*Indian Journal of Agricultural Sciences* **83** (12): 1368–71, December 2013/Article

# Forecasting volatiletime-series data through Stochasticvolatility model

BISHAL GURUNG[1], PRAJNESHU[2] and HIMADRI GHOSH[3]

*Indian Agricultural Statistics Research Institute, Library Avenue, New Delhi 110 012*

Received: 14 June 2012; Revised accepted : 15 October 2013

ABSTRACT

Forecasting of volatile data is generally carried out using Generalized autoregressive conditional heteroscedastic (GARCH) model.However, there are some limitations of this methodology, such as its inability to capture empirical properties observed in time-series data. Further, the GARCH assumption that volatility is driven by past observable variables only sometimes becomes a constraint. Accordingly, in this paper, a promising methodology of Stochasticvolatility (SV) model, in which the time-varying variance is not restricted to follow a deterministic process, is considered.The estimation of parameters of this model is carried out using a powerful technique of Kalman filter (KF) in conjunction with Quasi-maximum likelihood (QML) method.As an illustration, volatile dataset of Month-wise total exports of fruits and vegetables seeds from India during the period April 2004 to January 2012 are considered. It is concluded that SV model performs quite well for modelling as well as forecasting of the volatile data under consideration.

**Key words**: Fruits and vegetables seeds export, Heteroscedasticity, Kalman filter, Stochastic volatility model

In agriculture, data are usually collected over time. In the early stages of time-series analysis, main interest was to find a model which could explain effectively the mean behaviour of data (Box *et al*. 2008). Subsequently, concerns about volatility or variance in the data have been raised because changes or patterns in volatility are observed in real data. As emphasized by Jaffee (2005), volatility seems to be the norm rather than exception in international markets for agricultural commodities due to structure of trade, climatic conditions, and rapidity with which producers can respond to price changes. The exports of many agricultural commodities show a great degree of fluctuations, caused by delays between production decisions and delivery to the market. Deo *et al*. (2008) empirically examined the implied volatility function for selected individual equity call options from Indian Stock Market. The authors also evaluated the implied volatilities of in-the-money option which were higher than implied volatility of out-of-the-money option. Bauwens *et al*. (2012) have given an excellent description of various aspects of volatility models and their applications.

A huge amount of fruits and vegetables seeds are exported from India worldwide. The capital gained from export of seeds fluctuates due to various reasons, like variable market prices, production, time lags between production decisions and delivery to the market, and area under seed production. Moreover, in India, exports are allowed only after domestic requirements are met, which is also a cause of severe fluctuations leading to volatility in exports. Efficient forecasting of such volatile datasets is useful not only to the exporters, but to the policy makers also. For some periods of time, data may experience small variances while for other periods of time, these may show large variances. Hence, it is not reasonable to assume that data have a constant variance as in the classical time-series analysis. Furthermore, if it is possible to predict future variance, it would be very useful to control the risk.

To handle this changing conditional variance, Engle (1982) in a path-breaking research work, put forward the Autoregressive conditional heteroscedastic (ARCH) families of parametric nonlinear time-series models. However, conditional variance of ARCH model has the property that unconditional autocorrelation function (acf) of squared residuals, if it exists, decays very rapidly, unless the maximum lag is large. To overcome this limitation, Bollerslev (1986) proposed the Generalized ARCH (GARCH) model, in which unconditional acf of squared residuals has slow decay rate, giving a more parsimonious model of the conditional variance. However, GARCH model cannot capture in an appropriate way the main empirical properties, like high skewness and kurtosis, which are often observed involatile time-series data. Moreover, the GARCH assumption that volatility is driven by past observable variables only can become a constraint.

[1]Scientist (e mail: vsalrayan@gmail.com), [2]Principal Scientist and Head (e mail: prajneshu@yahoo.co.in), [3]Senior Scientist (e mail: hghosh@gmail.com), Division of Statistical Genetics

Consequently, Stochastic volatility (SV) parametric nonlinear time-series model, to describe time-varying volatility, was proposed by Taylor (1994). Here, variance is an unobserved component following a particular stochastic process, and not restricted to follow a deterministic process. This way, SV model is more attractive and closer to the empirical properties observed in realtime-series data. Although its estimation is more complicated because volatility cannot be observed one-step ahead, it gives parsimonious models. Jordi and Josep (2012) applied a maximum likelihood method to study the performance in terms of log-price probability, volatility probability, and its mean first-passage time.

