Physica A 391 (2012) 5546–5556

Contents lists available at SciVerse ScienceDirect

Physica A

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

Forecasting volatility of fuel oil futures in China: GARCH-type, SV or realized volatility models?

Yu Wei

School of Economics and Management, Southwest Jiaotong University, The First Section of the Northern Second Ring Road, Chengdu 610031, China

ARTICLE INFO

Article history:
Received 24 May 2011
Received in revised form 23 July 2011
Available online 15 June 2012

Keywords:
Oil futures
Volatility forecasting
Realized volatility

ABSTRACT

In most previous works on forecasting oil market volatility, squared daily returns were taken as the proxy of unobserved actual volatility. However, as demonstrated by Andersen and Bollerslev (1998) [22], this proxy with too high measurement noise could be perfectly outperformed by a so-called realized volatility (RV) measure calculated by the cumulative sum of squared intraday returns. With this motivation, we further extend earlier works by employing intraday high-frequency data to compare the performance of three typical volatility models in the daily out-of-sample volatility forecasting of fuel oil futures on the Shanghai Futures Exchange (SHFE): the GARCH-type, stochastic volatility (SV) and realized volatility models. By taking RV as the proxy of actual daily volatility and then computing forecasting errors, we find that the realized volatility model based on intraday high-frequency data produces significantly more accurate volatility forecasts than the GARCH-type and SV models based on daily returns. Furthermore, the SV model outperforms many linear and nonlinear GARCH-type models that capture long-memory volatility and/or the asymmetric leverage effect in volatility. These results also prove that abundant volatility information is available in intraday high-frequency data, and can be used to construct more accurate oil volatility forecasting models.

1. Introduction

Modeling and forecasting of volatility in the oil markets is a key issue in such fields as oil derivative product pricing, portfolio allocation and risk measurement. Engle's [1] seminal paper kicked off the development of a large number of so-called historical volatility models in which a time-varying volatility process is extracted from financial return data. Most such models can be regarded as variants of the generalized autoregressive conditional heteroskedasticity (GARCH) models developed by Bollerslev [2] and others [3–9]. An important competitive class of GARCH-type models is associated with the stochastic volatility (SV) model [10–20]. Both the GARCH-type and SV models are commonly employed in analysis of daily, weekly and monthly returns. However, the recent widespread availability of intraday high-frequency prices for financial assets, and the research thereon, has shed new light on the concept of volatility. The availability of intraday data has not only led to the development of improved volatility measurements but has also inspired research into the potential value of these data as an information source for volatility forecasting [21].

The majority of research carried out to date regards daily squared returns as a daily volatility measurement, although Andersen and Bollerslev [22] and Barndorff-Nielsen and Shephard [23] demonstrated that the measurement noise in daily squared returns is too high for observing the true underlying volatility process. They showed that realized volatility (a daily volatility measure), as calculated by the cumulative sum of squared intraday returns, is less subject to measurement error and thus less noisy. In fact, under certain assumptions, realized volatility (RV) is a consistent estimator of the quadratic

E-mail addresses: ywei@home.swjtu.edu.cn, weiyusy@126.com.

0378-4371/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2011.08.071

variation in the underlying diffusion process [22]. In principle, the volatility measures derived from intraday high-frequency data should prove to be more accurate, thus allowing forecast efficiency gains [24–27].

