Energy Economics 65 (2017) 411–423

Contents lists available at ScienceDirect

Energy Economics

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

Comparison of data-rich and small-scale data time series models generating probabilistic forecasts: An application to U.S. natural gas gross withdrawals

Kannika Duangnate [a,*], James W. Mjelde [b]

[a] Faculty of Economics, Chiang Mai University, Chiang Mai, Thailand
[b] Department of Agricultural Economics, Texas A&M University, College Station, TX 77843-2124, United States

ARTICLE INFO

Article history:
Received 18 July 2016
Received in revised form 3 March 2017
Accepted 24 April 2017
Available online 3 May 2017

JEL classification:
Q43
Q47

Keywords:
FAVAR
Prequential forecasting
Probability forecasting
Model selection
Energy forecasting

ABSTRACT

Time series models derived from using data-rich and small-scale data techniques are estimated to examine: 1) if data-rich methods forecast natural withdrawals better than typical small-scale data, time series methods; and 2) how the number of unobservable factors included in a data-rich model influences the model's probabilistic forecasting performance. Data rich technique employed is the factor-augmented vector autoregressive (FAVAR) approach using 179 data series; whereas the small-scale technique uses five data series. Conclusions drawn are ambiguous. Exploiting estimated factors improves the forecasting ability, but including too many factors tends to exacerbate probabilistic forecasts performance. Factors, however, may add information about seasonality for forecasting natural gas withdrawals. Results of this study indicate the necessity to examine several measures and to take into account the measure(s) that best meets the purpose of the forecasts.

1. Introduction

Given the importance of both the energy sector and forecasting for decision making, it is not surprising that there is an enormous volume of literature on forecasting concerning energy. Forecasting some form of price dominates this literature (He and Casey, 2015; Noel and Chu, 2015; Wang et al., 2015; Ziel et al., 2015). Consumption (demand) and generation (supply) follow a distant second and third in terms of number of articles (Castelli et al., 2015; El-Shazly, 2013; Moreno, 2009). Despite the importance of other aspects in the energy sector, one only occasionally sees articles on forecasting these aspects such as construction costs (Keng, 1985), petroleum trade (Lee and Das, 1989), CO_2 emissions (Christodoulakis et al., 2000), alternative fuel vehicles (Ahn et al., 2008), natural gas production from shale (Cueto-Felgueroso and Juanes, 2013), and adoption of energy efficient technology (Hlavinka et al., 2016).

Although price is often the ultimate issue, knowledge of these other aspects are important. Physical quantities, for example, determine infrastructure and labor needs, along with capital requirements. Natural gas gross withdrawals, the variable of interest here, have implications for the size and operation of the gathering, processing, and transportation systems for transporting natural gas from the field to distribution systems for delivery to consumers. When natural gas is processed, useful byproducts known as natural gas liquids are separated from the gas (NaturalGas.org, 2013). The amount and quality of the raw natural gas determine the amount of these byproducts; therefore, determining capital and labor necessary to handle the byproducts. The use of horizontal drilling and hydraulic fracturing has prominently increased the capability of producers to commercially recover natural gas and oil from low-permeability geologic formations, mainly shale formations (U.S. Energy Information Administration, 2011). Increasing shale gas production has led to transmission bottleneck issues within the industry (Grimes, 2014; American Petroleum Institute, 2014).

Because almost all if not all sectors of the economy rely on energy, it is not surprising that the sector is a complex interaction among financial, macroeconomic, climate, and regulatory factors (Barsky and Kilian, 2004; Zagaglia, 2010; Duangnate, 2015). Incorporating richer data sets in time series analysis to capture these complex linkages has caught the attention of academics (Sargent and Sims, 1977; Stock and Watson, 2002; Bernanke and Boivin, 2003; Bernanke et al., 2005; Moench, 2008; Zagaglia, 2010; Naser, 2016). These studies assume that variation in economic time series are captured by a small set of

* Corresponding author.
E-mail addresses: kannika.du@cmu.ac.th (K. Duangnate), j-mjelde@tamu.edu (J.W. Mjelde).

http://dx.doi.org/10.1016/j.eneco.2017.04.024
0140-9883/

Abbreviations

FAVAR factor-augmented vector autoregressive
FAVAR(*h*) factor-augmented vector autoregressive with *h* factors included
VAR vector autoregressive (VAR)
VAR(w) vector autoregressive with w variables included
AR autoregressive
AR(w) autoregressive with w lags
RMSE root mean-squared error
PC_{p1} and PC_{p2} two panel criteria 1 and 2 (PC_{p1} and PC_{p2})
IC_{p1} and IC_{p2} two information criteria 1 and 2
AIC_3 Akaike information criteria
BIC_3 Bayesian information criteria
RPS ranked probability score
GDP gross domestic product
PSM multiple probability score

Brier score decompositions
MinVar(*f*) the dispersion of probability forecasts, which cannot be explained by the conditional dispersion
Scat(*f*) the weighted average of the conditional variances
$Bias^2$ calibration error regardless of the direction (positive or negative) of the error
Cov(*f*,*d*) the ability of a model in distinguishing whether events occur or not
Var(*d*) captures out-of-model factors affecting forecasts

influencing variables; these variables are considered the set of common factors (Sargent and Sims, 1977).

