



Munich Personal RePEc Archive

A Comparison of VaR Estimation Procedures for Leptokurtic Equity Index Returns

Bhattacharyya, Malay and Madhav R, Siddarth

Indian Institute of Management Bangalore

2012

Online at <https://mpra.ub.uni-muenchen.de/54189/>
MPRA Paper No. 54189, posted 07 Mar 2014 19:26 UTC

A Comparison of VaR Estimation Procedures for Leptokurtic Equity Index Returns*

Malay Bhattacharyya¹, Siddarth Madhav R²

¹Indian Institute of Management Bangalore, Bangalore, India

²Barclays Capital, New York, USA

Email: malayb@iimb.ernet.in, rsmadhav@gmail.com

Received July 11, 2011; revised August 1, 2011; accepted August 29, 2011

ABSTRACT

The paper presents and tests Dynamic Value at Risk (VaR) estimation procedures for equity index returns. Volatility clustering and leptokurtosis are well-documented characteristics of such time series. An ARMA (1, 1)-GARCH (1, 1) approach models the inherent autocorrelation and dynamic volatility. Fat-tailed behavior is modeled in two ways. In the first approach, the ARMA-GARCH process is run assuming alternatively that the standardized residuals are distributed with Pearson Type IV, Johnson S_U , Manly's exponential transformation, normal and t -distributions. In the second approach, the ARMA-GARCH process is run with the pseudo-normal assumption, the parameters calculated with the pseudo maximum likelihood procedure, and the standardized residuals are later alternatively modeled with Mixture of Normal distributions, Extreme Value Theory and other power transformations such as John-Draper, Bickel-Doksum, Manly, Yeo-Johnson and certain combinations of the above. The first approach yields five models, and the second approach yields nine. These are tested with six equity index return time series using rolling windows. These models are compared by computing the 99%, 97.5% and 95% VaR violations and contrasting them with the expected number of violations.

Keywords: Dynamic VaR; GARCH; EVT; Johnson S_U ; Pearson Type IV; Mixture of Normal Distributions; Manly; John Draper; Yeo-Johnson Transformations

1. Introduction

VALUE AT RISK (VaR) is a popular measure of risk in a portfolio of assets. It represents a high quantile of loss distribution for a particular horizon, providing a loss threshold that is exceeded only a small percentage of the time.

Traditional methods of calculating VaR include historical simulation and the analytic variance-covariance approach. However, these models fall short when tested against actual market conditions. The historical simulation approach assumes constant volatility of stocks over an extended period of time. It fails to account for the phenomenon of volatility clustering, when periods of high and low volatility occur together. This leads to underestimation of VaR during periods of high volatility, and overestimation in times of calm. The analytic variance-covariance approach assumes that returns are jointly normally distributed. However, the fat-tailed non-normal behaviour of returns would mean that this methodology tends to underestimate VaR as well.

Fama [1] and Mandelbrot [2] report the failure of the

*This work was carried out when Siddarth Madhav R was a graduate student at the Indian Institute of Management Bangalore.

normal distribution to model asset returns, sparking a slew of papers addressing the issue of accurately modeling leptokurtic time series with volatility clustering. The approaches can be roughly divided in two, the first assuming that returns are independent and modeling unconditional distribution of returns. In this approach, numerous distributions have been proposed, Fama [1] and Mandelbrot [2] use the stable Paretian distribution, Blattberg and Gonedes [3] suggest the use of Student t -distribution. The mixture of normal distributions is used by Ball and Torous [4] and Kon [5] and the logistic distribution, the empirical power distribution and the Student t -distributions have been compared by Gray and French [6]. The Pearson type IV distribution is used by Bhattacharyya, Chaudhary and Yadav [7] for dynamic VaR estimation and by Bhattacharyya, Misra and Kodase [8] for dynamic MaxVaR estimation. Bhattacharyya and Ritolia [9] use EVT for dynamic VaR estimation.

The second approach considers returns to be serially correlated and uses conditional variance models or stochastic volatility models to model asset returns. Engle [10] and Bollerslev [11] use ARCH and GARCH models to account for volatility clustering. GARCH models have

been shown to be more suited to this purpose by various studies such as Poon and Granger [12]. The GARCH (1, 1) model performs well for most stock returns and this paper adopts this approach.

The following model has been extensively used to model dynamism in forecasts of returns and volatility of returns.

$$X_t = \mu_t + \sigma_t Z_t \tag{1}$$

where X_t is the actual return on day t , μ_t is the expected return on day t , σ_t is the volatility estimate on day t and Z_t is the standardized residual, having a normal distribution with zero mean and unit standard deviation.

ARMA processes are useful for modeling μ_t , the predicted mean of the time series data, and GARCH processes are good models for σ_t , the predicted volatility. However, the inherent leptokurtic behaviour of asset returns makes the ARMA-GARCH model insufficient for the purpose of calculating VaR.

In this paper, ARMA (1, 1) model is used for the calculation of predicted mean and GARCH (1, 1) model is used for modeling the observed volatility clustering. Models are developed using two approaches. In the first one, consisting of five models, ARMA-GARCH model parameters are calculated assuming that standardized residuals alternatively follow Pearson Type IV distribution, Johnson S_U distribution, Manly's exponential transformation, normal and Student t -distributions. In the second approach, the ARMA-GARCH parameters are calculated using the pseudo-normal assumption, *i.e.*, assuming that standardized residuals are normally distributed, and they are later modeled using the mixture of normal distributions, Extreme Value Theory, and other power transformations such as John-Draper, Bickel-Doksum, Manly, Yeo-Johnson and certain combinations of the above. The second approach yields nine models.

While developing and testing VaR models, the authors find it important to develop those that are applicable in real world scenarios. This translates to certain simplicity in execution and fast run-times for calculations, as time can be a critical issue. At the same time, the importance of creating an accurate measure of risk cannot be understated, given how the stock market crash of 2008 bankrupted firms and individuals alike, and sent the world spiraling into recession.

2. Leptokurtic Density Functions

2.1. Pearson Type IV Distribution

The Pearson family of curves, a generalized family of frequency curves developed by Karl Pearson, embodies a wide range of commonly observed distributions. The Pearson curves are a solution to the differential equation

$$\frac{1}{f(x)} \frac{df(x)}{dx} = \frac{x - \alpha}{c_0 + c_1x + c_2x^2} \tag{2}$$

The system of curves which arise from the above differential equation cover a wide spectrum of skewness and kurtosis (**Figure 1**). The type of distribution obtained post-integration is dictated by the roots of the quadratic equation $c_0 + c_1x + c_2x^2 = 0$.

The Type IV curve is obtained when the roots of the quadratic equation $c_0 + c_1x + c_2x^2 = 0$ are complex, *i.e.*, when $c_1^2 < 4c_0c_2$. It is suitable for those distributions which have high excess kurtosis and moderate skewness. Financial return data fall in this category. The probability density function (PDF) of the Type IV curve (Heinrich, [13]) is

$$f(x) = k \left[1 + \left(\frac{x - \lambda}{a} \right)^2 \right]^{-m} \exp \left[-v \tan^{-1} \left(\frac{x - \lambda}{a} \right) \right] \tag{3}$$

where λ , a , v and m are real parameters (functions of α , c_0, c_1 and c_2), $m > 1/2$, $-\infty < x < \infty$ and k is a norma-



Figure 1. The diagram of the Pearson curve family. It shows the type of curve to be used for each range of skewness and kurtosis. The x-axis is $\beta_1 = \text{skewness}^2$, and the y-axis is β_2 , the traditional kurtosis.

lizing constant, dependent on λ, a, ν and m .

The PDF gives rise to a bell shaped curve, where λ is the location parameter, a is the scale parameter, ν and m can be interpreted as the skewness and kurtosis parameters respectively.

The type of Pearson curve to use for a particular situation is dictated by the skewness and kurtosis. **Table 1** shows the observed skewness and excess kurtosis for the six equity indices. Cross-referencing them with **Figure 1**, we can see that Pearson Type IV curve is the model to be used.

For a standardized Pearson Type IV curve, *i.e.*, with zero mean and unit standard deviation, we need to add the following constraints.

$$a_s = \sqrt{\frac{r^2(r-1)}{r^2 + \nu^2}} \tag{4}$$

$$\lambda_s = a\nu/r \tag{5}$$

2.2. Johnson S_U Distribution

The Johnson family of distributions (Johnson, [14]) consists of three distributions, which cover all possible average, standard deviation, skewness and kurtosis values, excluding the impossible region. These consist of the S_U, S_B and the lognormal curves. The transformations have the general form

$$Z = \gamma + \delta \cdot g\left(\frac{X - \xi}{\lambda}\right) \tag{6}$$

where the transformation parameters ξ is the location, λ is the scale and γ and δ are shape parameters. Z is the resulting normal distribution. $g(\cdot)$ is one of the following functions:

$$g(y) = \begin{cases} \ln(y) & \text{Lognormal distribution} \\ \sinh^{-1} y & \text{S}_U \text{ distribution} \\ \ln(y/(1-y)) & \text{S}_B \text{ distribution} \\ y & \text{Normal distribution} \end{cases}$$

Since we are modeling the innovations of the ARMA-

ARCH process, we require a transformation function which can accept arguments that may be positive or negative. Hence we need to use the Johnson S_U distribution, as the sine hyperbolic inverse function has a domain all over the real line.

So we have

$$Z = \gamma + \delta \cdot \sinh^{-1}\left(\frac{X - \xi}{\lambda}\right) \tag{7}$$

where γ and δ are assumed to be positive.