A great deal of research has focused on evaluating the forecasting performance of different volatility models in the oil markets by using the GARCH-class ones [28–46]. For example, Sadorsky [45] employed the closing futures prices for WTI crude oil, heating oil, unleaded gasoline and natural gas to compare the volatility forecasting accuracy of the GARCH, GARCH-in-mean and TGARCH (or GJR) models with that of a number of others under three loss functions. The empirical results by DM test of Diebold and Mariano [47] indicated that the TGARCH model fitted well for heating oil and natural gas volatility, and the GARCH model fitted well for crude oil and unleaded gasoline volatility. Despite their greater complexity, such models as the state space, vector autoregression and bivariate GARCH did not appear to perform as well as the single equation GARCH model. Narayan and Narayan [42] employed the exponential GARCH (EGARCH) model to examine whether shocks had asymmetric and persistent effects on crude oil price volatility. They found inconsistent evidence of such asymmetry and persistence in different sub-samples, although the evidence over their full sample period suggested that shocks had both permanent and asymmetric effects on volatility. Kang et al. [38] evaluated the out-of-sample forecasting accuracy of four GARCH-class models (GARCH, IGARCH, CGARCH, FIGARCH) using the DM test under two loss functions. They found that the CGARCH and fractionally integrated GARCH (FIGARCH) models were able to capture the long-memory volatility of three crude oil markets (WTI, Brent and Dubai) and exhibited superior performance to their GARCH and IGARCH counterparts. In a study complementing those of earlier works [38,42,45], Cheong [32] investigated the out-of-sample forecasting performance of four GARCH-class models (GARCH, APARCH, FIGARCH, FIAPARCH) under three loss functions, finding that the simplest and most parsimonious GARCH model fitted the Brent crude oil data better than the other models examined, although the FIAPARCH out-of-sample WTI forecasts exhibited superior performance. Agnolucci [29] compared the predictive ability of GARCH-type models with the implied volatility (IV) obtained by inverting the Black equation in WTI crude oil futures contracts quoted on the New York Mercantile Exchange (NYMEX). His empirical results indicated that the GARCH-type models seemed to perform better than the implied volatility. Employing the weekly crude oil spot prices in 11 international markets, Mohammadi and Su [40] compared the forecasting accuracy of four GARCH-class models (GARCH, EGARCH, APARCH, FIGARCH) under two loss functions. The DM test showed the APARCH model to exhibit the best performance. In a recent investigation, Nomikos and Pouliasis [43] concluded the superiority of Markov Regime Switching GARCH (MRS-GARCH) model among the GARCH-type ones.

It is clear from the foregoing summary of the extant literature that the empirical evidence is mixed. Which volatility model is superior in modeling and forecasting the volatility of oil markets remains obscure. Furthermore, all of the conclusions in the earlier papers are based on a method that actual volatility is assessed by daily or weekly squared returns. Although Chevallier [33] used intraday tick-by-tick data of carbon prices to measure realized volatility, his emphasis was to detect the outliers and instability in the volatility of carbon prices instead of volatility forecasting. To the best of the authors' knowledge, no previous research evaluating the volatility forecasting performance of different models in the oil markets has adopted the realized volatility measurement as a proxy for actual daily volatility. Hence, the conclusions reached in that research may be unreliable and deserving of further investigation.

As discussed above, there are two aspects which deserve to be further extended in the earlier researches on forecasting oil volatility. For one thing, most papers focused on the GARCH-type models, for another, noisy squared daily/weekly returns were treated as actual volatility to evaluate forecasting accuracy. However, Hansen and Lunde [48] stated that the substitution of a noisy proxy such as squared daily returns for the true but unobservable conditional variance could result in an inferior model being chosen as the best one. The use of RV as a proxy variable, in contrast, did not lead to the favoring of an inferior model.

Motivated by this, the aim of this paper is to evaluate which model commonly used in financial research, GARCH-type, SV or RV, is more suitable for the volatility forecasting of fuel oil futures on the Shanghai Futures Exchange (SHFE). This paper is *different* from the existing research on volatility forecasting of oil markets in the following respects. First, a scaled RV measurement based on intraday high-frequency data proposed by Hansen and Lunde [49] is used as a proxy for actual daily volatility to calculate forecasting error. Second, a larger scope of forecasting models is adopted, i.e., the GARCH-type models commonly employed in earlier studies are compared not only with a SV model based on daily data, but also with a RV model based on intraday high-frequency data. Finally, to obtain robust conclusions, this paper incorporates most of the GARCH-type models proved to perform well in earlier studies, and all the evaluation criteria adopted by earlier literature [29,32,38,40,42,46]. We also employ a technique for model comparison as in Ref. [46], namely, the superior predictive ability (SPA) test by Hansen [50]. Hansen states that the stationary bootstrap procedure employed in the SPA test gives it good power properties and makes it more robust than the approaches used in previous studies, such as the DM test and White's [51] reality check test. Furthermore, at the end of the empirical analysis, we also use squared daily returns as an alternative benchmark of volatility forecasting to check the performance of different volatility models.

The remainder of this paper is organized as follows. Section 2 introduces the sample data and discusses the construction of the daily and intraday returns. Section 3 shows how RV is derived from intraday returns and the autoregressive fractionally integrated moving average (ARFIMA) model, then Section 4 briefly describes the historical volatility models, i.e., the GARCH-type and SV models. The methodology of out-of-sample forecasting and the SPA test are discussed in Section 5, and the empirical forecasting results are presented in Section 6. Section 7 concludes the paper.