linear specification. We name this model the VARVOLX as it allows for both stochastic volatility and exogenous regressors.

One way to deal with *high-frequency* variations in volatility is to consider the Student's t -distribution in the shock structure. As these variations are often very large, this specification can deal efficiently with outliers and extreme events.¹⁶ We thus estimate a tVARX model allowing for t -distributed errors but not for stochastic volatility.

We then enlarge our set of model with the estimation of a tVARVOLX model that permits to capture both the transient (low-frequency) and the persistent (high-frequency) shifts in volatility¹⁷ and a BVARX allowing for neither fat tails nor stochastic volatility, as a benchmark model. Hence, comparing all these four models enables to assess which specification best characterize the data.

Finally, because the number of parameters to estimate is large, we choose to make use of the variable selection technique. Specifically, variable selection can help reduce the risk of overfitting in our out-of-sample forecasting exercise. We implement variable selection for all our specifications, choosing as in [Korobilis \(2013\)](#) from three types of priors: Ridge regression prior, Minnesota (Litterman) prior, and Hierarchical Bayes Shrinkage prior for the penalization of coefficients. To ensure compatibility of the results with those from models without variable selection, we choose the Minnesota (Litterman) prior.

¹⁶[Jacquier et al. \(2004\)](#) first provided a detailed analysis of this issue in a univariate framework.

¹⁷The VARVOLX and the tVARX models allowing for stochastic volatility only and Student's errors only, respectively, are then restricted version of the tVARVOLX model. Details on how to deal with both stochastic volatility and fat tails at the same time are given in [Chiu et al. \(2017\)](#).

4 Estimation algorithms and forecasting procedure

We estimate all our models using the Markov Chain Monte Carlo (MCMC) Gibbs sampler. The VAR parameter priors are calibrated as in [Bańbura et al. \(2010\)](#).¹⁸ Our reported results are based on 20,000 Gibbs replications discarding the first 15,000 as burn-in. To gauge the convergence of the Gibbs algorithm for the selected draws, we report in [Figure B.1](#) in the [Appendix B](#) the recursive means of the estimated coefficients for some selected parameters. As little fluctuation is observed, convergence is reached.

We next assess the forecasting performance of our above models considered by producing pseudo out-of-sample forecasts.¹⁹ The four models are estimated recursively from August 2013 to August 2018 which will be our evaluation period. At each iteration, we construct the forecast density for the models.

$$P(\hat{y}_{t+h} | y_t) = \int P(\hat{y}_{t+h} | y_t, \Psi_{t+h}) P(\hat{\Psi}_{t+h} | \Psi_t, y_t) P(\hat{\Psi}_t | y_t) d\Psi \quad (3)$$

where $h = 1, 2, \dots, 12$ and Ψ denotes the model parameters. $P(\hat{\Psi}_t | y_t)$ represents the posterior density of the parameters that is obtained via the MCMC simulation. $P(\hat{y}_{t+h} | y_t, \Psi_{t+h})$ and $P(\hat{\Psi}_{t+h} | \Psi_t, y_t)$ denote the density forecast of the data and the parameters that can be obtained by simulation. The point forecast is therefore the mean of the forecast density. The density forecasts, which are evaluated using log scores (LS), are defined as follows:

$$LS_t = \ln P(y_{t+h}) \quad (4)$$

where $P(y_{t+h})$ denotes the forecast density evaluated at data observations. A higher value of LS_t suggests a more accurate density forecast. As in [Chiu et al. \(2017\)](#) we employ kernel methods to estimate the density and distribution function of the forecasts. This addresses potential non-linearity in the forecast distribution.

As a measure of the accuracy of point forecasts, we use *Mean Square Prediction Errors* (*MSPEs*). The *MSPE* for $h = 1, \dots, 12$ is computed as:

$$MSPE_h = \frac{1}{T - R} \sum_{t=R}^{T-1} (\hat{y}_{h,t+1|t} - y_{h,t+1})^2 \quad (5)$$

where T is the number of observations, R is the length of the rolling window and $\hat{y}_{h,t+1|t}$ are individual forecasts of the monthly average of real Henry Hub natural gas price forecasts and $y_{h,t+1}$ are monthly averages of real Henry Hub natural gas price forecasts. For this purpose we report the *MSPE* of each specification as well as the *MSPE* ratio relative to no-change forecasts.

Moreover, in order to statistically assess and identify the best forecasting model, we perform both conditional and unconditional pairwise model comparisons based on the forecasting performance test procedure of [Giacomini and White \(2006\)](#). This procedure is particularly relevant for both point and density forecasts when using recursive forecasting on real-time data.

¹⁸Again, more details about prior and hyperparameter choice for different VAR specifications (tVAR, VARVOL and tVARVOL) can be found in [Chiu et al. \(2017\)](#).

¹⁹All the estimations and computations presented in this paper were carried out using Julia. As the recursive estimation of different models can be very time-consuming, we take advantage of the computing clusters by using parallel computing in Julia.

5 Empirical results

In this section, we first compare the forecasting performance of real-time data to the one based on post-revised data and confirm the findings in [Baumeister and Kilian \(2012\)](#) for crude oil, namely that real-time data help in forecasting real prices. Second, we provide evidence of the additional informational content of temperature data in various VAR settings. Third, we show that improvement in forecasting accuracy mostly comes from including the variable on temperature conditions despite considering stochastic volatility and fat tails also enhance forecasts.

5.1 Do real-time data improve forecast accuracy?

To assess the advantages of using real-time data, we start performing the forecasting exercise, similar to what we do for real-time data, but using post-revised data. We then compute MSPE ratios relative to the no-change for both types of data.

