WP/16/2015

NSE WORKING PAPER

How Good is the Transformation-Based Approach to Estimate Value at Risk?

Simulation and Empirical Results

G. P. Samanta

National Stock Exchange of India Limited

February 2015

NSE Working Paper

How Good is the Transformation-Based Approach to Estimate Value at Risk? Simulation and Empirical Results

Prepared by G. P. Samanta*
February 2015

Abstract

This paper examines the performance of the indirect transformation-based approach for the measurement of value at risk (VaR) suggested by Samanta (2008). A technical problem usually encountered in practice comes from the departure of the observed return distribution from a specific form of distribution, viz., normal distribution. Traditional approaches tackle the problem by identifying a suitable non-normal distribution for returns. However, Samanta (2008) addressed the issue indirectly by transforming the non-normal return distribution to normality. The simulation exercise carried out in this paper shows that the transformation to normality provides a sensible alternative to the measurement of VaR. Further, the empirical assessment of the accuracy of the VaR estimates with respect to selected exchange rates reveals that the transformation-based approach outperforms the method based on the normality assumption for return distribution; moreover, the former produces VaR accuracy that is usually better than that of a more advanced tail-index-based approach.

Keywords: asset price behaviour, tail-index, transformation to normality, value at risk, Kupiec's test, loss-functions

* G. P. Samanta is Director and Member of Faculty, Reserve Bank Staff College, Chennai, India (email: rbigps@yahoo.com). The views expressed in the paper are those of the author and do not necessarily reflect the opinion of the National Stock Exchange of India Ltd or the Reserve Bank Staff College.

How Good is the Transformation-Based Approach to Estimate Value at Risk?
Simulation and Empirical Results

1. Introduction

The concept of value at risk (VaR) gained importance in the banking and finance literature over the past two decades. This was originally proposed as a measure of market risk exposure, thereby serving as the basis for calculating related risk capital (Basel Committee, 1996a, 1996b). Over time, VaR has emerged as a unified tool to measure other risk categories, such as, credit and operational risks. Further, the domain of application of this measure has widened from being the basis for determining the risk capital at banks, to the calculation of the margin requirement for traders/investors at stock exchanges, and to the design of the so-called 'macro markets' (Majumder and Majumder, 2002). The concept of macro markets involves a new set of markets for non-financial income; it was pioneered by Shiller (1993a, 1993b) and Shiller and Wincoop (1999), among others. The concept of VaR suffers from a major limitation of not being a coherent risk measure (Artzner, et al., 1999). Further, it cannot assess the complete risk profile. For instance, VaR cannot assess the magnitude of excess loss. On the other hand, concepts such as excess shortfall (ES) (which is helpful in assessing the average magnitude of losses over VaR), loss severity (Jorian, 2001), and conditional-VaR (Co-VaR), which is useful in assessing the financial stability of an economy (Acharya, et al., 2010; Acharya, et al., 2012; Adrian and Marcus, 2008), all depend on VaR. The growing application of VaR for diverse purposes warrants further improvements to or simplifications of the task of measuring VaR.

A particularly well-identified problem in connection with VaR estimation stems from the observed deviation of return distribution from normality. The normality assumption brings in theoretical convenience because a normal distribution is fully characterised by its first two moments, i.e., mean and variance. However, in reality, the observed return distributions are usually far from normal. Conventionally, the issue of non-normality is addressed by directly fitting suitable non-normal distributions, either parametrically or non-parametrically. This task faces two challenges. First, one has to identify a suitable form of return distribution from the set of all relevant non-normal distributions, such as the *t*-distribution, mixture of two or more normal distributions (van den Goorbergh and Vlaar, 1999), Laplace distribution (Linden, 2001; Puig and Stephens, 2007), hyperbolic distribution (Bauer, 2000), and Auto-Regressive Conditional Heteroscedasticity (ARCH) or the Generalised-ARCH (GARCH) (Engle, 1982; Bollerslev, 1986; Wong et al., 2003). The search domain is very heterogeneous and wide, and one is always exposed to the risk of choosing a wrong distribution/model as the best one. Second, each class of distributions/models in the search set is unique in its own way, which calls for specific conceptual understanding and computational requirements. Eventually, this gives rise to complexity in the task, particularly for practitioners, who are already grappling with several business activities.

The observed non-normality can be sensibly handled indirectly by transforming the non-normal returns into normality. In an empirical exercise, Samanta (2008) experimentally adopted one such approach and found quite encouraging results. In the present paper, we assess the performance of the transformation-based indirect approach on two counts. First, we conduct a simulation exercise to examine how accurately the transformation induces normality when applied to random observations drawn from potential forms of non-normality. In this exercise, the Student-t distribution, skewed-Laplace, and ARCH and GARCH models are transformed using normality-transformation. Second, we check the robustness of Samanta's (2008) empirical results by employing real exchange rate data covering the period after the recent global financial crisis.

The rest of the paper is organised as follows. Section 2 presents a broad outline of the issues related to VaR estimation through direct distribution fitting approaches and summarises the transformation-based indirect approach. The results of the simulation exercise for assessing the performance of the normality transformation are presented in Section 3. In Section 4, the empirical results are discussed. Section 5 concludes the paper.

2. Value-at-Risk: Concept and Estimation Issues

Let W_t denote the total value of the underlying assets in a portfolio at time t. The change in total value from time t to t+k is $\Delta W_t(k) = (W_{t+k} - W_t) = (1+r)W_t$, where r represents the proportional change from time t to time (t+k). An individual with a long financial position on the asset portfolio will incur a loss if $r < 0$, but a short position will see a loss when $r > 0$. Thus, a rise (fall) in the value of r would indicate a profit to someone holding a long (short) position. However, at time t, r is unknown; it can be thought of as a random variable. Let $f(r,\beta)$ denote the probability density function of r; β is the vector of the unknown parameters. At time t, the VaR over time horizon k and given probability p ($0 < p < 1$), i.e., $100 \times (1 - p)$% confidence level would be estimated as

Long-position: $\text{Prob}(r < -\text{VaR}) = p$, i.e., $\text{Prob.}(r \geq -\text{VaR}) = 1 - p$ (2.1)

Short-position: $\text{Prob}(r > \text{VaR}) = p$, i.e., $\text{Prob.}(r \leq \text{VaR}) = 1 - p$ (2.2)

where Prob(.) denotes the probability measure.

Thus, for a long-position holder, the VaR at $100 \times (1 - p)$% confidence level would be the p^{th} percentile of the distribution represented by $f(r,\beta)$. For the short position, this would be the $(1 - p)$ percentile, i.e., the threshold value of r with right-tail probability p.

2.1 Conventional Approaches of VaR Estimation: Direct approaches

If r followed a normal distribution, VaR estimation would be very simple because the normal distribution is completely characterized by its first two moments, viz., mean and variance. For a normal distribution $N(\mu,\sigma^2)$ with mean μ and variance σ^2, the VaR for the long position and short position would be calculated as:

$$VaR = \begin{cases} |\mu - z_p\sigma| & \textit{for Long Position} \\ |\mu + z_p\sigma| & \textit{for Short Position} \end{cases} \tag{2.3}$$

where |.| denotes the absolute value and z_p denotes the absolute value of p^{th} percentile for the N(0,1) distribution. The property of the N(0,1) distribution indicates that $z_{0.01} = -2.33$ and $z_{0.05} = -1.65$.

However, the practical problem is that the distribution of the observed r is usually far from normal, which may be detected through significant skewness, or kurtosis different from 3 (i.e., excess kurtosis different from zero), or both. Thus, a deviation of the return distribution from normality leads to huge complexity and computational burden because one has to select an appropriate distributional form depending on the significance of skewness and excess kurtosis since the VaR is no longer a simple linear function of the mean and variance. The accuracy of the VaR estimates depends on how well the chosen functional form of $f(r,\beta)$ fits the observed returns. The function $f(r,\beta)$ may be identified either non-parametrically or parametrically. Parametric modelling covers several different types of approaches that (1) model the complete portion of the distribution, such as fitting a suitable standard form of non-normal distribution (the Student-*t* distribution, Laplace distribution, hyperbolic distribution, etc.), fitting a mixture distribution (such as a mixture of two or more normal distributions)[1], identifying ARCH/GARCH types of model, or (2) model only the tails of the observed return distribution, such as the tail-index approach. Thus, conventional approaches of VaR estimation differ in terms of the strategy and functional form adopted in identifying the appropriate form of $f(r,\beta)$. A wrongly identified $f(r,\beta)$ would lead to inaccuracy in the VaR estimates.

In contrast, the transformation-based approach does not focus on identifying the appropriate form of $f(r,\beta)$. Instead, this indirect approach looks for a suitable continuous monotonic (one-to-one) function of r, $g(r,\theta)$, where θ is the vector of the transformation parameters such that the probability distribution of $g(r,\theta)$ given θ is (approximately) normal.

