International Review of Financial Analysis 34 (2014) 166–176

Contents lists available at ScienceDirect

International Review of Financial Analysis

Modeling and forecasting the additive bias corrected extreme value volatility estimator

Dilip Kumar [a,*], S. Maheswaran [b,1]

[a] Institute for Financial Management Research, 24, Kothari Road, Nungambakkam, Chennai 600034, India
[b] Centre for Advanced Financial Studies, Institute for Financial Management and Research, 24, Kothari Road, Nungambakkam, Chennai 600034, India

ARTICLE INFO

Article history:
Received 7 January 2014
Received in revised form 28 May 2014
Accepted 17 June 2014
Available online 25 June 2014

JEL classification:
C32
C52
C53
G10

Keywords:
Volatility modeling
Volatility forecasting
Forecast evaluation
Bias corrected extreme value estimator

ABSTRACT

In this paper, we provide a framework to model and forecast daily volatility based on the newly proposed additive bias corrected extreme value volatility estimator (the Add RS estimator). The theoretical framework of the additive bias corrected extreme value volatility estimator is based on the closed form solution for the joint probability of the running maximum and the terminal value of the random walk. Using the opening, high, low and closing prices of S&P 500, CAC 40, IBOVESPA and S&P CNX Nifty indices, we find that the logarithm of the Add RS estimator is approximately Gaussian and that a simple linear Gaussian long memory model can be applied to forecast the logarithm of the Add RS estimator. The forecast evaluation analysis indicates that the conditional Add RS estimator provides better forecasts of realized volatility than alternative range-based and return-based models.

1. Introduction

Precise volatility estimation and forecasting play a critical role in financial markets because of their importance for portfolio management, derivatives pricing and risk management. It is well known in the finance literature that volatility is time varying, exhibits persistent characteristics and is in general latent, that is, not directly observable. Interest in this has led to the development of various models to describe the evolution of volatility over time. The most popular of them include various extensions of the generalized autoregressive conditional heteroskedasticity (GARCH) models and the stochastic volatility (SV) models. The popularity of the GARCH class of models has its roots in capturing many stylized facts such as volatility clustering, in its ability to account for dynamic changes in conditional volatility over various horizons and also in providing good in-sample estimates. However, Pagan and Schwert (1990), Figlewski (1997), Andersen and Bollerslev (1998) and Andersen, Bollerslev, and Lange (1999) find that the GARCH model forecasts based on squared return as the ex post volatility measure can be highly unsatisfactory and controversial. In addition, the return-based volatility estimators suffer from inefficiency when compared with extreme value volatility estimators.

Another extension in the literature of forecasting volatility deals with forecasting the realized volatility as a measure. Realized volatility can be estimated using intraday high frequency data and its forecasts can be generated using a linear Gaussian model, such as the long memory ARFIMA model (Andersen, Bollerslev, Diebold, & Labys, 2003; Pong, Shackleton, Taylor, & Xu, 2004). However, intraday high frequency data are plagued by non-negligible market microstructure issues which make the estimation of volatility highly complex. Moreover, high frequency data are usually expensive and may not be available for tradable assets for various emerging markets or may be available only for a shorter duration. In addition, working with high frequency data requires substantial computational resources. Furthermore, as argued by Rogers and Zhou (2008), from the perspective of a practicing quant who is interested in pricing an option or in hedging it, the amount of effort and resources required to implement realized volatility estimators may appear excessive.

Volatility estimators based on the high and the low, also known as extreme value volatility estimators have been acknowledged as being highly efficient volatility estimators in the finance literature. In addition, the daily opening, high, low and closing prices of most of the tradable assets are easily available. The different variants of the extreme value volatility estimators can be categorized as: method of moments estimators (see Garman & Klass, 1980; Kunitomo, 1992; Parkinson, 1980; and

* Corresponding author. Tel.: +91 44 2830 3542; fax: +91 44 2827 9208.
E-mail addresses: dksic212@gmail.com, dilip.kumar@ifmr.ac.in (D. Kumar), mahesh@ifmr.ac.in (S. Maheswaran).
[1] Tel.: +91 44 2830 3443; fax: +91 44 2827 9208.

http://dx.doi.org/10.1016/j.irfa.2014.06.002
1057-5219/

Rogers & Satchell, 1991) and maximum likelihood (ML) estimators (see Ball & Torous, 1984; Horst, Rodriguez, Gzyl, & Molina, 2012; and Magdon-Ismail & Atiya, 2003). The ML estimators are efficient under ideal conditions; however, from a practical viewpoint, they suffer from a serious disadvantage. As highlighted by Maheswaran and Kumar (2013), due to the complexity of the joint density, the ML estimator cannot be expressed in closed-form, and consequently it is difficult to correct for the bias in the estimator arising from a potential misspecification of the data generating process. Furthermore, since the ML estimator does not have a time-separable form, something that is usually available in method of moments estimators, it is difficult to assess the sensitivity of the ML estimator to outlier observations. On the other hand, among the method of moments estimators, the one proposed by Rogers and Satchell (1991), hereafter referred as the RS estimator, stands out because it is the only one that is unbiased regardless of the drift parameter whereas all others are biased in one way or another if the mean return (drift) is non-zero. It needs to be noted here that the RS estimator is unbiased under the assumption that the intraday price process follows a Brownian motion. All these extreme value volatility estimators are unconditional and lack the ability of capturing the dynamics of financial markets and are of less interest to practitioners and quant specialists.