Table 1 summarizes the real gas price forecast accuracy of both a univariate specification and multivariate Bayesian autoregressive models estimated with some various combinations of the potential endogenous variables. The univariate model is a simple Bayesian autoregressive (BAR) model whose results can serve as a point of reference when forecasts are made using the gas price series only. For both the BAR and all BVAR models, the best forecast accuracy is unsurprisingly obtained using 12 lags for the exogenous variable.

BVAR models include market real gas price (HH) itself, and its main drivers: the U.S. Natural Gas production (P) to proxy for supply, the U.S. Natural gas consumption (C) to proxy for demand, an estimate of U.S. Natural Gas Inventories (S) for the sake of comparison with the existing literature and the real West Texas Intermediate oil prices (WTI).²⁰ As is done in [Baumeister and Kilian \(2012\)](#), we compare the estimated recursive MSPE and to those of the no-change forecast for 1-month, 3-month, 6-month, 9-month, and 12-month forecasts.

We choose to evaluate the models over the August 2013–August 2018 period thereby leaving the January 1997–July 2013 period for estimation of the models’ parameters. Point forecasts are produced via recursive estimation. Each model is estimated using 20,000 iterations, with 15,000 burns.

Not surprisingly, and in line with the results in [Baumeister and Kilian \(2012\)](#) for the oil market, results from Table 1 Panel B show that models incorporating real-time data offer a statistically significant improvement over models incorporating post-revised data; for all models and for all horizons except the one-month horizon, MSPE ratios are lower than one.²¹ The improvement reaches 47% for the 12-month horizon. These substantial differences in forecasts support our hypothesis that using real-time data goes some way toward solving the well-known issues that arise with data revisions.

Besides, results from Table 1 (Panel A and B) provide strong evidence that BVAR models give higher forecast accuracy than a univariate model, thereby supporting the adopted

²⁰The use of Cholesky decomposition dictates our choice of variable order. WTI is ordered first in the VAR, in the light of the findings in [Kilian and Vega \(2011\)](#). Then, for robustness, we estimate different models with different variable ordering and results remain unchanged.

²¹For the sake of completeness, Table C.1 in Appendix C provides MSPE ratios relative to the no-change forecasts.

Table 1 – Recursive MSPE of forecasting accuracy

	Panel A: Real-Time				Panel B: Real-time over Post-Revised			
	BAR	BVAR			BAR	BVAR		
	P,C,HH,S	WTI,P,C,HH	WTI,P,C,HH,S		P,C,HH,S	WTI,P,C,HH	WTI,P,C,HH,S	
1M	1.1934	1.1265	1.0530	1.0307	1.0000	1.1412	1.0621	1.0341
3M	1.3037	1.0029	0.9289**	0.9121*	1.0000	0.9602*	0.9410*	0.9466*
6M	1.4487	0.8563**	0.7785***	0.7723**	1.0000	0.9049**	0.8044**	0.8274**
9M	1.6103	0.6988***	0.6371***	0.6368***	1.0000	0.7591***	0.6736***	0.6732***
12M	1.5902	0.5878***	0.5288***	0.5295***	1.0000	0.6583***	0.5732***	0.6009***

Notes: Panel A presents ratios of each model’s MSPE to the baseline no-change forecast successively evaluated for the entire evaluation period (August 2013–August 2018). Panel B presents MSPE ratios of the models including real-time data over those including post-revised data. Entries lower than 1 indicate that forecasts from the model are more accurate than forecasts from the associated baseline model. ***, ** and * indicate MSPE ratios are significantly different from 1 at 1%, 5% and 10%, according to both the conditional and unconditional [Giacomini and White \(2006\)](#) test.

multivariate framework. Apart from results at the one-month horizon, the forecast accuracy gains using BVAR models with respect to no-change are between 7% and 47%.

A comparison across BVAR models incorporating either real-time or revised data reveals the benefit of including WTI in terms of forecast accuracy. Including WTI leads to a relative reduction in MSPE of 21% on average (for all horizons). The relative reduction in MSPE from including natural gas inventories is 14%. Moreover, the five-variable BVAR model including both WTI and gas inventories performs roughly as well on forecasting as the four-variable BVAR including WTI only. In view of parsimony, our benchmark model in the rest of the paper will be the latter one.

Table D.2, in Appendix D, reports estimates of the log score of the BVARX model including nCHDD relative to the BVAR model, so that positive values indicate an improvement over our benchmark. Results point to significant improvements in most cases.

To summarize, linear Bayesian VAR models including variables quite similar to those in [Baumeister and Kilian \(2012\)](#) deliver highly competitive forecasts relative to the no-change prediction. This important result does confirm the reliability of our new real-time dataset and the ability of real-time fundamental variables related to the natural gas market to generate competitive forecasts of the future gas price in a multivariate framework.

5.2 The informational content of temperature data

We now wish to take advantage of the strong relationship between temperature and gas consumption as exemplified in [Timmer and Lamb \(2007\)](#). In other words, might models incorporating temperature as an additional regressor provide more accurate real gas price forecasts when considered in our Bayesian VAR specifications?

Gaussian Bayesian VAR

In Table 2 we report the performance of BARX and BVARX models incorporating either a measure of Heating Degree Days (HDD) or nCHDD, our measure of temperature as specified in Section 2.1, in the BVAR (WTI, P, C, HH) benchmark model.²²

²²Note that in the columns dedicated to BVAR (columns 1 and 5), only the MSPE is reported. Also note that the BARX model is estimated using the nCHDD measure only.

