Short Communication

Indian Journal of Agricultural Sciences **88** (10): 1644–7, October 2018/Short Communication

Stochastic volatility in mean model for capturing the conditional variance in volatile time series data

RAVINDRA SINGH SHEKHAWAT[1], K N SINGH[2], AJAY KUMAR[3], KRISHNA PADA SARKAR[4], RIPI DONI[5] and BISHAL GURUNG[6]

ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110 012

Received: 01 June 2018; Accepted: 11 July 2018

Key word: Particle filter, Stochastic volatility in mean model, Time-series, Volatility

Researchers possess a well-established methodology based on linear time-series models, called the Box-Jenkins methodology (Box *et al*. 2008) which is used for modelling and forecasting of data collected sequentially in time. It consists of fitting the Autoregressive integrated moving average (ARIMA) models by model selection and estimation of parameters followed by significance tests to check the adequacy of fitted model. Popularity of these ARIMA models is certainly due to their relative simplicity and also due to the existence of various computer softwares incorporating the same. The ARIMA model, however, is insufficient as it is not able to capture many important characteristic features of the datasets and also it requires stationary time-series data. For over three decades, time-series analysis has moved towards the nonlinear domain, which is commonly more appropriate for accurately recounting the dynamics of a time-series and also for making superior multi-step-ahead forecasts (Fan and Yao 2003). In agriculture, data are usually collected sequentially over time. In the early stages of time-series analysis, main interest was to find a model which could explain efficiently the mean behaviour of data (Box *et al*. 2008). Consequently, concerns about volatility or variance in the data have been raised because changes or patterns in volatility are observed in real datasets. The exports and imports of many agricultural commodities show a great degree of sudden 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 exceptional report of various aspects of volatility models and their applications.

Discrete-time model due to Koopman and Uspensky (2002) for regularly spaced data on India's monthly spices export is applied for modelling and forecasting of unobserved volatility process. Spices are the most important commercial crops in India and have great medicinal value. The important spices extensively grown in India are: Cardamom, pepper, chillies, turmeric, and ginger. With respect to export, consumption, and production of spices, India stands at number one in the whole World. The total area under these spices in India is over one million ha, and these accounted for an annual export of about ₹ 11843.97 crores during the year 2011. However, it is observed to be highly volatile with changes in successive months' values ranging from ₹ 148 to 128 crores. Moreover, the LM test statistic is found to be significant as well.

To handle this changing conditional variance, Engle (1982) in a seminal research work, put forward the Autoregressive conditional heteroscedastic (ARCH) family 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 incredibly rapid, except when 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 suitable way the main empirical properties, like high skewness and kurtosis, which are often observed in volatile time-series data. Moreover, the GARCH assumption that volatility is driven by past observable variables only can become a constraint. Consequently, Stochastic volatility (SV) parametric nonlinear time-series model, to describe time-varying volatility, was proposed by Taylor (1994).

[1]Scientist (e mail: ravindra.shekhawat@icar.gov.in), [2]Principal Scientist and Head (e mail: knsingh@iasri.res.in), Division of Forecasting and Agriculture Systems modelling, [3]Research Associate (e mail: ajayyadav62063@gmail.com), NAHEP KAB-II, [4]Research Scholar (e mail: krishnapadasarkar07@gmail.com) [5]Research Scholar (e mail: ripidoni14@gmail.com), ICAR-IARI, New Delhi. [6]Scientist (e mail: vsalrayan@gmail.com). ICAR-IASRI.

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 real time-series data. 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. Another popular version of the SV model is the one where the explanatory variable in the model is the variance process itself. This type of SV model is called the Stochastic Volatility in Mean (SVM) model. The estimation of such an intricate model is not straight forward since volatility now appears in both the mean and the variance equation. Fortunately, to this end, the Particle filter (PF) methodology may be employed. PF is a sophisticated simulation based technique used for estimation of unobservable state.