The density function of Johnson S_U distribution can be easily found in closed-form from variable transformation:

$$f(x; \gamma, \delta; \xi; \lambda) = \frac{\delta}{\lambda \sqrt{1 + \left(\frac{x - \xi}{\lambda}\right)^2}} \phi\left[\gamma + \delta \sinh^{-1}\left(\frac{x - \xi}{\lambda}\right)\right] \tag{8}$$

where $x \in R$, ϕ is the density function of $N(0,1)$, ξ and $\lambda > 0$ are location and scale parameters respectively, γ can be interpreted as a skewness parameter, and $\delta > 0$ can be interpreted as a kurtosis parameter. The distribution is positively or negatively skewed according to whether γ is negative or positive. Holding γ constant and increasing δ reduces the kurtosis. However, γ and δ cannot be viewed purely as skewness or kurtosis parameters, respectively. The mean and the variance of Johnson S_U distribution are given as:

$$\mu = \xi + \lambda \omega^{\nu^2} \sinh \Omega \tag{9}$$

$$\sigma^2 = \frac{\lambda^2}{2} (\omega - 1) (\omega \cosh 2\Omega + 1) \tag{10}$$

where $\omega = \exp(\delta^{-2})$ and $\Omega = \gamma/\delta$.

2.3. Extreme Value Theory

Extreme value theory provides a framework to formalize the study of behavior in the tails of a distribution. According to the Fisher-Tippet theorem, there can be three possible extreme value distributions for the standardized variable.

Table 1. Comparison of moments for each stock index return series.

Index	Sensex	NIFTY	DJI	FTSE	HSI	Nikkei
Observations	1500	1500	1500	1500	1500	1500
Dates	Mar 03 - Feb 09	Mar 03 - Feb 09	Mar 03 - Feb 09	Mar 03 - Feb 09	Mar 03 - Feb 09	Mar 03 - Feb 09
Mean	0.0009	0.0009	0.0001	0.0002	0.0004	0.0001
Std. Deviation	0.0178	0.0181	0.0124	0.0127	0.0169	0.0163
Skewness	-0.4276	-0.5130	0.2624	0.1409	0.3876	-0.2730
Kurtosis	7.2358	8.6112	17.3926	14.5248	15.5344	12.8085

2.3.1. Gumbel Distribution

As with the normal and gamma distributions, the tail can be unbounded, have finite moments and decay exponentially. The distribution function is given by:

$$G(x) = \exp(e^{-x}) \quad \text{for } -\infty < x < \infty \quad (11)$$

2.3.2. Frechet Distribution

The tail can be unbounded, and decay by a power as with the Cauchy and Student t -distribution. The distribution function is given by

$$G(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ \exp(-x^{-\alpha}) & \text{for } x > 0 \end{cases} \quad (12)$$

Moments exist only up to the integer part of α , higher moments do not exist, as the tails are fat, they are not integrable when weighted by tail probabilities.

2.3.3. Weibull Distribution

The tails are constant-declining, and all moments exist. They are thin, and have upper bounds. The distribution function is:

$$G(x) = \begin{cases} \exp(-(-x)^\alpha) & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases} \quad (13)$$

Now, since the financial returns data are fat-tailed and unbounded, we must clearly use the Frechet distribution for modeling extreme value distributions.

2.3.4. Generalized Extreme Value Distribution

The Generalized Extreme Value Distribution (GEVD) unifies the above three distributions. Here the tail index (τ) is the inverse of the shape parameter (α). In this equation given below, if $\tau = 0$, it is a Gumbel distribution, if $\tau < 0$, it is a Frechet distribution else if $\tau > 0$ it is a Weibull distribution.

$$F_x(x) = \begin{cases} \exp(-(1+\tau x)^{-1/\tau}) & \text{for } \tau \neq 0 \\ \exp(-e^{-x}) & \text{for } \tau = 0 \end{cases} \quad (14)$$

To build the series of maxima or minima, there are two methods:

2.3.5. Block Maxima

This approach consists of splitting the series into equal non-overlapping blocks. The maximum from each block is extracted and used to model the extreme value distribution. As volatility clustering is a well observed phenomenon in financial data, very high or very low observations tend to occur together. Thus, this technique runs the risk of losing extreme observations.

2.3.6. Peak over Threshold

The second approach consists of sampling maxima by selecting those that exceed a chosen threshold. A low threshold would give rise to a larger number of observations, running the risk of including central observations in the extremes data. The tail index computed has lesser variance but is subject to bias. A high threshold has few observations, and the tail index is more imprecise, but unbiased. The choice of the threshold is thus a trade-off between variance and bias. For the analysis in this paper, we use the Peak over Threshold method.

2.4. Mixture of Normal Distributions

The mixture of normal distributions, used to model fat-tailed distributions, assumes that each observation is generated from one of N normal distributions. The probability that it is generated from a distribution " i " is p_i , with

$$\sum_{i=1}^N p_i = 1.$$

The resultant density function

$$\eta(x; p_1\phi_1 p_2\phi_2 \dots p_N\phi_N) = \sum_{i=1}^N p_i\phi_i(x) \quad (15)$$

where ϕ_i is a normal distribution with mean μ_i and standard deviation σ_i . For the special case of $N = 2$, we have

$$\eta(x; \theta) = p\phi_1(x) + (1-p)\phi_2(x) \quad (16)$$

where $\theta = (p, \mu_1, \mu_2, \sigma_1, \sigma_2)$ is the parameter vector.

For a mixture of N normal distributions, the first four moments are:

$$\mu_\eta = \sum_{i=1}^N p_i\mu_i \quad (17)$$

$$\sigma_\eta^2 = \sum_{i=1}^N p_i\sigma_i^2 + \sum_{i=1}^N p_i\mu_i^2 - \mu_\eta^2 \quad (18)$$

$$\tau_\eta = \frac{1}{\sigma_\eta^3} \left[\sum_{i=1}^N (3\mu_i\sigma_i^2 + \mu_i^3) - 3\mu_\eta\sigma_\eta^2 - \mu_\eta^3 \right] \quad (19)$$

$$\kappa_\eta = \frac{1}{\sigma_\eta^4} \left[\sum_{i=1}^N p_i (3\sigma_i^4 + 6\mu_i^2\sigma_i^2 + \mu_i^4) - 4\tau_\eta\mu_\eta\sigma_\eta^3 - 6\mu_\eta^2\sigma_\eta^2 - \mu_\eta^4 \right] \quad (20)$$

A mixture of more than two normal distributions may provide a better fit to the series, but Tucker [15] reports that the improvement by increasing the number of normal distributions in the mixture from two is not too significant. Estimation of parameters for the mixture of normal distribution is problematic. This is because, although we have a well defined distribution function in a closed form, using maximum likelihood techniques for parameter estimation leads to convergence issues (Hamilton, [16]).

Using method of moments is another option, but even for the simplest case of $N = 2$, we need five moment equations to find the five parameters,

$p, \mu_1, \mu_2, \sigma_1, \sigma_2$, and there may not be a solution at all (Titterton, Smith and Makov, [17]). Alternate methods have been suggested, such as fractile-to-fractile comparisons (Hull and White, [18]) and Bayesian updating schemes (Zangari, [19]).

This paper uses the fractile-to-fractile comparison technique along with a simplifying assumption that one of the means of the mixture of normal distributions is zero. This is a reasonable assumption, in the data set, as most observations (about 95%) lie in the zero-mean normal distribution, and it simplifies calculations considerably.

2.5. Power Transformations

Box and Cox [20] propose one of the first power transformations converting a non-normal distribution into a normal one. In its original form, the transformation function is:

$$y(\lambda) = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \text{if } \lambda \neq 0 \\ \log y, & \text{if } \lambda = 0 \end{cases} \quad (21)$$

However, as it can be seen, the power transformation cannot be applied to negative values of y . Since then, many modifications of the original Box-Cox power transformation have been proposed.

2.5.1. Manly’s Exponential Distribution

Manly [21] proposed the exponential distribution given below.

$$y(\lambda) = \begin{cases} \frac{e^{\lambda y} - 1}{\lambda}, & \text{if } \lambda \neq 0 \\ y, & \text{if } \lambda = 0 \end{cases} \quad (22)$$

Negative values of y are permitted. This transformation is useful for transforming skewed distributions to normal (Li, [22]).

2.5.2. Bickel-Doksum Transformation

Bickel and Doksum [23] transform the original Box-Cox transformation to

$$y(\lambda) = \frac{|y|^\lambda \text{sign}(y) - 1}{\lambda}, \text{ for } \lambda > 0 \quad (23)$$

where

$$\text{sign}(y) = \begin{cases} +1, & \text{if } y \geq 0 \\ -1, & \text{if } y < 0 \end{cases} \quad (24)$$

The addition of the *sign* function makes this transformation compatible for negative values of y as well.

2.5.3. John-Draper Modulus Transformation

John and Draper [24] propose the modulus transformation given below:

$$y(\lambda) = \begin{cases} \text{sign}(y) \frac{(|y|+1)^\lambda - 1}{\lambda}, & \text{if } \lambda \neq 0 \\ \text{sign}(y) \log(|y|+1), & \text{if } \lambda = 0 \end{cases} \quad (25)$$

where

$$\text{sign}(y) = \begin{cases} +1, & \text{if } y \geq 0 \\ -1, & \text{if } y < 0 \end{cases} \quad (26)$$

The modulus transformation works best on those distributions which are approximately symmetric about some central point (Li, [22]). It reduces the kurtosis of the series, while introducing some degree of skewness to a symmetric distribution.

2.5.4. Yeo-Johnson Transformation

Yeo and Johnson [25] propose the following transformation in 2000:

$$y(\lambda) = \begin{cases} \frac{(y+1)^\lambda - 1}{\lambda}, & \lambda \neq 0, y \geq 0 \\ \log(y+1), & \lambda = 0, y \geq 0 \\ \frac{(1-y)^{2-\lambda} - 1}{\lambda - 2}, & \lambda \neq 2, y < 0 \\ -\log(1-y), & \lambda = 2, y < 0 \end{cases} \quad (27)$$

In their original paper, Yeo and Johnson [25] find the value of λ by minimizing the Kullback-Leibler distance between the normal and transformed distributions. In this paper however, we have found λ by maximizing log-likelihoods. This transformation, like Manly, reduces skewness of the distribution and makes the transformed variable more symmetric.