A remarkable finding, which is central to our analysis, is that BVARX models incorporating nCHDD exhibit the best forecasting performance, always dominating the benchmark model with respect to MSPE. This is true for all horizons and for both real-time and post-revised data. More specifically, the gains generated by including nCHDD are significant, between 5% and 15%, depending on the horizon.²³ The highest improvement in MSPE for real-time data is for the six-month horizon with a reduction of the MSPE higher than 15%. This amelioration of the forecast accuracy remains even when post-revised data are used as shown in Panel B of Table 2 where the BVARX outperforms the benchmark model by between 3% and 12%.²⁴

Another noteworthy conclusion from Table 2 is that including HDD instead of nCHDD leads to less accurate forecasts at all horizons, suggesting that the noise added from using this variable does more damage than good. This may be due to the high correlation between HDD and natural gas consumption (about 60% for our sample), meaning that HDD contains information already present in gas consumption data. Only extraordinary weather conditions are therefore likely to explain real gas price formation.

Overall, our results confirm that the inclusion of temperature data is fruitful at all horizons, thereby supporting the idea that considering demand in real-time through an adequate measure of temperature conditions significantly help in forecasting natural gas prices. While there is extensive evidence that temperature is a major determinant of natural gas price and consumption, our forecasting evidence is new to the literature and is likely to be helpful for all actors in the U.S. gas markets as well as for policy-makers whose decisions duly rely on energy prices.

Table 2 – Recursive MSPE ratio relative to BVAR as benchmark

X	Panel A: Real-Time				Panel B: Post-Revised			
	BVAR (MSPE)	BARX nCHDD	BVARX HDD nCHDD		BVAR (MSPE)	BARX nCHDD	BVARX HDD nCHDD	
1M	0.0756	3.4404	1.5058	0.9516*	0.0711	3.6542	1.6155	0.9682*
3M	0.2663	2.5093	1.3310	0.8997*	0.2833	2.3613	1.5026	0.9495*
6M	0.3985	1.8531	1.4328	0.8479**	0.4954	1.4908	1.1487	0.9266*
9M	0.4105	2.5401	1.4875	0.8711**	0.6093	1.7110	1.1252	0.9028*
12M	0.3878	2.0271	1.6064	0.9271*	0.6765	1.1620	1.1242	0.8808*

BARX is estimated using 12 lags. Panel A (resp. B) presents MSPE ratios of model including real-time data (resp. including post-revised data). For BVAR, only MSPE is provided. Entries lower than 1 indicate that forecasts from the model are more accurate than forecasts from the associated baseline model. ***, ** and * indicate MSPE ratios are significantly different from 1 at 1%, 5% and 10%, according to both the conditional and unconditional [Giacomini and White \(2006\)](#) test.

²³To complement our analysis, we also consider the density forecast accuracy as an additional tool to gauge the additional informational content of temperature conditions when forecasting real gas prices. All results for the density forecast analysis are provided in Appendix D. We first show that density forecasts using real-time data favorably compares to those from post-revised data (see Table D.1). Then, Table D.2 reports estimates of the log score of the BVARX model including nCHDD relative to the BVAR model, so that positive values indicate an improvement over our benchmark. Results point to significant improvements in most cases.

²⁴For exhaustiveness, MSPE ratios relative to the no-change – and not the benchmark model – are reported in Table C.2 of Appendix C.

Stochastic volatility and fat tails: Does it further help and how?

In light of the forecasting performance of the BVARX model incorporating nCHDD as an exogenous variable, we choose it as the new benchmark. Our aim is now to account for stochastic volatility and fat tails. Because the number of parameters to estimate is large, we allow for variable selection keeping in mind that variable selection may further improve forecasting accuracy.

To this end, we first present a comparison of their fitting properties (adequacy) based on marginal likelihood. Table 3 gives the estimated values of the log marginal likelihood for each model on the full sample. VS indicates whether variable selection is used. We observe that the combination of fat tails and stochastic volatility within the tVARVOLX model has the highest estimated marginal likelihood, thereby delivering the best fit. We also note that the simple (most restricted) BVARX has the lowest log marginal likelihood. Importantly, however, allowing for stochastic volatility improves the fit of the model much more than allowing for fat tails, which is in line with the properties of our time-series. In addition, variable selection always increases the fit of the model in all four specifications.

Table 3 – Log-likelihood

	BVARX		tVARX		VARVOLX		tVARVOLX	
VS	No	Yes	No	Yes	No	Yes	No	Yes
Log-likelihood	-1988	-1934	-1461	-1458	-1042	-1040	-1039	-1034

The log marginal likelihood values are estimated for each model using the full sample with 20,000 iterations. Estimation methodology follows [Chiu et al. \(2017\)](#)

We now turn to the forecasting performance itself. How do models allowing for stochastic volatility and fat tails compare with their restricted version? Recall that the reference model is a (WTI, P,C,HH) BVARX model with nCHDD as an additional regressor. Table 4 reports recursive MSPE of alternative models relative to the benchmark. All models are estimated with and without variable selection (VS).²⁵ Again, figures in the first column are MSPE.

Table 4 – Recursive MSPE ratio relative to BVARX

	BVARX		tVARX		VARVOLX		tVARVOLX	
VS	No	Yes	No	Yes	No	Yes	No	Yes
1M	0.0720	0.9931*	0.9929	0.9600*	0.9625*	0.9699*	0.9685*	0.9631*
3M	0.2396	0.9766*	0.9395*	0.9191*	0.8797**	0.8800**	0.8754**	0.8725**
6M	0.3379	0.9612*	0.9213*	0.9209*	0.8735**	0.8734**	0.8659**	0.8630**
9M	0.3576	0.9521*	0.8922*	0.9141*	0.8632**	0.8630**	0.8524**	0.8496**
12M	0.3596	0.9487*	0.8588**	0.9034*	0.8511**	0.8509**	0.8374**	0.8344**

Notes: This table compares each model's MSPE ratio relative to the baseline (WTI, P,C,HH) BVARX model evaluated for the (August, 2013 - August, 2018) period. The first column reports MSPE. Entries lower than 1 indicate that forecasts from the model are more accurate than forecasts from the associated baseline model. Yes (resp. No) in line VS (Variable selection) means (resp. No) Bayesian shrinkage in the VAR coefficient. ***, ** and * indicate MSPE ratios are significantly different from 1 at 1%, 5% and 10%, according to both the conditional and unconditional [Giacomini and White \(2006\)](#) test.