The objective is to investigate if data rich method forecasts natural gas withdrawals better than typical small-scale data, time series methods that usually limit the analysis to eight or fewer variables. In addition, how the number of unobservable factors included in a data-rich model influences the model's probabilistic forecasting performance is examined. The factor-augmented vector autoregressive (FAVAR) approach, proposed by Bernanke et al. (2005), is compared to standard vector autoregressive (VAR) and univariate models. Prequential forecasting approach introduced by Dawid (1984) is applied to evaluate predictive distributions. Two FAVAR models differing in their number of factors (five and ten factors) based on the range of optimal number of factors derived from Bai and Ng's (2002) criteria, along with a five variable VAR and a univariate model are compared. Bai and Ng (2002) propose several criteria to determine the appropriate number of common factors to include. In empirical applications, Bai and Ng's (2002) various criteria, however, usually lead to differing numbers of factors (Moench, 2008; Zagaglia, 2010).

Natural gas withdrawals are the full well-stream volume of produced natural gas, excluding condensate separated at the lease, which is roughly the quantity of natural gas supplied from U.S. gas wells. As previously noted, the energy sector including natural gas quantity supplied is a function of the complex interactions of among financial, macroeconomic, climate, and regulatory variables. The FAVAR approach allows for the use of many more of these variables than other methodologies. Factors for the use in the FAVAR models are derived from a data set containing 178 data series on the energy sector and U.S. macroeconomic conditions. Because of limitations of empirical models in considering large data set, the scope of energy studies using time series methods is usually restricted such that either variables from only a specific sub-sector or a couple selected variables from assorted sub-sectors are considered. In the time series literature, few studies focus on the entire energy system interacting with the economy. To our knowledge, this is the first study to generate probabilistic forecasts for a non-price energy variable using the FAVAR framework. The study contributes to both the forecasting and the energy sector literature. The FAVAR framework allows for a more complete interaction of the economic variables influencing natural gas withdrawals. Specification of the number of factors is central to the empirical validity of the FAVAR model (Bai and Ng, 2002), but few studies have examined whether the number of factors influence forecasting performance. Thus, this study adds to our knowledge of the influence of the number of factors and the influence of modeling using data rich instead of small-scale methodologies.

2. Literature review

Studies using data-rich environments suggest that the use of large data sets improves inferences and forecast precision over small data sets (Stock and Watson, 2002; Bernanke and Boivin, 2003; Bernanke et al., 2005; Moench, 2008; Zagaglia, 2010; Bupta and Kabundi, 2011). Stock and Watson (2002) extract common factors from a large data set using principal components methods. They show that forecasting models, which include these common factors, outperform models such as univariate autoregressive, traditional vector autoregressive, Bayesian vector autoregressive, and leading indicator models. Bernanke and Boivin (2003) employ the factor-model approach developed by Stock and Watson (2002) to estimate and forecast the Federal Reserve's policy reaction function. Their findings are in line with Stock and Watson's (2002) results that including systematic information found in large data sets summarized by a relatively few estimated factors improves forecasting performance.

Bernanke et al. (2005) propose a FAVAR model in which both unobservable factors and observable economic variables (such as a policy indicator) characterize the common forces that determine the dynamics of the macroeconomic economy. They apply the model to measure the effects of monetary policy; exploiting information derived from the FAVAR model significantly increases the ability of identifying the monetary transmission mechanism. Bernanke et al. (2005, p. 406) claim, "The FAVAR approach is successful at extracting pertinent information from a large data set of macroeconomic indicators."

Employing the FAVAR approach in a data-rich environment improves forecasting performance (Moench, 2008; Zagaglia, 2010). Moench (2008) uses short-term interest rates as policy instrument and factors obtained from a large number of macroeconomic variables to forecast the yield curve under a no-arbitrage restriction. He finds that macroeconomic variables explain most of the variation in interest rates. The no-arbitrage FAVAR model forecasts the yield curve better than the Duffee (2002) or the Nelson-Seigel model modified by Diebold and Li (2006). Zagaglia (2010) extracts common factors from a large data set including global macroeconomic indicators, financial market indices, and quantities and prices on energy products to study the dynamics of oil futures' prices traded at NYMEX using a FAVAR model. He finds that the estimated factors can be categorized into energy prices, energy quantities, and macroeconomic and financial data. Combining these factors with oil returns improves the forecasting performance of oil futures' prices over a VAR model of returns only, a factor-included VAR model, and a random walk model.

Studies such as Banerjee et al. (2014), Li and Chen (2014), and Barnett et al. (2014) have extended the FAVAR methodology to include error correction and forecast combination techniques concluding FAVAR or its extensions generally forecast better than small-scale data, benchmark models. In addition, the extensions tend to forecast better than the FAVAR models. However, few studies, if any, have examined how the inclusion of a different number of factors in FAVAR influences the performance of the FAVAR.

A search of the literature revealed no study using a natural gas variable as the augmenting variable in a FAVAR. Further, Arora and Lieskovsky (2012) state that little is known about natural gas impact on the U.S. economy. Although there is a multitude of natural gas issues