In this paper, SVM model and procedure for estimation of its parameters using Particle filter (PF) is discussed. The changing conditional variance is also estimated. Subsequentially, the methodology is illustrated using All-India data of monthly export of spices during the period January, 2006 to January, 2012.

Description of the SVM model

The SVM model is more versatile and pragmatic than the GARCH-type models, since it essentially involves two random processes, one for the observation equation, and one for the latent volatility equation. In a way, the model assumes a form of state space representation. The SV-M model proposed by Koopman and Uspensky (2002) can be written as

$$y_t = \delta\sigma^{*2}\sigma_t^2 + \sigma^*\varepsilon_t\sigma_t\,, \quad t = 1, \ldots, T \ldots (1)$$

where y_t denotes the observation, σ_t is the latent volatility and δ is the parameter measuring the volatility-in-mean effect and σ^* is a scale parameter that removes the need for a constant term in the stationary first-order log-volatility autoregressive process.

Following the convention usually considered in literature we write $h_t \equiv \log\sigma_t^2$.

$$h_{t+1} = \varphi h_t + \eta_t\,,\ \eta_t \sim IID(0,\sigma_\eta^2)\,,\ |\varphi| < 1 \tag{2}$$

where, $|\varphi| < 1$ implies stationarity of h_t. The parameter φ is often considered as a measure of the persistence of shocks of the changing volatility. When φ is close to *1* and σ_η^2 is close to *0*, the evolution of volatility over time is very smooth, but becomes haphazard if not satisfied. The variance of the log-volatility process, σ_η^2 measures the uncertainty about future volatility.

Estimation of parameters

Estimation of parameters of SVM model can be carried out through the promising and powerful technique of Particle filtering (PF) by representing the model in state space form. The PF performs sequential Monte Carlo (SMC) estimation based on point mass representation of probability densities. The main idea of PF method is to represent the probability distribution, which is usually hard to obtain directly, using particles and associated weights, so that the empirical distribution of particles replaces the conditional density function. PF is applied recursively through time to construct forecasts and forecast variances. Each stage of the progression allows the subsequent observation to be forecast based on preceding observation and forecast of preceding observation, a characteristic feature of state space methodology.

Testing for ARCH effects

Let ε_t be the series of residuals. The squared series (ε_t^2) is considered to check for conditional heteroscedasticity, also known as the ARCH effects. The usual Ljung-Box statistic Q(*m*) is applied to the series (ε_t^2), where the null hypothesis is that the first *m* lags of autocorrelation functions of the series (ε_t^2) are zero. The other test for conditional heteroscedasticity is the ARCH-Lagrange multiplier (ARCH-LM) test of Engle (1982). This test is equivalent to usual *F*-statistic for testing H_0: a_i = 0, 1, 2,....*q* in the linear regression

$$\varepsilon_t^2 = a_0 + a_1\varepsilon_{(t-1)}^2 + \varepsilon + a_q\varepsilon_{(t-q)}^2 + e_t, \quad t = q+1, \ldots, T \tag{3}$$

where e_t denotes the error term, *q* is the pre-specified positive integer, and *T* is the sample size.

Denote

$$SSR_0 = \sum\nolimits_{t=q+1}^{T}\left(\varepsilon_t^2 - \varpi\right)^2 \tag{4}$$

where

$$\varpi = \sum_{=q+1}^{T} \varepsilon_t^2 \big/ T$$

is the sample mean of (ε_t^2), and

$$SSR_1 = \sum\nolimits_{t=q+1}^{T}\hat{e}_t^2\ ,$$

where $\hat{e}_t$ is the least square residual.

Then, under H_0,

$$F = \frac{\left(SSR_0 - SSR_1\right)/q}{SSR_1(T-q-1)} \tag{5}$$

is symptotically distributed as chi-squared distribution with *q* degrees of freedom. The decision rule is to reject H_0 if $F>X_q^2(\alpha)$, , where $X_q^2(\alpha)$ is the upper 100 (1– α)th percentile of X_q^2 or, alternatively, the *p*-value of *F* is less than α.