Three important points emerge from Table 4. First, allowing for fat tails appears to be

²⁵As we did for estimation without variable selection, we report in Figure B.2 in the Appendix B the recursive means of the estimated coefficients showing as little fluctuation than before. Hence, convergence is obtained.

beneficial: the tVARX model offers an MSPE reduction as large as 14% at the 12-month horizon without variable selection (10% when variable selection is applied).

Second, considering low-frequency movements in volatility through stochastic volatility appears to be useful at all horizons, especially longer horizons. It is clear, however, that the VARVOLX provides greater improvements than the tVARX, suggesting that gas market variables are more characterized by persistent than by transient shifts in volatility.

A third interesting observation from Table 4 is that the combination of fat tails and stochastic volatility (tVARVOLX) does deliver the best forecasting performance. The tVARVOLX dominates our benchmark significantly, with an MSPE reduction going from 13% to 16% for horizons higher than three months. Note, though, that the marginal improvement of the tVARVOLX over the VARVOLX is rather limited supporting the idea that high-frequency movements in volatility are of minor help to better forecast gas prices.

Further robustness analysis, regarding density forecasts, is included in Table D.3. The results corroborate previous findings. Log score comparisons are in line with the relative MSPE ratios, confirming that BVARX forecasting is strictly dominated when fat tails and stochastic volatility are taken into account. Improvements are highly significant at longer horizons, but not significant at the three-month horizon. Similarly, one month ahead, the improvement is quite limited. Once again, the VARVOLX specification delivers better forecasting performance than the tVAR model, both in magnitude and significance. The largest forecast accuracy gain is provided by tVARVOLX, with improvement reaching 58% at the 12-month horizon.

Confirming the source of improved forecasting performance

At this stage, we cannot dismiss the hypothesis that the improvement in forecasting ability of the BVAR models allowing for stochastic volatility and/or fat tails might be due to their own specifications only and not, or very few, related to the measure of temperature conditions. To investigate this issue, we compare the forecasting performance of these models to similar models with no exogenous variable.

Table 5 – Recursive MPSE ratio relative to each specification without exogenous variable

	BVARX		tVARX		VARVOLX		tVARVOLX	
	/BVAR		/tVAR		/VARVOL		/tVARVOL	
VS	No	Yes	No	Yes	No	Yes	No	Yes
1M	0.9516*	1.0189	1.003	0.9960	0.9941	0.9623	1.0146	1.0210
3M	0.8997*	0.8901*	0.8999*	0.8980*	0.9021*	0.8956*	0.9481*	0.9473*
6M	0.8479**	0.8657**	0.9997	0.9992	0.9999	0.9968	0.9247*	0.9209*
9M	0.8711**	0.8459**	0.8634**	0.8592**	0.9167*	0.8955*	0.8944*	0.8971*
12M	0.9271*	0.9303*	0.9547*	0.9293*	0.8960*	0.9001*	0.8623**	0.8664**

Notes: This table compares each model's MSPE ratio to the same model without exogenous variable evaluated for the (August, 2013 - August, 2018) period. Yes (resp. No) in line VS (Variable selection) means (resp. No) Bayesian shrinkage in the VAR coefficient.***, ** and * indicate MSPE ratios are significantly different from 1 at 1%, 5% and 10%, according to both the conditional and unconditional [Giacomini and White \(2006\)](#) test.

Relative MSPE ratios from Table 5 unambiguously confirms the relevance of using temperature as an additional regressor in our empirical analysis. Indeed, almost all models incorporating nCHDD exhibit better forecasting accuracy than their counterparts with no

exogenous variable. A few exceptions at the one-month horizon can be observed from the Table but associated estimates are insignificant and are therefore not likely to cast doubt on the usefulness of using temperature conditions as an additional regressor.²⁶

As pointed out by [Korobilis \(2013\)](#), the more variables in the model, the more useful penalization within variable selection will become. In our study, variable selection in many cases improved forecasting performance, despite the small number of variables included in the VAR models. We estimate our models with different values for the hyper-parameter (π_{0j}) which parameterizes the proportion of the predictors that should be in the final model. We thus select values between 0.5 (non-informative), meaning that half of the BVAR coefficients should be restricted, and 0.8, meaning only 20% of coefficients will be restricted. The final value used for the reported estimation results is 0.8, which gives the best forecasting performance gain. This is also consistent with the assertion of [Korobilis \(2013\)](#) that, for small BVAR, choosing a non-informative prior for this hyper-parameter probably yields too many restricted BVAR coefficients, thereby worsening the model’s forecasting performance.

6 Concluding remarks

The U.S. natural gas price and its evolution are vital to the U.S. economy. This makes reliable forecasts of the real price of gas important to market participants, as well as to consumers, regulators, and central bankers.

In this paper, we estimate various Bayesian VAR models following the [Baumeister and Kilian \(2012\)](#) seminal contribution recently extended in [Baumeister et al. \(2022\)](#) with an emphasis on the demand-side. We provide evidence that considering demand is highly relevant in the case of the U.S. natural gas market. We further show that a measure of temperature conditions that is, by nature, truly available in real-time does improve significantly the forecasting accuracy at most horizons from 1 to 12 months. We also contribute to the literature in proving that allowing for stochastic volatility is critical to the performance of the models under consideration.

While we do provide evidence of significant statistical improvements in terms of forecasting, we do not evaluate the economic value of such an enhancement. Forecasts may thus be used to price option on natural gas futures as quoted on the NYMEX (CME Group).