3. Dynamic VaR Models

This section describes the methods used to calculate dynamic Value at Risk for equity index returns.

3.1. Model for Conditional Mean and Variance

To calculate conditional mean μ_t given the time series data until time $t - 1$, we use an ARMA (1, 1) process.

$$X_t = C + \phi_1 X_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t \quad (28)$$

We use the GARCH (1,1) process to model the volatility of the innovation term.

$$\sigma_t^2 = K + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \quad (29)$$

3.2. Models for Innovations

In Equation (1), the forecasted mean and variance are calculated by an ARMA (1, 1)-GARCH (1, 1) model. As mentioned in the introduction, there are two approaches followed to model innovations. In the first approach, ARMA (1, 1)-GARCH (1, 1) model parameters are calculated assuming that standardized residuals alternatively follow Pearson Type IV distribution, Johnson S_U distribution, Manly's exponential transformation, normal and Student t -distributions. In the second approach, ARMA (1, 1)-GARCH (1, 1) parameters are calculated assuming that standardized residuals are normally distributed. The extracted standardized residuals are then modeled using the mixture of normal distributions, Extreme Value Theory, and other power transformations such as John-Draper, Bickel-Doksum, Manly, Yeo-Johnson and certain combinations of the above.

Method 1

The first approach consists of five models, whose designs are outlined below.

Model 1.1 GARCH-N Model

In Equation (1), Z_t is assumed to be a standard normal distribution. Therefore, the innovations term, ϵ_t , has zero mean and the standard deviation of h_t .

$$Z_t \sim N(0,1) \Rightarrow \epsilon_t \sim N(0, h_t) \tag{30}$$

$$f(\epsilon_t|F_{t-1}) = \frac{1}{\sqrt{2\pi h_t}} \exp\left(-\frac{\epsilon_t^2}{2h_t}\right) \tag{31}$$

Therefore, the log likelihood function, which is maximized to find the parameters of the ARMA-GARCH model for the series of length T is given by

$$LLF = -\frac{1}{2} \sum_{t=1}^T \left[\log(2\pi h_t) + \frac{\epsilon_t^2}{2h_t} \right] \tag{32}$$

The maximum likelihood estimates for the ARMA (1, 1)-GARCH (1, 1) parameters are found by minimizing the negative of the above function using the *fmincon* function in MATLAB.

Model 1.2 GARCH-t Model

In Equation (1), Z_t is assumed to be a Student t -distribution with zero mean and unit standard deviation. Therefore, the log likelihood function, the logarithm of the density function of the innovations term, ϵ_t , for the series of length T is given by

$$LLF = \sum_{t=1}^T \left\{ \log \left[\frac{\Gamma\left[\frac{\nu+1}{2}\right]}{\Gamma\left[\frac{\nu}{2}\right] \sqrt{\pi(\nu-2)}} \right] - \frac{1}{2} \log(h_t) - \frac{(\nu+1)}{2} \log \left[1 + \frac{\epsilon_t^2}{h_t(\nu-2)} \right] \right\} \tag{33}$$

where ν represents the degrees of freedom in the t -distribution.

The maximum likelihood estimates for the ARMA (1, 1) GARCH (1, 1) parameters are found by minimizing the negative of the above function using the *fmincon* function in MATLAB.

Model 1.3 GARCH-PIV Model

In Equation (1), Z_t is assumed to be a Pearson Type IV distribution. The standardized innovations series has unit variance, but not necessarily a zero mean. This was justified by Newey and Steigerwald [26], who proved that an additional location parameter is needed to satisfy the identification condition for the consistency of parameter estimates when conditional innovation distribution in the GARCH model is asymmetric. Hence Equation (4) holds, but Equation (5) does not. Therefore,

$$E(X_t|F_{t-1}) = \mu_t + \sqrt{h_t} \left(\lambda - \frac{a_s \nu}{r} \right) \tag{34}$$

Hence, for modeling innovations, we need to change the location and scale parameters to $\lambda\sqrt{h_t}$ and $a_s\sqrt{h_t}$ respectively. The normalizing parameter is inversely proportional to the scale parameter, so it changes to $k/\sqrt{h_t}$.

$$Z_t \sim PIV(k, m, \nu, a_s, \lambda) \Rightarrow \epsilon_t \sim PIV\left(k/\sqrt{h_t}, m, \nu, a_s\sqrt{h_t}, \lambda\sqrt{h_t}\right) \tag{35}$$

The distribution function of the innovation series is given by

$$f(\epsilon_t|F_{t-1}) = \frac{k}{\sqrt{h_t}} \left[1 + \left(\frac{\epsilon_t - \lambda\sqrt{h_t}}{a_s\sqrt{h_t}} \right)^2 \right]^{-m} \exp \left[-\nu \tan^{-1} \left(\frac{\epsilon_t - \lambda\sqrt{h_t}}{a_s\sqrt{h_t}} \right) \right] \tag{36}$$

The log likelihood function to be maximized is given by

$$LLF = \sum_{t=1}^T \left\{ \log k - \frac{1}{2} \log h_t - \left(\frac{r+2}{2} \right) \log \left[1 + \frac{(\epsilon_t - \lambda\sqrt{h_t})^2 (r^2 + \nu^2)}{r^2 (r-1) h_t} \right] - \nu \tan^{-1} \left(\frac{\epsilon_t - \lambda\sqrt{h_t} \sqrt{r^2 + \nu^2}}{\sqrt{r^2 (r-1) h_t}} \right) \right\} \tag{37}$$

We use Equation (4) and the relation $r = 2(m-1)$ to write a_s and m in terms of r . The log likelihood function is maximized (by minimizing $-LLF$) using the *fmin*

con function in MATLAB. The maximum likelihood estimates from the GARCH-N model and the Pearson Type IV parameters calculated from the first four moments of the resulting standardized innovations series (under the pseudo-normal assumption) are used as initial estimates for the optimization function. The normalizing constant k is computed by the technique used by Heinrich [13].

Model 1.4 GARCH-JSU Model

In this model, the standardized innovations in Equation (1), Z_t is assumed to be a Johnson S_U distribution. As with the GARCH-PIV model, the standardized innovations have unit variance, but not necessarily zero mean. Therefore, from Equation (10), the scale parameter λ is constrained.

$$\lambda_s = \sqrt{\frac{2}{(\omega-1)(\omega \cosh \cosh 2\Omega + 1)}} \tag{38}$$

where $\omega = \exp(\delta^{-2})$ and $\Omega = \gamma/\delta$.

Note that Equation (9) does not hold, and the parameter ξ has to be estimated during optimization. The predicted future value of the time series is given by

$$E(X_t|F_{t-1}) = \mu_t + \sqrt{h_t} (\xi + \lambda \omega^{1/2} \sinh \Omega) \tag{39}$$

Now, for modeling the innovations series ϵ_t , the location and scale parameters must be changed to $\xi\sqrt{h_t}$ and $\lambda_s\sqrt{h_t}$.

$$\begin{aligned} Z_t &\sim JSU(\gamma, \delta, \xi, \lambda_s) \\ \Rightarrow \epsilon_t &\sim JSU(\gamma, \delta, \xi\sqrt{h_t}, \lambda_s\sqrt{h_t}) \end{aligned} \tag{40}$$

$$\begin{aligned} f(\epsilon_t|F_{t-1}) &= \frac{\delta}{\lambda_s\sqrt{h_t} \sqrt{1 + \left(\frac{\epsilon_t - \xi\sqrt{h_t}}{\lambda_s\sqrt{h_t}}\right)^2}} \phi \\ &\left[\gamma + \delta \sinh^{-1} \left(\frac{\epsilon_t - \xi\sqrt{h_t}}{\lambda_s\sqrt{h_t}} \right) \right] \end{aligned} \tag{41}$$

where $\phi \sim N(0,1)$.

The log likelihood function to be maximized is given by

$$\begin{aligned} LLF = \sum_{t=1}^T &\left\{ \log \delta - \log \lambda_s - \frac{1}{2} \log 2\pi h_t \right. \\ &\left. - \frac{1}{2} \log \left(1 + \eta_t^2 \right) - \frac{1}{2} \left(\gamma + \delta \sinh^{-1} \eta_t \right)^2 \right\} \end{aligned} \tag{42}$$

The maximum likelihood estimates are calculated by minimizing the negative of the above function using the *fmincon* function in MATLAB.

Model 1.5 GARCH-Manly Model

In this model, the standardized innovations in Equation (1), it is assumed that when Z_t is put through Manly's exponential transformation (Equation (22)), it becomes nor-

mally distributed. Assuming that the transformed normal function has zero mean and unit standard deviation, Z_t has the following closed form probability distribution function

$$f(Z_t|F_{t-1}) = \frac{\sqrt{2/\pi}}{1 + \operatorname{erf}\left(\frac{1}{\lambda\sqrt{2}}\right)} \exp\left[\lambda Z_t - \frac{M(\lambda, Z_t)^2}{2}\right] \tag{43}$$

where $M(\lambda, Z_t)$ is the exponentially transformed (Equation (22)) value of Z_t and *erf* is the error function.

Therefore the standardized innovations (ϵ_t) have the following distribution

$$\begin{aligned} f(\epsilon_t|F_{t-1}) &= \frac{\sqrt{\frac{2}{\pi}}}{h_t \left(1 + \operatorname{erf}\left(\frac{1}{\lambda\sqrt{2}}\right) \right)} \\ &\exp\left\{ -\frac{1}{2} \left(\frac{\exp\left(\frac{\lambda x}{h_t}\right) - 1}{\lambda} \right)^2 \right\} \exp(\lambda x/h_t) \end{aligned} \tag{44}$$

The log likelihood function to be maximized is given by

$$\begin{aligned} LLF = \sum_{t=1}^T &\left\{ -\log \left(\frac{2}{1 + \operatorname{erf}\left(\frac{1}{\lambda\sqrt{2}}\right)} \right) \right. \\ &\left. - \frac{1}{2} \log 2\pi h_t^2 + \lambda \frac{\epsilon_t}{h_t} - \frac{M\left(\lambda, \frac{\epsilon_t}{h_t}\right)^2}{2} \right\} \end{aligned} \tag{45}$$

The maximum likelihood estimates are calculated by minimizing the negative of the above function using the *fmincon* function in MATLAB. The above equations are derived in detail in the Appendix.