An illustration

As an illustration, All-India data of monthly export of spices during the period January, 2006 to January, 2012 is considered. Data has been collected from indiastat.com website. Out of total 73 data points, first 63 data points corresponding to the period January, 2006 to March, 2011 are used for training purpose while the remaining 10 data points, i.e. from April, 2011 to January, 2012 are used for validation the fitted model. A perusal of Fig 1, 2 indicates presence of volatility at several time-epochs.

A high volatility is noticed in March, 2007 where export suddenly jumped almost 140% to the level of ₹ 402 crores and then an abrupt dip in the very next month to ₹ 301

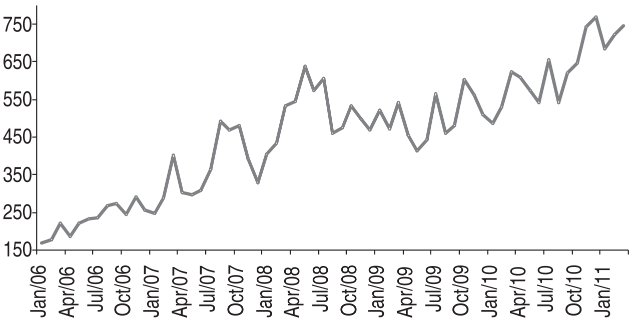

Fig 1 All-India data of monthly export of spices.

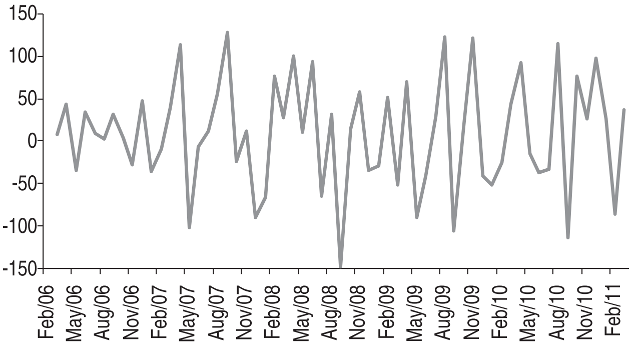

Fig 2 First differenced data of All-India monthly export of spices.

crores. Volatility can also be seen in many other time-points, like August, 2007, March, 2008, October, 2009, March, 2010, and December, 2010. Further, the range of the export value is exorbitantly high which makes it hard for ARIMA to capture features of the data.

As the time-series data clearly indicates presence of trend, first differenced series is shown to be stationary with significant autocorrelation at lag one, (Fig 2). To test nonlinearities in error variance structure estimated from autoregressive modelling, appropriate ARIMA model is chosen on the basis of minimum Akaike information criterion (AIC) and Bayesian information criterion (BIC) values given as

$$AIC = Tlog\ (\sigma^2) + 2(p + 1)$$
$$BIC = Tlog\ (\sigma^2) + 2(p + q + 1)logT$$

where p and q are the orders of autoregressive and moving average coefficients respectively.

On the basis of aforementioned criteria, the ARIMA (1,1,0) model is selected for modelling of the monthly export of spices.

The acf of the squared residuals of the fitted ARIMA(1,1,0) model is found to be reasonably high at lag *6*, which is *-0.22*. Consequently, the ARCH-LM test is carried by first squaring the estimated residuals from fitted ARIMA(1,1,0) model. The ARCH-LM test statistic at lag 6 is computed and is found to be significant at 5% level.