Another strand of the literature initiated in [Alquist and Kilian \(2010\)](#) deals with the information content of futures prices to make predictions about commodity prices. While we do not elaborate on this possibility for natural gas in the present paper, a related approach in [Thomas \(2020\)](#), relying on a non-causal bivariate VAR for both crude oil and natural gas, deserves further attention. In his model, the only two variables under consideration are the energy price and the convenience yield. The latter variable is included on the economic grounds that it proxies for expectations in the derivatives market. Surprisingly, this very parsimonious model exhibits highly accurate forecasts at horizons of up to several months.

²⁶Log scores in Table D.4 provide qualitatively similar results for density forecasts. Models incorporating nCHDD remain the best-performing competitors, though forecasting performance gains are not statistically significant in this case.

References

- Alquist, R., Kilian, L., 2010. What do we learn from the price of crude oil futures? *Journal of Applied Econometrics* 25 (4), 539–573.
- Alquist, R., Kilian, L., Vigfusson, R. J., 2013. Chapter 8 - forecasting the price of oil. In: Elliott, G., Timmermann, A. (Eds.), *Handbook of Economic Forecasting*. Vol. 2 of *Handbook of Economic Forecasting*. Elsevier, pp. 427 – 507.
- American Gas Association, 2019.
URL <https://www.aga.org/globalassets/2019-aga-overview-updated.pdf>
- Auffhammer, M., 2007. The rationality of eia forecasts under symmetric and asymmetric loss. *Resource and Energy Economics* 29 (2), 102–121.
- Bachmeier, L. J., Griffin, J. M., 2006. Testing for market integration crude oil, coal, and natural gas. *Energy Journal* 27 (2), 55–71.
- Baumeister, C., Guérin, P., Kilian, L., 2015. Do high-frequency financial data help forecast oil prices? the midas touch at work. *International Journal of Forecasting* 31 (2), 238 – 252.
- Baumeister, C., Hamilton, J. D., 2019. Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks. *American Economic Review* 109 (5).
- Baumeister, C., Kilian, L., 2012. Real-time forecasts of the real price of oil. *Journal of Business & Economic Statistics* 30 (2), 326–336.
- Baumeister, C., Kilian, L., Lee, T., 2017. Inside the crystal ball: New approaches to predicting the gasoline price at the pump. *Journal of Applied Econometrics* 32 (2), 275–295.
- Baumeister, C., Kilian, L., Zhou, X., 2018. Are product spreads useful for forecasting oil prices? an empirical evaluation of the verleger hypothesis. *Macroeconomic Dynamics* 22 (3), 562–580.
- Baumeister, C., Korobilis, D., Lee, T. K., 2022. *Energy Markets and Global Economic Conditions*. *Review of Economics and Statistics* 104 (4), 828–844.
- Bañbura, M., Giannone, D., Reichlin, L., 2010. Large Bayesian vector auto regressions. *Journal of Applied Econometrics* 25 (1), 71–92.
- Bello, A., Bunn, D. W., Reneses, J., Munoz, A., 2017. Medium-term probabilistic forecasting of electricity prices: A hybrid approach. *IEEE Transactions on Power Systems* 32 (1), 334–343.
- Berrisford, H. G., 1965. The relation between gas demand and temperature: A study in statistical demand forecasting. *Journal of the Operational Research Society* 16 (2), 229–246.
- Bessembinder, H., Chan, K., 1992. Time-varying risk premia and forecastable returns in futures markets. *Journal of Financial Economics* 32 (2), 169 – 193.
- Brehm, P., 2019. Natural gas prices, electric generation investment, and greenhouse gas emissions. *Resource and Energy Economics* 58, 101106.
- Brown, S. P., Yücel, M. K., 2008. Deliverability and regional pricing in U.S. natural gas markets. *Energy Economics* 30, 2441–2453.

- Bunn, D., Chevallier, J., Le Pen, Y., Sévi, B., 2017. Fundamental and financial influences on the co-movement of oil and gas prices. *Energy Journal* 38 (2), 201–228.
- Caporin, M., Fontini, F., 2017. The long-run oil–natural gas price relationship and the shale gas revolution. *Energy Economics* 64, 511 – 519.
- Chen, Y.-C., Rogoff, K. S., Rossi, B., 08 2010. Can Exchange Rates Forecast Commodity Prices?*. *The Quarterly Journal of Economics* 125 (3), 1145–1194.
- Chiu, C.-W. J., Mumtaz, H., Pintér, G., 2017. Forecasting with VAR models: Fat tails and stochastic volatility. *International Journal of Forecasting* 33 (4), 1124–1143.
- Cogley, T., Sargent, T. J., April 2005. Drift and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S. *Review of Economic Dynamics* 8 (2), 262–302.
- Doyle, M., Fell, H., 2018. Fuel prices, restructuring, and natural gas plant operations. *Resource and Energy Economics* 52, 153–172.
- Energy Information Administration, 2014. Short-term energy outlook supplement: Weather sensitivity in natural gas markets.
URL https://www.eia.gov/outlooks/steo/special/pdf/2014_sp_03.pdf
- Energy Information Administration, 2019. Natural gas annual report.
URL <https://www.eia.gov/naturalgas/annual/pdf/nga19.pdf>
- Ferrari, D., Ravazzolo, F., Vespignani, J., 2021. Forecasting energy commodity prices: A large global dataset sparse approach. *Energy Economics* 98, 105268.
- Gao, S., Hou, C., Nguyen, B., 2020. Forecasting natural gas prices using highly flexible time-varying parameter models. Tasmanian School of Business and Economics, University of Tasmania.
- Geweke, J., 1993. Bayesian treatment of the independent student-t linear model. *Journal of Applied Econometrics* 8 (S1), S19–S40.
- Giacomini, R., White, H., 2006. Tests of conditional predictive ability. *Econometrica* 74 (6), 1545–1578.
- Hong, H., Yogo, M., 2012. What does futures market interest tell us about the macroeconomy and asset prices? *Journal of Financial Economics* 105 (3), 473 – 490.
- Hou, C., Nguyen, B. H., 2018. Understanding the U.S. natural gas market: A Markov switching VAR approach. *Energy Economics* 75, 42–53.
- Huurman, C., Ravazzolo, F., Zhou, C., 2012. The power of weather. *Computational Statistics & Data Analysis* 56 (11), 3793 – 3807.
- Jacquier, E., Polson, N. G., Rossi, P., 2004. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Journal of Econometrics* 122 (1), 185–212.
- Jadidzadeh, A., Serletis, A., 2017. How does the U.S. natural gas market react to demand and supply shocks in the crude oil market? *Energy Economics* 63, 66–74.
- Kilian, L., 2016. The impact of the shale oil revolution on U.S. oil and gasoline prices. *Review of Environmental Economics and Policy* 10 (2), 185–205.

