Modeling and Forecasting Volatility – How Reliable are modern day approaches?

Mehta, Anirudh and Kanishka, Kunal

8 November 2014

Modeling and Forecasting Volatility – How Reliable are modern day approaches?

Anirudh Mehta

Kunal Kanishka, FRM

Abstract

This study explores the volatility models and evaluates the quality of one-step ahead Forecasts of volatility constructed by (1) GARCH, (2) TGARCH, (3) Risk metrics And (4) Historical volatility. Volatility forecasts suggest that TGARCH performs Relatively best in term of MSPE, followed by GARCH, Risk metrics and historical Volatility. In terms of VaR, we test for correct unconditional coverage and index-Dependence of violations using Likelihood Ratio tests. The tests suggest that VaR forecasts at 90 % and 95% have desirable properties. Regarding 99% VaR forecasts, We find significant evidence that suggests none of the models can reliably predict at this confidence level.

Number of Pages in PDF File: 11

Keywords: Asset pricing, Volatility Forecasting, GARCH, T-GARCH, Risk metrics, LR ratio, VaR

JEL Classification: C10, C12, C15, C19, C51, C53, C58

If you have any comment or suggestion, you can reach to the authors at:
Anirudh.mehta@outlook.com
kunalhansy@gmail.com

1 Dataset and stylized facts

We use daily returns, realized variance (based on 5-minute intra-day returns) and Parkinsons (1980) variance estimator based on the high-low range for the Amsterdam Exchange index (AEX) over the period January 3, 2000 - September 28, 2012 (3242 observations). Figure 1a shows a histogram of the returns plotted together with a theoretical density line of the normal distribution having the same mean and variance as the observed data. Figure 1b shows the autocorrelation of the first 100 lags of the returns, the absolute value of the returns and the squared returns. In addition, we plot the autocorrelation of the realized variance based on intra-day data and based on the Parkinson estimator in figures 1c - 1d , respectively.