The SVM model is represented in state space form and the model is fitted by making use of PF to obtain the volatility as state at each time epoch. The parameters are estimated by maximizing the "Prediction error" form of the likelihood in conjunction of Particle filter. Subsequently, SVM models fitted to the data to capture the volatility using MATLAB, Ver. 7.2 software package, yielding

$$y_t = 0.69 \times 0.81\sigma_t^2 + 0.9\varepsilon_t\sigma_t$$
$$h_{t+1} = 0.94h_t + \eta_t,\ \ \sigma_\eta^2 = 0.04$$

Since estimate of φ, viz. 0.94 is near to unity and estimate of σ_η^2, viz. 0.04 is near to zero, therefore there is presence of high persistence of shocks to volatility throughout the dataset.

Table 1 Estimates of parameters along with their standard errors for fitted ARIMA model

Parameter	Estimate	Standard error
Intercept	11.56	6.51
AR1	-0.25	0.12

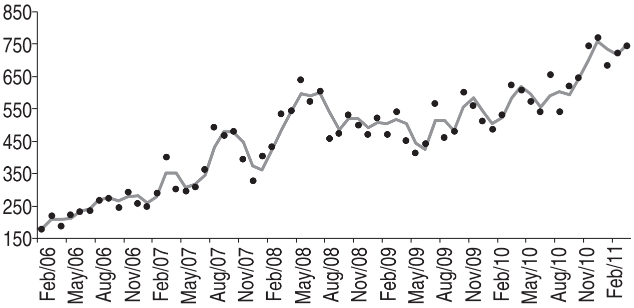

Fig 3 Fitted Stochastic volatility in Mean(SVM) model along with data.

Further, to study appropriateness of the fitted SVM 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 graph of fitted SVM model along with data is exhibited in Fig 3, which indicates that fitted SVM 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 spices during the period 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.A perusal of Table 2 shows that forecast values obtained by SVM model are very close to actual values.

To sum up, for the data under consideration, SVM model is found to be reasonably good for modelling as well as forecasting the mean and the voaltility. The volatilities in the time-series data for spices export 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 spread over the globe. It is hoped that agricultural scientists would start employing SVM methodology rather than the usual ARIMA methodology for modelling and forecasting of their volatile datasets.

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

Months	Actual	Forecast
Apr ‘11	758.45	760.29 (23.31)
May ‘11	890.10	833.64 (72.12)
Jun ‘11	876.86	885.89 (39.23)
Jul ‘11	1007.94	948.95 (67.67)
Aug ‘11	1222.66	1189.81 (56.89)
Sep ‘11	1248.52	1228.49 (37.88)
Oct ‘11	1266.68	1259.811(25.54)
Nov ‘11	1160.27	1225.77 (52.98)
Dec ‘11	1256.98	1229.34 (65.76)
Jan ‘12	1071.73	1177.65 (78.32)

SUMMARY

In the early years of time-series data analysis in both linear as well as nonlinear models, researchers were mainly focussed in modelling the mean equation. But, recently, concerns about conditional variance or volatility in the data have increased interest in the area of time-series modelling. Spices are an indispensable crop of India as it contributes substantially to agriculture in terms of farm income, employment and export earnings, which directly affect the GDP. It has been seen that there are sudden and high fluctuations in the spices export time series data. It is known that knowledge of volatility can be a good piece of information for a decision-making process. To this end, Stochastic Volatility in mean (SVM) model was proposed. In this paper, a methodology for estimation of SVM using Particle filter is carried out. Further, illustration of the developed methodology is also carried out using volatile dataset. Statistical measures are also computed to validate the importance of the developed methodology.

REFERENCES

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

Bollerslev Tim. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**(3): 307–27.

Box G, Jenkins G M and Reinsel G. 2008. *Time Series Analysis: Forecasting and Control*, 4 edn, p 746. John Wiley & Sons, Hoboken, New Jersey.

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.

Engle R F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**(4): 987–1007.

Fan J and Yao Q. 2003. Nonlinear Time Series: Nonparametric and parametric Methods. Springer, New York.

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

Koopman S J and Hol Uspensky E. 2002. The stochastic volatility in mean model: empirical evidence from international stock markets. *Journal of Applied Econometrics* **17**: 667–89.

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