- Kilian, L., Murphy, D. P., 2014. The role of inventories and speculative trading in the global market for crude oil. *Journal of Applied Econometrics* 29 (3), 454–478.
- Kilian, L., Vega, C., May 2011. Do Energy Prices Respond to U.S. Macroeconomic News? A Test of the Hypothesis of Predetermined Energy Prices. *The Review of Economics and Statistics* 93 (2), 660–671.
- Korobilis, D., March 2013. VAR Forecasting Using Bayesian Variable Selection. *Journal of Applied Econometrics* 28 (2), 204–230.
- Müller, J., Hirsch, G., Müller, A., 2015. Modeling the price of natural gas with temperature and oil price as exogenous factors. In: *Innovations in Quantitative Risk Management*. Springer International Publishing, p. 109–128.
- Nguyen, B. H., Okimoto, T., 2019. Asymmetric reactions of the U.S. natural gas market and economic activity. *Energy Economics* 80, 86–99.
- Nick, S., Thoenes, S., 2014. What drives natural gas prices? — a structural var approach. *Energy Economics* 45, 517–527.
- Panagiotidis, T., Rutledge, E., 2007. Oil and gas markets in the U.K.: Evidence from a cointegrating approach. *Energy Economics* 29 (2), 329–347.
- Primiceri, G. E., 2005. Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies* 72 (3), 821–852.
- Thomas, A., 2020. The role of expectations in predicting the real price of oil: A non causal analysis. Mimeo.
- Timmer, R. P., Lamb, P. J., 2007. Relations between temperature and residential natural gas consumption in the central and eastern United States. *Journal of Applied Meteorology and Climatology* 46 (11), 1993–2013.
- Wang, Y., Liu, L., Wu, C., 2020. Forecasting commodity prices out-of-sample: Can technical indicators help? *International Journal of Forecasting* 36 (2), 666 – 683.

Appendix

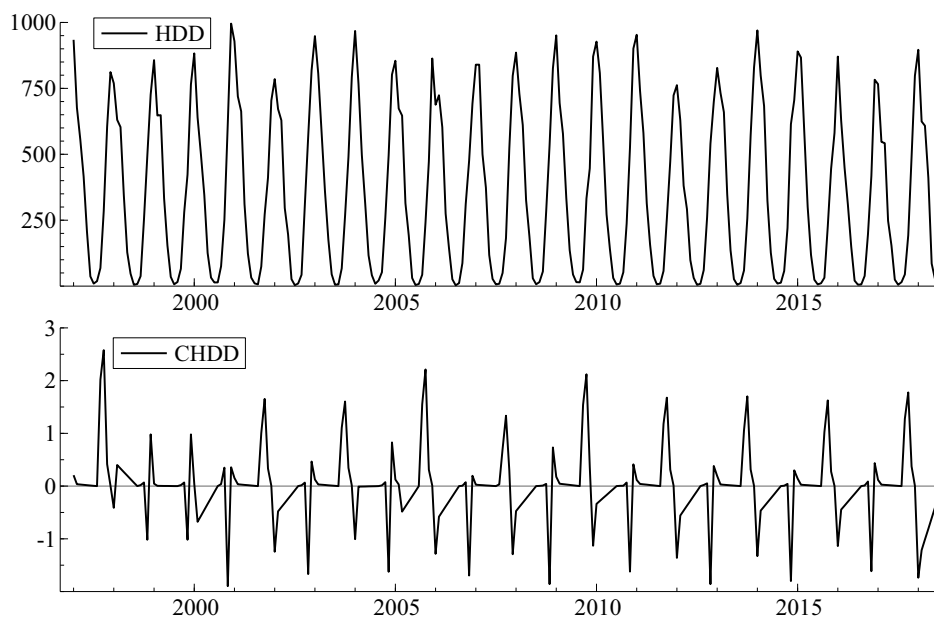
A Data

Following Müller et al. (2015), we use the normalized Cumulated Heating Degree Days (CHDD). More explicitly, we cumulate heating degree days over a winter, from which we subtract the average of cumulated heating degree days over previous winters. Our measure of the normalized cumulated heating degree days is :

$$nCHDD_{d,w} = CHDD_{d,w} - \frac{1}{w-1} \sum_{l=1}^{w-1} CHDD_{d,l} \text{ for } 1 \leq d \leq 182$$

