A study on volatility estimation in soybean price using GARCH model

Ashima Talwar & Dr. C.K. Goyal*
IBMR, IPS Academy, Indore-452012
*Email: ashimatalwar@ipsacademy.org

Abstract

This study is aimed at identifying the best GARCH model for risk estimation of soybean price. Here GARCH model parameters are estimated using normal distribution, student-t distribution and Generalized error distribution (GED). For this study, the daily soybean price data of NCDEX from Indore market is taken up. The soybean price data for return series exhibits high kurtosis and volatility clustering. The characteristics of the data captured by all three error distributions improve the efficiency of the predictive model. The results demonstrate that volatility is highly persistent in the price data of the commodity. The result also suggested that the GARCH model with GED distribution satisfy the diagnostic phase and gives optimum accuracy measures for a future price estimation.

Key Words: GARCH, Error distribution, Soybean, Forecasting, Volatility

Introduction

Commodity price movements have an impact on overall macroeconomic indicators of a country. The study on conditional volatility of agriculture market gives widespread anticipation towards the upcoming market environment linked to a selected commodity. Forecasting agriculture commodity price is essential for policymakers and stockholders as well as producers. Due to the existence of volatility in price behavior; risk and uncertainty occur in the market. A price forecasting system reduces the risk associated with price variability and help policymaker, economist and producer (farmer) to make their optimum decisions. Agriculture commodity prices are normally random or stochastic. This leads to substantial uncertainty and risk in the process of price forecasting and modeling. In a market-oriented economy, the better volatility estimation system helps to monitor aspects that have an impact on price oscillation. Therefore, the volatility forecasting system act as a key input to market intelligence for macroeconomic policy planning.

In case economic variables, it is commonly observed kurtosis and lack of symmetry. Econometrician have developed GARCH family of models to capture the asymmetry in the series (Engle 1982, Bollerslev 1986). The aim of this paper is to select the best GARCH model using three error distribution viz. normal distribution, student-t distribution and GED distribution for daily closing price data of soybean.

Literature review

Engle (1982) ARCH model and Bollerslev (1986) GARCH model describes time-varying volatility. These models lack in capturing the standard features of return series i.e time-varying skewness and kurtosis. Hensen (1994) established that the error distribution model can acquire the presence of both time-varying skewness and kurtosis.

Galeano & Tsay (2009) describes that excess kurtosis and volatility clustering of return series is not effectively interned by GARCH models and suggests that change in parameter has significant effect on kurtosis and volatility level. Zhang, (2009) observes that if data does not follow normal distribution but skewness exist in the data, then the GARCH model with conditional error distribution gives improved volatility forecasting. Giller (2005) demonstrated probability distribution for univariate series. Wilhelmsson, (2006) analyzes intraday data having leptokurtic distribution by applying GARCH model and concludes that error distribution measure gives better volatility forecast. Koksal (2009) in their empirical study mention that fat tail in return series is well captured by T-GARCH then other error distributions models. Jondeau & Rockinger, (2003) designate T-GARCH model with varying movement and discover that occurrence of skewness is more frequent related to kurtosis. Hamilton and Susmel, (1994) estimated the accuracy of the ex-post volatility by mean square error and found no improvement in error estimation by using constant variance. Kelkey and Emmanuel (2014) modeled and analyzed the price volatility of agriculture commodity. The return series exhibits leptokurtic distribution, volatility clustering and asymmetric effect in the data. They concluded that the GARCH model with error distribution as the best-fitted model amongst ARCH family of models.

Purpose of Research

The study is based on the following objectives

1. To estimate the volatility of soybean return
2. To develop the appropriate GARCH model for forecasting soybean pricing in Indore market.

Methodology

Volatility Modeling

For modeling the commodity price volatility, various ARCH family of models can be used (Tsay, 2005). Whenever market volatility shows clustering behavior means high volatility period continues for some time before the market returns to its steady level. Then, generalized autoregressive conditional heteroskedastic (GARCH) approach is used to build more accurate and reliable volatility models.

