18th World IMACS / MODSIM Congress, Cairns, Australia 13-17 July 2009
http://mssanz.org.au/modsim09

Forecasting Value-at-Risk using Maximum Entropy Density

Felix Chan [1]

[1]*School of Economics and Finance, Curtin University of Technology, Perth, Western Australia*
Email: F.Chan@curtin.edu.au

Abstract: Despite its shortcoming, Value-at-Risk (VaR) remains as one of the most important measures of risk for financial assets. Although it is used widely by regulatory authority in assessing risk of the financial markets, the robust construction of VaR forecasts remains a controversial issue. This paper proposes a new method to construct VaR forecasts based on Maximum Entropy Density, along with the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model of Bollerslev (1986).

Using the result in Ling and McAleer (2003), the Quasi-Maximum Likelihood Estimator (QMLE) with the normal density for ARMA-GARCH model is consistent and asymptotically normal under mild assumptions. This implies that it is possible to obtain consistent estimates of the standardized residuals even when the underlying distribution of returns is non-normal. Given this, the distribution of the standardized residuals can then be approximated using Maximum Entropy Density (MED) which allows different characteristics of the distribution, such as excess kurtosis, to be accommodated. The one-day-ahead VaR forecasts can then be constructed by using the estimated ARMA-GARCH model and the MED.

The practical usefulness of the proposed method is evaluated empirically against ARMA-GARCH and ARMA-GJR models with different distributional assumptions using daily S&P 500 data. The empirical results show promising sign of the proposed method.

Keywords: Value-at-Risk, Maximum Entropy Density, GARCH, GJR, Volatility.

INTRODUCTION

Value-at-Risk (VaR) has been used widely as a measure of risk for portfolio of financial assets and it is a requirement of the Basel Capital Accord (see Basel Committee on Banking Supervision (1988, 1995, 1996)). Essentially, VaR is a procedure designed to forecast the maximum expected negative return over a target horizon, given a statistical confidence limit (see for example, Jorion (2000) for an insightful discussion). However, robust construction of VaR forecast remains a controversial issue. Given the importance of VaR in risk management for finance and banking industries, it is important to develop a robust and accurate method to forecast VaR.

There are two general approaches to forecast VaR, namely the parametric approach and nonparametric approach. The parametric approach generally requires an assumption on the probability distribution of the underlying return. Given the assumed distribution, the parameters in the distribution will then be estimated using historical data. Under the parametric approach, it is possible to allow the parameters in the distribution to change over time and therefore allows the shape and the location of the distribution to change over time. Two of the most popular models used in this approach are the Autoregressive-Moving Average Generalized Autoregressive Conditional Heteroskedasticity (ARMA-GARCH) model of Bollerslev (1986) and the ARMA-Glosten, Jagannathan and Runkle (ARMA-GJR) model of Glosten, Jagannathan and Runkle (1992). These two models provide convenience framework to allow the mean and the variance of the conditional distribution to change over time. Although these models are flexible in terms of accommodating the time-varying nature of the distribution, the distributional assumption is crucial in producing accurate forecast of VaR. However, it is difficult to identify the appropriate distribution based on existing data.

This paper proposes a new method to forecast VaR with maximum entropy density (MED). Using the results in Ling and McAleer (2003), the parameters in an ARMA-GARCH model can be consistently estimated by Quasi-Maximum Likelihood Estimator (QMLE) with normal density even when the underlying distribution of returns is non-normal. This implies that it is possible to obtain consistent estimates of the standardized residuals which can be used to approximate the true distribution using Maximum Entropy Density. This also allows other characteristics of the underlying distribution, such as asymmetry and excess kurtosis to be accommodated. Given the MED and the estimated ARMA-GARCH model, the one-day-ahead VaR forecasts can then be constructed.

The practical usefulness of the proposed method will be evaluated empirically using daily data of S&P 500 from 13 January 1989 to 14 January 2009 with a total of 5219 observations. In addition to the proposed method, VaR forecasts will also be constructed using ARMA-GARCH and ARMA-GJR models with different distributional assumptions. The results show promising sign for the proposed method.

The paper is organized as follows: Section 2 provides a summary of VaR and its construction using the GARCH and GJR models, this is followed by a summary on Maximum Entropy Density in Section 3. Section 4 contains the empirical results and Section 5 contains some concluding remarks.