Method 2

The second approach consists of nine models, whose designs are outlined below.

Model 2.1 GARCH-EVT Model

In this model, the ARMA (1, 1)-GARCH(1, 1) parameters are found under the pseudo-normal assumption, i.e., that the standardized innovations in Equation (1) Z_t is a standard normal function. Now, the assumption made is that the values of Z_t considered for calculation of VaR, i.e., the 99th, 97.5th and 95th percentiles are part of an extreme value distribution. This assumption is theoretically justified, as the ARMA-GARCH process gets rid

of the serial correlation between terms, and the Fisher-Tippet theorem is applicable.

We use the Peak over Threshold (POT) method to observe the number of values which exceed a high threshold. The distribution of conditional excess losses over a certain high threshold follows a Generalized Pareto Distribution (GPD).

$$G_{\xi, \beta}(x) = \begin{cases} 1 - \left(1 + \frac{\xi x}{\beta}\right)^{-1/\xi}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{x}{\beta}\right), & \xi = 0 \end{cases} \quad (46)$$

where ξ is the shape parameter (positive in our specific case, as this yields a heavy tailed GPD) and β is the scaling parameter.

The formula for conditional excess losses above the threshold u (We consider the negative of the return series, thereby ensuring that the threshold is positive, and mean excess return is positive) is given by

$$F_u(y) = P(X < y + u | X > u) \quad (47)$$

$$F_u(y) = (F(y + u) - F(u)) / (1 - F(u)) \quad (48)$$

Since $F_u(y)$ is a GPD with positive ξ , we need to back-calculate $F(y + u)$. $F(u)$ is given by N_u/N , where N is the total number of observations and N_u

is the number of observations above the threshold u . Therefore, the tail estimator becomes

$$F(x) = 1 - \frac{N_u}{N} \left(1 + \xi \frac{x - u}{\beta}\right)^{-\frac{1}{\xi}}, \text{ for } x > u \quad (49)$$

$$VaR_q = u + \frac{\beta}{\xi} \left\{ \left(\frac{N}{N_u} (1 - q) \right)^{-\xi} - 1 \right\} \quad (50)$$

The Value at Risk is now calculated by the formula

$$VaR_t^q = \mu_t + \sigma_t VaR_q \quad (51)$$

Choosing the threshold to be used in the calculations is a subjective process. In this paper, we calculate the mean excess returns for various values of thresholds and plot them. For a GPD, the mean excess return is given by:

$$e(u) = \frac{\sigma + \xi u}{1 - \xi} \quad (52)$$

The threshold is calculated by observing the graphs and identifying the point from which the conditional excess return increases linearly with the threshold values. It is possible to consider any larger value as a threshold as well, but this way, the maximum number of data points gets accommodated in the extreme value distribution, thus reducing the variance of the obtained parameters. In **Figures 2(a)** and **(b)**, we observe that the thresh old

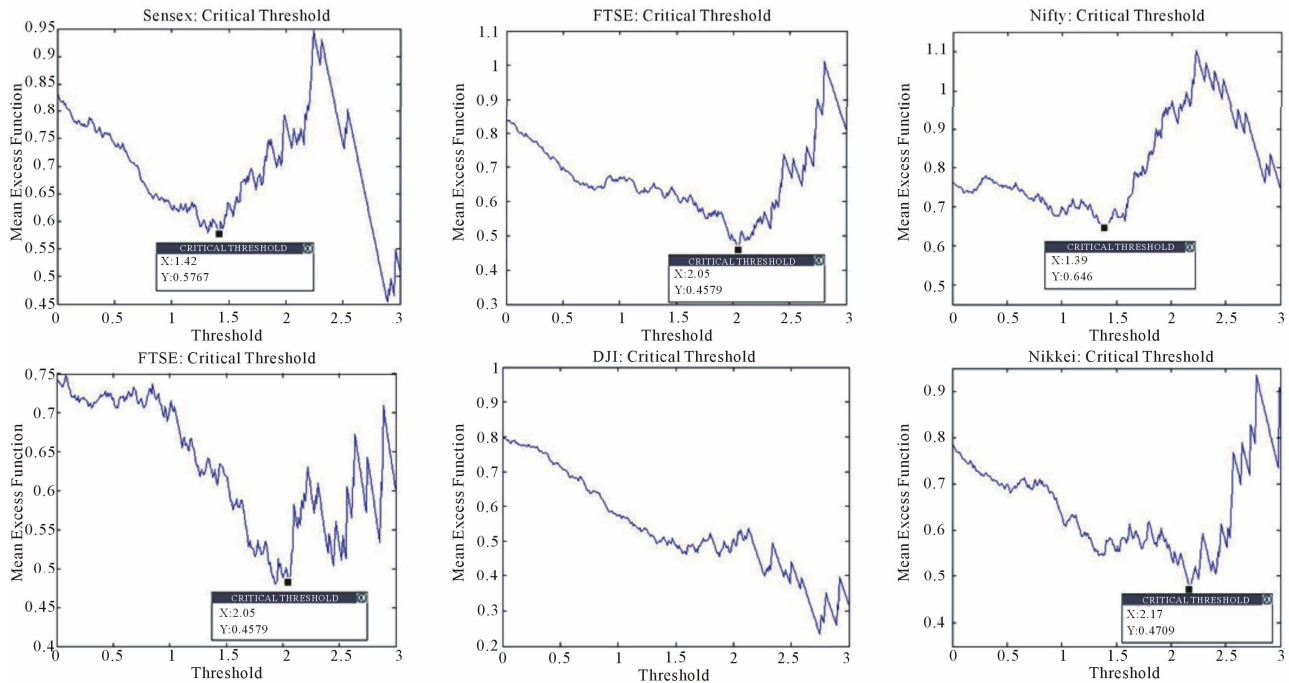


Figure 2. The optimal threshold is calculated by plotting the mean excess function of the six time series. The point is chosen at the point where the graph begins to slope upwards. As can be seen, the DJI graph is an anomaly, where no such clear point is present.

value for Sensex returns is at 1.4, and for Nifty, it is at 1.5. Note that in the graphs, we consider the negative of the return series, which is why the threshold values are positive.

For certain time series, the graph obtained is not very useful for finding the threshold. Consider the mean excess return for DJI in **Figure 2** for instance. In such cases, we consider an appropriately high value for the threshold, such as the 95th percentile of negative returns.

Model 2.2 GARCH-MixNorm Model

This model also makes use of the pseudo-normal assumption to calculate the ARMA (1, 1)-GARCH (1, 1) parameters. The standardized innovations are assumed to have a mixture of two normal distributions. We calculate the mean, standard deviation, skewness and kurtosis of these standardized innovations.

The mean of one of the two normal distributions in the mixture is assumed to be zero. This assumption is reasonable, as results show that the probability that the standardized residuals lie in this normal distribution is very high. A small percentage lies in the other distribution, with the non-zero mean and higher variance, these yield the very high and very low values observed in the data.

Thus, the parameter vector is of size four:

$\theta = (p, \mu_1, \sigma_1, \sigma_2)$. p is the probability that the data point lies in the first (non-zero mean) distribution, μ_1 is the mean of the first distribution, σ_1 and σ_2 are the standard deviations of the first and second distributions respectively. The mean of the second distribution is assumed to be zero.

The parameter vector components must satisfy the four moment constraints.

$$\mu_e = p\mu_1 \tag{53}$$

$$\sigma_e^2 = p\sigma_1^2 + (1-p)\sigma_2^2 + \mu_e^2 \left(\frac{1}{p} - 1 \right) \tag{54}$$

$$\tau_e = \frac{\mu_e}{\sigma_e^3} \left[3(\sigma_1^2 - \sigma_e^2) + \mu_e^2 \left(\frac{1}{p^2} - 1 \right) \right] \tag{55}$$

$$\begin{aligned} \kappa_e = \frac{1}{\sigma_e^4} & \left[3p\sigma_1^4 + 3(1-p)\sigma_2^4 + 6\mu_e^2 \left(\frac{\sigma_1^2}{p} - \sigma_e^2 \right) \right. \\ & \left. - 4\tau_e\mu_e\sigma_e^3 + \mu_e^4 \left(\frac{1}{p^3} - 1 \right) \right] \end{aligned} \tag{56}$$

An obtained solution is feasible if it satisfies the constraints $\sigma_1^2 > 0, \sigma_2^2 > 0$ and $0 < p < 1$.

To calculate the parameters through the method of moments, we need five moment equations. It is possible that there may not be a solution even if the first five moments were calculated. So we employ a fractile-to-fractile comparison test in addition to using certain moment equations.

We employ a modified version of the technique used by Perez [27]. The data (standardized residuals) is divided into seven sets; less than 0.5 standard deviations, 0.5 – 1, 1 – 1.5, 1.5 – 2, 2 – 2.5, 2.5 – 3, and greater than 3 standard deviations. The actual number of residuals in each category (α_k) is compared with the predicted number of residuals for the solution each obtained from the moment equations (ϱ_k). The solution considered is the one obtained by maximizing the log likelihood function

$$L(\alpha, \varrho) = \sum_{k=1}^7 \alpha_k \log(\varrho_k) \tag{57}$$

and satisfying the constraint Equations (49), (50), (52) and (53). As it turns out in most cases, there is no solution which satisfies all of them, in such cases, constraint Equation (52) is dropped. The minimization is carried out using the *fmincon* function in MATLAB. It turns out that the optimum values of the parameter are dependent on the initial values considered, so the parameters obtained for the previous data point are used as initial values in the optimization for the next one.

