International Journal of Forecasting 31 (2015) 238–252

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Do high-frequency financial data help forecast oil prices? The MIDAS touch at work

Christiane Baumeister [a], Pierre Guérin [a], Lutz Kilian [b,c,*]

[a] Bank of Canada, Canada
[b] University of Michigan, United States
[c] CEPR

ARTICLE INFO

Keywords:
Mixed frequency
Real-time data
Oil price
Forecasts

ABSTRACT

In recent years there has been an increased interest in the link between financial markets and oil markets, including the question of whether financial market information helps to forecast the real price of oil in physical markets. An obvious advantage of financial data in forecasting monthly oil prices is their availability in real time on a daily or weekly basis. We investigate the predictive content of these data using mixed-frequency models. We show that, among a range of alternative high-frequency predictors, cumulative changes in US crude oil inventories in particular produce substantial and statistically significant real-time improvements in forecast accuracy. The preferred MIDAS model reduces the MSPE by as much as 28% compared with the no-change forecast and has a statistically significant directional accuracy as high as 73%. This MIDAS forecast is also more accurate than a mixed-frequency real-time VAR forecast, but is not systematically more accurate than the corresponding forecast based on monthly inventories. We conclude that there is not typically much lost by ignoring high-frequency financial data in forecasting the monthly real price of oil.

1. Introduction

The substantial variation in the real price of oil since 2003 has renewed interest in the question as to how monthly and quarterly oil prices should be forecast.[1] The links between financial markets and the price of oil have received particular attention, including the question of whether financial market information may help forecast the price of oil in physical markets (e.g., Fattouh, Kilian, & Mahadeva, 2013). One obvious advantage of financial data is their availability in real time at high frequency. Financial data are not subject to revisions and are available on a daily or weekly basis. Existing forecasting models for the monthly real price of oil do not take advantage of these rich data sets. Our objective is to assess whether there is useful predictive information for the real price of oil in high-frequency data from financial and energy markets, and to identify which predictors are most useful. The incorporation of daily or weekly data into monthly oil price forecasts requires the use of models for mixed-frequency data.

The development of models for variables sampled at different frequencies has attracted a substantial amount of interest in recent years. A comprehensive review is provided by Foroni, Ghysels, and Marcellino (2013). A large and growing body of literature has documented the

* Correspondence to: Department of Economics, University of Michigan, 309 Lorch Hall, 611 Tappan Street, Ann Arbor, MI 48104, United States.

E-mail address: lkilian@umich.edu (L. Kilian).

[1] A comprehensive review of this body of literature is provided in the handbook chapter by Alquist, Kilian, and Vigfusson (2013). More recent contributions not covered in that review include the studies by Baumeister and Kilian (2014a, 2014b), Baumeister, Kilian, and Zhou (2013), Bernard, Khalaf, Kichian, and Yelou (2013), and Chen (2014).

http://dx.doi.org/10.1016/j.ijforecast.2014.06.005
0169-2070/

benefits of combining data of different frequencies for forecasting macroeconomic variables such as real GDP growth and inflation. One approach has been to construct mixed-frequency vector autoregressive (MF-VAR) forecasting models (e.g., Schorfheide & Song, 2014). An alternative approach involves the use of univariate mixed-data sampling (MIDAS) models (e.g., Andreou, Ghysels, & Kourtellos, 2011). The MIDAS model employs distributed lag polynomials to ensure a parsimonious model specification, while allowing for the use of data sampled at different frequencies. The original MIDAS model requires nonlinear least squares estimation (see Andreou, Ghysels, & Kourtellos, 2010). Foroni, Marcellino, and Schumacher (in press) propose a simplified version of the MIDAS model (referred to as unrestricted MIDAS or U-MIDAS) that may be estimated by ordinary least squares, and it has been shown to produce highly accurate out-of-sample forecasts in many applications, provided that the data frequencies to be combined are not too different.