2.2 Indirect Approach of VaR Estimation

2.2.1 Transformations, percentiles, and VaR

The transformation-based approach is outlined in detail in Samanta (2008). We summarise this approach for ready reference. Let a continuous random variable r be transformed as $y = g(r, \theta)$, where θ is a vector of the constant parameters. For given any value of θ, $g(r,\theta)$ is a continuous, monotonically increasing one-to-one function of r. For any real valued number γ, the events $\{ g(r,\theta) < \gamma \}$ and $\{ r < g^{-1}(\gamma, \theta) \}$ are equivalent. This means that

[1] The Laplace distribution can also be thought of (derived) as a mixture distribution. However, without deviating from the main estimation issue, we consider the Laplace distribution as a standard form of distribution.

$$\text{Prob}[\, g(r,\theta) < \gamma \,] = \text{Prob}[r < g^{-1}(\gamma,\theta)] \tag{2.4}$$

where Prob[.] denotes the probability measure.

By replacing γ in Eq. (2.4) with the p^{th} percentile of the distribution of $g(r,\theta)$, i.e., γ_p, we get the p^{th} percentile of the distribution of r as

$$\xi_p = g^{-1}(\gamma_p,\theta) \tag{2.5}$$

For estimating the VaR for r, we essentially need to estimate ξ_p for a given probability level p. If it were relatively easier to fit/approximate the distribution for $g(r,\theta)$ instead of the original r, one could first estimate γ_p via $g(.,\theta)$ and invert it to get ξ_p. In fact, the popular log-normal distribution adopts exactly the same strategy. It is known that the logarithm transformation of a log-normal variable follows a normal distribution. Therefore, we fit the distribution for the log-transformation of the random variable, instead of identifying the actual distribution of the original random variable.

2.2.2 Transformation to normality and VaR estimates

An additional degree of simplicity would be obtained if $g(r,\theta)$ were (approximately) a normal variable. By applying the properties of normal distribution, we would get $\gamma_p = \{\mu_g + z_p\, \sigma_g\}$, where z_p is the p^{th} percentile of standard normal distribution, and μ_g and σ_g are the mean and standard deviation of $g(r, \theta)$, respectively. The major advantage here is that z_p for a given p is known; therefore, the estimation of γ_p effectively requires only the estimation of the mean and standard deviation of $g(.,\theta)$. As stated earlier, it is known that $z_{0.01} = -2.33$ and $z_{0.05} = -1.65$. Further, as the standard normal distribution is symmetric about zero, the values of $z_{0.99}$ and $z_{0.95}$ are 2.33 and 1.65, respectively. Conventionally, the VaR for market risk is estimated for $p = 0.01$, which means the 99% VaR for the given portfolio (corresponding to the return series r) is as follows:

$$\xi_p = g^{-1}(\mu_g + z_p\, \sigma_g,\, \theta), \text{ and VaR} = |\xi_p| \tag{2.6}$$

In reality, the transformation parameter θ is seldom known; therefore, it has to be estimated from the data. Whether θ is known or is estimated from the data, $g(.,.)$ may not be perfect enough to transform any possible return distribution to exact normal distribution. However, as long as the transformed return is reasonably approximated by the normal distribution, the relationship in Eq. (2.6) holds in approximation. In other words:

$$\text{VaR} = |\xi_p| \approx |\, g^{-1}(\mu_g + z_p\, \sigma_g,\, \theta)\, | \tag{2.7}$$

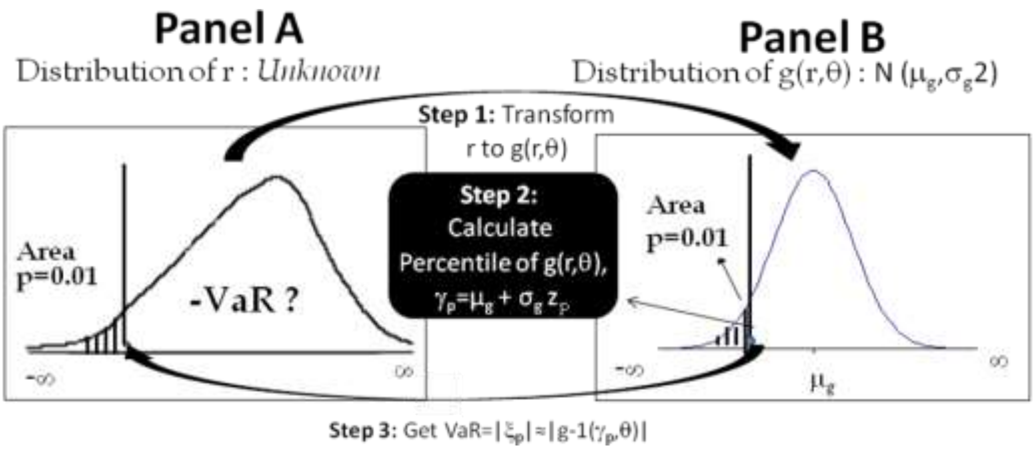
where the symbol $\approx$ indicates "approximately equal", and |.| represents absolute value.

This idea is intuitively appealing, is easy to understand, and requires simple computational efforts for implementation. However, we need to know the functional form of $g(.,\theta)$, and we also need to estimate the unknown transformation parameter θ. The accuracy of the VaR estimation depends on the power of $g(.,.)$ to induce normality in the transformation. The theoretical literature on the families of transformations to normality/symmetry is quite vast; in the following section, we discuss some significant developments that are relevant to this study.

2.2.3 Indirect approach vs. direct approach

The broad idea of the transformation-based indirect approach of VaR estimation with respect to the long position is diagrammatically presented in Fig. 2.1.

Fig. 2.1: Diagrammatic Representation of Indirect Approach of VaR Estimation



Panel A in Fig. 2.1 represents the unknown distribution of r; the probability level for VaR is fixed at 0.01 (i.e., the 99% confidence level). The observed returns are transformed through the one-to-one monotonically increasing function g(r,θ) (where θ is the transformation parameter) such that g(r,θ) follows (near) normal distribution with the mean (μ_g) and standard deviation (σ_g). This transformed (near) normal distribution is shown in Panel B. Since this transformed distribution is approximately normal, its percentile is $\gamma_p \approx \mu_g + \sigma_g z_p$, where $0 < p < 1$, and z_p represents the p^{th} percentile of the N(0,1) distribution, which can be computed easily from the tabulated values for the N(0,1) distribution. In particular, $z_{0.01}$= -2.33 and $z_{0.05}$ = -1.65.

Unlike the indirect approach, the direct approach does not require Panel B (Fig. 2.1). Instead, the direct approach focuses on approximating the unknown distribution in Panel A (Fig. 2.1) with a suitable distribution, either parametric or non-parametric, and subsequently determining the appropriate percentile from the fitted distribution.

2.2.4 Transformation of a random variable to normality

Since the pioneering work by Box and Cox (1964), the research on transformation to normality has grown into a vast body of literature. Several families of transformation for improving the normality or symmetry of the distribution have been proposed, with varying degrees of success. We discuss only a few of these transformations that are found useful for the task at hand (Samanta, 2008). Of particular use to us are the modulus transformation proposed by John and Draper (1980) and the more recent transformation class proposed by Yeo and Johnson (2000).

For transforming a symmetric distribution to near normality, John and Draper (1980) suggest the following modulus transformation for the original variable x,

$$g^{JD}(x,\delta) = \begin{cases} \mathrm{sign}(x)\{(1+|x|)^{\delta}-1\}/\delta & \text{if}\ \delta \neq 0 \\ \mathrm{sign}(x)\log(1+|x|) & \text{if}\ \delta = 0 \end{cases} \tag{2.8}$$

The transformation parameter δ in Eq. (2.8) may be estimated by maximizing the likelihood function. The extant literature suggests that the transformation $g^{JD}(x,\delta)$ is suitable for dealing with the kurtosis problem. However, it has serious drawbacks when applied to skewed distribution. In particular, if the distribution of x is a mixture of standard normal and gamma densities, the distributions of $g^{JD}(x,\delta)$ would be bimodal and would be far from normal. To circumvent the significant skewness, Yeo and Johnson (2000) proposed a new family of transformations:

$$g^{YJ}(x,\lambda) = \begin{cases} \{(1+x)^{\lambda}-1\}/\lambda & \text{if} \quad x \geq 0, \lambda \neq 0 \\ \log(1+x) & \text{if} \quad x \geq 0, \lambda = 0 \\ -\{(1-x)^{2-\lambda}-1\}/(2-\lambda) & \text{if} \quad x < 0, \lambda \neq 2 \\ -\log(1-x) & \text{if} \quad x < 0, \lambda = 2 \end{cases} \tag{2.9}$$

The parameter λ of $g^{YJ}(x,\lambda)$ can be estimated using the maximum-likelihood technique (Yeo and Johnson, 2000).

2.2.5 Implementation of the transformation-based approach

A departure of the observed return r from normality can be identified via three possible scenarios: (i) if the measure of skewness ($\sqrt{\beta_1}$) is non-zero; (ii) if the measure of kurtosis (β_2) is significantly different from 3 (i.e., if excess kurtosis is non-zero); (iii) if both the previous reasons hold true. Denoting the measure of skewness $\sqrt{\beta_1} = \mu_3/\mu_2^{(3/2)}$ and the measure of kurtosis $\beta_2 = \mu_4/\mu_2^2$, where μ_j denotes the j^{th} order central moment ($j \geq 2$), and noting that for normal distribution, $\sqrt{\beta_1} = 0$ and $\beta_2 = 3$, the following hypotheses are usually tested for normality.