The Value at Risk is now calculated by the formula in Equation (48), where $VarR_t^q$ is calculated from inserting the calculated parameters in the mixture of normals probability density function given by Equation (16) and cumulating it by numerical methods.

Model 2.3 GARCH-Bickel-Doksum Model

We calculate the ARMA (1, 1)-GARCH (1, 1) parameters under the pseudo-normal assumption. The standardized residuals obtained (ϵ_i) are put through the transformation suggested by Bickel and Doksum [23] to normalize them (Equations (23) and (24)). If we assume that for some value of the parameter λ , the transformed observations $T(\lambda, \epsilon_i)$ are normally distributed with mean μ and standard deviation σ . The parameter is estimated by maximizing the log likelihood function

$$\begin{aligned} l_t(\theta|\epsilon) = & -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^t \{T(\lambda, \epsilon_i) - \mu\}^2 \\ & + (\lambda - 1) \sum_{i=1}^t \log|\epsilon_i| \end{aligned} \tag{58}$$

where $\theta = (\lambda, \mu, \sigma)$. The maximum likelihood estimate for the mean and variance is given by

$$\hat{\mu}(\lambda) = \frac{1}{t} \sum_{i=1}^t T(\lambda, \epsilon_i) \tag{59}$$

$$\hat{\sigma}^2(\lambda) = \frac{1}{t} \sum_{i=1}^t \{T(\lambda, \epsilon_i) - \hat{\mu}(\lambda)\}^2 \tag{60}$$

The estimate for λ can, therefore, be obtained by simply maximizing the likelihood function

$$l_i(\theta|\epsilon) = -\frac{n}{2} \log(2\pi\hat{\sigma}^2) + (\lambda-1) \log|\epsilon_i| \quad (61)$$

As shown in **Table 2(a)**, the Bickel-Doksum transformation does not handle skewed distributions well, as it only reduces kurtosis. Hence, this model must be modified to fix this drawback.

The Value at Risk is now calculated by the formula in Equation (48), where VaR_t^q is calculated from the inverse Bickel-Doksum formula

$$VaR_t^q = -\left\{-1 - N^{-1}[(1-q), \hat{\mu}(\lambda), \hat{\sigma}^2(\lambda)]\right\}^{1/\lambda} \quad (62)$$

where $N^{-1}[q, \hat{\mu}(\lambda), \hat{\sigma}^2(\lambda)]$ is the inverse normal function for probability $(1-q)$, mean $\hat{\mu}(\lambda)$ and variance $\hat{\sigma}^2(\lambda)$.

Model 2.4 GARCH-John-Draper Model

We calculate the ARMA (1, 1)-GARCH (1, 1) parameters under the pseudo-normal assumption. The standardized residuals obtained (ϵ_i) are transformed with the modulus transformation proposed by John and Draper [24] (Equations (25) & (26)). By using similar arguments as the previous model, the parameter λ is estimated by maximizing the log likelihood function

$$l_i(\theta|\epsilon) = -\frac{n}{2} \log(2\pi\hat{\sigma}^2) + (\lambda-1) \sum_{i=1}^t \log(|\epsilon_i|+1) \quad (63)$$

where $\hat{\sigma}^2$ is given by Equation (56) with $T(\lambda, \epsilon_i)$ representing the modulus transformation.

As with the Bickel-Doksum transformation, **Table 2(a)** shows that the modulus transformation is not a skew-corrector, it reduces kurtosis. Hence, this model must be modified to correct this.

The Value at Risk is now calculated by the formula in Equation (48), where VaR_t^q is calculated from the inverse John-Draper formula

$$VaR_t^q = 1 - \left\{1 - \lambda N^{-1}[(1-q), \hat{\mu}(\lambda), \hat{\sigma}^2(\lambda)]\right\}^{1/\lambda} \quad (64)$$

where $N^{-1}[q, \hat{\mu}(\lambda), \hat{\sigma}^2(\lambda)]$ is the inverse normal function for probability $(1-q)$, mean $\hat{\mu}(\lambda)$ and variance $\hat{\sigma}^2(\lambda)$.

Model 2.5 GARCH-Yeo-Johnson Model

We calculate the ARMA (1, 1)-GARCH (1, 1) parameters under the pseudo-normal assumption. The standardized residuals obtained (ϵ_i) are transformed with the Yeo-Johnson [25] transformation (Equations (27)). By using similar arguments as the previous models, the parameter λ is estimated by maximizing the log likelihood function

$$l_i(\theta|\epsilon) = -\frac{n}{2} \log(2\pi\hat{\sigma}^2) + (\lambda-1) \sum_{i=1}^t \text{sign}(\epsilon_i) \log(|\epsilon_i|+1) \quad (65)$$

where $\hat{\sigma}^2$ is given by Equation (56) with $T(\lambda, \epsilon_i)$ re-

presenting the Yeo-Johnson transformation.

Tables 2(a) and **(b)** show that the Yeo-Johnson transformation is a skew-correcting transformation. The model must be modified to enable kurtosis-handling as well.

The Value at Risk is now calculated by the formula in Equation (48), where VaR_t^q is calculated from the inverse Yeo-Johnson formula

$$VaR_t^q = 1 - \left\{1 + (\lambda-2) N^{-1}[(1-q), \hat{\mu}(\lambda), \hat{\sigma}^2(\lambda)]\right\}^{1/(2-\lambda)} \quad (66)$$

where $N^{-1}[q, \hat{\mu}(\lambda), \hat{\sigma}^2(\lambda)]$ is the inverse normal function for probability $(1-q)$, mean $\hat{\mu}(\lambda)$ and variance $\hat{\sigma}^2(\lambda)$.

Model 2.6 GARCH-Manly-John-Draper Model

We calculate the ARMA (1, 1)-GARCH (1, 1) under the pseudo-normal assumption. The innovations are initially transformed through the Manly exponential transformation to rid it of skewness. The symmetric data is now transformed with the John-Draper modulus transformation, which reduces kurtosis. The doubly-transformed data obtained is now roughly normally distributed (**Tables 2(a)** and **(b)**).

To obtain the parameter for the Manly transformation, the following log-likelihood function is maximized.

$$l_i(\theta|\epsilon) = -\frac{n}{2} \log(2\pi\hat{\sigma}^2) + \sum_{i=1}^t \lambda \epsilon_i \quad (67)$$

The parameter for the John-Draper transformation is obtained by maximizing the log-likelihood function in Equation (60).

The inverse Manly transformation is given by

$$VaR_t^q = \frac{1}{\lambda} \log\left[1 + \lambda N^{-1}[(1-q), \hat{\mu}(\lambda), \hat{\sigma}^2(\lambda)]\right] \quad (68)$$

The Value at Risk is calculated in two steps. First, the low quantile value is subjected to the inverse John-Draper transformation in Equation (61) and this value is back-transformed with the inverse Manly transformation in Equation (65).

Model 2.7 GARCH-Manly-Bickel-Doksum Model

We calculate the ARMA (1, 1)-GARCH (1, 1) under the pseudo-normal assumption. The innovations are initially transformed through the Manly exponential transformation remove skewness, and then with the Bickel-Doksum transformation, which reduces kurtosis. The skewness and kurtosis of the doubly-transformed insample data is given in **Tables 2(a)** and **(b)**.

The parameters for the Manly and Bickel-Doksum transformations are calculated by maximizing log-likelihoods in Equations (64) and (58). After the two parameters are obtained, the VaR is calculated from the inverse transformations in Equations (59) and (65) carried out serially in that order.

Table 2. (a) Skewness comparison of std. residuals after power transformation; (b) Kurtosis comparison of std. residuals after power transformation.

(a)

	Sensex	NIFTY	DJI	FTSE	HSI	Nikkei
Initial skewness	-0.4764	-0.5311	-0.1024	-0.3660	-0.1836	-0.3209
Manly	0.0134	0.0615	0.0035	0.0162	0.0126	0.0072
John-Draper	-0.3694	-0.2842	-0.0801	-0.2763	-0.0815	-0.2299
Yeo-Johnson	0.0037	0.0442	0.0094	0.0055	-0.0175	-0.0306
Bickel-Doksum	-0.4379	-0.3769	-0.0858	-0.3045	-0.1069	-0.2638
Manly-Yeo-Johnson	0.0052	0.0356	0.0093	0.0068	-0.0221	-0.0287
Manly-John-Draper	0.0070	-0.0026	-0.0026	0.0052	0.0226	0.0150
Manly-Bickel-Doksum	0.0126	0.0281	-0.0015	0.0068	0.0208	0.0123
John-Draper-Yeo-Johnson	-0.0063	-0.0087	0.0051	-0.0022	-0.0151	-0.0284
Yeo-Johnson-John-Draper	-0.0002	0.0028	0.0028	-0.0002	0.0088	-0.0130
Yeo-Johnson-Bickel-Doksum	0.0029	0.0233	0.0041	-0.0003	0.0026	-0.0193

The standardized residuals for the in-sample data are transformed with various power transformations. The skewness of each transformed output is compared to check their normalizing effect. For double-transformations, the data is first transformed with the transformation mentioned first, and then subjected to the second transformation.

