International Review of Economics and Finance 49 (2017) 276–291

Contents lists available at ScienceDirect

International Review of Economics and Finance

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

Realized volatility forecast of agricultural futures using the HAR models with bagging and combination approaches

Ke Yang[a], Fengping Tian[b,*], Langnan Chen[c], Steven Li[d]

[a] School of Economics and Commerce, South China University of Technology, China
[b] International School of Business and Finance, Sun Yat-sen University, China
[c] Lingnan (University) College, Sun Yat-sen University, China
[d] Graduate School of Business and Law, Royal Melbourne Institute of Technology University, Australia

ARTICLE INFO

Keywords:
Realized volatility
Forecast
Agricultural commodity futures
Bagging approach
Combination approaches

ABSTRACT

In order to reduce the uncertainty associated with a single predictor model, we incorporate the bagging and combination approaches into a HAR model with the lags of realized volatility and other potential predictors to forecast the realized volatility of agricultural commodity futures in China. We evaluate the performances of the two approaches by employing the mean square forecast error (MSFE) loss function, the modified DM test and the model confidence set (MCS) test at the multiple horizons over the three out-of-sample periods. We find that the realized forecasts from the HAR model with bagging and principal component (PC) combination approaches produce the lowest MSFE at relatively longer forecast horizons. We also find that the simple average of the forecasts from the HAR models with bagging and PC combination methods leads to a further reduction in MSFE, suggesting that they are the effective methods to forecast the realized volatility of agricultural commodity futures in China.

1. Introduction

Forecasting the volatility of the agricultural commodity futures is crucial for agricultural production, resource allocation, and risk management. Traditionally, the GARCH-type models initially proposed by Engle (1982) and Bollerslev (1986) are widely used to forecast the volatility of agricultural commodity futures (Baillie, Han, Myers, & Song, 2007; Chang, McAleer, & Tansuchat, 2012; Coakley, Dollery, & Kellard, 2008; Crato & Ray, 2000; Jin & Frechette, 2004; Hyun-Joung, 2008; Sephton, 2009). However, these G ARCH-type models are based on the daily price returns without taking into account the intraday trading information, and treat the volatility as a latent process, which makes the latent volatility difficult to be forecasted.

Nevertheless, due to the advent of high-frequency data containing more intraday trading information, the realized volatility based on the sum of squared intraday high-frequency returns opens a new era of forecasting the financial market volatility (Andersen & Bollerslev, 1998; Andersen, Bollerslev, Diebold, & Labys, 2003). This nonparametric measure of volatility allows us to treat the volatility as an observed variable, and can be modeled directly, rather than be treated as a latent process in the GARCH-type models. Therefore, it is not surprising that a wide-range of volatility models is proposed to forecast the realized volatility (see, for example, Andersen, Bollerslev, & Meddahi, 2005, 2011; Bollerslev et al., 2016; Corsi, 2009; Corsi, Pirino & Reno, 2010; Patton & Sheppard, 2015). Among the models of the realized volatility, the heterogeneous autoregressive (HAR) realized volatility model proposed by Corsi (2009) is one of the most popular models. This model is a predictive regression that considers lagged daily, weekly and monthly

* Correspondence to: International School of Business and Finance, Sun Yat-sen University, No.135, Xingangxi Road, Haizhu, 510275, Guangzhou, China. Tel.: +86 13430283956.

E-mail addresses: yangkedc@foxmail.com (K. Yang), tfengp@mail.sysu.edu.cn (F. Tian), lnscln@mail.sysu.edu.cn (L. Chen), steven.li@rmit.edu.au (S. Li).

http://dx.doi.org/10.1016/j.iref.2017.01.030
Received 6 October 2016; Received in revised form 23 December 2016; Accepted 23 January 2017
Available online 11 February 2017
1059-0560/

realized volatility as the predictor variables for the future realized volatility. Although the HAR model does not formally belong to the class of long memory models, it can accommodate some 'stylized facts' in the financial volatility such as long memory and multiscaling in a very simple and parsimonious way, and is widely used in the literature on the financial volatility forecast due to its tractable estimation and superior performance.