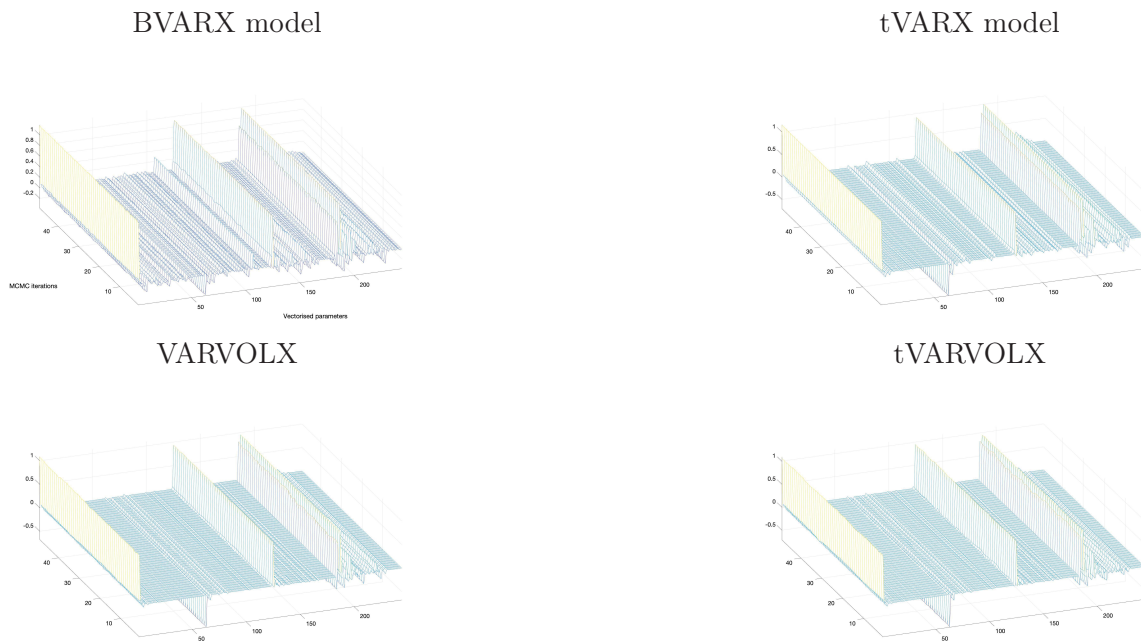
where 182 is the number of days in a winter w (from the 1st of October to the 31st of March). Then, Λ returns to zero during next summer.

Figure A.1 – The nCHDD measure of temperature conditions



B Empirical convergence results

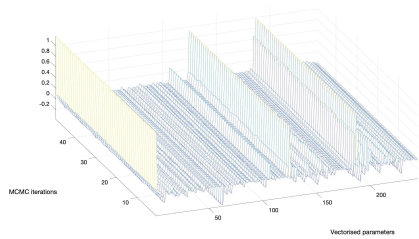
Figure B.1 – Mean recursive estimations of VAR coefficients for different specifications



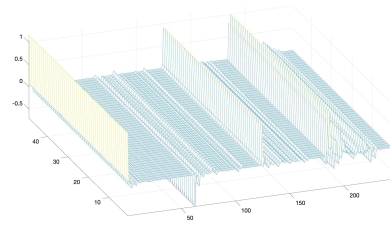
Figures in [B.1](#) represent the recursive mean of the variation in the coefficients of the autoregressive part of the Bayesian model along the iterations selected from the Gibbs sample (5,000 iterations). According to these figures, there is minimal variation in the estimated value for each coefficient, which ensures the convergence of the model. [B.2](#) is the counterpart with variable selection and provide similar results with respect to convergence.

Figure B.2 – Mean recursive estimations of VAR coefficients for different specifications

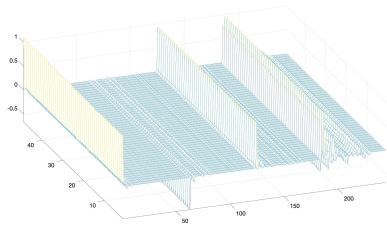
BVARX model with variable selection



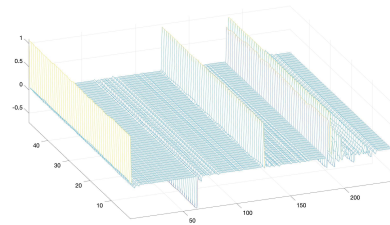
tVARX model



VARVOLX with variable selection



tVARVOLX with variable selection



C Temperature and post-revised data

Table C.1 – Recursive MSPE ratio relative to the no-change forecast with post-revised data

	BAR		BVAR		
			P,C,HH,S	WTI,P,HH,C	WTI,P,C,HH,S
1M	1.1934	0.9871		0.9914	0.9967
3M	1.3037	1.0443		0.9871	0.9635
6M	1.448	0.9462		0.9678	0.9334*
9M	1.6103	0.9205*		0.9458	0.9459
12M	1.5902	0.8929**		0.9225*	0.8811**

Notes: The table compares each model's MSPE ratio to the baseline no-change forecast successively evaluated for the entire evaluation period (August 2013–August 2018). Entries lower than 1 indicate that forecasts from the model are more accurate than forecasts from the associated baseline model. ***, ** and * indicate MSPE ratios are significantly different from 1 at 1%, 5% and 10%, according to both the conditional and unconditional [Giacomini and White \(2006\)](#) test.

Table C.2 – Recursive MSPE ratio relative to the no-change forecast (Temperature)

	Panel A: Real-Time				Panel B: Post-Revised			
	BVAR	BARX	BVARX		BVAR	BARX	BVARX	
X		CHDD	HDD	CHDD		CHDD	HDD	CHDD
1M	1.0530	3.6228	1.5857	1.0021	0.9914	3.6228	1.6017	0.9606*
3M	0.9289*	2.3309	1.3310	0.8339**	0.9871**	2.3309	1.4833	0.9373*
6M	0.7785***	1.4427	1.0898	0.6601***	0.9678*	1.4427	1.1041	0.8967*
9M	0.6371***	1.6183	0.9477*	0.5550***	0.9458*	1.6183	1.0643	0.8539**
12M	0.5288***	1.0720	0.8495**	0.4903***	0.9225*	1.0720	1.0371	0.8124**

Notes: Panel A (resp. Panel B) compares MSPE ratios of models including real-time data (resp. including post-revised data) to the baseline no-change model forecast successively evaluated for the entire evaluation period (August 2013–August 2018). Entries lower than 1 indicate that forecasts from the model are more accurate than forecasts from the associated baseline model. ***, ** and * indicate MSPE ratios are significantly different from 1 at 1%, 5% and 10%, according to both the conditional and unconditional [Giacomini and White \(2006\)](#) test.