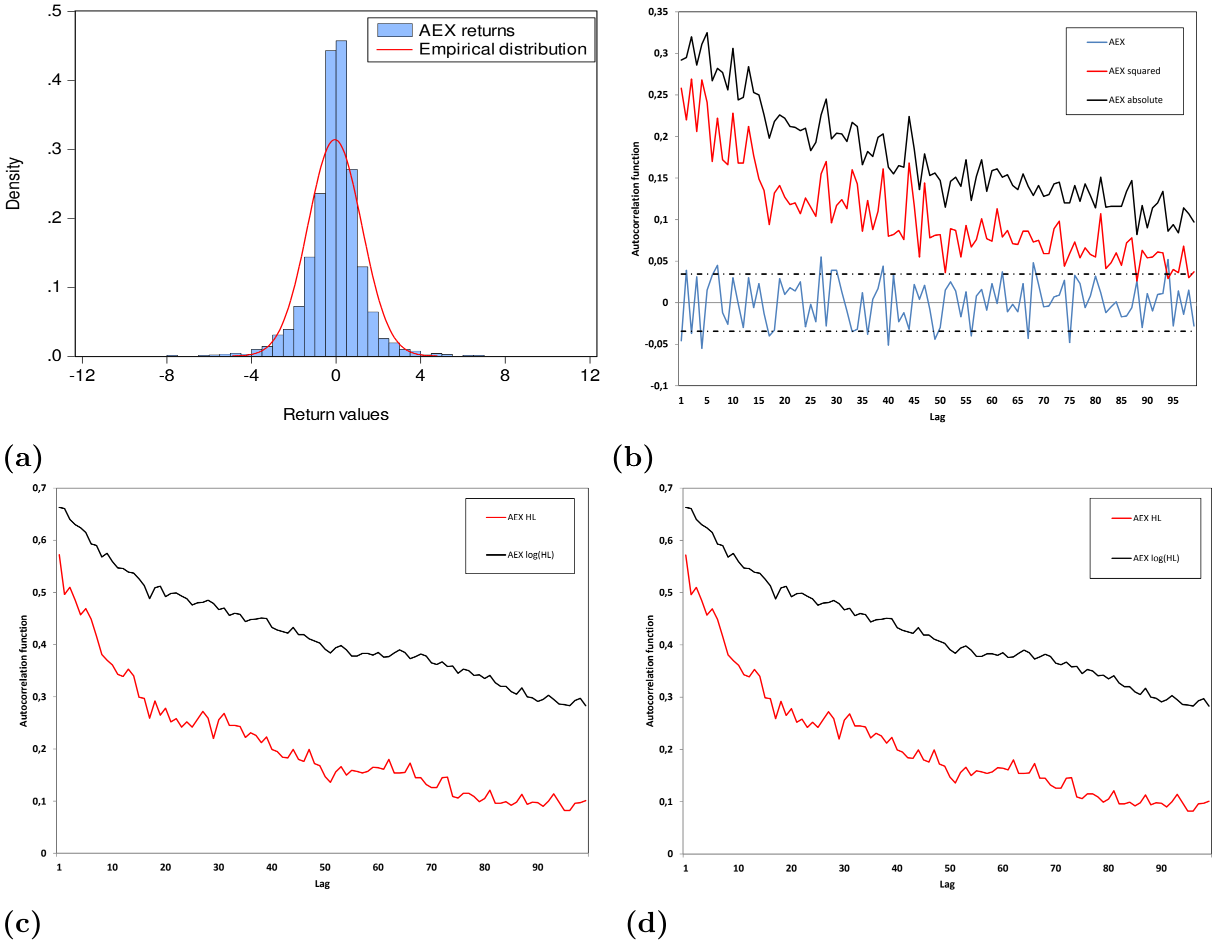


Figure 1: The first graph shows a histogram of the AEX returns plotted together with an empirical normal normal density. The second graph shows autocorrelations (ACF) of returns, absolute returns and squared returns for AEX stock returns. The dash-dotted (-.-) straight lines are confidence lower and upper bounds based on a normal distribution and 3242 observations. The last two graphs show autocorrelations of intra-day 5 minute based "realized" variance and variance calculated using the Parkinson method.

Figures 1a - 1b clearly show that the AEX returns are characterized by the so-called stylized facts of asset returns. We find high peaks that are unlikely for a normal distribution, as well as some observations in the tails that have a very low empirical probability in figure 1a. Also, the autocorrelation functions in figure 1b display significant positive autocorrelations only in the absolute and squared returns, whereas the returns themselves only show a few spikes outside the confidence interval. The latter few violations are not surprising considering the fact that 5 percent of the correlation should lie outside the interval by theory. The autocorrelations of the realized variance and the estimator by Parkinson show the same behavior, as the squared and absolute returns of AEX. In fact, autocorrelations have even a higher magnitude compared to the AEX squares and absolute values. Next we look at some some descritive statistics of the data in table 1.

	AEX	AEX RV	LRV	AEX HL	LHL
Mean	-0.057	1.447	-0.218	1.449	-0.427
Max	9.237	36.240	3.590	48.985	3.892
Min	-8.410	0.038	-3.267	0.020	-3.932
St. dev	1.270	2.309	1.020	2.793	1.212
Skewness	-0.170	5.700	0.434	6.267	0.272
Kurtosis	9.033	52.537	3.038	62.663	2.961
Jarque-Bera	4933	349007	93	502071	40
p-value	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**
Observations	3242	3242	3242	3242	3242

Table 1: This table shows descriptive statistics of AEX returns, realized variance based on 5-minute intra-day intervals, the log of realized variance, variance estimated using Parkinson's estimator and the log of the Parkinson estimator, denoted by AEX, AEX RV, LRV, AEX HL and LHL, respectively

AEX seems to differ significantly from a normal distribution, by judging the Jarque-Bera p-values. Moreover non-normality, the realized variance and the estimator proposed by Parkinson are characterized by excessive kurtosis values exceeding 50. These high numbers are caused by excessive peakedness as well as heavy tail mass. The range of AEX also resembles more than six standard deviations, again supporting the idea that we are dealing with non-normal series. Also worth noting, realized variance climbs up to very high levels, whereas the log operator smooths the series out in this context, and also smooths the peakedness and mass measured by the kurtosis.