1. PARAMETRIC VOLATILITY MODELS AND VALUE-AT-RISK

This section provides a summary of VaR and its construction using parametric volatility models, such as the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model of Bollerslev (1986) and the Glosten, Jagannathan and Runkle (GJR) model of Glosten et al. (1992).

The literature on modeling volatility has become increasingly popular in the last two decades. Since the introduction of the Autoregressive Conditional Heteroskedasticity (ARCH) model of Engle (1982), many different variations and extensions have been proposed in the literature aiming to capture the dynamic of the conditional second moments. Two of the simplest and most popular variations are the Generalized ARCH (GARCH) model of Bollerslev (1986) and the Glosten, Jagannathan and Runkle model of Glosten et al. (1992). The former extended the ARCH model by incorporating the past conditional variance and thus, allows long-run persisting behavior of shocks. The latter extended the GARCH model by incorporating asymmetric behavior of shocks, that is, the model allows future conditional variance to react differently between past positive and negative shocks.

Consider the following model for financial returns:

$$\begin{aligned} (1-\phi(L))\, r_t &= (1-\theta(L))\varepsilon_t \\ \varepsilon_t = \eta_t \sqrt{h_t} \qquad & \eta_t \sim iid(0,1) \end{aligned} \tag{2.1}$$

where r_t is the return at time t, L denotes the lag operator so that $Lr_t = r_{t-1}$, $\phi(L)$ and $\theta(L)$ are two polynomials in lags so that $\phi(L)=1-\sum_{i=1}^{r}\phi_i L^i$ and $\theta(L)=1+\sum_{i=1}^{s}\theta_i L^i$. h_t denotes the conditional variance and the standardized residuals, η_t, is a sequence of independently, identically distributed random variables with zero mean and unit variance.

Obviously, the specification of h_t governs the dynamics of the conditional variance. One of the most popular specifications is the GARCH (p,q) model, which is defined to be

$$h_t = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_i h_{t-i}. \tag{2.2}$$

However, both positive and negative shocks have the same impact on the conditional variance, which is not realistic. Glosten et al. (1992) proposed a further extension to accommodate the asymmetric behavior, the GJR model is defined to be

$$h_t = \omega + \sum_{i=1}^{p} (\alpha_i + \gamma_i I_{t-i})\, \varepsilon_{t-i}^2 + \sum_{i=1}^{q} \beta_i h_{t-i} \tag{2.3}$$

where

$$I_t = \begin{cases} 0, & \varepsilon_t \geq 0 \\ 1, & \varepsilon_t < 0. \end{cases}$$

The parameters in the models as defined in equations (2.1), (2.2) and (2.3) are typically estimated by the Quasi Maximum Likelihood Estimator with the normal density, which is defined as:

$$\hat{\theta} = \arg\max_{\theta\in\Theta} -\frac{1}{2}\sum_{t=1}^{T}\left(\log h_t + \frac{\varepsilon_t^2}{h_t}\right) \tag{2.4}$$

where the conditional variance, h_t, follows either the GARCH process or the GJR process as defined in equations (2.2) and (2.3), respectively. For the ARMA-GARCH model $\theta=(\phi_1,\ldots\phi_r,\theta_1,\ldots\theta_s,\omega,\alpha_1,\ldots\alpha_p,\beta_1,\ldots\beta_q)$ is the $p+q+1+r+s$ parameter vector and $\theta=(\phi_1,\ldots\phi_r,\theta_1,\ldots\theta_s,\omega,\alpha_1,\ldots\alpha_p,\beta_1,\ldots\beta_q,\gamma_1,\ldots\gamma_1)$ is the $2p+q+1+r+s$ parameter vector for the ARMA-GJR process. Note that equation (2.4) defines the QMLE with normal density. If η_t follows a normal distribution then equation (2.4) defines the Maximum Likelihood Estimator (MLE), otherwise it is a QMLE.