Recently, a series of meaningful extensions for the HAR model has been proposed by incorporating other exogenous variables into model such as the jump components (Anderson & Vahid, 2007; Corsi et al., 2010; Sévi, 2014), lagged returns (Fernandes, Medeiros, & Scharth, 2014; Scharth & Medeiros, 2009), and days-of-the-week effects (Martens, Van Dijk, & De Pooter, 2009; Yang & Chen, 2014). However, it is important to control the predictor variables for model uncertainty and parameter instability given that there are a plethora of potential predictors and the predictive ability of individual variables changes over time, as discussed in Rossi, Elliott and Timmermann (2012) and Tian, Yang and Chen (2017). According to Inoue and Kilian (2008), the bagging method originally proposed by Breiman (1996) and developed by Buhlmann and Yu (2002), can account for both model uncertainty and parameter instability explicitly by allowing the predictive model to change not only over time but also across bootstrap samples for the same time period and effectively averaging the parameter estimates across bootstrap samples in each time period. However, to the best of our knowledge, the literature on the applications of bagging to the realized volatility forecast is limited, and there is no empirical evidence on the effectiveness of combining the HAR model with bagging approach for the realized volatility forecast in agricultural commodity futures.

In addition, a more sizable strand of the recent studies suggests the usefulness of combination method as an alternative mean to forecast the macroeconomic and financial variables in the presence of wealth as a potential predictor whose predictive power can vary markedly over time. Stock and Watson (2004) suggest that the combination approaches with a weighted average of the individual forecasts are the effective approaches to overcome the substantial model uncertainty and structural instability. Except for simple average, the combination approach such as principal component combination (Stock & Watson, 2004) and cluster combination (Aiolfi & Timmermann, 2006) have a superior forecast performance in forecasting the macroeconomic variables. Becker and Clements (2008) find that the combination of short and long memory models of realized volatility is superior to the individual models considered. Yang, Chen and Tian (2015) investigate the realized volatility forecast of stock indices under the structural breaks, and find that the combination approaches with the time-varying weights across individual forecast models outperform the competing models.

Both bagging and combination methods offer ways of generating more reliable forecasts in the realistic environments where there are a plethora of potential predictors and the predictive ability of individual variables changes over time. It is natural to ask which method outperforms better. However, up to now, there has not been the literature on the construction and comparisons between the bagging and combination approaches for the realized volatility forecast of agricultural commodity futures. This paper intends to fill the gap in the literature by provide a rigorous and detailed analysis on the use of the HAR model with bagging and combination approaches for forecasting the realized volatility in the under-researched Chinese agricultural commodity futures market, the biggest market in the world. We incorporate the bagging and several combination approaches into the extended HAR models and evaluate the performances of these approaches based on the mean square forecast error (MSFE) loss function, the modified Diebold and Mariano (1995) test, and the model confidence set (MCS) test at horizons of 1, 5, 10 days over three out-of-sample periods, i.e. the financial crisis period, the post-crisis period and the long period.

The remainder of this paper is organized as follows. Section 2 presents the data and summary statistics of the realized volatility series. Section 3 constructs the HAR with bagging approach and various combination approaches. Section 4 reports the forecast and evaluation results. Section 5 concludes.

2. Data and realized volatility

2.1. Data

Chinese agricultural commodity futures market is selected for this study for a few reasons. Agriculture is of significantly economic and political importance to China due to its large population. Further, the trading volume in Chinese agricultural commodity futures was around 292 trillion Yuan at the end of 2014, and ten out of the top 20 agricultural contracts by volume were traded in the Chinese exchanges according to the Futures Industry Association (FIA) statistics in the USA. These figures show that Chinese agricultural commodity futures market has become the world's biggest market and has an increasing influence on the global pricing and the decisions on the portfolio diversification (e.g., Cheung & Miu, 2010). There are three commodity futures exchanges in China: Zhenzhou Commodity Exchange (ZCE), Dalian Commodity Exchange (DCE) and Shanghai Futures Exchange (SFE). SFE specializes in metal, energy and chemical-related futures while ZCE and DCE specialize in agricultural commodity futures. We employ the intraday data of the 4 agricultural commodity futures contracts collected from Wind Financial Terminal in Chinese markets, including soybean, cotton, gluten wheat, corn. Soybean and corn are traded in DCE while cotton and gluten wheat are traded in ZCE. The selected agricultural commodities are the major agricultural products in the world.

We eliminate transactions executed during the weekends, public holidays and the days with low trading activity that could lead to estimation bias. After applying this elimination rule, the sample period extends from 1 August 2003 to 30 June 2014 for Soybean futures and Gluten wheat futures, from 28 May 2004 to 30 June 2014 for Cotton futures, and from 22 September 2004 to 30 June 2014 for Corn futures. The total number of trading days is 2636 for soybean futures, 2439 for cotton futures, 2638 for gluten wheat futures, and 2370 for gluten wheat futures.

We follow the algorithm detailed in Barndorff-Nielsen, Hansen, Lunde, and Shephard (2009) to clean the high frequency data, and