2 Methodology

We try to model volatility of the AEX daily returns, denoted by r_t at time t, using four different models: (1) a GARCH(1,1) model, (2) a Treshold GARCH(1,1), (3) riskmetrics and (4) historical volatility. Respectively, the formulas for these models are given below by 1 - 2 ; 1 and 3 ; 4 ; and 5, respectively.

$$r_t = \mu + \varepsilon_t \tag{1}$$

$$\sigma^2_{t,\text{GARCH(1,1)}} = \omega + \alpha\varepsilon^2_{t-1} + \beta\sigma^2_{t-1,\text{GARCH(1,1)}} \tag{2}$$

$$\sigma^2_{t,\text{TGARCH(1,1)}} = \omega + \alpha\varepsilon^2_{t-1} + \gamma\varepsilon^2_{t-1}\text{I}[\varepsilon_{t-1} < 0] + \beta\sigma^2_{t-1,\text{TGARCH(1,1)}} \tag{3}$$

$$\sigma^2_{t,\text{RM}} = \lambda\sigma^2_{t-1,\text{RM}} + (1-\lambda)r^2_{t-1} \tag{4}$$

$$\sigma^2_{t,\text{HV}} = \frac{1}{L}\sum_{j=1}^{L}(r_{t-j} - \overline{r})^2 \tag{5}$$

Where $\lambda = 0.94$ for riskmetrics and $\overline{r}$ is simply the average return up to period $t-1$. We fit data using the first 1514 observations, ranging from January 3, 2000 until December 30, 2005 for the first two models. For the historical volatility we compute the values based on moving windows of length 63, 125 and 250. Also, the riskmetrics model is computed with an initialization of the the first 63 length historical volatility. The latter two volatilities are computed using the entire period. Regarding the evaluation of the model fit, we undertake two steps. First, we examine volatility estimates and compare these with other volatility measures. Second, we study the distributional and autocorrelation properties of the standardized residuals of both the GARCH(1,1) and TGARCH(1,1) model.

As a next step, we perform an expanding one-step ahead forecasting exercise for models (1)-(4) on the out-of-sample period of 1728 observations ranging from January 2, 2006 until September 28, 2012. We will compare the forecasts over the different models using MSPE based on comparison with both the squared returns and the realized variance. The latter variance is computed based on 5 minute intervals and also using the estimator proposed by Parkinson (1980). From these forecasts, we also construct value at risk (notation: $\mathrm{VaR}_t(1-q,1)$) estimates for our four models as described below:

$$\mathrm{P}[r_{t+1} \leq \mathrm{VaR}_t(1-q,1)|\Omega_t] = F_{t+1|t}(\mathrm{VaR}_t(1-q,1)) = q \tag{6}$$

Where Ω_t denotes the information set at time t and $F(\cdot)$ denotes the cumulative distribution function. In addition, we perform Likelihood Ratio tests for correct unconditional coverage and independence, as suggested by Christoffersen (1998). The approach is as follows. First define violations I_t according to the following indicator, and define an estimate of the probability of a violation $\widehat{\pi}$ as follows:

$$\mathrm{I}_t = \begin{cases} 1 & \text{if } r_{t+1} \in (L_{t+1|t}(q), U_{t+1|t}(q)) \\ 0 & \text{if } r_{t+1} \notin (L_{t+1|t}(q), U_{t+1|t}(q)) \end{cases}$$

$$\widehat{\pi} = \widehat{\mathrm{P}}[\mathrm{I}_t = 1] = \frac{T_1}{T_0 + T_1} \tag{7}$$

$$\mathrm{LR}_{uc} = -2\log\left(\frac{\mathcal{L}(q : \mathrm{I}_T, \mathrm{I}_{T-1}, ..., \mathrm{I}_1}{\mathcal{L}(\widehat{\pi} : \mathrm{I}_T, \mathrm{I}_{T-1}, ..., \mathrm{I}_1}\right) \overset{\text{asy}}{\sim} \chi^2(1) \tag{8}$$

Where $T_1 = \sum_{t=1}^{T} i_t$, and $T_0 = T - T_1$. $T_{ij} = 1$. We can then derive a distribution for the "ratio" given by 8, which turns out to be $\chi^2(1)$. In addition, we calculate proportions of subsequent violations and test their significance using again a likelihood ratio test.