The Generalized autoregressive conditional heteroskedasticity (GARCH) Model

The autoregressive conditional heteroskedasticity (ARCH) model was proposed by Robert Engle (1982) for modeling and forecasting stochastic variance. ARCH model uses variance rather than standard deviation for forecasting. If in a linear time series ARMA model is selected for error variance, the exhibited model is called GARCH model. In 1986, T. Bollerslev developed parsimonious representations of conditional heteroskedasticity described by simultaneous mean and variance equations given as follows:

Mean equation: $r_t = \mu + \varepsilon_t\ ,\ \ \varepsilon_t = \sqrt{h_t}\, z_t$

Variance equation: $h_t = \bar{\omega} + \sum_{i=1}^{p} \beta_i\, h_{t-i} + \sum_{J=1}^{q} \alpha_J \varepsilon_{t-j}^2$

Where p and q are order of GARCH and ARCH term respectively.

The estimated parameter α (ARCH effect) represents symmetric or magnitude of shocks or spillover effect and β (GARCH effect) implies the persistent in conditional volatility regardless of other variables of the market, while $\alpha + \beta$ defines the die out rate of ARCH and GARCH effect. With GARCH (p, q) model, multi-period volatility forecast can be made (Alexander, 2009). To deal with the characteristics of the errors, the parameters of the models are measured by these distributions.

Error Distribution

Normal (Gaussian) distribution:

The normal distribution has zero value of kurtosis and skewness. The density function for normal distribution is given as (Ghalanos, 2013)

$$f(\varepsilon) = \frac{1}{\sigma\sqrt{2\Pi}} e^{\frac{(\varepsilon_t - \mu)^2}{2\sigma^2}}$$

Where μ and σ are denoted as mean and standard deviation of the equation

Student-t distribution:

When return series exhibits flatter tail, it is rational to use student's t distribution Bollerslev (1987)

$$f(\varepsilon) = \frac{\Gamma(v + 1/2)}{\Gamma\left(\frac{v}{2}\right)\sqrt{\pi(v-2)\sigma_t^2}}\left(1 + \frac{\varepsilon_t^2}{(v-2)\sigma_t^2}\right)^{-\frac{(v+1)}{2}}$$

The symbol v express the degree of freedom and Γ denotes the gamma function given as

$$\Gamma(\varepsilon) = \int_0^\infty t^{\varepsilon - 1} e^t \cdot dt$$

For $\rightarrow \infty$, student t distribution value tends to normal distribution.

Generalized error distribution (GED)

GED is a symmetrical distribution (Ghalanos, 2013) that transform the shape of normal density function by changing the value of β. The standard GED density function is given by

$$f(\varepsilon) = \frac{\beta}{2\sigma\Gamma\left(\frac{1}{\beta}\right)} exp\left\{-\left(\frac{|\varepsilon - u|}{\sigma}\right)^{\beta}\right\}$$

For β=2, GED shape parameter reduces to normal distribution.

Result and discussion

The sample contains daily data of closing soybean prices obtained from NCEDX. To make forecast, the full sample is divided into two parts, in sample observations covering the period from 1st 2006 to 31st December 2017 (including 3337 observations excluding public holidays) and out of sample forecast covering the period from 1st j January 2018 to 30th April 2018 (total 80 observations). The time plot of soybean prices from 2nd January 2006 to 31st December 2017 are given in fig 1

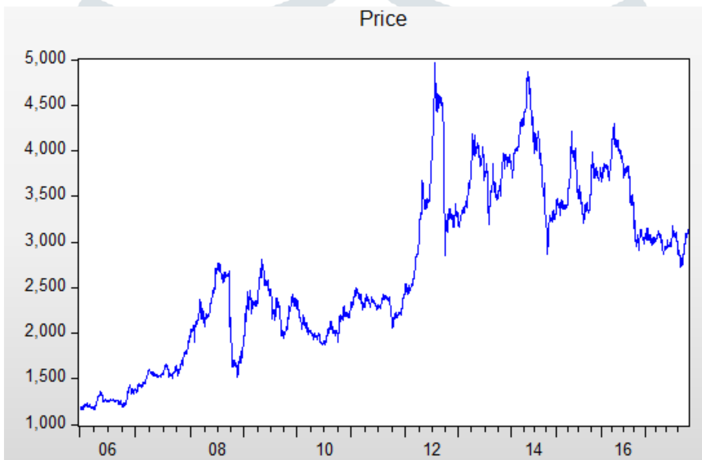

Fig 1: Graphical presentation of daily soyben prices

From the graph, we depict that soybean prices fluctuated around 1161.150 to 4958 Rs./quintal and time series have random walk pattern.