The structural and statistical properties of GARCH and GJR models have been established. For the ARMA-GARCH model as defined in equations (2.1) and (2.2), Ling and McAleer (2003) derived the structural and statistical properties for the GARCH(p,q) model. Under the existence of the second moment, namely $\sum_{i=1}^{p}\alpha_i + \sum_{i=1}^{q}\beta_i < 1$, $\hat{\theta} \xrightarrow{p} \theta_0$ for ARMA-GARCH model, where θ_0 denotes the true parameter vector (see Ling and McAleer (2003)). McAleer, Chan and Marinova (2007) provided the structural and statistical properties of GJR(1,1) model. The importance of these properties, especially in the case of ARMA-GARCH, is that the Quasi Maximum Likelihood Estimator (QMLE) for GARCH(p,q) is consistent and asymptotically normal when the second moment exists. This implies that it is possible to obtain consistent estimation of the parameters in the GARCH model without any distributional assumption.

Under the parametric approach, the VaR with the confidence level α at time t is defined to be

$$\mathrm{VaR}_t^{\alpha} = \hat{r}_t - \kappa\sqrt{h_t} \tag{2.5}$$

where $\hat{r}_t$ is the forecast of return for time t, h_t is the conditional variance and κ is a real number so that $P(\eta_t \le \kappa) = 1 - \alpha$. Under model (2.1), both $\hat{r}_t$ and h_t can be easily obtained, but κ depends on the distribution of the standardized residuals, η_t, which is generally not available. Often the distribution of η_t is assumed to be normal which is not adequate for financial returns given the excess kurtosis. Many studies have assumed more flexible distribution for η_t. For example, Bollerslev (1987) proposed a Maximum Likelihood Estimator (MLE) for the GARCH model based on the t-distribution. The problem with assuming different distribution is that the statistical properties of QMLE are only known under the normal density. Therefore, the statistical properties of QMLE under the t-distribution or the Generalized Error Distribution (GED) are still unknown, which casts doubts on the validity of subsequence inferences. The results from Ling and McAleer (2003) offers an alternative, under the appropriate assumption, $\hat{\theta} \xrightarrow{P} \theta_0$, and it is possible to show that the estimated standardized residuals will converge in distribution to the true standardized residuals, that is $\hat{\eta}_t \xrightarrow{d} \eta_t$. Therefore, it is possible to estimate the density of η_t using the estimated standardized residuals. The next section proposes to estimate the density of η_t using maximum entropy density. Once the distribution of η_t is obtained, κ can then be computed.

2. MAXIMUM ENTROPY DENSITY

The concept of entropy was first introduced in Shannon (1948) and popularized by Jaynes (1957a, 1957b). Consider the Shannon's Entropy for continuous random variable:

$$H(p) = \int_{-\infty}^{\infty} p(\eta_t)\log p(\eta_t)\,d\eta_t.$$

This entropy aims to measure the difference in the information content between the density, $p(\eta_t)$ and the uniform distribution. The motivation of the Shannon's entropy relies on the fact that the uniform distribution is used in the absence of any information and therefore, the distance between $p(\eta_t)$ and the uniform distribution should provide a measure of information content in $p(\eta_t)$. This interpretation provides an insight into the estimation of $p(\eta_t)$, that is, $p(\eta_t)$ can be estimated by maximizing $H(p)$ subject to a set of moment constraints. Mathematically,

$$\begin{aligned} &\hat{p}(\eta_t) = \max_p H(p) \\ &\text{s.t.} \\ &\int_{-\infty}^{\infty} \hat{p}(\eta_t)\, d\eta_t = 1, \\ &\int_{-\infty}^{\infty} \eta_t^i \hat{p}(\eta_t)\, d\eta_t = m_i, \qquad i = 1,\dots,k. \end{aligned} \tag{2.6}$$

The idea is that the information contained in the data is summarized by the first k sample moments. Therefore, the information content in the data is transferred into $\hat{p}(\eta_t)$ by restricting its first k moments to be the same as that of the sample.