$$\widehat{\pi}_{01} = \frac{T_{01}}{T_{00} + T_{01}} \tag{9}$$

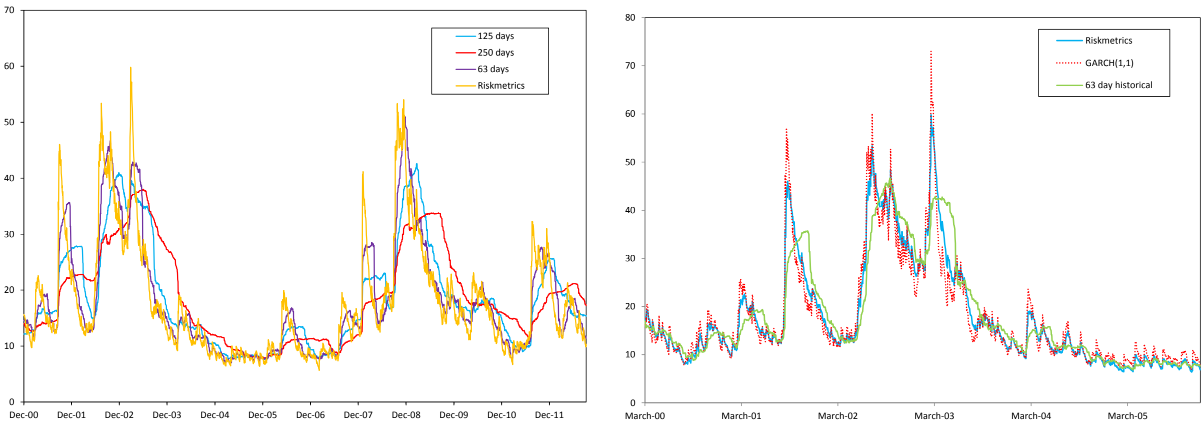
$$\widehat{\pi}_{11} = \frac{T_{11}}{T_{10} + T_{11}} \tag{10}$$

$$\widehat{\pi}_2 = \frac{T_{01} + T_{11}}{T_{00} + T_{10} + T_{01} + T_{11}} \tag{11}$$

Where T_{ij} denotes the number of observations for which $\mathrm{I}_{t+1} = j$ and $\mathrm{I}_t = i$. There is also an LR test in this case. Details can be found in Christoffersen (1998). In this report we will display values of 7, 9, 10 and 11, including the LR test statistics.

3 Results

This section shows the results of the model fit using all models noted in the previous section and evaluates the results of the constructed forecasts. We start with a graph of the volatility movements over time as displayed in 2a and 2a.



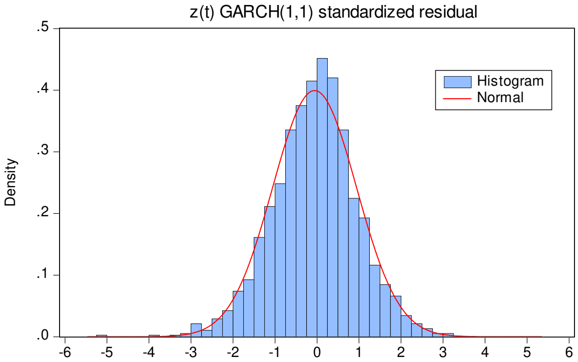
(a) (b)

Figure 2: (a) Graphs of annualized volatility estimates of daily returns of AEX index based on Riskmetrics model and historical moving window method for window lengths of 63, 125 and 250 days. Estimation period is from 12/22/2000 until 9/28/2012. (b) Graph of annualized volatility estimates of daily returns of AEX index based on Riskmetrics, GARCH(1,1) and 63 day historical moving window method. Estimation period is from 3/31/2000 until 12/30/2005.

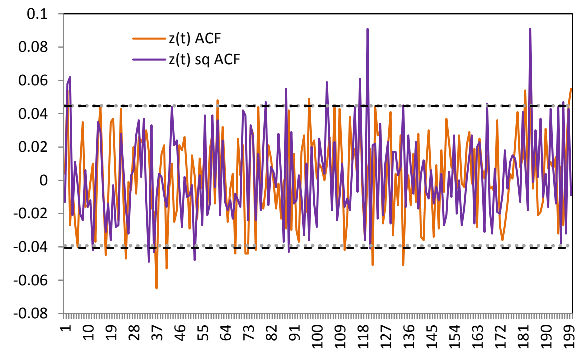
Figure 2a clearly shows that the historical volatility estimate is smoothed out, whenever the size of the moving window is increased. It also suggests that Riskmetrics estimates decline more rapidly compared to the latter. This can be explained by the fact that the Riskmetrics 1-step ahead estimates are calculated by taking the residual effect from previous day's volatility into account, while historical volatility is characterized by slow adaptation of estimates if abnormally high or low returns do occur, since each new observation gets an equal weight.