In this paper, SV model and procedure for estimation of its parameters using Kalman filter (KF) is discussed. Finally, the methodology is illustrated using All-India monthly fruits and vegetables seeds export time-series data.

## MATERIALS AND METHODS

The SV model, proposed by Taylor (1994), is given by

$$y_t = \sigma_* \varepsilon_t \sigma_t, \qquad t = 1, 2, \ldots, T, \qquad \ldots(1)$$

where $y_t = \beta x_{t-1}$, $x_t$ being the observations, $\beta$ is AR(1) coefficient, $\sigma_*$ is scale parameter and $\varepsilon_t$ is a white noise process with unit variance. The corresponding volatility $\sigma_t^2$ is defined as $h_t = \log(\sigma_t^2)$, which follows an AR(1) process given by

$$y_{t+1} = \varphi h_t + \eta_t,\ \eta_t \sim \text{IID}(0, \sigma_\eta^2),\ |\varphi| < 1 \qquad \ldots(2)$$

Where $\sigma_\eta^2$ is variance of the log-volatility process. Eq (1) can be written in linear state space form as

$$\log(y_t^2) = \log(\sigma_*^2) + h_t + \log(\varepsilon_t^2) \qquad \ldots(3)$$

Estimation of parameters of SV model can be carried out through the promising and powerful technique of Kalman filtering (KF). The KF is a set of mathematical equations that provides an efficient and powerful computational (recursive) means to estimate the state of a process in a way that minimizes the Mean squared error. It can be applied to any time-series model which can be written in "state space" form. A State space model includes two classes of variables, the state variable and the observation variable, which are modelled as:

$$\alpha_{t+1} = F_t \alpha_t + G_t \varepsilon_t \qquad \ldots(4)$$

$$y_t = H'_t \alpha_t + \nu_t \qquad \ldots(5)$$

Kalman filter is applied recursively through time to construct forecasts and forecast variances. Each step of the process allows the next observation to be forecast based on previous observation and forecast of previous observation (Durbin and Koopman 2001).

The parameters of SV model can be estimated using KF technique in conjunction with Quasi-maximum likelihood (QML) principle. The QML estimator is based on linearizing SV model by taking logarithms of squares of the observations and maximizing Gaussian log-likelihood function even if assumption of normality is violated (Harvey *et al*. 1994). Andersen *et al*. (2001) also showed that the log-volatility process can be well approximated by normal distribution. A good description of various aspects of Kalman filter is provided by McMillan *et al*. (2013).

## RESULTS AND DISCUSSION

As an illustration, Month-wise total exports of fruits and vegetables seeds from India during the period April 2004 to January 2012 available at the website of Indiastat (*www. indiastat. com*) are considered. Out of total 94 data points, first 84 data points corresponding to the period April 2004 to March 2011 are used for model building and remaining 10 data points, i.e. from April 2011 to January 2012 are used for validation purpose. A perusal of data indicates presence of volatility at several time-epochs.

In the first instance, ARIMA model is fitted to the data using EViews, Ver. 4 software package. On the basis of minimum Akaike information criterion (AIC) and Bayesian information criterion (BIC) values, ARIMA (2, 1, 0) model is selected and the results are reported in Table 1. However, autocorrelation functions (acfs) of squared residuals of fitted ARIMA (2, 1, 0) model reveal that the acfat lag 12 is 0. 29, which is quite high. Consequently, ARCH-LM test is applied to the squared residuals to test for autocorrelation. The test statistic is calculated as 2.82, which is found to be significant at 5% level, indicating presence of volatility in the data.