The optimization problem as stated in equation (2.6) can be solved by defining the following Hamiltonian:

$$L(p) = H(p) + \lambda_0 \left[\int_{-\infty}^{\infty} p(\eta_t)\, d\eta_t - 1 \right] + \sum_{i=1}^{k} \lambda_i \left[\int_{-\infty}^{\infty} \eta_t^i p(\eta_t)\, d\eta_t - m_i \right]$$

and by applying Calculus of Variations yields

$$\hat{p}(\eta_t) = Q^{-1} \exp\left(\sum_{i=1}^{k} \lambda_i \eta_t^i \right) \tag{2.7}$$

where $Q = \int \exp\left(\sum_{i=1}^{k} \lambda_i \eta_t^i\right) d\eta_t$. Equation (2.7) is known as the Maximum Entropy Density where the shape and the location of the distribution is governed by the parameters, $\lambda_1, ..., \lambda_k$, which depends on the sample moments (see Chan (2009) for a detailed discussion). Given the distribution as defined in equation (2.7), κ can be computed by solving

$$\int_{-\infty}^{\kappa} \hat{p}(\eta_t)\, d\eta_t = 1 - \alpha. \tag{2.8}$$

The procedure of constructing the VaR forecast using MED can be summarized as follows:

Step 1. Estimate the parameters in the ARMA-GARCH model as defined in equations (2.1) and (2.2) using the QMLE as defined in equation (2.4).

Step 2. Given the parameter estimates, compute the estimated standardized residuals as

$$\hat{\eta}_t = \frac{r_t - \hat{r}_t}{\hat{h}_t}.$$

Step 3. Estimate the first k sample moments for $\hat{\eta}_t$.

Step 4. Construct the MED based on the sample moments calculated in Step 3 using equation (2.7) (see Chan (2009) and Rockinger and Jondeau (2002) for efficient algorithms to calculate $\lambda_1,...,\lambda_k$ given the moments).

Step 5. Calculate κ by solving equation (2.8).

Step 6. The one-day-ahead VaR forecast is then $\hat{r}_{t+1} - \kappa\sqrt{\hat{h}_{t+1}}$.

3. EMPIRICAL RESULTS

The empirical usefulness of the proposed method will be evaluated empirically in this section using daily S&P 500 data from 13 January 1989 to 14 January 2009 with a total of 5219 observations. In addition to the proposed method, the performance of VaR forecasts based on GARCH and GJR models with normal, Student-t and GED distributions will also be provided as a comparison to the proposed method.

The first 4000 observations will be used to estimate the parameters in the GARCH and GJR models under differential distributional assumption. The one-day-ahead VaR forecast with 99% confidence level will then be constructed for the remaining sample using equation (2.5), with different values of κ based on MED, normal, Student-t and GED distributions. The performance of each model can then be compared by calculating the percentage of violation relative to the total number of forecasts, namely, 1219.

All the estimations in this paper were conducted using Scilab 4.1. In terms of model specification, all models estimated in this paper were assumed to follow an ARMA process in the conditional mean with conditional heteroskedastic errors following either GARCH or GJR process with $r = p = q = 1$ and $s = 0$. The parameter estimates for both GARCH and GJR models under different distributional assumptions can be found in Table 1.

As shown in Table 1, the parameter estimates under different distribution assumptions are very similar, with the exception of ARMA-GARCH model under the Student-t distribution where convergence could not be reached. Since estimates can not be obtained for this particular case, the evaluation on the VaR performance will not be conducted for the ARMA-GARCH model with the Student-t distribution.

	GARCH			GJR		
	Normal	Student-t*	GED	Normal	Student-t	GED
ϕ_0	0.048 (3.631)	0.060 (5.404)	0.002 (15.000)	0.030 (3.023)	0.043 (3.763)	0.029 (2.760)
ϕ_1	0.012 (0.705)	-0.018 (-3.357)	0.007 (17.450)	0.023 (1.429)	0.003 (0.275)	0.003 (2.742)
ω	0.005 (1.99)	0.034 (5.616)	0.003 (1.728)	0.011 (2.421)	0.007 (2.328)	0.008 (2.717)
α_1	0.045 (3.721)	0.530 (22.160)	0.042 (4.238)	0.004 (0.583)	0.007 (1.331)	0.006 (1.228)
β_1	0.950 (70.538)	0.935 (65.642)	0.956 (90.913)	0.937 (65.108)	0.944 (76.572)	0.943 (79.351)
γ_1				0.094 (4.427)	0.085 (3.901)	0.087 (4.211)
Shape Parameters		2.116	1.627 (21.971)		6.336 (8.519)	1.571 (19.931)
Second Moment	0.995		0.998	0.988	0.994	0.993

* No convergence was achieved for the ARMA(1,0)-GARCH(1,1) under the Student-t distribution.