Volatility estimates from Riskmetrics and GARCH(1,1) model are plotted in figure 2b with a 63 day historical moving window. The movement of the historical volatility estimates show a considerable 'Ghosting effect', since a period in which the historically estimated volatility rises and stays high for some time, is often followed by a sudden drop. This behavior can be attributed to the fact that observations of abnormally high returns can stay inside the estimation period for multiple periods giving rise to peaks, and once the moving window passes through that period it will go down to a lower level. Also, GARCH(1,1) volatility estimates are more peaked and decline more sharply compared to the Riskmetrics model. This can be explained by the fact that GARCH estimation does not fix the weighing of shocks to a new period a priori, compared to 0.94 in Riskmetrics, but will pick the sensitivity to shocks in such a way, that the likelihood of observing the data is maximized.

To evaluate in-sample fit, we look at some desirable and implied properties by our model setup. In this context we retrieve the standardized residuals from the estimated parameters and check whether these are normal and test for significant autocorrelation in values and absolute values..



(a)



(b)

Figure 3: (a) Freguency distribution of GARCH(1,1) standardized residuals calculated from estimated model parameters under normality assumption of the standardized unexpected returns. (b) Auto-correlation function for GARCH(1,1) standardized residuals and squared residuals calculated for 200 lags.

Mean	-0.057
Median	-0.015
Maximum	3.111
Minimum	-5.001
Std. Dev.	0.998
Skewness	-0.169
Kurtosis	3.669
Jarque-Bera	35.411
p-value	**0.000**
Observations	1514

Table 2: Descriptive statistics for standardised residuals calculated for GARCH(1,1) model using 3 different distribution assumption for standardized unexpected return namely Normal dist., Generalized Error dist. and Student's t distribution.

Figure 3a and Table 2 indicates that the distribution of the standardized residuals is close to normal, though it has slightly fatter tail (Kurtosis > 3) and negative Skewness compared to a normal distribution. Even if the normality assumption for distribution of standardized unexpected returns is slightly misspecified, according to theory of Quasi Maximum Likelihood we still get consistent estimates for parameters, though the standard errors may vary a bit. Figure 3b clearly shows that there is no significant auto-correlation in the residual and squared residual terms.

The conditional variance expression for Threshold GARCH(1,1) model in Equation 3 can be rewritten as below where $\alpha + \gamma$ is the coefficient for negative unexpected return and α is the coefficient for positive unexpected return.

$$\sigma^2_{t,\mathrm{TGARCH(1,1)}} = \omega + \alpha\varepsilon^2_{t-1}\mathrm{I}[\varepsilon_{t-1} > 0] + (\alpha+\gamma)\varepsilon^2_{t-1}\mathrm{I}[\varepsilon_{t-1} \leq 0] + \beta\sigma^2_{t-1,\mathrm{TGARCH(1,1)}} \quad (12)$$

	GARCH(1,1)	TGARCH(1,1)
ω	0.012	0.013
α	0.113	-0.020
β	0.880	0.925
γ		0.162
$(\alpha + \gamma)$		0.142

Table 3: Coefficient estimation for GARCH(1,1) and TGARCH(1,1) models as per Equation 2 and 12.