Numerous studies have documented the ability of MIDAS regressions to improve the accuracy of quarterly macroeconomic forecasts based on monthly predictors, and the accuracy of monthly forecasts based on daily or weekly predictors (e.g., Andreou, Ghysels, & Kourtellos, 2013; Armesto, Engemann, & Owyang, 2010; Clements & Galvao, 2008, 2009; Ghysels & Wright, 2009; Hamilton, 2008). In practice, the use of high-frequency financial data is of particular interest, because financial asset prices embody forward-looking information. Another reason for this interest is that financial data are measured accurately and are available in real time, while lower-frequency macroeconomic data tend to be subject to revisions and are available only with a delay.

These differences in informational structure are particularly evident when forecasting oil prices. Commonly used predictors of the real price of oil, such as global oil production, global oil inventories, global real activity, or the US refiners' acquisition cost for crude oil, only become available with considerable delays and are subject to potentially large, but unpredictable, revisions that may persist for up to two years (see Baumeister & Kilian, 2012). Despite these drawbacks, several recent studies have shown that it is possible to systematically beat the no-change forecast of the monthly real price of oil in real time (e.g., Baumeister & Kilian, 2012, 2014a, 2014b).

The current paper investigates whether the accuracy of oil price forecasts can be improved by utilizing high-frequency information from financial markets and from US energy markets. The set of high-frequency predictors includes (1) the spread between the spot prices of gasoline and crude oil; (2) the spread between the oil futures price and the spot price of crude oil; cumulative percentage changes in (3) the Commodity Research Bureau (CRB) index of the price of industrial raw materials, (4) US crude oil inventories, and (5) the Baltic Dry Index (BDI); (6) returns and excess returns on oil company stocks; (7) cumulative changes in US nominal interest rates (LIBOR, Fed funds rate); and (8) cumulative percentage changes in the US trade-weighted nominal exchange rate.

Our starting point is a MIDAS model for the monthly real price of oil. For reasons discussed in Section 2, we focus initially on predictors measured at weekly intervals and constructed from daily observations. As is standard in the oil price forecasting literature, we assess all forecasts based on their mean-squared prediction errors and directional accuracy. We consider forecast horizons, h, ranging from 1 month to 24 months. Our MIDAS models nest the no-change forecast of the real price of oil, allowing us to compare the accuracy of MIDAS regressions with those of competing models evaluated against the same benchmark. We also compare the MIDAS model forecasts to real-time forecasts from the corresponding model based on the same predictors measured at monthly frequency.

Our results reinforce and strengthen recent evidence that the monthly real price of oil can be forecast in real time. We find that the most accurate h-month-ahead forecasts are obtained based on the percentage change in US crude oil inventories over the preceding h months. For example, the preferred MIDAS forecast has a statistically significant directional accuracy as high as 56% at the 12-month horizon, and as high as 69% at the 24-month horizon. It also produces mean-squared prediction error (MSPE) reductions relative to the no-change forecast of 8% at the 12-month horizon and of 28% at the 24-month horizon. These improvements in forecast accuracy are large by the standard of previous work on forecasting oil prices. However, at horizons shorter than 12 months, the MSPE reductions of this MIDAS model are quite modest or nonexistent.

The way in which the MIDAS model is implemented matters to some extent. While there is typically little difference in accuracy between the MIDAS model with equal weights and the MIDAS model with estimated weights, the unrestricted MIDAS model tends to be slightly less accurate than the other specifications. The success of these MIDAS forecasts based on US crude oil inventories prompted us to also investigate the accuracy of the MF-VAR model obtained by including the same weekly inventory data in a monthly oil market VAR forecasting model of the type examined by Baumeister and Kilian (2012). We found that the latter specification did not perform systematically better than the original VAR model, and was clearly worse than the MIDAS model. The MIDAS model for US crude oil inventories does not have systematically lower MSPEs than the corresponding forecasting model based on monthly US inventory data, however, and has comparable directional accuracy.

While the improvements in forecast accuracy are less substantial for other weekly financial predictors, the pattern of results is similar. Although MIDAS models often significantly outperform the no-change forecast, the corresponding forecasts from models based on monthly financial predictors do too, and there is little to choose between these models. Examples include models based on oil futures prices, returns on oil company stocks and gasoline price spreads. In some cases, the MIDAS model forecasts are actually inferior to the forecasts from the corresponding monthly model, or fail to improve on the no-change forecast.

These conclusions are robust to whether the MIDAS models are estimated based on daily or weekly data. Even when MIDAS models work well, therefore, not much is lost