Stationarity test:

A suitable model is fitted on stationary time series i.e. the observed series must have constant mean. For this the hypothesis that the data has unit root (non-stationarity) are checked by parametric (Augmented Dickey-Fuller) and non-parametric (Phillip- Perron) test statistics. The test indicates the existence of unit root in daily soybean price index as shown in table1.

	Price index		Return series	
	ADF test	PP test	ADF test	PP test
t-statistics	-1.827114	-1.982164	-47.40791	-48.66054
p-value	0.3676	0.2949	0.0001	0.0001

Table 1: Results of unit root test

To make price series stationary the series is tranform into log diffrencing denoted as return series .By doing this the series became stationary in mean and variance.The return series is generated to stabilized the series (Fang Xu,2009). The log difference transformation is given as

$$r_t = log\left(\frac{p_{t+1}}{p_t}\right) *100$$

Again, the stationarity of return soybean price series data is checked. ADF and PP test statistics are used to determine the presence of unit root in the series (in table 1). The ADF test statistics for return series of soybean have considerable small t-statistics value i.e-47.40791, p-value is approximate zero specifies that the ADF t-statistics value is significant. Hence null hypothesis is rejected that the first difference for daily soybean price series has a unit root. The pp test for return series of soybean confirm the existence of stationarity. Thus, we make the same inference from parametric and non -parametric test.

Mean	0.029639	Skewness	-3.762246
Median	0.061888	Kurtosis	67.33389
Maximum	4.852460	Jarque-Bera	583344.8
Minimum	-24.17465	JB test p-value	0.0000
Std. dev.	1.197755		

Table 2: Descriptive Statistics

Descriptive statistics for log return series shows (in table 2) that the mean is close to zero with a negative skewness and Kurtosis greater value than 3 indicating price return are leptokurtic and fat-tailed. The soybean price return indicates the low level of unconditional standard deviation. The Jarque -Bera statistics applied for the hypothesis that log return series exhibits normal distribution is rejected at 5% significance level. The index describes random variation and reflect volatility clustering is present.

Mean model specification

Once the return series become stationary, conditional mean model is specified by different AR and MA combination to choose the best ARMA structure. The appropriate model selection is based on the values of Akaike information criteria (AIC) and Schwarz information criteria (SIC). Among various selected models ARMA (1,2) is found as better fitted model in the soybean return series.

After estimate ARMA model for conditional mean, the residuals of the ARMA model is tested for serially correlation or the presence of ARCH effect. ARCH family of models are applied when ARCH effect is present in the residual diagnostic. The presence of remaining ARCH effect in variance is tested by Lagrange Multiplier (LM) test (Engle,1982).

ARCH-LM test

Heteroskedasticity Test: ARCH

F-statistic	10.43658	Prob. F(1,3334)	0.0012
Obs*R-squared	10.41025	Prob. Chi-Square(1)	0.0013

Table 3: *ARCH-LM test on ARMA (1,2) residuals*

In table 3, ARCH test for heteroskedasticity $obs^{*}\ R^2$ and corresponding p value. Since $p < 0.05$, it shows that the test statistics is highly significant. So, rejecting null hypothesis of no ARCH effect at 1% significance level, confirms the existence of ARCH effect.

Soybean price return series has analyzed for optimum value of lag specification of GARCH models. For modelling conditional heteroskedasticity, the estimated errors in ARMA is modeled by using GARCH specification.

Volatility model specification

GARCH model describes the coefficient of conditional mean and variance equation simultaneously. The suitable model selection is based on AIC and SIC criteria. The parameter of the GARCH models are estimated by maximum likelihood estimator using BFGS/ Marquardt algorithm (Ghalanos,2013). The conditional volatility model optimization is done by Comparing different error distribution GARCH models as given in table 4.