(i) H_{01}: $(\sqrt{\beta_1},\beta_2) = (0,3)$, which will be tested against the alternative hypothesis H_{11}: $(\sqrt{\beta_1},\beta_2) \neq (0,3)$.

(ii) H_{02}: $\sqrt{\beta_1} = 0$, which will be tested against the alternative hypothesis H_{12}: $\sqrt{\beta_1} \neq 0$.

(iii) H_{03}: $\beta_2 = 3$, which will be tested against the alternative hypothesis H_{13}: $\beta_2 \neq 3$.

The normality null hypothesis H_{01} can be tested against H_{11} using Jarque and Bera's (1987) test statistics given by $Q = n[\,(b_1)^2/6 + (b_2 - 3)^2/24]$, where b_1 and b_2 are sample estimates of $\sqrt{\beta_1}$ and β_2, respectively, and n is the number of observations used to derive these estimates. Under H_{01}, Q is the asymptotically χ^2 variable with 2 degrees of freedom. Further, under normality, b_1 and b_2 are both asymptotically normally distributed with mean zero and variances 6/n and 24/n, respectively, implying that both $[n\,(b_1)^2/6]$ and $[n\,(b_2-3)^2/24]$ are asymptotically χ^2 variables with 1 degree of freedom.

For implementing the transformation-based approach, we first need to check whether the underlying distribution deviates from normal. The acceptance of the null hypothesis H_{01} would indicate that the returns follow a normal distribution, which means that the VaR can be estimated easily via the expression given in Eq. (2.3). However, if H_{01} could not be accepted, the transformation-based approach may come in handy. In such cases, three possible scenarios could arise. First, if the distribution of r turns out to be skewed (signalled by the simultaneous non-acceptance of H_{02} and acceptance of H_{03}), we may transform the original return r via $g^{YJ}(r,\lambda)$ to near normality (for a suitably chosen value for the constant λ) and estimate VaR via the formula given in Eqs. (2.6)–(2.7). Second, if the distribution of r is symmetric with significant excess kurtosis (signalled by the simultaneous acceptance of H_{02} and non-acceptance of H_{03}), we may transform r to normality via the $g^{JD}(x,\delta)$ transformation (for a suitably chosen value for the constant δ) and estimate VaR using Eqs. (2.6)–(2.7). However, the distribution of r may be skewed with significant excess kurtosis (when neither H_{02} nor H_{03} could be accepted); in such cases, we use a composite transformation $g^{JD}(g^{YJ}(r,\lambda), \delta)$ for suitably chosen values for (λ,δ).

From a practical perspective, it would be convenient to transform non-normal r through the composite transformation $g^{JD}(g^{YJ}(r,\lambda), \delta)$ for suitably chosen values for (λ,δ). In such cases, one has to examine whether the individual transformations $g^{YJ}(r,\lambda)$ and $g^{JD}(r,\delta)$ do not distort the distributional properties of r if it truly follows a normal distribution, i.e., the transformations should preserve normality even if were applied on truly normally distributed variables. For the sake of simplicity, we employ this convenient strategy throughout the paper, i.e., we transform the observed return r through the transformation $g^{JD}(g^{YJ}(r,\lambda), \delta)$ for suitably chosen values for (λ,δ).

3. Simulation Exercise to Evaluate Transformations to Normality

It is difficult to theoretically compare and assess the power of the different families of transformation for achieving normality. In this paper, we attempt to address this issue using a simulation study. We generate observations randomly from various non-normal distributions (i.e., skewed and/or heavy-tailed, including mixture distributions), and we examine how the chosen transformations perform in terms of converting these random observations to normality. Further, we examine how efficient the transformation is in preserving normality when applied to truly normally distributed random observations. The alternative classes of distributions/models considered in this simulation study are:

- **Student-*t* distribution:** Symmetric but fat-tailed distribution.
- **Skewed-Laplace:** Skewed and fat-tailed distribution (which can be seen as a mixture distribution).
- **ARCH/GARCH:** Model observed phenomenon of volatility clustering in returns, which leads to a fat-tailed unconditional return distribution.
- **Normal Distribution:** This ensures that the transformation does not convert truly normally distributed random observations to some distribution other than normal.

> In practice, it is quite possible that true normality remains undetected for various reasons or the transformation is applied on truly normal observations by chance. In such cases, the normality property should not be disturbed by the transformation.

The alternative forms of distributions/models considered in the simulation study and the strategies adopted in drawing random observations from each of the chosen distributions/models are discussed in the following section.

3.1 Data Generation from Alternative Distributions/Models

3.1.1. Student-*t* distribution (symmetric but heavy-tailed distribution)

The Student-*t* distribution is symmetric and heavy-tailed. A random variable following a Student-*t* distribution with ν degrees of freedom (denoted by t_ν-*distribution*) has the expected value 0 and variance $\nu/(\nu - 2)$, if $\nu > 2$. The skewness of the distribution is 0 if $\nu > 3$, and the excess kurtosis is $6/(\nu - 4)$ if $\nu > 4$. It may be noted that as the parameter ν becomes larger, the excess-kurtosis approaches zero, implying that the distribution t_ν-*distribution* tends towards normal distribution as the degree of freedom ν tends to infinity. Further, the maximum value of excess kurtosis of a t_ν-distribution when it exists (i.e., when $\nu > 4$) is obtained when $\nu = 5$; the maximum value of excess kurtosis here is 6.

A random observation from t_ν-*distribution* can be obtained using a sample of (ν+1) observations drawn randomly from normal distribution (with known mean μ and unknown standard deviation σ) as $t = \frac{(\bar{x} - \mu)\sqrt{(\nu+1)}}{S}$, where $\bar{x}$ is the sample mean and S is the sample standard deviation. In our simulation exercise, we consider $\nu = 5$, which rendered the maximum possible excess kurtosis value (i.e., 6) for t_ν-*distribution*.

3.1.2. Skewed-Laplace distribution

The skewed-Laplace distribution is sometimes used to model the return distribution for financial portfolios. For example, Linden (2001) showed that if the stock returns conditional on the risk (variance) σ^2 follows a $N(0,\sigma^2)$ distribution, and if the risk σ^2 of the stock returns follows an exponential distribution with the probability density function $g(\sigma^2)=\alpha \exp[-\alpha\sigma^2]$ (where $\alpha > 0$ is a constant parameter), the mixture (unconditional) distribution of returns would follow a Laplace distribution of the form $f(x,\lambda) = (\alpha/2)\exp(-\alpha|x|)$, where |x| denotes the absolute value of x, $-\infty < x < \infty$. Alternatively, if we consider $\lambda = 1/\alpha$, the distribution can be expressed as $f(x,\lambda) = \exp(-x/\lambda)/(2\lambda)$, $\lambda>0$ and $-\infty < x < \infty$. This distribution is symmetric about mean zero, but it has excess kurtosis 6.

In a more general situation, a skewed-Laplace distribution can be obtained by considering different values for λ in positive and negative regions ($\lambda = \lambda_1$ for the region $x \leq 0$, and $\lambda = \lambda_2$ for the region $x > 0$) in support of the variable x. The probability density function of a skewed-Laplace distribution (Linden, 2001; Puig and Stephens, 2007) can be expressed as:

$$
f(x, \lambda_1, \lambda_2) = \begin{cases} \dfrac{1}{(\lambda_1 + \lambda_2)} e^{-\frac{|x|}{\lambda_1}} & \text{if } x \le 0 \\[2ex] \dfrac{1}{(\lambda_1 + \lambda_2)} e^{-\frac{|x|}{\lambda_2}} & \text{if } x > 0 \end{cases} \tag{3.1}
$$

The coefficients of skewness ($\sqrt{\beta_1}$) and kurtosis (β_2) of the skewed-Laplace distribution are as follows:

$$
\sqrt{\beta_1} = \frac{2(\lambda_2^3 - \lambda_1^3)}{(\lambda_1^2 + \lambda_2^2)^{3/2}} \text{ and } \beta_2 = 3 + \frac{6(\lambda_2^4 + \lambda_1^4)}{(\lambda_1^2 + \lambda_2^2)^2} \tag{3.2}
$$

Thus, the Laplace distribution is skewed positively (negatively) if $\lambda_2 > \lambda_1$ ($\lambda_2 < \lambda_1$). In particular, if $\lambda_1 = \lambda_2$, the distribution is symmetric around zero. Further, the Laplace distribution always has positive excess kurtosis, meaning that it is a heavy-tailed distribution.

Puig and Stephens (2007) discuss three alternative approaches for drawing random numbers from a skewed-Laplace distribution. Of these three, we use the classical inverse distribution method using the random observations from a uniform distribution over the range (0,1). If u ε (0,1) represents a random observation from the uniform (0,1) distribution, a random observation x from the skewed-Laplace distribution may be obtained by equating the cumulative distribution function (c.d.f.) of the skewed-Laplace distribution with u, i.e., by using the following expression:

$$
x = \begin{cases} \lambda_1 \log_e\left(\dfrac{\lambda_1 + \lambda_2}{\lambda_1} u\right) & \text{if } 0 < u \le \dfrac{\lambda_1}{\lambda_1 + \lambda_2} \\[2ex] \lambda_2 \log_e\left(\dfrac{\lambda_2}{(\lambda_1 + \lambda_2)(1-u)}\right) & \text{if } \dfrac{\lambda_1}{\lambda_1 + \lambda_2} < u < 1 \end{cases} \tag{3.3}
$$

The daily returns on financial assets are usually negatively skewed (see Linden, 2001 for a discussion of the returns on stocks). Accordingly, in our simulation study, we considered $\lambda_2 < \lambda_1$. For our simulation exercise, we fix $\lambda_1 = 3$ and $\lambda_2 = 2$, which resulted in $\sqrt{\beta_1} = -0.8107$ and $\beta_2 = 6.4438$ (i.e. excess kurtosis$= 3.4438$).