The SV model given in Eqs (1) and (2) is first represented in state space form. Then, the KF is applied to obtain the volatility as state at each time epoch. The parameters are now estimated by maximizing the "Prediction error" form of the likelihood, which was derived using Eqs (4) and (5) of KF. Subsequently, SV model is fitted to the data to capture the volatility using MATLAB, Ver. 7. 2 software package, yielding

$$\log(y_t^2) = \underset{(0.35)}{4.01} + h_t + \log(\varepsilon_t^2),\ h_{t+1} = \underset{(0.05)}{0.97} h_t + \eta_t,\ \beta = 0.30, \text{ and } \sigma_\eta^2 = 0.01,$$

where the entries within brackets () indicate corresponding standard errors. Since estimate of $\varphi$, viz. 0. 97 is near to unity and estimate of $\sigma_\eta^2$, viz. 0. 01 is near to zero, therefore there is presence of high persistence of shocks to volatility.

Further, to study appropriateness of the fitted SV model, acfs of standardized residuals and squared standardized residuals are computed. It is found that, in both situations, acfs are not significant at 5% level, thereby confirming that mean and variance equations are correctly specified. The

Table 1 Estimates of parameters along with their standard errors

| Parameter | Estimate | Standard error |
|---|---|---|
| Intercept | 0. 24 | 0.09 |
| AR1 | –0.41 | 0.11 |
| AR2 | –0.24 | 0.11 |

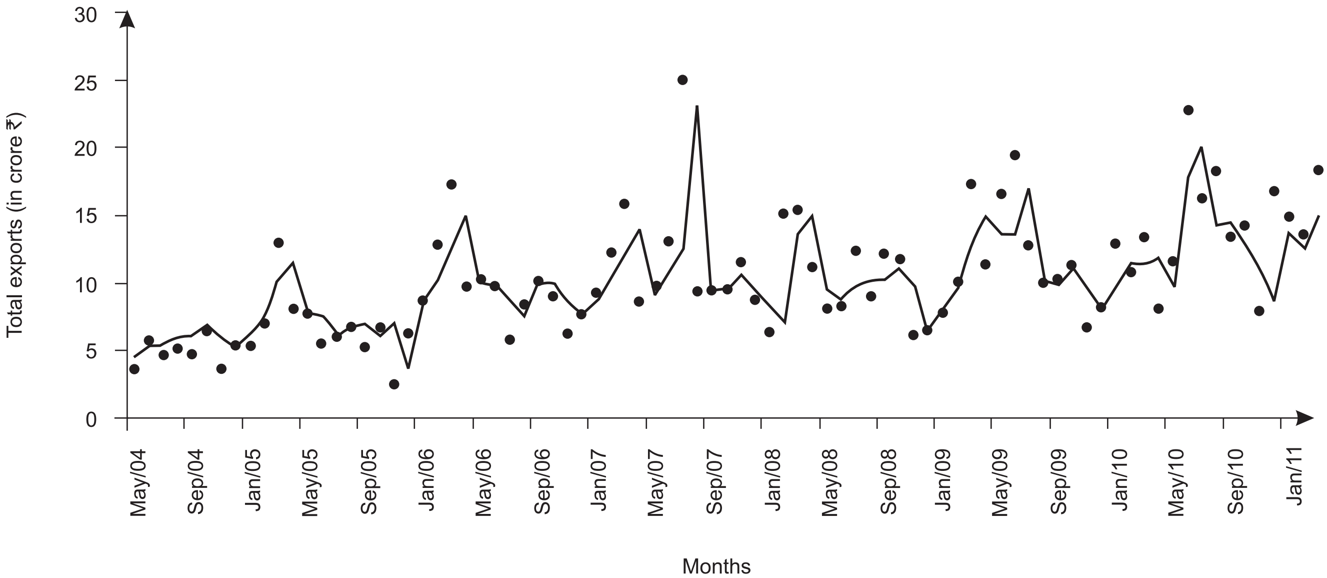


Fig. 1 Fitted Stochastic volatility model along with data.

graph of fitted SV model along with data is exhibited in Fig 1, which indicates that fitted SV model is able to capture volatilities present in the data satisfactorily.