Model	Error distribution	AIC value	SIC value
GARCH (1,1) model	Normal	2.844887	2.857713
	Student-t	2.685597	2.700255
	GED	2.716212	2.730870
GARCH (1,2) model	Normal	2.844995	2.859653
	Student-t	2.686627	2.703118
	GED	2.716784	2.733275
GARCH (2,1) model	Normal	2.845467	2.860125
	Student-t	2.686448	2.702938
	GED	2.716779	2.733269
GARCH (2,2) model	Normal	2.844938	2.861429
	Student-t	2.681143	2.699466
	GED	2.713308	2.731631

Table 4: *Parameter estimation of ARMA (1,2)-GARCH models for different error distributions*

From the estimated models, the best fitted models are N-GARCH (1,1), T-GARCH (2,2) and GED-GARCH (2,2). Further, Investigation of the adequacy of GARCH model by accuracy measures criteria RMSE, MAE and Theil U-statistics, is given below in table 5

Selected model	Error distribution	Accuracy measures indicators		
		RMSE	MAE	Theil U-statistics
GARCH (1,1) model	Normal Distribution	1.179383	0.767744	0.774899
GARCH (2,2) model	t- distribution	1.173991	0.760466	0.804167
GARCH (2,2) model	GED distribution	1.173635	0.759197	0.820268

Table 5. *Performance of different error distribution GARCH models*

Accuracy measures shows that GARCH (2,2,) model with GED distribution has lowest values in most of the statistical measures. Thus, for estimation of volatility in return series GED-GARCH (2,2,) is the best fitted model.

Forecasting volatility model

Dependent Variable: RETURN
Method: ML ARCH - Generalized error distribution (GED) (BFGS / Marquardt steps)
Date: 02/04/19 Time: 15:16
Sample (adjusted): 1/04/2006 12/29/2017
Included observations: 3336 after adjustments
Failure to improve likelihood (non-zero gradients) after 144 iterations
Coefficient covariance computed using outer product of gradients
MA Backcast: 1/02/2006 1/03/2006
Presample variance: backcast (parameter = 0.7)
GARCH = C(5) + C(6)*RESID(-1)^2 + C(7)*RESID(-2)^2 + C(8)*GARCH(-1) + C(9)*GARCH(-2)

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.058815	0.015521	3.789492	0.0002
AR(1)	0.628294	0.321653	1.953329	0.0508
MA(1)	-0.434498	0.322793	-1.346059	0.1783
MA(2)	-0.112302	0.067057	-1.674727	0.0940
Variance Equation				
C	0.000274	0.000122	2.242077	0.0250
RESID(-1)^2	0.197895	0.014441	13.70404	0.0000
RESID(-2)^2	-0.196743	0.014459	-13.60650	0.0000
GARCH(-1)	1.740872	0.000566	3076.232	0.0000
GARCH(-2)	-0.742168	0.000675	-1098.705	0.0000
GED PARAMETER	1.194442	0.017541	68.09384	0.0000

R-squared	0.036488	Mean dependent var	0.029421
Adjusted R-squared	0.035620	S.D. dependent var	1.197868
S.E. of regression	1.176340	Akaike info criterion	2.713308
Sum squared resid	4610.744	Schwarz criterion	2.731631
Log likelihood	-4515.798	Hannan-Quinn criter.	2.719863
Durbin-Watson stat	2.003210		
Inverted AR Roots	.63		
Inverted MA Roots	.62	-.18	

Table 6: *Result of ARMA (1,2) - GARCH (2,2) model with GED distribution*

Analyzing result of ARMA (1,2)-GARCH (2,2) models (in table 6) clearly demonstrate that parameters for GARCH model are statistically significant. The constant of the GARCH model (ω) and coefficient viz. ARCH term (α_1 *and* α_2) and GARCH term (β_1 *and* β_2) are found to be highly significant. The estimated coefficients in the conditional variance equation describe the $\beta_1 + \beta_2 > \alpha_1 + \alpha_2$ indicating the presence of volatility in the market and current volatility is sensitive to its lagged value. Here, α and β coefficients determine the volatility and the sum of the coefficients of α and β is close to unity represents that the shocks will continue for many future periods.

Conclusions:

The study observed the GARCH model under the three-error distribution (normal, student t, GED) to capture the common features (volatility) of commodity market. The daily return series exhibit negative skewness, high kurtosis, volatility clustering and ARCH effect. GARCH model is suitable to capture these standard features of commodity market. For best model selection comparative analysis done on the basis of AIC, SIC, MAE, RMSE and Theil inequality coefficients and find that for GED-GARCH (2,2) model all statistical measures have low value. The estimated parameter of GARCH (2,2) model lies within the normal ranges. The reaction parameter α_1 is about 0.19 shows that the market is jumpy or nervous and the persistent parameter β_1 is 1.74. For, GARCH model, the value of the coefficients ($\alpha + \beta$) is close to unity. which demonstrates that volatility shocks are highly persistent. The results of the study show that volatility shocks are persistent in soybean prices.