3.1.3. ARCH/GARCH model

The volatility clustering phenomenon and the risk-return trade-off in financial market returns are sometimes modelled through Auto-Regressive Conditional Heteroscedastic (ARCH) or Generalised-ARCH (GARCH) models, or some of their more advanced forms. These models can lead to heavy-tailed unconditional distributions of the returns. The broad structures of the ARCH and GARCH models considered in the simulation study[2] are given here:

ARCH model: $X_t = \theta_0 + \theta_1 X_{t-1} + \theta_2 X_{t-2} + \ldots\ldots + \theta_l X_{t-l} + \varepsilon_t$ (3.4)

$$\varepsilon_t \mid \psi_{t-1} \sim N(0, h_t) \text{ and } h_t = \alpha_0 + \alpha_1 e_{t-1}^2 + \ldots.. + \alpha_p e_{t-p}^2 \quad (3.5)$$

where Ψ_{t-1} is the information set available at time t-1; ε_t is the usual noise; h_t is the conditional variance of ε_t; l and p are positive integers; and $\alpha_0 > 0$, $\alpha_j \geq 0$, j = 1,2, ….,p, θ_k's, k = 1,2,…..,l are constant parameters.

GARCH model: $X_t = \theta_0 + \theta_1 X_{t-1} + \theta_2 X_{t-2} + \ldots\ldots + \theta_l X_{t-l} + \varepsilon_t$ (3.6)

$$\varepsilon_t \mid \psi_{t-1} \sim N(0, h_t) \text{ and } h_t = \alpha_0 + \sum_{j=1}^{p} \alpha_j e_{t-j}^2 + \sum_{k=1}^{q} \beta_k h_{t\text{-}k} \quad (3.7)$$

where Ψ_{t-1} is the information set available at time t-1; ε_t is the usual noise; h_t is the conditional variance of ε_t; p and q are positive integers; θ_i's, i = 1,2,…..,l $\alpha_0 > 0$, $\alpha_j \geq 0$, j = 1,2, ….,p, and $\beta_k \geq 0$, k = 1,2,….,q, are constant parameters.

Eq. (3.5) is known as the ARCH(p) process, and Eq.(3.7) is known as the Generalised-ARCH(p,q) process, i.e., the GARCH(p,q) process. The unconditional/stationary variance of the ARCH(p) process exists if $\sum_{j=1}^{p} \alpha_j < 1$; the unconditional variance is $\alpha_0 / \left(1 - \sum_{j=1}^{p} \alpha_j\right)$. The GARCH(p,q) process requires $\left(\sum_{j}^{p} \alpha_j + \sum_{k=1}^{q} \beta_k\right) < 1$ for unconditional variance to exist; the expression for the variance is $\alpha_0 / \left(1 - \sum_{j=1}^{p} \alpha_j - \sum_{k=1}^{q} \beta_k\right)$.

If e_t in the ARCH process (including the GARCH process) is conditionally normal, its unconditional distribution is symmetric (Bera and Higgins, 1993). Further, the ARCH model is known to have greater kurtosis than that of normal distribution (i.e., excess kurtosis), though the close-form expression for kurtosis of the GARCH process is not known. For special cases, however, Engle (1982) provided the expression for kurtosis for the ARCH(1) process as $\frac{E(e_t^4)}{[Var(e_t)]^2} = 3\left(\frac{1-\alpha_1^2}{1-3\alpha_1^2}\right) = 3 + \frac{6\alpha_1^2}{(1-3\alpha_1^2)}$, if $3\alpha_1^2 < 1$; Bollerslev (1986) provides the expression for kurtosis of the GARCH(1,1) process as $\frac{E(e_t^4)}{[Var(e_t)]^2} = 3 + \frac{6\alpha_1^2}{(1-\beta_1^2 - 2\alpha_1\beta_1 - 3\alpha_1^2)}$, if $(\beta_1^2 + 2\alpha_1\beta_1 + 3\alpha_1^2) < 1$. If kurtosis does exist for

[2] Bera and Higgins (1993) provide a detailed account of the properties of ARCH models (including the original Generalised-ARCH models and subsequent developments).

the ARCH(1) and GARCH(1,1) processes, it is greater than 3 (i.e., greater than the kurtosis of normal distribution), meaning that the unconditional distribution of ε_t is heavy-tailed.

In our simulation exercise, we generated random observations from the ARCH(1) process with the specification

$$\varepsilon_t \mid \psi_{t-1} \sim N(0, h_t) \text{ and } h_t = 0.9 + 0.4\varepsilon_{t-1}^2 \tag{3.8}$$

The unconditional variance of ε_t in this case is 0.9/(1-0.4) = 1.5, and the kurtosis is $\left[3 + \frac{6\alpha_1^2}{\left(1 - 3\alpha_1^2\right)}\right] = 4.8462$ (i.e., excess kurtosis = 1.8462). Following Bera and Higgins (1993), the data-generating process adopted for drawing random observations from the ARCH(1) process is

$$\varepsilon_t = \eta_t \sqrt{h_t} \text{ and } h_t = 0.9 + 0.4\varepsilon_{t-1}^2, \text{ where } \eta_t \sim N(0,1) \tag{3.9}$$

For drawing a random sample of size n from the ARCH(1) process in Eq. 3.8, we assume the process began at time t = 0, and we fix the initial value to be $h_0 = 1.5$, which is the unconditional variance of the chosen ARCH(1) process. The initial value of the process ε_0 is fixed at $\varepsilon_0 = \eta_0 \sqrt{h_0}$, where η_0 is a randomly drawn observation from N(0,1). Thereafter, we repeatedly generate the observations of ε_t (500 + n) times; in each repetition, we first draw a random observation of η_t from N(0,1), and subsequently derive the random observation for ε_t following the generating process given in Eq. (3.9). Finally, the last n-observations (of the 500 + n observations thus generated) are considered for the simulation exercise; the first 500 observations are excluded in an attempt to eliminate possible initial value effect.

The specification of the GARCH model considered in the simulation exercise is given in Eq. (3.10).

$$\varepsilon_t \sim N(0, h_t) \text{ and } h_t = 0.9 + 0.2\varepsilon_{t-1}^2 + 0.4\,h_{t-1} \tag{3.10}$$

The unconditional variance of ε_t in Eq. (3.10) is [0.9/(1-0.2-0.4)] = 3, and the kurtosis is $\left[3 + \frac{6\alpha_1^2}{\left(1 - \beta_1^2 - 2\beta_1\alpha_1 - 3\alpha_1^2\right)}\right] = 4.6364$ (i.e., excess kurtosis = 1.6364). Similar to the case of the ARCH(1) process, the data-generating process for the chosen GARCH(1,1) process is

$$\varepsilon_t = \eta_t \sqrt{h_t} \text{ and } h_t = 0.9 + 0.2\varepsilon_{t-1}^2 + 0.4\,h_{t-1}, \text{ where, } \eta_t \sim N(0,1) \tag{3.11}$$

The strategy adopted for drawing n observations from the GARCH process is similar to that for the ARCH process, except that in this context, the expression for h_t depends on both ε_{t-1} as well as h_{t-1}, as given in Eq. (3.11).

3.2 Simulation strategy

For the simulation study, we considered several alternative skewed and/or heavy-tailed distributions for returns (log-returns). As the returns can take positive as well as negative values, we considered the distributions to have real-line as support. For each chosen distribution/model, the simulation study was carried out in the following steps:

- **Step 1:** Draw n random observations from the given distribution/model. Let these observations be represented by $r_1, r_2, \ldots , r_n$. In our simulation exercise, we chose n = 500.

- **Step 2:** Apply the transformation to normality on these observations. For simplicity, the normality transformation is conducted in two phases irrespective of the data-generating distribution/process.

  First, the original observations $r_1, r_2, \ldots , r_n$ are passed through the $g^{YJ}(.,\lambda)$ transformation to reduce/remove possible skewness. Let λ^* be the estimated value of the parameter λ.

  Second, the observations $g^{YJ}(r_t,\lambda^*)$, $t=1,2,\ldots,n$ are passed through the $g^{JD}(.,\delta)$ transformation to eliminate/cure possible excess kurtosis. Let δ^* be the estimated value of the parameter δ.

  Thus, for the original observation r_t, the final transformed observation y_t, is obtained as $y_t = g^{JD}[\, g^{YJ}(r_t, \lambda^*), \delta^*]$, $t=1,2,\ldots,n$. The transformation parameters λ and δ may be estimated by maximizing the likelihood function (maximum-likelihood method). Further, we estimated the parameter heuristically by minimizing the magnitude of skewness/excess kurtosis, which we call the heuristic approach.