(b)

	Sensex	NIFTY	DJI	FTSE	HSI	Nikkei
Initial kurtosis	3.7840	5.0195	3.3459	3.8574	3.9326	3.5752
Manly	3.3038	4.9318	3.3380	3.4958	3.8097	3.1979
John-Draper	3.1862	3.1718	2.8817	3.1754	2.8691	3.0347
Yeo-Johnson	3.3475	4.7952	3.3385	3.5389	3.8324	3.2611
Bickel-Doksum	3.5532	3.8147	2.9518	3.3488	3.0716	3.2118
Manly-Yeo-Johnson	3.3032	4.8972	3.3388	3.4952	3.8182	3.2126
Manly-John-Draper	3.0932	3.2107	2.8838	3.0976	2.8498	2.9224
Manly-Bickel-Doksum	3.2771	3.8375	2.9502	3.1660	3.0251	2.9903
John-Draper-Yeo-Johnson	2.9229	3.0349	2.8775	3.0082	2.8458	2.8674
Yeo-Johnson-John-Draper	3.1005	3.1869	2.8840	3.1073	2.8531	2.9431
Yeo-Johnson-Bickel-Doksum	3.2998	3.7688	2.9565	3.1907	3.0305	3.0275

The standardized residuals for the in-sample data are transformed with various power transformations. The kurtosis of each transformed output is compared to check their normalizing effect. For double-transformations, the data is first transformed with the transformation mentioned first, and then subjected to the second transformation.

Model 2.8 *GARCH-Yeo-Johnson-John-Draper Model*

We calculate the ARMA (1, 1)-GARCH (1, 1) under the pseudo-normal assumption. The innovations are initially transformed through the Yeo-Johnson transformation to rid it of skewness. The symmetric data is now transformed with the John-Draper modulus transformation, which reduces kurtosis. The doubly-transformed data obtained is now roughly normally distributed (**Tables 2(a)**

and **(b)**).

After the parameters for the two transformations are obtained from Equations (62) and (60), the VaR is computed from the inverse transformations in Equations (61) and (63).

Model 2.9 *GARCH-Yeo-Johnson-Bickel-Doksum Model*

We calculate the ARMA (1, 1)-GARCH (1, 1) under the pseudo-normal assumption. The innovations are initially

transformed through the Yeo-Johnson transformation to remove skewness, and then transformed with the Bickel-Doksum transformation, which removes excess kurtosis. The skewness and kurtosis of the doubly-transformed data obtained are given in **Tables 2(a)** and **(b)**.

Parameters for the Yeo-Johnson and Bickel-Doksum transformations are calculated by maximizing log-likelihoods in Equations (62) and (58). The VaR is calculated

from the inverse transformations in Equations (59) and (63) carried out serially in that order.

4. Testing

The data series are of length 1500; these are divided into the in-sample series (length 1000) and out-of-sample series (length 500). For each data point in the out-of-sample region, we estimate model parameters using the pre-

Table 3. (a) 99% VaR violations comparisons for model 1 series; (b) 97.5% VaR violations comparisons for model 1 series; (c) 95% VaR violations comparisons for model 1 series.

(a)

99% VaR	Model 1.1 Normal	Model 1.2 T	Model 1.3 Pearson Type IV	Model 1.4 Johnson S _U	Model 1.5 Manly	Expected Violations
Sensex	16	16	7	7	8	5
Nifty	16	14	8	8	13	5
DJI	22	20	9	9	11	5
FTSE	19	21	13	13	17	5
H S I	15	13	6	6	10	5
Nikkei	11	13	7	7	9	5

This table shows the VaR violation comparisons for the Model 1 series. The expected number of violations is given in the last column, 99% VaR is expected to be violated 5 times for a 500 point out-of-sample data set. As can be seen, Models 1.3 and 1.4 are the best performing ones.

(b)

97.5% VaR	Model 1.1 Normal	Model 1.2 T	Model 1.3 Pearson Type IV	Model 1.4 Johnson S _U	Model 1.5 Manly	Expected Violations
Sensex	28	27	16	15	21	12.5
Nifty	24	24	16	15	21	12.5
DJI	34	27	22	22	23	12.5
FTSE	29	30	25	25	24	12.5
H S I	24	23	20	19	22	12.5
Nikkei	29	29	14	14	18	12.5

This table shows the VaR violation comparisons for the Model 1 series. The expected number of violations is given in the last column, 97.5% VaR is expected to be violated 12.5 times for a 500 point out-of-sample data set. As can be seen, Models 1.3 and 1.4 are the best performing ones.

(c)

95% VaR	Model 1.1 Normal	Model 1.2 T	Model 1.3 Pearson Type IV	Model 1.4 Johnson S _U	Model 1.5 Manly	Expected Violations
Sensex	38	38	29	28	33	25
Nifty	40	38	29	32	36	25
DJI	55	49	42	41	42	25
FTSE	38	39	36	36	41	25
H S I	37	34	33	33	33	25
Nikkei	46	43	33	33	38	25

This table shows the VaR violation comparisons for the Model 1 series. The expected number of violations is given in the last column, 95% VaR is expected to be violated 25 times for a 500 point out-of-sample data set. As can be seen, Model 1.3 is the best performing one.

vious 1000 data points, *i.e.*, for finding VaR on day t , we consider data points from day $t-1000$ to day $t-1$.

The tests are run in MATLAB version 7.2 on a Windows XP operating system with 1.6 GHz processing speed. While running the program to calculate VaR for a single day, the results are generated well within 30 seconds for

most cases.

5. Results

5.1. Data and Model Parameters

The models are tested on six equity indices, Sensex, Nifty,

Table 4. (a) 99% VaR violations comparisons for model 2 series; (b) 99% VaR violations comparisons for model 2 series; (c) 95% VaR violations comparisons for model 2 series.

(a)

99% VaR	Model 2.1 EVT	Model 2.2 Mixture of Normals	Model 2.3 Bickel-Doksum	Model 2.4 John-Draper	Model 2.5 Yeo-Johnson	Model 2.6 Manly-John- Draper	Model 2.7 Manly-Bickel- Doksum	Model 2.8 Yeo-Johnson- John-Draper	Model 2.9 Yeo-Johnson- Bickel-Doksum	Expected Violations
Sensex	7	7	15	13	13	7	7	7	7	5
Nifty	9	9	13	12	13	9	10	9	10	5
DJI	11	15	17	14	15	10	11	10	11	5
FTSE	14	16	18	18	16	14	15	13	15	5
H S I	6	13	9	8	10	6	6	6	6	5
Nikkei	8	9	10	12	10	8	8	8	8	5

This table shows the VaR violation comparisons for the Model 2 series. The expected number of violations is given in the last column, 99% VaR is expected to be violated 5 times for a 500 point out-of-sample data set. As can be seen, Model 2.6 is the best performing one.

(b)

97.5% VaR	Model 2.1 EVT	Model 2.2 Mixture of Normals	Model 2.3 Bickel-Doksum	Model 2.4 John-Draper	Model 2.5 Yeo-Johnson	Model 2.6 Manly-John- Draper	Model 2.7 Manly- Bickel- Doksum	Model 2.8 Yeo-Johnson- John-Draper	Model 2.9 Yeo-Johnson- Bickel-Doksum	Expected Violations
Sensex	22	20	24	23	22	19	19	19	19	12.5
Nifty	22	18	22	21	21	18	18	18	18	12.5
DJI	33	25	28	28	25	22	22	22	22	12.5
FTSE	25	23	24	24	24	23	23	23	23	12.5
H S I	23	19	21	21	23	19	19	19	19	12.5
Nikkei	24	19	23	23	19	17	17	18	18	12.5

This table shows the VaR violation comparisons for the Model 2 series. The expected number of violations is given in the last column, 97.5% VaR is expected to be violated 12.5 times for a 500 point out-of-sample data set. As can be seen, Models 2.6 and 2.7 are the best performing ones.

(c)

95% VaR	Model 2.1 EVT	Model 2.2 Mixture of Normals	Model 2.3 Bickel-Doksum	Model 2.4 John-Draper	Model 2.5 Yeo-Johnson	Model 2.6 Manly-John- Draper	Model 2.7 Manly- Bickel-Doksum	Model 2.8 Yeo-Johnson- John-Draper	Model 2.9 Yeo-Johnson- Bickel-Doksum	Expected Violations
Sensex	34	38	36	36	34	34	32	32	32	25
Nifty	34	43	36	36	34	34	33	33	32	25
DJI	51	57	52	52	49	47	45	45	43	25
FTSE	35	38	37	37	34	34	34	33	32	25
H S I	31	37	35	37	34	34	33	34	33	25
Nikkei	37	44	43		38	35	34	34	34	25

This table shows the VaR violation comparisons for the Model 2 series. The expected number of violations is given in the last column, 95% VaR is expected to be violated 25 times for a 500 point out-of-sample data set. As can be seen, Model 2.9 is the best performing one.

DJI, FTSE, HSI and Nikkei. The data used is the closing value of these indices from the period March 2003 to February 2009. The data was obtained from www.finance.yahoo.com, and the time period includes the stock market crash of 2008. The details regarding the returns of the series, and the first four moments are given in **Table 1**.

5.2. VaR Violations and Comparison of Models

We test the effectiveness of each model by calculating the number of times the calculated VaR has been violated.

The expected number of violations for a q -percentile VaR is given by

$$\text{Expected } q\% \text{ VaR violations} = (1 - q\%)N \quad (69)$$

where N is the total number of VaR measurements.

We measure VaR for each out-of-sample data point, therefore, $N = 500$. We calculate 95%, 97.5% and 99% VaR for each data point. Therefore, the expected violations for each would be 25, 12.5 and 5 respectively.

Tables 3(a)-(c) compare the five models of the Model 1 series, comparing VaR violations for the six equity

Table 5. (a) LR Test for 99% VaR violations for model 1 series; (b) LR Test for 97.5% VaR violations for model 1 series; (c) LR Test for 95% VaR violations for model 1 series.

(a)

99% VaR	Model 1.1 Normal	Model 1.2 T	Model 1.3 Pearson Type IV	Model 1.4 Johnson S_U	Model 1.5 Manly
Sensex	15.47	15.47	0.72	0.72	1.54
Nifty	15.47	10.99	1.54	1.54	8.97
DJI	31.78	25.91	2.61	2.61	5.42
FTSE	23.13	28.80	8.97	8.97	17.90
H S I	13.16	8.97	0.19	0.19	3.91
Nikkei	5.42	8.97	0.72	0.72	2.61

This table shows the LR test statistic for the Model 1, 99% VaR violation observations. The numbers in bold indicate situations where the null hypothesis, *i.e.* the observed violations is equal to the predicted one, is rejected.