D Density forecast accuracy

Table D.1 – Recursive average percentage improvement in Log-Score

Real-time over Post-Revised			
BVAR			
	P,C,HH,S	WTI,P,HH,C	WTI,P,C,HH,S
1M	-3.62	-5.75	-5.39
3M	2.85	2.88	6.54
6M	4.77	1.49	1.03
9M	7.29*	5.52	1.41
12M	7.46*	18.00**	10.30

Notes: The table compares the percentage values of log predictive density scores of models incorporating real-time data to those incorporating post-revised data successively evaluated for the entire evaluation period (August 2013–August 2018). The notation BVAR refers to univariate Bayesian vector autoregression. Density forecasts are produced via recursive estimation of the models. Positive entries indicate that Log-Scores from the model are improved (in %) relative to the Log-Score from the associated baseline model. ***, ** and * indicate Log-Score variations are significantly different from 0 at 1%, 5% and 10%, according to both the conditional and unconditional [Giacomini and White \(2006\)](#) test.

Table D.2 – Recursive average percentage improvement in Log-Score relative to BVAR as the benchmark

	Panel A: Real-Time				Panel B: Post-Revised			
	BVAR	BARX	BVARX		BVAR	BARX	BVARX	
X		CHDD	HDD	CHDD		CHDD	HDD	CHDD
1M	0.7062	-86.024	-24.15	2.10	0.7417	-82.78	-27.98	-0.01
3M	0.5930	-765.26	-11.49	5.77	0.5186	-731.26	-12.07	4.20
6M	0.4399	-2826.83	-12.98	10.38*	0.4386	-2857.98	-10.43	9.99*
9M	0.3707	$-\infty$	-17.34	12.67**	0.3516	$-\infty$	-15.95	8.64*
12M	0.3314	$-\infty$	-28.66	8.20*	0.2806	$-\infty$	-20.05	8.35*

Notes: Panel A (resp. Panel B) compares the percentage values of log predictive density scores of models incorporating real-time data (resp. post-revised data) to BVAR. For BVAR, only Log-Score is provided. Positive entries indicate that Log-Scores from the model are improved (in %) relative to the Log-Score from the associated baseline model. ***, ** and * indicate Log-score variations are significantly different from 0 at 1%, 5% and 10%, according to both the conditional and unconditional [Giacomini and White \(2006\)](#) test.

Table D.3 – Recursive average percentage improvement in Log-Score relative to BVARX as the benchmark

	BVARX		tVARX		VARVOLX		tVARVOLX	
	No	Yes	No	Yes	No	Yes	No	Yes
VS								
1M	0.7210	1.22	-5.82	-5.78	-2.52	-3.16	1.4	-4.16
3M	0.6272	5.53	2.98	2.79	1.23	3.89	3.40	5.02
6M	0.4956	2.84	11.03*	9.50*	18.24***	21.11***	24.28***	23.87***
9M	0.4092	2.10	19.09***	18.65***	32.16***	33.28***	37.00***	40.42***
12M	0.3610	6.00	30.68***	31.20***	48.33***	48.85***	53.79***	58.22***

Notes: The table compares the percentage values of log predictive density scores of models incorporating t-distribution and/or stochastic volatility to the baseline models. The baseline model is a BVARX model without variable selection. For BVARX, only predictive density scores are provided. Positive entries indicate that Log-Scores from the model are improved (in %) relative to the Log-Score from the associated baseline model. The forecasts are produced via recursive estimation of the models successively evaluated for the entire evaluation period (August 2013–August 2018). Yes (resp. No) in line VS (Variable selection) means (resp. No) Bayesian shrinkage in the VAR coefficient. ***, ** and * indicate Log-score variations are significantly different from 0 at 1%, 5% and 10%, according to both the conditional and unconditional [Giacomini and White \(2006\)](#) test.

Table D.4 – Recursive average percentage improvement in Log-Score relative to each specification without exogenous variable

	BVARX		TVARX		VARVOLX		TVARVOLX	
	No	Yes	No	Yes	No	Yes	No	Yes
VS								
1M	2.10	2.89	0.43	-0.38	0.39	1.00	1.70	2.60
3M	5.77*	7.47*	1.19	1.00	2.29	0.47	6.61*	4.82
6M	12.67**	11.93**	0.13	1.85	3.61	4.00	17.44**	17.01**
9M	10.38**	14.33**	0.53	1.08	2.25	5.50	7.98*	18.09**
12M	8.93*	8.07*	2.06	1.29	4.08	2.95	24.15***	21.78***

Notes: The table compares the percentage values of log predictive density scores of models incorporating exogenous variables with the associated model without exogenous variable successively evaluated for the entire evaluation period (August 2013– August 2018). The forecasts are produced via recursive estimation of the models. Positive entries indicate that Log-Scores from the model are improved (in %) relative to the Log-Score from the associated baseline model. Yes (resp. No) in line VS (Variable selection) means (resp. No) Bayesian shrinkage in the VAR coefficient. ***, ** and * indicate Log-score variations are significantly different from 0 at 1%, 5% and 10%, according to both the conditional and unconditional [Giacomini and White \(2006\)](#) test.