- **Step 3:** Compute the measures of skewness and excess kurtosis based on the transformed observations $y_1, y_2, \ldots ,y_n$. At the i^{th} repetition, let S_i and K_i represent the measures of skewness and excess kurtosis thus calculated.

  If normality is achieved, both these values should be zero (or statistically insignificant). Statistical tests were performed on $y_1, y_2, \ldots ,y_n$ in each repetition for the null hypotheses of (i) skewness = 0; (ii) excess kurtosis = 0; and (iii) normality (Jarque-Bera test). Each test was performed for two alternative sizes (0.01 and 0.05).

- **Step 4:** Repeat steps 1–3 T times. In our simulation exercise, T was fixed at T = 10,000. Compute the average values of S_i and K_i for $i=1,2,\ldots,T$. If the transformation were successful in inducing normality, both these averages would be close to zero. Further, compute how frequently (proportion of T repetition) each of the three null hypotheses has been accepted: (i) skewness = 0; (ii) excess kurtosis = 0; and (iii) normality test, i.e., joint test of skewness = 0 and excess kurtosis = 0 (Jarque-Bera test). Compute this proportion separately under 1% and 5% levels of significance. A greater proportion of acceptance of these null hypotheses would indicate better performance of the transformation in achieving normality.

This 4-step simulation strategy was implemented on the random observations drawn from different non-normal distributions (i.e., skewed/leptokurtosis distributions such as the Student-*t* distribution and Laplace distribution) and data-generating processes (such as ARCH/GARCH that model volatility-clustering phenomenon of financial market returns and can capture heavy-tailed return distribution). The process of simulation was repeated 10,000 times for any given distribution/model, and in each repetition, 500 random observations were drawn from the given distribution/model.

3.3 Simulation Results

The simulation exercise was intended to examine how good the normality transformation was in transforming the random observations from different probability distributions or data-generating models/processes.

For a given distribution/model, we first drew 500 observations randomly and applied the normality transformation. The transformation parameters were chosen in two ways: maximum-likelihood method and heuristic approach.

Under each strategy of choosing the transformation parameters, we computed the measure of skewness and excess kurtosis for the transformed observations. If the transformations were good, both these measures would be zero or statistically insignificant. Therefore, we tested the significance of skewness, excess-kurtosis, and joint skewness-excess kurtosis (Jarque-Bera test of normality using skewness and excess kurtosis).

The simulation exercise was repeated 10,000 times separately for each class of probability distribution or model. The simulated average values of the measures of skewness and excess kurtosis of the transformed observations based on the 10,000 repetitions are reported in Table 3.1. The proportion of times (out of 10,000) each of the three hypothesis related to the normality (H_{01}, H_{02}, and H_{03}) of the transformed observations were accepted at the 1% and 5% levels of significance are presented in Table 3.2. If the transformation successfully converted a distribution to normality, the corresponding proportion of acceptance of the null hypothesis would be close to 0.95 for the 5%level of significance and 0.99 for the 1% significance level.

Table 3.1: Normality of Transformed Observations
(Measures of Skewness & Excess Kurtosis)

True Model/ Distribution of Original Observations (*which were transformed to Normality*)	Average Skewness#	Average Excess Kurtosis#
(A) Maximum-Likelihood Estimates of Transformation Parameters		
Normal	0.0004	-0.0187
Student-t distribution (ν= 5)	0.0001	0.1701
Laplace ($\lambda_1 = 3$, $\lambda_2 = 2$)	0.1526	0.1014
ARCH(1)	-0.0002	0.1513
GARCH(1,1)	0.0002	0.0580
(B) Heuristic Approach to Estimate Transformation Parameters		
Normal	0.0001	0.0001
Student-t distribution (ν= 5)	0.0001	0.0000
Laplace ($\lambda_1 = 3$, $\lambda_2 = 2$)	0.1624	0.0001
ARCH(1)	-0.0003	-0.0001
GARCH(1,1)	-0.0001	-0.0000

Based on 10,000 repetitions, with 500 observations in each repetition.

Table 3.2: Proportion of Acceptance of Null Hypotheses Related to Normality of Transformed Observations#

True Distribution/Model of Original Observations (*which were transformed to Normality*)	Null Hypothesis Tested (Test Size = 0.01)			Null Hypothesis Tested (Test Size = 0.05)		
	$\beta_1 = 0$	$\beta_2 = 0$	$(\beta_1, \beta_2) = (0,0)$	$\beta_1 = 0$	$\beta_2 = 0$	$(\beta_1, \beta_2) = (0,0)$
(A) Maximum-Likelihood Estimates of Transformation Parameters						
Normal	1.0000	0.9991	0.9994	1.0000	0.9965	0.9988
Student-t distribution (ν=5)	0.9998	0.9930	0.9969	0.9983	0.9637	0.9843
Laplace ($\lambda_1 = 3$, $\lambda_2 = 2$)	0.9991	0.9992	0.9985	0.9558	0.9931	0.9765
ARCH(1)	1.0000	0.9810	0.9908	1.0000	0.9422	0.9743
GARCH(1,1)	1.0000	0.9946	0.9975	1.0000	0.9820	0.9927
(B) Heuristic Approach to Estimate Transformation Parameters						
Normal	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Student-t distribution (ν=5)	0.9997	1.0000	0.9999	0.9954	1.0000	0.9995
Laplace ($\lambda_1 = 3$, $\lambda_2 = 2$)	0.9731	1.0000	0.9959	0.8290	1.0000	0.9599
ARCH(1)	1.0000	1.0000	1.0000	0.9998	1.0000	1.0000
GARCH(1,1)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Based on 10,000 repetitions, with 500 observations in each repetition.

The simulated results presented in Table 3.1 show that the transformation strategy was able to transform the observations drawn from alternative non-normal distributions/models (the Student-t distribution, skewed-Laplace distribution, and ARCH and GARCH models) to normality reasonably well. This is indicated by the relatively low average values of the measures of skewness and excess kurtosis (although for the Laplace distribution, the skewness does not appear to be removed completely). The simulation results presented in Table 3.2 are quite interesting. When the transformation parameters are estimated through the maximum-likelihood method (the grid search method was adopted in this case), the proportion of acceptance of the null hypotheses is consistent with the size of the test. Even in the case of the heuristic approach to parameter estimation, the performance is equally good in all aspects except when testing $\sqrt{\beta_1} = 0$ at the 5% level of significance with the data originally drawn from the Laplace distribution (the corresponding proportion of acceptance 0.8290 is quite low as compared to the expected value of 0.95).

4. Empirical Analysis

The simulation exercise points out that the transformation considered above performs reasonably well in terms of converting the chosen non-normal distributions to (approximate) normality. Given that the chosen class of distributions cover the typically observed return distributions, the estimation of VaR through the transformation-based method appears to be quite sensible. Indeed, the simulated results provide an explanation for the empirical results reported by Samanta (2008) in support of the transformation-based VaR measurement. However, it is imperative to examine the robustness of such empirical findings over time, particularly in the years following the global financial crisis. In this section, we report the results of the empirical analysis.

4.1 Data

The performance of a VaR measurement technique may be examined with respect to certain real-life portfolios, i.e., the portfolios held by banks or investors. However, such portfolios are held privately, and hardly any information about their composition and other details are made public. This situation resulted in scarce reporting of empirical results based on real-life portfolios.[3] Most of the prior empirical studies relied on publicly available historical data, such as asset prices or indices.[4] Similarly, we employ daily data on the exchange rate of the Indian Rupee (INR) with respect to four major international currencies in India, viz., US Dollar, British Pound Sterling, Euro, and Japanese Yen; these currencies were covered in Samanta (2008) as well. The choice of four common exchange rates allows for the comparison and robustness check of the empirical results over time.

[3] Berkowitz and O'Brien (2002) is one such rare empirical study using proprietary data of banks.
[4] See Bauer (2000); Christoffersen et al. (2001); Mikael (2001); Sarma et al. (2003); Samanta (2008).

The daily exchange rates covering the period from January 1, 2009 to March 31, 2014 were collected from the database on the Indian economy available on the RBI Website (http://www.rbi.org.in). For this period, we obtained 1268 daily observations for each of the four exchange rates considered. For our analysis, "return" refers to "log-return". For any given exchange rate, the return (i.e., log-return) on a particular day, say the t^{th} day, is computed as follows:

$$R_t = 100^* [\log_e(P_t) – \log_e(P_{t-1})] \tag{4.1}$$

where P_t and R_t denote the values of the given exchange rate and the daily return on t^{th} day, respectively.

We assess the performance of the transformation-based VaR model using widely used techniques: the normal (variance-covariance) method and the extreme value approach using tail-index. In the latter approach, the tail-index is estimated via two alternative techniques, viz., Hill's estimator (Hill, 1975) and the ordinary least squares (OLS) estimator discussed in van den Goorbergh (1999). Thus, we evaluate the performance of four competing VaR models:

(i) Normal method: VaR was estimated under the assumption of the normality of log-return.
(ii) Extreme value theory : tail-index was estimated via Hill's estimator
(iii) Extreme value theory: tail-index was estimated via OLS regression
(iv) Transformation-based approach

The description of the direct VaR estimation techniques considered here are available in the standard literature. Appendix A provides a summary for ready reference.