(b)

97.5% VaR	Model 1.1 Normal	Model 1.2 T	Model 1.3 Pearson Type IV	Model 1.4 Johnson S_U	Model 1.5 Manly
Sensex	14.66	13.02	0.92	0.48	4.94
Nifty	8.59	8.59	0.92	0.48	4.94
DJI	26.01	13.02	6.06	6.06	7.28
FTSE	16.38	18.16	9.98	9.98	8.59
H S I	8.59	7.28	3.92	3.00	6.06
Nikkei	16.38	16.38	0.18	0.18	2.19

This table shows the LR test statistic for the Model 1, 97.5% VaR violation observations. The numbers in bold indicate situations where the null hypothesis, *i.e.* the observed violations is equal to the predicted one, is rejected.

(c)

95% VaR	Model 1.1 Normal	Model 1.2 T	Model 1.3 Pearson Type IV	Model 1.4 Johnson S_U	Model 1.5 Manly
Sensex	6.18	6.18	0.64	0.37	2.46
Nifty	8.08	6.18	0.64	1.90	4.51
DJI	28.67	19.18	10.19	9.11	10.19
FTSE	6.18	7.10	4.51	4.51	9.11
H S I	5.32	3.08	2.46	2.46	2.46
Nikkei	15.04	11.33	2.46	2.46	6.18

This table shows the LR test statistic for the Model 1, 95% VaR violation observations. The numbers in bold indicate situations where the null hypothesis, *i.e.* the observed violations is equal to the predicted one, is rejected.

indices. **Tables 4(a)-(c)** compare the same for the nine models of the Model 2 series. The expected violations for 99%, 97.5% and 95% VaR are given in the last column of each table. The mean violation for each model is computed, and the best model for each percentile VaR is found.

It can be seen that Models 1.3 and 1.4 are best per-

forming models across all indices. Amongst those of the Model 2 series (where ARMA (1, 1)-GARCH (1, 1) parameters are calculated with the pseudo-normal assumption) however, Models 2.6, 2.7, 2.8 and 2.9 perform the best. This is expected from the skewness-kurtosis **Table 2**, where the most normalized transformations are shown to be Manly-John-Draper, Manly-Bickel-Doksum, Yeo-John-

Table 6. (a) LR Test for 99% VaR violations for model 2 series; (b) LR Test for 97.5% VaR violations for model 2 series; (c) LR Test for 95% VaR violations for model 2 series.

(a)

99% VaR	Model 2.1 EVT	Model 2.2 Mixture of Normals	Model 2.3 Bickel-Doksum	Model 2.4 John-Draper	Model 2.5 Yeo-Johnson	Model 2.6 Manly- John-Draper	Model 2.7 Manly-Bickel- Doksum	Model 2.8 Yeo-Johnson- John-Draper	Model 2.9 Yeo-Johnson- Bickel-Doksum
Sensex	0.72	0.72	13.16	8.97	8.97	0.72	0.72	0.72	0.72
Nifty	2.61	2.61	8.97	7.11	8.97	2.61	3.91	2.61	3.91
DJI	5.42	13.16	17.90	10.99	13.16	3.91	5.42	3.91	5.42
FTSE	10.99	15.47	20.46	20.46	15.47	10.99	13.16	8.97	13.16
H S I	0.19	8.97	2.61	1.54	3.91	0.19	0.19	0.19	0.19
Nikkei	1.54	2.61	3.91	7.11	3.91	1.54	1.54	1.54	1.54

This table shows the LR test statistic for the Model 2, 99% VaR violation observations. The numbers in bold indicate situations where the null hypothesis, *i.e.* the observed violations is equal to the predicted one, is rejected.

(b)

99% VaR	Model 2.1 EVT	Model 2.2 Mixture of Normals	Model 2.3 Bickel-Doksum	Model 2.4 John-Draper	Model 2.5 Yeo-Johnson	Model 2.6 Manly- John-Draper	Model 2.7 Manly- Bickel-Doksum	Model 2.8 Yeo-Johnson- John-Draper	Model 2.9 Yeo-Johnson- Bickel-Doksum
Sensex	3.92	6.06	8.59	7.28	6.06	3.00	3.00	3.00	3.00
Nifty	2.19	6.06	6.06	4.94	4.94	2.19	2.19	2.19	2.19
DJI	9.98	23.95	14.66	14.66	9.98	6.06	6.06	6.06	6.06
FTSE	7.28	9.98	8.59	8.59	8.59	7.28	7.28	7.28	7.28
H S I	3.00	7.28	4.94	4.94	7.28	3.00	3.00	3.00	3.00
Nikkei	3.00	8.59	7.28	7.28	3.00	1.50	1.50	2.19	2.19

This table shows the LR test statistic for the Model 2, 97.5% VaR violation observations. The numbers in bold indicate situations where the null hypothesis, *i.e.* the observed violations is equal to the predicted one, is rejected.

(c)

99% VaR	Model 2.1 EVT	Model 2.2 Mixture of Normals	Model 2.3 Bickel-Doksum	Model 2.4 John-Draper	Model 2.5 Yeo-Johnson	Model 2.6 Manly- John-Draper	Model 2.7 Manly-Bickel- Doksum	Model 2.8 Yeo-Johnson- John-Draper	Model 2.9 Yeo-Johnson- Bickel-Doksum
Sensex	3.08	6.18	4.51	4.51	3.08	3.08	1.90	1.90	1.90
Nifty	3.08	11.33	4.51	4.51	3.08	3.08	2.46	2.46	1.90
DJI	22.17	32.16	23.73	23.73	19.18	16.37	13.75	13.75	11.33
FTSE	3.77	6.18	5.32	5.32	3.08	3.08	3.08	2.46	1.90
H S I	1.41	5.32	3.77	5.32	3.08	3.08	2.46	3.08	2.46
Nikkei	5.32	12.52	11.33	11.33	6.18	3.77	3.08	3.08	3.08

This table shows the LR test statistic for the Model 2, 95% VaR violation observations. The numbers in bold indicate situations where the null hypothesis, *i.e.* the observed violations is equal to the predicted one, is rejected.

son-John-Draper and Yeo-Johnson-Bickel-Doksum. Model 2.1 performs well too, especially for higher VaR estimation.

In order to test the observed VaR numbers, we use Kupiec's test to determine if the observed VaR violations are significantly different from their expected values. The test is based on the fact that the number of violations N in a sample of size T is binomially distributed as $N \sim B(T, p)$. Thus, the probability of N excesses occurring over a T day period is given by $p^N (1-p)^{T-N}$ where p is the probability of exceeding VaR on a given day. Under the null hypothesis that $N/T = p$, we calculate the Likelihood Ratio (LR) test statistic

$$2 \ln \left[\frac{(1-N/T)^{T-N} (N/T)^N}{(1-p)^{T-N} p^N} \right] \sim \chi^2(1) \quad (70)$$

The test statistics for the VaR violation observations are given in **Tables 5(a)-(c)** for the Model 1 series, and in **Tables 6(a)-(c)** for the Model 2 series. The values in bold are those where the observed VaR violations are significantly different from expected ones.

6. Conclusions

In this work, we build different models for accurate measurement of Value at Risk. We use an ARMA (1, 1) process to model conditional expectation, and a GARCH (1, 1) process to model conditional variance. Models 1.x calculate parameters for the above processes without the pseudo-normal assumption, while Models 2.x calculate them with the pseudo-normal assumption. The following conclusions can be made from the results.

- Models 1.3 and 1.4 (GARCH-PIV and GARCH-JSU) are far and away the best performers among all the models. Their consistency can be seen across indices and VaR percentiles.
- Among the models which use the pseudo-normal assumption, Models 1.6, 1.7, 1.8 and 1.9 perform the best. These use two transformations to normalize the standardized innovations, the first one makes the distribution symmetric, while the second one reduces the kurtosis.
- Model 2.1 (GARCH-EVT) performs well for high percentile VaR estimates.

Computationally, Model 2.x series are slightly faster than Models 1.3 and 1.4, but the difference of a few seconds does not mandate using them in the place of the more accurate GARCH-PIV and GARCH-JSU models.