References:

Akaike, H., 1969, "Fitting autoregressive models for prediction", *Ann. Ins. Statist. Math*.21, pp.243-247.

Bollerslev, T. 1986. "Generalized Autoregressive Conditional Heteroscedasticity", *Journal of Econometrics*, 31, pp.307-327.

Bollerslev, T., 1987. "A conditionally heteroscedastic time series model for speculative prices and rates of return", *The review of economics and statistics*, 69(3), pp. 542-547.

Cryer, D.J. and Chan, K-S,1998 "Time series analysis with application in R." *2nd Edition. Springer*.

Engle, R.F. (1982), "Autoregressive Conditional Heteroscedasticity with the Estimates of the variance of United Kingdom Inflation", *Econometrica*,50(4), pp.986-1007.

Dempster, A.P., Laird, N.M., Rubin, D.B.,1977, "Maximum likelihood from incomplete data via. The EM algorithm", *Journal of the royal statistical society series B- Statistical methodology*, Vol. 9 (1), pp.1-38.

Dickey, D.A., "Distribution of the estimators for autoregressive time series with a unit root", *Journal of the American statistical association*, Volume 74, Issue 366(Jun., 1979), pp. 427-431.

Durbin J. and Watson G.S., 1951 "Testing for serials correlation in least-Square Regression", *Biometrika*, Vol 37, pp. 409-428 and vol. 38, pp.159-178.

Engle, R.F. 1982, "Autoregressive Conditional Heteroscedasticity with the Estimates of the variance of United Kingdom Inflation", *Econometrica*,50(4), pp 986-1007.

Galeano, P, Tsay,R.S, 2009, "Shifts in individual parameter of a GARCH model.", *Journal of financial econometrics*,8(1),122-153.

Glejser, H., “A new test for heteroskedasticity”, *Journal of the American Statistical Association*, Vol.64, 1969, pp. 316-23.

Goldfield, S.M. and Quandt, R.E. 1965. “Some tests for Homoscedasticity”, *Journal of the American Association*, Vol. 60, pp. 539-47.

Ghalanos, A.,2013, “Introduction to the rugarch package,r package (version 1.0-14)”.

Hamilton J., Sumel, R.,1994, “Autoregressive conditional heteroskedasticity and change in regime”, *Journal of econometricas*,64, pp 307-333.

Hansen B., 1994, “Autoregressive conditional density estimation”, *International economic review*, 35, pp. 705-730.

Jarque, C.M., and Bera, A.K.,“Time series analysis forecasting and control” *san Franciso: Holden-Day*,1970.

Jondeau, E., Rockinger, M.,2003, “Conditional volatility, skewness, and kurtosis existence, persistence, and co movements”, *Journal of economic dynamics & control*, 27(2003), pp. 1699-1737.

Kelkay, B.D., Emmanuel, G./Yohannes, 2014, “The application of Garch family models to some agriculture crop products in Amhara National Regional State”, *Journal of Economics and sustainable development*, ISSN 2222-1700, Vol 5, No 11,2014.

Ljung, G.M. and Box, G.E.P.1978, “On a measure of lack of fit in the time series models”. *Biometrika*, 65, pp. 297-303.

Koksal, B.,2009, “A comparison of conditional volatility estimators for ISE national 100 Index returns”, *Journal of economics and social researcg*, 11(2), pp. 1-28.

Makridakis, S. and Hibbon M. (1979), “Accuracy of forecasting: An empirical investigation”, *Journal of Royal Statistical Society of Australia*, vol 142(120), pp. 97-145.

Schwarz G. 1978, “Estimating the dimension of a model, London U.K.”, *The Annals of Statistics*, 6, pp. 461-464.

Giller, L.G.,2005. “A generalized error distribution”, *SSRN electronic journal*.

Tsay, R.S.,2005 “Modelling financial time series”, *John Wiley and sons ltd*.

Wilhemsson, A., “Garch forecasting performance under different error distribution assumptions” *Journal of forecasting*, 25, pp. 561-578.

Zhang,Z., 2009 “Analysis skewness in GARCH model”,School of economics and social science, Hogskolan Dalarna.SE 781 88.