Samanta (2008) assessed the performance of the transformation-based approach along with two competing VaR methods: the normal/covariance method (which assumes normality of return); and an approach based on tail-index estimated using Hill's estimator. Thus, we undertake our empirical assessment against a broader set of competing techniques (the approach measuring tail-index through regression analysis is an additional alternative in this study). Competing methods are applied for univariate series of daily portfolio/asset returns. The observed phenomenon of volatility clustering could be modelled through the classes of conditional heteroscedastic models. Alternatively, since we know that conditional heterescedasticity induces heavy-tails in unconditional distribution, we could put efforts into modelling the fat-tailed unconditional distribution of returns.

We consider 1-day VaR, expressed in percentage form at 99% confidence level. This means that for a good estimate of VaR, the theoretical probability of the realized daily return exceeding VaR equals 0.01; i.e., VaR exception/violation may occur in one out of 100 days. For underestimated (overestimated) VaR, the observed frequency of VaR exception would be significantly higher (lower) than 1%.

On any given date (say the t^{th} day), we estimate VaR for the $(t+1)^{th}$ day or future dates by adopting two alternative databases: full-sample estimates, obtained using historical returns from the starting time point of the database till the estimation date; and rolling-sample estimates, which are computed based on a fixed number of most recent returns (i.e., the returns on the date of computation as well as a pre-specified number of immediate preceding days). The number of latest returns considered for the rolling-sample estimate is called the rolling-sample/window size; in this study, the rolling sample included returns for the last 500 days.

4.2 Testing for Normality of Returns

The presence of volatility clustering in the market indicates that asset returns would seldom follow normal distribution unconditionally. Therefore, our empirical study begins by testing for the normality of returns (i.e., log-returns). The empirical results related to the normality hypotheses are given in Table 4.1. Table 4.1 shows that the Jarque-Bera test could not accept the normality hypothesis at any conventional level of significance (the *p*-values of the test statistics corresponding to the null hypotheses were much lower than 0.01). Further, significant excess kurtosis appears to be the main source of deviation from normality for all the return series except for US Dollar, where skewness is also statistically significant.

Table 4.1: Testing Normality of Returns on Exchange Rates

Asset/ Portfolio	Measure of Skewness	χ_1^2 for Skewness (Testing H_{02})	Excess Kurtosis	χ_1^2 for Excess Kurtosis (Testing H_{03})	Jarque-Bera Statistics (Testing H_{01})
US Dollar	0.2342**	11.5781** (0.0000)	3.2987**	574.4640** (0.0000)	586.0420** (0.0000)
Pound Sterling	-0.0819	1.4170 (0.2339)	2.3183**	283.7227** (0.0000)	285.1397** (0.0000)
Euro	0.0575	0.6978 (0.4035)	3.0266**	483.5957** (0.0000)	484.2935** (0.0000)
Japanese Yen	-0.0101	0.0214 (0.8838)	1.5208**	122.0979** (0.0000)	122.1192** (0.0000)

Figures in parentheses indicate significance level (i.e., *p*-value).
* and ** indicate significance at 5% and 1% levels, respectively.

4.3 Empirical Results: Transformations of log-returns to normality

The transformation parameters (λ,δ) are estimated using two alternative approaches: maximum likelihood and heuristic. For each of the alternatives, the optimization is done (in two stages, as discussed above) through a grid search, i.e., by looking for optimal values of λ and δ from a set of potential alternatives. Based on the empirical assessment, the set of potential values for λ is {-2.000, -1.999, ………, 1.999, 2.000} and that for δ is {0.000, 0.001, ……, 1.999, 2.000}. In simulation exercises, the estimation of these parameters using either the maximum-likelihood approach or the heuristic approach produces rather similar results; therefore, we adopted the maximum-likelihood approach in empirical analysis. Table 4.2 presents the maximum likelihood estimates of (λ,δ) for transforming each log-return to (near) normality.

Table 4.2 also presents the results of the normality tests for all the transformed returns. The normality transformation could cure the skewness/kurtosis problem with respect to almost all the log-return series except that of Euro, where some bit of kurtosis persists. The interesting point here is that the degree of excess kurtosis for the transformed returns on Euro is a lot milder, while H_{03} for the transformed series is accepted at the 1% level of significance (though not at 1% level) as evident from the corresponding *p*-value of 0.0382; the same hypothesis for original returns on Euro could not be accepted at any of these conventional levels.

Table 4.2: Testing Normality of Transformation of Returns

Asset/ Portfolio	**Transform-ation Parameters**	**Measure of Skewness**	**χ_1^2 for Skewness (Testing H_{02})**	**Excess Kurtosis**	**χ_1^2 for Excess Kurtosis (Testing H_{03})**	**Jarque-Bera Statistics (Testing H_{01})**
US Dollar	$\hat{\lambda} = 0.917$ $\hat{\delta} = 0.124$	-0.0201	0.0853 (0.7702)	0.1234	0.8033 (0.3701)	0.8886 (0.6413)
Pound Sterling	$\hat{\lambda} = 1.036$ $\hat{\delta} = 0.265$	-0.0289	0.1766 (0.6743)	0.1241	0.8132 (0.3672)	0.9898 (0.6096)
Euro	$\hat{\lambda} = 0.993$ $\hat{\delta} = 0.305$	-0.0436	0.4006 (0.5268)	0.2853*	4.2958* (0.0382)	4.6964 (0.0955)
Japanese Yen	$\hat{\lambda} = 0.999$ $\hat{\delta} = 0.475$	0.0041	0.0035 (0.9525)	0.0769	0.3125 (0.5761)	0.3161 (0.8538)

Figures in parentheses indicate significance level (i.e. p-value). The symbol ^ indicates estimates.
* and ** indicate significance at 5% and 1% levels, respectively.

4.4 Empirical Evaluation of VaR Estimates/Models

The VaR numbers estimated through a particular technique may be evaluated using different criteria based on the frequency of VaR-exception, the magnitude of VaR-exception (i.e., the excess loss over VaR at the instance of a VaR-exception), or both. The frequency-based evaluation of a VaR model can be done through a suitable test of the proportion of VaR-exception, such as statistical backtesting (suggested by regulators or the Basel Accord, more standard statistical tests (such as those suggested by Kupiec, 1995), or further sophisticated tests (such as that proposed by Christoffersen, 1998; Christoffersen et al., 2001). The severity of loss depends on the magnitude of excess loss, which is incorporated in the frequency-based evaluation criteria. Several assessment criteria incorporating the frequency as well as the magnitude of excess-losses were proposed by Lopez (1998) and Sarma et al. (2003).

In our empirical evaluation, the backtesting period covers the last 500 days in the database. On every backtesting day, the 1-day VaR at 99% confidence level was estimated based on all competing models/techniques using the rolling-sample (with size 500 days) and the full-sample strategies. Further, VaR was estimated separately for the left-tail and the right-tail of the return distributions (i.e., for long and short financial positions, respectively, on the asset). The alternative assessment/evaluation criteria involved two frequency-based assessments as well as two loss functions.

4.4.1 Frequency of VaR-exception

A simple criterion is to examine whether the proportion of VaR-exception over a number of days (i.e., the backtesting period) is consistent with the confidence level of the VaR numbers. For example, for a good VaR number representing maximum loss at 99% confidence level, the theoretical probability of loss being above the VaR estimate would be 0.01. Thus, ideally, for about 1% of the backtesting days, one may expect VaR exception. However, VaR-exception significantly higher than 1% would indicate the under-estimation of VaR, a situation that worries regulators.[5] Regulators may like to see a VaR model that does not generates VaR-exception at a frequency higher than 1%. Therefore, in order to assess a VaR model, one may assign a score of 1 at the instances of a VaR-exception, and 0 otherwise. Finally, the total score over a period of backtesting days (i.e., the sum of the daily scores over all the backtesting days) can be expressed in percentage form, and we can examine whether the observed percentage is close to the expected 1% level. A lower observed score or percentage of VaR-exception would be preferred by the regulators. Thus, on any backtesting day (say the t^{th} day), each VaR model is assigned a score Z_t as follows:

$$Z_t = \begin{cases} 1 & \text{if } L_t > V_{t|t-1} \\ 0 & \text{if } L_t \leq V_{t|t-1} \end{cases} \tag{4.2}$$

where $V_{t|t-1}$ and L_t are the VaR estimated for the t^{th} day given the data/information up to time (t-1) and the observed loss on the t^{th} day, respectively.

The total score over n backtesting days would be $Z = (Z_1 + Z_2 + \ldots.. + Z_n)$, i.e., the observed percentage of VaR-exception would be $(100 \times Z/n)$. By construction, $0 \leq Z \leq n$; therefore, $0 \leq 100 \times Z/n \leq 100$. For a good VaR model, the observed value of $(100 \times Z/n)$ should be close to the theoretical value, i.e., 1 (or less from the regulators' perspective) for VaR estimates at the 99% confidence level.