REFERENCES

- [1] E. Fama, "The Behavior of Stock Prices," *Journal of Business*, Vol. 47, No. 1, 1965, pp. 244-280.
- [2] B. B. Mandelbrot, "The Variation of Certain Speculative Prices," *Journal of Business*, Vol. 36, No. 4, 1963, pp. 394-419. [doi:10.1086/294632](https://doi.org/10.1086/294632)
- [3] R. Blattberg and N. Gonedes, "A Comparison of Stable and Student Distributions as Statistical Models of Stock Prices," *Journal of Business*, Vol. 47, 1974, pp. 244-280. [doi:10.1086/295634](https://doi.org/10.1086/295634)
- [4] C. A. Ball and W. N. Torous, "A Simplified Jump Process for Common Stock Returns," *Journal of Financial and Quantitative Analysis*, Vol. 18, No. 1, 1983, pp. 53-65. [doi:10.2307/2330804](https://doi.org/10.2307/2330804)
- [5] S. J. Kon, "Models of Stock Returns: A Comparison," *Journal of Finance*, Vol. 39, No. 1, 1984, pp. 147-165. [doi:10.2307/2327673](https://doi.org/10.2307/2327673)
- [6] J. B. Gray and D. W. French, "Empirical Comparisons of Distributional Models for Stock Index Returns," *Journal of Business Finance and Accounting*, Vol. 17, No. 3, 1990, pp. 451-459. [doi:10.1111/j.1468-5957.1990.tb01197.x](https://doi.org/10.1111/j.1468-5957.1990.tb01197.x)
- [7] M. Bhattacharyya, A. Chaudhary and G. Yadav, "Conditional VaR Estimation Using Pearson Type IV Distribution," *European Journal of Operational Research*, Vol. 191, No. 1, 2008, pp. 386-397. [doi:10.1016/j.ejor.2007.07.021](https://doi.org/10.1016/j.ejor.2007.07.021)
- [8] M. Bhattacharyya, N. Misra and B. Kodase, "Max VaR for Non-Normal and Heteroskedastic Returns," *Quantitative Finance*, Vol. 9, No. 8, 2009, pp. 925-935. [doi:10.1080/14697680802595684](https://doi.org/10.1080/14697680802595684)
- [9] M. Bhattacharyya and G. Ritolia, "Conditional VaR using EVT—Towards a Planned Margin Scheme," *International Review of Financial Analysis*, Vol. 17, No. 2, 2008, pp. 382-395. [doi:10.1016/j.irfa.2006.08.004](https://doi.org/10.1016/j.irfa.2006.08.004)
- [10] R. F. Engle, "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom inflation," *Econometrica*, Vol. 50, No. 4, 1982, pp. 987-1007. [doi:10.2307/1912773](https://doi.org/10.2307/1912773)
- [11] T. Bollerslev, "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, Vol. 31, No. 3, 1986, pp. 307-327. [doi:10.1016/0304-4076\(86\)90063-1](https://doi.org/10.1016/0304-4076(86)90063-1)
- [12] S. Poon and C. Granger, "Forecasting Volatility in Financial Markets," *Journal of Economic Literature*, Vol. 41, No. 2, 2003, pp. 478-539. [doi:10.1257/002205103765762743](https://doi.org/10.1257/002205103765762743)
- [13] J. Heinrich, "A Guide to the Pearson Type IV Distribution," 2004. http://www-cdf.fnal.gov/publications/cdf6820_pearson4.pdf.
- [14] N. L. Johnson, "Systems of Frequency Curves Generated by Methods of Translation," *Biometrika*, Vol. 36, No. 1-2, 1949, pp. 149-176. [doi:10.1093/biomet/36.1-2.149](https://doi.org/10.1093/biomet/36.1-2.149)
- [15] A. L. Tucker, "A Reexamination of Finite and Infinite Variance Distributions As Models of Daily Stock Returns," *Journal of Business & Economic Statistics*, Vol. 10, No. 1, 1992, pp. 73-81. [doi:10.2307/1391806](https://doi.org/10.2307/1391806)
- [16] J. D. Hamilton, "A Quasi-Bayesian Approach to Estimating Parameters for Mixtures of Normal Distributions," *Journal of Business and Economic Statistics*, Vol. 9, No. 1, 1991, pp. 27-39. [doi:10.2307/1391937](https://doi.org/10.2307/1391937)

- [17] D. M. Titterton, A. F. M. Smitha and U. E. Makov, "Statistical Analysis of Finite Mixture Distributions," John Wiley & Sons, Chichester, 1992.
- [18] J. Hull and A. White, "Value at Risk When Daily Changes in Market Variables Are Not Normally Distributed," *Journal of Derivatives*, Vol. 5, No. 3, 1998, pp. 9-19. [doi:10.3905/jod.1998.407998](https://doi.org/10.3905/jod.1998.407998)
- [19] P. Zangari, "An Improved Methodology for Measuring VaR," *Risk Metrics Monitor*, Reuters/JP Morgan, 1996.
- [20] G. E. P. Box and D. R. Cox, "An Analysis of Transformations," *Journal of the Royal Statistical Society*, Vol. 26, No. 2, 1964, pp. 211-252.
- [21] B. F. J. Manly, "Exponential Data Transformations," *The Statistician*, Vol. 25, No. 1, 1976, pp. 37-42. [doi:10.2307/2988129](https://doi.org/10.2307/2988129)
- [22] P. Li, "Box Cox Transformations: An Overview," University of Connecticut, Storrs, 2005.
- [23] P. J. Bickel and K. A. Doksum, "An Analysis of Transformations Revisited," *Journal of the American Statistical Association*, Vol. 76, 1981, pp. 296-311. [doi:10.2307/2287831](https://doi.org/10.2307/2287831)
- [24] J. A. John and N. R. Draper, "An Alternative Family of Transformations," *Applied Statistics*, Vol. 29, No. 2, 1980, pp. 190-197. [doi:10.2307/2986305](https://doi.org/10.2307/2986305)
- [25] I.-K. Yeo and R. Johnson, "A New Family of Power Transformations to Improve Normality or Symmetry," *Biometrika*, Vol. 87, No. 4, 2000, pp. 954-959. [doi:10.1093/biomet/87.4.954](https://doi.org/10.1093/biomet/87.4.954)
- [26] W. K. Newey and D. G. Steigerwald, "Asymptotic Bias for Quasi-Maximum-Likelihood Estimators in Conditional Heteroskedasticity Models," *Econometrica*, Vol. 65, No. 3, 1997, pp. 587-599. [doi:10.2307/2171754](https://doi.org/10.2307/2171754)
- [27] P. G. Perez, "Capturing Fat Tail Risk in Exchange Rate Returns Using SU-Curves: A Comparison with Normal Mixture and Skewed Student Distributions," *Journal of Risk*, Vol. 10, No. 2, 2007-2008, pp. 73-100.

Appendix

In the model 1.5 used, the returns series r_t is modeled as follows

$$r_t = \mu_t + \sigma_t X_t \tag{71}$$

We assume that Z_t is a distribution such that when transformed through Manly's exponential transformation (Equation (22)) it becomes normal.

$$Z = T(X), Z \sim N(\mu, \sigma) \tag{72}$$

$$P(z < a) = \frac{1}{\sqrt{2\pi}} \int_{-1/\lambda}^a \exp\left\{-\frac{1}{2}\left(\frac{m-\mu}{\sigma}\right)^2\right\} dm \tag{73}$$

The lower limit is given by $-1/\lambda$ since $Z = T(X)$. As X varies from $-\infty$ to $+\infty$, Z varies from $-1/\lambda$ to $+\infty$. In other words, it is impossible for Z to take on a value less than $-1/\lambda$.

$$P(z < a) = P(x < T^{-1}(a)) \tag{74}$$

This arises since the Manly's transformation is one-to-one and monotonically increasing. We name $b = T^{-1}(a)$ and proceed

$$P(x < b) = \frac{1}{\sqrt{2\pi\sigma}} \int_{T(-\infty)}^{T(b)} \exp\left\{-\frac{1}{2}\left(\frac{m-\mu}{\sigma}\right)^2\right\} dm \tag{75}$$

We name $m = T(n)$, and $dm = T'(n)dn$ follows.

$$P(x < b) = \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^b \exp\left\{-\frac{1}{2}\left(\frac{T(n)-\mu}{\sigma}\right)^2\right\} T'(n) dn \tag{76}$$

We need to add a normalizing constant k to the equation, such that $P(x < -\infty) = 1$.

$$P(x < -\infty) = \frac{k}{\sqrt{2\pi\sigma}} \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2}\left(\frac{T(n)-\mu}{\sigma}\right)^2\right\} T'(n) dn \tag{77}$$

Substituting $w = \frac{T(n)-\mu}{\sqrt{2\sigma}}$, $dw = \frac{T'(n)}{\sqrt{2\sigma}} dn$, and

changing limits from $(-\infty, +\infty)$ to $\left(-\left(\frac{\mu+1/\lambda}{\sqrt{2\sigma}}\right), +\infty\right)$.

$$P(x < -\infty) = \frac{k}{\sqrt{\pi}} \int_{-\left(\frac{\mu+1/\lambda}{\sqrt{2\sigma}}\right)}^{+\infty} \exp\{-w^2\} dw \tag{78}$$

$$P(x < -\infty) = \frac{k}{\sqrt{\pi}} \int_{-\left(\frac{\mu+1/\lambda}{\sqrt{2\sigma}}\right)}^{+\infty} \exp\{-w^2\} dw \tag{79}$$

$$+ \frac{2k}{\sqrt{\pi}} \int_0^{+\left(\frac{\mu+1/\lambda}{\sqrt{2\sigma}}\right)} \exp\{-w^2\} dw$$

$$k = \frac{2}{1 + \operatorname{erf}\left(\frac{\mu + \frac{1}{\lambda}}{\sqrt{2\sigma}}\right)} \tag{80}$$

The innovations are related to the standardized innovations by $\epsilon_t = h_t \eta_t$. We also assume that the transformed standardized residuals have zero mean and unit standard deviation. Therefore

$$f(\eta_t | F_{t-1}) = \frac{\sqrt{\frac{2}{\pi}}}{1 + \operatorname{erf}\left(\frac{1}{\lambda\sqrt{2}}\right)} \tag{81}$$

$$\exp\left\{-\frac{1}{2}\left(\frac{\exp(\lambda x) - 1}{\lambda}\right)^2\right\} \exp(\lambda x)$$

Since $P(\eta_t < a) = P(\int_t < ah_t)$,

$$f(\epsilon_t | F_{t-1}) = \frac{\sqrt{\frac{2}{\pi}}}{h_t \left(1 + \operatorname{erf}\left(\frac{1}{\lambda\sqrt{2}}\right)\right)} \exp\left\{-\frac{1}{2}\left(\frac{\exp(\lambda x/h_t) - 1}{\lambda}\right)^2\right\} \exp(\lambda x/h_t) \tag{82}$$

The log likelihood function to be minimized, is therefore

$$LLF = \sum_{t=1}^T \left\{ -\log \left(\frac{2}{1 + \operatorname{erf}\left(\frac{1}{\lambda\sqrt{2}}\right)} \right) - \frac{1}{2} \log 2\pi h_t^2 + \lambda \frac{\epsilon_t}{h_t} - \frac{M\left(\lambda, \frac{\epsilon_t}{h_t}\right)^2}{2} \right\} \tag{83}$$