** The robust t-ratios are in the parenthesis.

Table 1 Parameter Estimates for ARMA-GARCH and ARMA-GJR under Different Distributional Assumptions

	GARCH			GJR		
	MED	Normal	GED	Normal	Student-t	GED
κ	-2.723	-2.326	-2.628	-2.326	-3.628	-2.605
No of Violation	13	31	15	20	2	9
Percentage	1.067%	2.543%	1.231%	1.641%	0.164%	0.738%

Table 2 Out-of-Sample VaR Forecasts Performance

Table 2 contains the calculated κ under each distribution as well as the out-of-sample performance of the VaR forecasts for each model with different distribution. The percentage of violation, that is, the ratio between the numbers of violations relative to the total number of forecasts should be close to the significance level, which is 1%. As shown in Table 2, MED performed the best, follows by GARCH with GED then GJR with Normal distribution. Interestingly, both GJR with Student-t and GJR with GED produced extremely conservative VaR forecasts which explained their low percentages. Overall, the results support the importance of distribution in forecasting VaR through the calculation of κ and that approximating the distribution of the standardized residuals using Maximum Entropy Density seemed to be a feasible method to improve VaR forecast.

4. CONCLUSION

This paper proposed a new method to forecast VaR threshold using Maximum Entropy Density (MED). Given the results in Ling and McAleer (2003), consistent estimates of the standardized residuals in an ARMA-GARCH model can be obtained by using the QMLE with normal density. The distribution of the standardized residuals can then be approximated using MED with the estimated standardized residuals. This method allows other characteristics of the underlying distribution, such as excess kurtosis, to be accommodated and thus produce a more accurate forecast for VaR thresholds. Empirical evidence suggested the proposed method is promising in make more accurate VaR forecasts.

ACKNOWLEDGEMENT

The author would like to thank Hiroaki Suenaga for insightful discussions and Sylvia Soltyk for excellent research assistance. Financial supports from the School of Economics and Finance, Curtin University of Technology are gracefully acknowledged.

REFERENCES

Basel Committee on Banking Supervision (1988). Interntional convergence of capital measurement and capital standards, BIS, Basel, Switzerland.

Basel Committee on Banking Supervision (1995). An Internal Model-Based Approach to Market Risk Capital Requirements, BIS, Basel, Switzerland

Basel Committee on Banking Supervision (1996). Supervisory framework for the use of backtesting, BIS, Basel, Switzerland.

Bollerslev, T. (1986). "Generalized Autoregressive Conditional Heteroskedasticity." Journal of Econometrics **31**: 307-327.

Bollerslev, T. (1987). "A conditionally heteroskedastic time series model for speculative prices and rates of return." The Review of Economics and Statistics **69**(3): 542-547.

Chan, F. (2009). "Modelling time-varying higher moments with maximum entropy density." to appear in Mathematics and Computers in Simulation **doi:10.1016/j.matcom.2008.11.016**.

Engle, R. F. (1982). "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation." Econometrica **50**: 987-1007.

Glosten, L., R. Jagannathan and D. Runkle (1992). "On the Relation Between the Expected Value and Volatility of Nominal Excess Return on Stocks." Journal of Finance **46**: 1779-1801.

Jaynes, E. T. (1957a). "Information theory and statistical mechanics." The Physical Review **106**(4): 620-630.

Jaynes, E. T. (1957b). "Information theory and statistical mechanics II." The Physical Review **108**(2): 171-190.

Jorion, P. (2000). Value at Risk: The New Benchmark for Managing Financial Risk, McGraw-Hill, New York.

Ling, S. and M. McAleer (2003). "Asymptotic Theory for a Vector ARMA-GARCH Model." Econometric Theory **19**: 278-308.

McAleer, M., F. Chan and D. Marinova (2007). "An Econometric Analysis of Asymmetric Volatility: Theory and Application to Patents." Journal of Econometrics **139**: 259 - 284.

Rockinger, M. and E. Jondeau (2002). "Entropy densities with an application to autoregressive conditional skewness and kurtosis." Journal of Econometrics **106**: 119-142.

Shannon, C. (1948). "The mathematical theory of communication " Bell Systems Technical Journal **27**: 349-423.