With the work of Alizadeh, Brandt, and Diebold (2002), the practical usefulness of the simple extreme value volatility estimator, the log of the trading range, was formally established. They propose the use of range-based volatility measures in the estimation of stochastic volatility models. They show theoretically, numerically, and empirically that range-based volatility measures are highly efficient and approximately Gaussian and also that they are robust to microstructure noise. Chou (2005) proposes the Conditional Autoregressive Range Model (CARR) to capture the dynamics of range-based volatility. In particular, he finds that the CARR model can forecast return-based volatility more effectively than can the GARCH model. Brandt and Jones (2006) propose yet another model to capture the dynamics in range-based volatility by combining Exponential Generalized Autoregressive Conditional Heteroskedasticity (EGARCH) models with data on the range and find that the new models effectively forecast return-based volatility. They find that the range-based conditional volatility models can better forecast volatility over longer horizons up until 1 year in comparison to similar forecasts made by GARCH models. Li and Hong (2011) propose the range-based autoregressive volatility model and their findings are also in line with that of Chou (2005) and Brandt and Jones (2006), in that range-based conditional volatility models exhibit good performance in forecasting future volatility. All these conditional extreme value volatility estimator models are based on the range which can be noisy and biased.

Maheswaran and Kumar (2013) and Kumar and Maheswaran (2014) find that the original RS estimator is severely downward biased when implemented in the data because of the random walk effect. Maheswaran and Kumar (2013) propose an automatic bias correction (ABC) procedure which approximately corrects for the downward bias in the RS estimator that is observed in the data. Inspired by the works of Kou and Wang (2003, 2004), Kumar and Maheswaran (2014) propose an additive bias correction for the original RS estimator, called herein the Add RS estimator, that exactly corrects for the downward bias in the RS estimator. However, the Add RS estimator is an unconditional estimate of volatility.

In this paper, we address the issue of modeling and forecasting volatility based on the Add RS estimator. First, we explore the statistical properties of the logarithm of the Add RS (Log(Add RS)) estimator and find that the distribution of Log(Add RS) is approximately Gaussian and, hence, a linear Gaussian model can be applied to model the logarithm of the Add RS estimator. The inspiration to model the conditional Add RS estimator has its roots in the works of Andersen, Bollerslev, Diebold, and Labys (2001); Andersen et al. (2003) . We apply ARFIMA(p,d,q) by choosing appropriate orders, based on the Schwarz information criterion (SIC), to model the conditional Log(Add RS) estimator. This model hereafter is referred as the ARFIMA-Add RS model. To explore the superiority of the ARFIMA-Add RS model, we evaluate the forecasting performance of the ARFIMA-Add RS model based on the error statistics approach, the regression approach and the superior predictive ability (SPA) approach and compare the corresponding results with the alternative models that include the range-based conditional autoregressive range (CARR) model and several return-based models (Generalized autoregressive conditional heteroskedasticity (GARCH) model, Fractionally integrated generalized autoregressive conditional heteroskedasticity (FIGARCH) model, Exponential generalized autoregressive conditional heteroskedasticity (EGARCH) model, Fractionally integrated exponential generalized autoregressive conditional heteroskedasticity (FIEGARCH) model and the RiskMetrics model). Our findings indicate that the ARFIMA-Add RS model performs much better than the alternative models in forecasting realized volatility.

The remainder of this paper is organized as follows: Section 2 describes the data, methodology and the preliminary analysis. Section 3 reports our empirical findings and Section 4 concludes with a summary of our main findings.

2. Data, methodology and preliminary analysis

2.1. Data

To explore the forecasting performance of the conditional Add RS estimator, we use the daily opening, high, low and closing prices of four global stock indices: Standard & Poor 500 (S&P 500), a free-float capitalization-weighted index of prices of 500 large cap stocks actively traded on United States stock exchanges; CAC 40, a capitalization-weighted index of the prices of 40 highest market cap stocks listed on the Paris Bourse (Euronext, Paris); IBOVESPA, an accumulation index of about 50 stocks traded on the São Paulo Stock, Mercantile & Futures Exchange covering 70% of the value of the stocks traded; and S&P CNX Nifty, the broad-based benchmark of the Indian capital market. This covers the major developed markets (S&P 500 and CAC 40) from the United States and Europe and major emerging markets (IBOVESPA and S&P CNX Nifty) from South America and Asia. The sample period for all the indices is from January 1996 to June 2013. All the data have been collected from the Bloomberg database. In the following tables, we use Nifty to represent S&P CNX Nifty index.

2.2. Constructing the Add RS estimator

Suppose O_t, H_t, L_t and C_t are the opening, high, low and closing prices of an asset on day t. Define:

$$b_t = \log\left(\frac{H_t}{O_t}\right)$$

$$c_t = \log\left(\frac{L_t}{O_t}\right)$$

$$x_t = \log\left(\frac{C_t}{O_t}\right)$$

Let $u_t = 2b_t - x_t$ and $v_t = 2c_t - x_t$. Hence, the bias corrected extreme value estimators are given by:

$$Add\ \ ux = \frac{1}{2}\left(u_t^2 - x_t^2\right) + x_t^2 . 1_{\{b_t=0 \ \ or \ \ x_t=b_t\}} \tag{1}$$