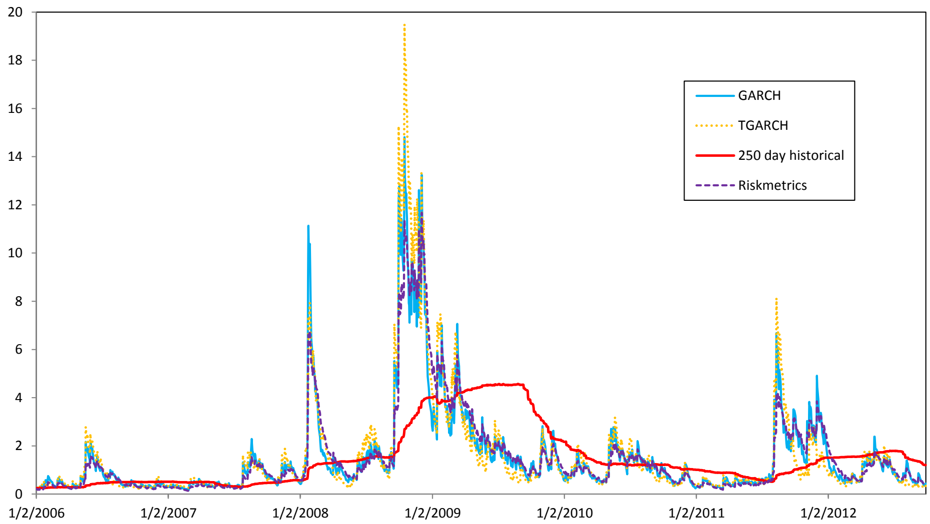


Figure 4: Conditional 1-step ahead variance forecasts for AEX index from GARCH(1,1) and Threshold GARCH(1,1) models made between 1/2/2006 and 9/7/2012 shown with Historical 250 day moving window variance estimates as well as Riskmetrics 1-step ahead variance forecasts within same out sample time period.

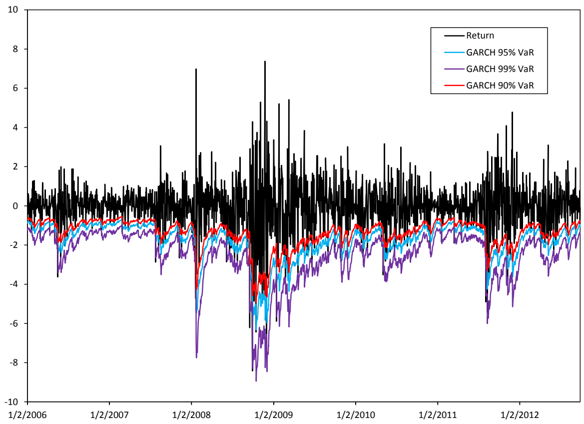
The estimated coefficients for GARCH(1,1) and Threshold GARCH(1,1) true model parameters from Equations 2 and 3 are given in Table 3. Both the models turn out to be covariance stationary as expected due to $\alpha + \beta < 1$ for the GARCH(1,1) model and $(2\alpha + \gamma)/2 + \beta < 1$ for the TGARCH(1,1) model. Also, both the models have a high persistence (high values of β), implying that modeled volatility is highly correlated. The TGARCH(1,1) setup allows for asymmetric reactions of upside and downside risk, which is often found in practice. We find that the effect of negative shocks to returns (0.142) is almost 7 times bigger than the effect of a positive shock (-0.020), confirming the formerly noted finding.

Figure 4 indicates that GARCH(1,1) and TGARCH(1,1) conditional variance forecasts are more peaky than Riskmetrics which assumes constant variance over past T days. A historical moving window estimate is much flatter and smoothed out than both GARCH models and the historical volatility. Now zoom in on the performance of these models to see which of these models captures the variance of returns best.

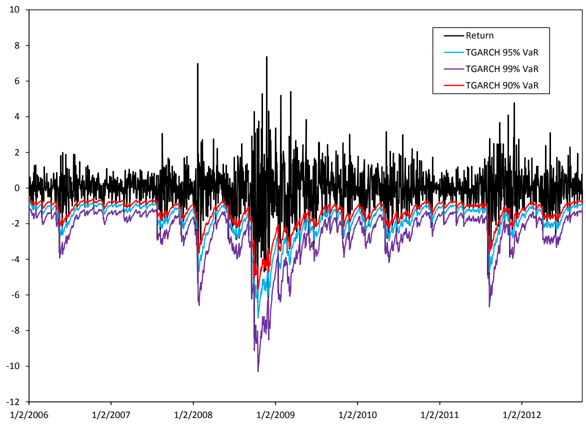
	MSPE		R^2	
	sqaured ret.	realized var.	sqaured ret.	realized var.
GARCH(1,1)	15.806	2.530	0.142	0.222
TGARCH(1,1)	15.732	2.466	0.160	0.264
250 day historical	17.937	5.310	0.032	0.012
Riskmetrics	15.840	2.876	0.138	0.169

Table 4: Mean Square Prediction Error of AEX index variance forecasts based on GARCH(1,1), Threshold GARCH(1,1), Historical 250 day moving window and Riskmetrics models. Squared daily returns and realized variance based on high frequency intra-day return data are used as proxies for actual observed variances to calculate MSPE.