For the empirical implementation of this scoring mechanism, the strategy adopted in this study to compare loss and estimated VaR is similar to that used in Bauer (2000). The backtesting period of 500 days is first partitioned into 50 blocks, each covering 10 days. Subsequently, the 99% daily VaR is estimated for day 1 of Block 1, which is assumed to remain unchanged throughout the block; we assigned scores to all 10 days in the block by

[5] Regulators may not be too concerned (in terms of the level of capital adequacy) if VaR-exception occurs at a frequency less than 1%, for instance, indicating the possible over-estimation of VaR. However, an individual bank or investor may have reasons to avoid too much of such VAR over-estimation from the perspective that maintaining additional capital (over the ideal amount) may adversely affect profitability. Accordingly, one can analyse such situations by selecting appropriate criteria in line with the particular form of loss-function suggested by Sarma et al. (2003) or even the more general form of loss function proposed by Lopez (1998).

comparing the daily loss/returns with the estimated VaR. Finally, repeating the process for all the 50 Blocks, we calculated the values of Z, which would provide the basis for evaluating a VaR model.[6] We report the values of Z expressed in percentage form (i.e., percentage of VaR-exception over 500 backtesting days) in Table 4.3. As can be seen from Table 4.3, the frequency of VaR exceptions varies considerably across the competing models; at times, it appears to be considerably higher than the 1% threshold. The normality assumption for return distribution is seen to produce relatively higher VaR-exception. The tail-index approaches (which are usually believed to capture fat-tails better than the normal approach does), generate relatively less frequent VaR-exceptions. Interestingly, the performance of the transformation-based approach is not bad compared to that of tail-index approaches. Despite being simple to understand, the performance of the transformation-based approach appears to be uniformly better than that of the other alternatives in almost all the cases, i.e., short and long positions on different assets except for long position on Japanese Yen (in both full-sample and rolling-sample results).

Table 4.3: Percentage of VaR Violations by Competing Models

Asset/ Portfolio	Left-Tail (Long Financial Position)				Right-Tail (Short Financial Position)			
	Normal Method	Tail-Index		Trans-based Method	Normal Method	Tail-Index		Trans-based Method
		Hill's	OLS			Hill's	OLS	
(A) Full-Sample Results								
US Dollar	2.4	2.0	2.0	1.6	2.6	2.0	1.8	1.8
Pound Sterling	1.0	0.4	0.4	0.4	1.4	1.2	1.2	1.2
Euro	1.2	0.6	0.6	0.6	1.2	1.2	1.2	1.2
Japanese Yen	1.4	0.6	0.6	0.8	1.8	1.6	1.6	1.4
(B) Rolling-Sample Results								
US Dollar	2.2	2.2	1.8	1.8	2.2	1.8	1.8	1.6
Pound Sterling	1.8	1.8	1.6	1.6	2.0	2.0	1.6	1.2
Euro	1.4	0.6	0.8	0.6	1.2	1.2	1.2	1.2
Japanese Yen	2.0	1.0	1.0	1.2	2.0	1.6	1.6	1.4

4.4.2 Results of Kupiec's test

A statistical test of whether the observed frequency of VaR-exception, i.e., the values of Z, can be considered equal to their theoretical counterparts is provided by Kupiec (1995). The statistics for Kupiec's test are given by

[6] The regulators' backtesting (as suggested in the Basel Accords) can be implemented by using the proportion Z/n (or by expressing it in percentage form). We report the values of Z obtained in the empirical exercise, which can be used easily for the regulators' backtesting, which is primarily concerned with the underestimation of VaR. In addition, we employ the test suggested by Kupiec (1995), which assesses model accuracy by taking in to account too low as well as too high frequencies of VaR-exception (i.e., both under-estimation and over-estimation of VaR).

$$LR = 2\left[\log\left(\left(\frac{z}{n}\right)^{z}\left(1-\frac{z}{n}\right)^{n-z}\right) - \log\left(p^{z}(1-p)^{n-z}\right)\right] \qquad (4.3)$$

where Z denotes the number of VaR-exceptions over n backtesting days, and *p* is the probability level of VaR (in this case, n = 500 and *p* = 0.01).

Under the null hypothesis of normality, the LR-statistic follows a χ^2-distribution with 1 degree of freedom. Ideally, greater closeness between Z/n and *p* would indicate greater accuracy of VaR estimates (i.e., the corresponding VaR model). The null hypothesis Z/n = *p* may be tested against the alternative hypothesis $(Z/n) \neq p$.

The results for Kupiec's test are given in Table 4.4. It is interesting to note that for some assets (such as US Dollar and Japanese Yen), the normal method estimates inaccurate VaR numbers. In all the other cases, the accuracy level appear to be not significantly different to that in the competing models. This indicates that even if a transformation-based indirect approach (which is easy to understand and intuitively appealing) is used, we would usually end up generating VaR estimates as accurate as those obtained from more complex methods (such as a tail-index-based approach, for instance). However, this assessment depends solely on the frequency of VaR-exceptions, which is only one component of risk or severity of loss. This assessment did not incorporate magnitude of excess loss, the other component of risk or severity of excess loss. Such evaluations of VaR models are done using two alternative loss-functions as described earlier.

Table 4.4: Kupiec's Test—Observed Value of Test Statistics

Asset/ Portfolio	Left-Tail (Long Position)				Right-Tail (Short Position)			
	Normal Method	Tail-Index		Trans.-based Method	Normal Method	Tail-Index		Trans.-based Method
		Hill's	OLS			Hill's	OLS	
(A) Full-Sample Results								
US Dollar	7.11**	3.91*	3.91*	1.54	8.97**	3.91*	2.61	2.61
	(0.0077)	(0.0479)	(0.0479)	(0.2149)	(0.0027)	(0.0479)	(0.1060)	(0.1060)
Pound Sterling	0.00	2.35	2.35	2.35	0.72	0.19	0.19	0.19
	(1.0000)	(0.1250)	(0.1250)	(0.1250)	(0.3966)	(0.6630)	(0.6630)	(0.6630)
Euro	0.19	0.94	0.94	0.94	0.19	0.19	0.19	0.19
	(0.6630)	(0.3315)	(0.3315)	(0.3315)	(0.6630)	(0.6630)	(0.6630)	(0.6630)
Japanese Yen	0.72	0.94	0.94	0.22	2.61	1.54	1.54	0.72
	(0.3966)	(0.3315)	(0.3315)	(0.6414)	(0.1060)	(0.2149)	(0.2149)	(0.3966)
(B) Rolling-Sample Results								
US Dollar	5.42*	5.42*	2.61	2.61	5.42*	2.61	2.61	1.54
	(0.0199)	(0.0199)	(0.1060)	(0.1060)	(0.0199)	(0.1060)	(0.1060)	(0.2149)
Pound Sterling	2.61	2.61	1.54	1.54	3.92*	3.92*	1.54	0.19
	(0.1060)	(0.1060)	(0.2149)	(0.2149)	(0.0479)	(0.0479)	(0.2149)	(0.6630)
Euro	0.72	0.94	0.22	0.94	0.19	0.19	0.19	0.19
	(0.3966)	(0.3315)	(0.6414)	(0.3315)	(0.6630)	(0.6630)	(0.6630)	(0.6630)
Japanese Yen	3.91*	0.00	0.00	0.19	3.91*	1.54	1.54	0.72
	(0.0479)	(1.0000)	(1.0000)	(0.6630)	(0.0479)	(0.2149)	(0.2149)	(0.3966)

Figures within parentheses denote significance level (i.e., *p*-value); * and ** denote significance at 5% and 1% levels of significance, respectively.

4.4.3 Evaluation of VaR models using loss functions

The idea of incorporating the magnitude of excess loss while evaluating VaR models was first implemented by Lopez (1998) by assigning certain scores/loss-function values to a VaR model. Further, Sarma et al. (2003) suggested a few specific forms of such evaluation criteria, including the regulator's loss-function and the firm's loss-function. We employ the regulators' loss-function proposed by Sarma et al. (2003), which assigns a score S_t to a model on the t^{th} backtesting day through the following formulation:

$$S_t = \begin{cases} (L_t - V_{t|t-1})^2 & \text{if } L_t > V_{t|t-1} \\ 0 & \text{otherwise} \end{cases} \qquad (4.4)$$

where $V_{t|t-1}$ and L_t represent the estimated VaR for the t^{th} day using data/information up to time (t-1) and the observed loss on t^{th} day, respectively.

The overall score S over n backtesting days is calculated as $S = S_1 + S_2 + \ldots\ldots + S_n$. A model with lower S value would be preferred over others. It may be noted here that a VaR model gets penalized for each instance of VaR-exceptions; however, the magnitude of the penalty score S_t depends on the magnitude of excess loss (instead of a constant penalty score value of 1 that would be assigned in case of a frequency-based model evaluation irrespective of the severity of excess loss). However, in one of the loss-functions proposed by Lopez (1998), excess loss is considered along with the frequency as follows:

$$S_t = \begin{cases} 1 + (L_t - V_{t|t-1})^2 & \text{if } L_t > V_{t|t-1} \\ 0 & \text{otherwise} \end{cases} \tag{4.5}$$

where $V_{t|t-1}$ and L_t represent the estimated VaR for the t^{th} day using data/information up to time (t-1) and the observed loss on the t^{th} day, respectively. Accordingly, lower overall score $S = S = S_1 + S_2 + \ldots\ldots + S_n$ over n back testing days would be preferred in selecting the VaR model.