*Forecasting performance*

Forecasting performance for 10 months corresponding to All-India data of monthly export of fruits and vegetables seeds from April 2011 to January 2012 as hold-out-data is studied. One-step ahead forecasts are computed and the same are reported in Table 2. The entries within brackets ( ) indicate corresponding standard errors. It may be pointed out that the realistic feature of varying standard errors for forecasts cannot be captured through ARIMA methodology in view of the assumption of homoscedasticity of error terms. A perusal of Table 2 shows that forecast values obtained by SV model are much closer to actual values than those obtained by ARIMA model. Further, for the months of June, August, and November, actual values lie outside the one standard error of the forecast values obtained through ARIMA methodology. In contrast to this, all actual values lie within one standard error of corresponding forecast values obtained by SV methodology. Performance of one-step ahead forecasts is assessed on the basis of Mean absolute prediction error (MAPE), Mean square prediction error (MSPE), and Relative mean absolute prediction error (RMAPE) criteria. These values for fitted ARIMA and SV models are computed and reported in Table 3. Evidently, lower values of all the criteria throughout reflect the superiority of SV model over ARIMA model.

Table 2 One-step ahead forecasts (in Crore ₹)

| Months | Actual | Forecast using SV | Forecast using ARIMA |
|---|---|---|---|
| April 2011 | 18.31 | 16.94 (4.24) | 16.64(4.20) |
| May 2011 | 20.38 | 18.84 (4.11) | 18.54(4.20) |
| June 2011 | 23.19 | 21.45 (3.98) | 18.32(4.20) |
| July 2011 | 25.71 | 23.77 (4.06) | 21. 77(4.20) |
| August 2011 | 18.55 | 19.70 (4.84) | 24.23(4.20) |
| September 2011 | 18.90 | 19.36 (3.97) | 21.11(4.20) |
| October 2011 | 23.81 | 22.01 (4.03) | 20.70(4.20) |
| November 2011 | 12.15 | 13.78 (4.27) | 21.94(4.20) |
| December 2011 | 20.41 | 18.37 (4.94) | 15.98(4.20) |
| January 2012 | 24.01 | 22.20 (4.79) | 20.05(4.20) |

Table 3 Performance of one-step ahead forecasts

| Criterion | SV Model | ARIMA Model |
|---|---|---|
| MAPE | 1. 55 | 4. 15 |
| MSPE | 2. 58 | 22. 35 |
| RMAPE | 7. 71 | 22. 87 |

To sum up, for the data under consideration, SV model is found to be reasonably good for modelling as well as forecasting purposes. The volatilities in the time-series data for fruits and vegetables seeds may be due to fluctuating patterns in the world economy and may also be attributed to the fact that there are fluctuations in the demand from many importing countries. It is hoped that agricultural scientists would start employing SV methodology rather than the usual ARIMA methodology for modelling and forecasting of their volatile datasets.


## ACKNOWLEDGEMENTS

Authors are grateful to the referee for valuable comments.

REFERENCES

Andersen T G, Bollerslev T, Diebold F X and Labys P. 2001. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association* **96** : 42–55.

Bauwens L, Hafner C M and Laurent S. 2012. *Handbook of Volatility Models and Their Applications*. Wiley, USA.

Bollerslev T. 1986. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* **31** : 307–27.

Box G E P, Jenkins G M and Reinsel G C. 2008. *Time Series Analysis: Forecasting and Control*, edn 4. Prentice Hall, USA.

Deo M, Devanadhen K and Srinivasan K. 2008. An empirical analysis of implied volatility in Indian options market. *International Research Journal of Finance and Economics* **18** : 108–26.

Durbin J and Koopman S J. 2001. *Time Series Analysis by State Space Models*. Oxford University Press, Oxford.

Engle R F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50** : 987–1 007.

Harvey A C, Ruiz E and Shephard N G. 1994. Multivariate stochastic variance models. *Review of Economic Studies* **61** : 247–64.

Jaffee S. 2005. *Delivering and Taking the Heat: Indian Spices and Evolving Product and Process Standards*. Agriculture and Rural Development Discussion Paper 19. The World Bank, USA.

Jordi C and Josep P. 2012 Maximum likelihood approach for several stochastic volatility models. *Journal of Statistical Mechanics: Theory and Experiment* **8** : 8–16.

Mc Millan H K, Hreinsson E Ö, Clark M P, Singh S K, Zammit C and Uddstrom M J. 2013. Operational hydrological data assimilation with the recursive ensemble Kalman filter. *Hydrology and Earth System Sciences* **17** : 21–38.

Taylor S J. 1994. Modelling stochastic volatility: A review and comprehensive study. *Mathematical Finance* **4** : 183–204.