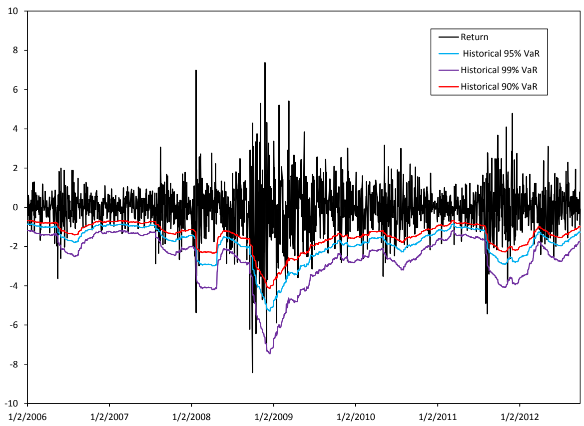
The Mean Squared Prediction Error (MSPE) values and R-squared values (from regression of true variance proxy on forecasted variance) shown in Table 4 clearly shows that a Threshold GARCH(1,1) model, which accounts for asymmetry effect of Positive and negative shocks to returns, performs the best in terms of forecasting volatility in the sense that it has the lowest MSPE and highest R-squared values both when squared return and realized variance are used as proxy for true variance. This is followed by GARCH(1,1) and Riskmetrics model respectively. Historical moving window estimate performs the worst in all respect and yields very low R-squared value compared to other volatility models, which is not unexpected since it has serious drawbacks. When realized variance based on high frequency intra-day return data is used as proxy for true daily variance, the MSPE values of conditional variance forecasts are much lower overall and the R-squared value also increases for all the models expect historical volatility. This confirms the notion that squared daily returns are, despite being unbiased, a very noisy estimate of actual volatility. Next, we plot VaR estimates at different confidence levels together with actual returns for the different model based approaches considered so far.



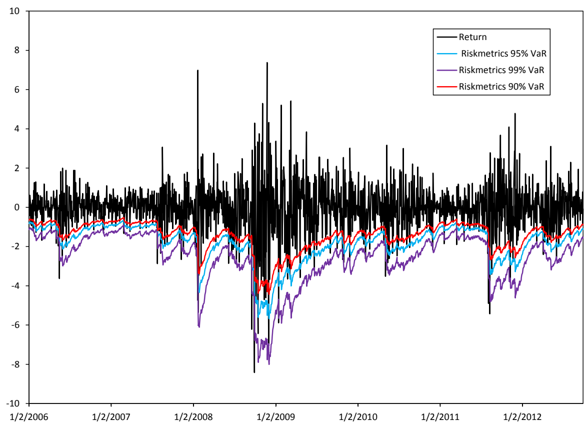
(a)



(b)



(c)



(d)

Figure 5: Daily 90%, 95% and 99% VaR estimates plotted with actual AEX index daily returns for the time period 1/2/2006 until 9/27/2012. VaR estimates are calculates based on daily volatility forecasts from (a) GARCH(1,1), (b) Threshold GARCH(1,1), (c) Historical 250 day moving average and (d) Riskmetrics.

On first sight, it seems as if the historical based VaR estimates are not always that accurate. We evaluate the VaR estimates in terms of correct unconditional coverage and independence, by showing LR test statistics which are presented in Table 5 and Table 6.

1-day VaR	q	$\hat{\pi}$	T_0	T_1	LR_{uc}
Panel A: GARCH(1,1) model					
90% confidence	0.10	0.107	1542	185	0.953
95% confidence	0.05	0.059	1625	102	2.829*
99% confidence	0.01	0.024	1685	42	25.550***
Panel B: Threshold GARCH(1,1) model					
90% confidence	0.10	0.103	1550	177	0.118
95% confidence	0.05	0.059	1626	101	2.487
99% confidence	0.01	0.023	1687	40	22.036***
Panel C: Historical volatility 65 days					
90% confidence	0.10	0.096	1562	165	0.387
95% confidence	0.05	0.059	1625	102	2.829*
99% confidence	0.01	0.025	1684	43	27.381***
Panel D: Riskmetrics					
90% confidence	0.10	0.096	1562	165	0.387
95% confidence	0.05	0.058	1627	100	2.166
99% confidence	0.01	0.024	1685	42	25.550***

Table 5: Testing correct unconditional coverage of daily VaR estimates for AEX index daily returns from different model based approaches using Christoffersen (1998) likelihood ratio test as formulated in Equation 8. VaR estimates are calculated for out of sample time period 1/2/2006 to 9/27/2012. 1, 2 or 3 star(s) denote(s) statistical significance at 10, 5 and 1 percent, respectively.

1-day VaR	q	$\hat{\pi}_2$	$\hat{\pi}_{01}$	$\hat{\pi}_{11}$	T_{00}	T_{01}	T_{10}	T_{11}	LR_{ind}
Panel A: GARCH(1,1) model									
90% confidence	0.10	0.107	0.110	0.081	1371	170	170	15	1.582
95% confidence	0.05	0.059	0.061	0.029	1525	99	99	3	2.064
99% confidence	0.01	0.024	0.025	0.000	1642	42	42	0	-
Panel B: Threshold GARCH(1,1) model									
90% confidence	0.10	0.103	0.106	0.073	1385	164	164	13	1.974
95% confidence	0.05	0.059	0.062	0.010	1525	100	100	1	6.781***
99% confidence	0.01	0.023	0.024	0.000	1646	40	40	0	-
Panel C: Historical volatility 65 days									
90% confidence	0.10	0.096	0.093	0.121	1416	145	145	20	1.297
95% confidence	0.05	0.059	0.059	0.059	1528	96	96	6	0.000
99% confidence	0.01	0.025	0.025	0.023	1641	42	42	1	0.005
Panel D: Riskmetrics									
90% confidence	0.10	0.096	0.097	0.085	1410	151	151	14	0.251
95% confidence	0.05	0.058	0.060	0.030	1529	97	97	3	1.808
99% confidence	0.01	0.024	0.025	0.000	1642	42	42	0	-

Table 6: Testing independence of daily VaR estimates for AEX index daily returns from different model based approaches using Christoffersen (1998) likelihood ratio test. VaR estimates are calculated for out of sample time period 1/2/2006 to 9/27/2012. 1, 2 or 3 star(s) denote(s) statistical significance at 10, 5 and 1 percent, respectively.

Table 5 and Table 6 show that the null hypothesis of correct unconditional coverage for 90% and 95% VaR estimates for all the models considered cannot be rejected even at a 99% confidence interval. For 99% VaR estimates, the null hypothesis of correct unconditional coverage is convincingly rejected for all the models under study. The null hypothesis of independence of VaR limit violations cannot be rejected for most of the VaR percentiles for all the models. This suggests that all 4 models considered here produce reasonably accurate VaR estimates in terms of correct unconditional coverage and independence of violations.

4 Conclusion

In this study, we have addressed model fit and evaluated forecasts of a few volatility models. Subsequently, we constructed 1-step ahead daily VaR estimates for the AEX index returns. We found that the Threshold GARCH(1,1) model performs the best in terms of MSPE for predicting volatility and confirms the fact that negative shocks on returns have a much higher effect on volatility compared to positive shocks. A symmetric GARCH(1,1) is preferred second in terms of Mean Square Prediction Error followed by Riskmetrics and historical moving window forecast. Historical estimates are characterized by relatively high values of prediction error compared to the other 3 volatility models. We think this due to the fact that GARCH models of volatility are based on conditional return distribution, which makes them superior compared to Riskmetrics where a static distribution of underlying returns is assumed over a certain period. Also worth noting, all the models perform reasonably well in predicting the downside risk in terms of 1-step ahead VaR limit at different confidence levels, unless the confidence level is set really high (99%).

References

Christoffersen, P.F. (1998) "Evaluating interval forecasts," *International economic review*, Vol. 39, No. 4, pp. 841–862.

Parkinson, M. (1980) "The extreme value method for estimating the variance of the rate of return," *Journal of Business*, Vol. 53, No. 1, pp. 61–65.