We present the values of the score based on the loss-function given by Sarma et al. (2003) in Table 4.5 and those based on Lopez's loss-function in Table 4.6. These tables show that for short positions on financial assets (i.e., right-tail of return distributions), the performance of the transformation-based method is uniformly the best across all the assets, whether it is a full-sample estimate or a rolling-sample estimate. In the case of long positions (left-tail), the result is mixed. In this case as well, the transformation-based approach always performs better than the normal method; on a few occasions, the former performs better than the tail-index approaches as well. Thus, the transformation-based approach appears to be the sensible choice for estimating VaR numbers.

Table 4.5: Value/Score of Regulatory Loss-Function proposed by Sarma et al. (2003) (Eq. 4.4)

Asset/ Portfolio	Left-Tail (Long Position)				Right-Tail (Short Position)			
	Normal Method	Tail-Index		Trans.-based Method	Normal Method	Tail-Index		Trans.-based Method
		Hill's	OLS			Hill's	OLS	
(A) Full-Sample Results								
US Dollar	3.53	3.20	3.20	**2.61**	12.90	10.99	10.15	**8.86**
Pound Sterling	0.40	**0.03**	0.04	0.09	6.64	6.02	5.85	**4.84**
Euro	0.64	0.30	**0.27**	0.30	9.98	8.94	8.43	**7.73**
Japanese Yen	1.28	0.38	**0.32**	0.56	8.04	6.99	7.03	**5.83**
(B) Rolling-Sample Results								
US Dollar	2.56	2.74	2.38	**1.75**	10.09	7.37	7.19	**5.60**
Pound Sterling	1.29	0.90	0.59	**0.49**	7.94	7.30	7.99	**5.57**
Euro	0.78	0.51	0.53	**0.38**	9.81	9.98	9.43	**7.12**
Japanese Yen	2.39	1.06	**0.56**	1.54	8.95	8.12	8.13	**6.43**

Table 4.6: Values/Scores of Lopez's Loss-Function (Eq. 4.5)

Asset/ Portfolio	Left-Tail (Long Position)				Right-Tail (Short Position)			
	Normal Method	Tail-Index		Trans.-based Method	Normal Method	Tail-Index		Trans.-based Method
		Hill's	OLS			Hill's	OLS	
(A) Full-Sample Results								
US Dollar	15.53	13.21	13.20	**10.61**	25.90	20.99	19.15	**17.86**
Pound Sterling	5.40	**2.03**	2.04	2.09	13.64	12.02	11.85	**10.84**
Euro	6.64	3.30	**3.27**	3.30	15.98	14.94	14.43	**13.73**
Japanese Yen	8.28	3.38	**3.32**	4.56	17.04	14.99	15.04	**12.83**
(B) Rolling-Sample Results								
US Dollar	13.56	13.74	11.38	**10.75**	21.09	16.37	16.19	**13.60**
Pound Sterling	10.29	9.90	8.60	**8.49**	17.94	17.30	15.99	**11.57**
Euro	7.78	3.51	4.53	**3.38**	15.81	15.98	15.43	**13.12**
Japanese Yen	12.39	6.06	**5.56**	7.54	18.95	16.12	16.13	**13.43**

5. Concluding Remarks

The concept of Value-at-Risk (VaR) has become a key tool not only for measuring various categories of financial risk (such as market risk, credit risk, and operational risk) and for computing the capital that needs to be maintained by banks for holding such risk exposures but also for other purposes such as determining the margin requirement at stock exchanges. The VaR suffers from the limitation of not being a coherent risk measure. Over time, it has gained importance in the context of risk management. It has been argued that excess shortfall (ES), defined as the average of the losses over VaR, would capture the severity of losses better. However, the accuracy of ES measures depends on the quality of the VaR numbers. Further, the VaR has been the basis for other new/related concepts, such as conditional-VaR (Co-VaR), which gained importance in assessing the stability of financial systems. The growing applicability of VaR for dealing with wider types of financial risks and for other purposes through related concepts such as ES and Co-VaR make a renewed case for improving the accuracy of the VaR measurement.

We consider a case involving the estimation of VaR when the historical returns on a portfolio become available. Traditional literature suggests modelling the distribution of returns or log-returns based on the historical values. If the returns were normally distributed, the VaR could be estimated simply by using the first two moments of the distribution and the tabulated values of the standard normal distribution. Thus, the normal method or the covariance approach of VaR estimation in homoscedastic situations as well as heteroscedastic cases has been overwhelmingly popular among practitioners. However, the extant empirical literature shows that the task is potentially difficult because the financial market returns seldom follow normal distribution. There is empirical evidence that the distributions of returns have thicker tails than normal and are skewed at times.

In order to handle the observed non-normality of returns, a number of techniques have been proposed in the literature. Most of the available techniques (parametric or non-parametric) aim to directly identify the best fitted return distribution (which is possibly not normal). An indirect approach would be to transform the possible non-normal returns to near (approximate) normal variables and use the properties of normal distribution to estimate the threshold tail-value. Samanta (2008) experimented with an indirect approach by transforming the observed returns (which possibly do not follow normal distribution) to approximate/near normality. The empirical results presented in this prior study are quite encouraging and show that the transformation-based approach is a sensible alternative for measuring VaR.

In this paper, we re-assess the performance of the indirect approach on two counts. First, we undertake a simulation exercise to examine how good the transformation is in transforming random observations drawn from potential classes of non-normal distributions (the student-t distribution, skewed Laplace distribution, ARCH/GARCH models) to normal variables. Our simulated results show that the transformation is quite useful in inducing normality to observations drawn from several heterogeneous classes of non-normal distributions. Interestingly, the transformation preserves normality in the sense that it does not distort the skewness and excess kurtosis to be different from zero when the original observations truly come from a normal distribution.

Second, we examine the robustness of the empirical results reported in Samanta (2008) based on real data pertaining to the years before the recent global financial crisis. To do so, we used the daily exchange rate data from the post-crisis period and compared the accuracy of the VaR estimates obtained through the indirect approach and a few other competing techniques (normal method and two forms of tail-index methods). Our empirical results are quite interesting. The indirect approach, despite being intuitively appealing and requiring simple practical computation, outperforms the normal method; it also produces VaR estimates that are no worse than those produced by more sophisticated and complex approaches (such as those based on tail-index).

The simulated and empirical results presented in this paper indicate that the simple transformation-based indirect approach of VaR estimation is a sensible one. The ease of understanding and simplicity of implementation of this approach are particularly useful to practitioners who are grappling with the demanding nature of decision-making under dynamic settings. Future extensions of this research could look for theoretical justifications as to why and when such transformations of returns would induce normality. Further, researchers could examine the robustness of the empirical results over time across markets and countries.

APPENDIX A

Direct Approaches of VaR Estimation:

Competing Techniques Considered in the Empirical Exercise

The central issue to any VaR measurement strategy has been the estimation of the quantiles/percentiles of change in the value or returns of the portfolio. If the distribution of the change in value or returns were normal, one would have simply estimated the mean and standard deviation of the normal distribution, thereby estimating the implied percentiles that would be the function of the mean and variance. However, the biggest practical problem in measuring VaR is that the returns generally do not follow a normal distribution.

A.1 Conventional Direct Approaches: Broad Categories

The conventional direct approaches dealing with non-normality can be classified into two broad categories: non-parametric and parametric. The non-parametric approaches (such as historical simulation) do not assume any specific parametric form of the underlying probability distribution; they attempt to discover the distribution non-parametrically from past data. The parametric category for non-normality is vast and includes all the relevant probability distributions other than the normal one. There are a number of alternative strategies under this category, some of which handle possible non-normality by fitting suitable non-normal distributions to past data directly, while others handle non-normality indirectly.

First, one may handle the observed non-normality by directly identifying suitable non-normal distributions such as the Student-t distribution, Laplace distribution, hyperbolic distribution, or a mixture of two or more distributions (van den Goorbergh and Vlaar, 1999; Bauer, 2000; Linden, 2001). Even by mixing two or more normal distributions, one may generate non-normal (fat-tail and/or asymmetrical) distribution.

Second, one may model only the fat-tails of the underlying distribution through the extreme value theory either by modelling the distribution of extreme returns or by estimating the tail-index that measures tail fatness (Tsay, 2002; van den Goorbergh and Vlaar, 1999). Under the tail-index approach, the focus is not to fit the complete portion of the underlying distribution. Rather, it models only the tails of the underlying distribution and identifies a suitable Pareto distribution to fit only the given tail portion. The possible asymmetry of the returns distribution is addressed indirectly as the Pareto distribution can be identified separately for each tail.

Third, for building models that capture volatility clustering, the choice ranges from the exponential weighted average method (J.P. Morgan/Reuters, 1996) to the ARCH/GARCH models or similar more general models for portfolio returns (Engle, 1982; Bollerslev, 1986; Wong et al., 2003). Even conditionally normal variables can be non-normal unconditionally.

A.2 Normal (Covariance) Method

The simplest possible VaR method is the normal (covariance) method. If μ and σ are the mean and standard deviation, respectively, for the returns at a future date, the VaR would be calculated from the expression $(\mu + \sigma z_\alpha)$, where z_α represents the percentile corresponding to the left-tail probability α of the standard normal distribution, and α is the probability level attached to the VaR numbers. This approach is static in the sense that it models the unconditional returns distribution (van den Goorbergh and Vlaar, 1999).

The unconditional distribution of returns generally shows fatter tails (leptokurtosis or excess kurtosis) than normal. This means that the normality assumption for unconditional returns distribution is not realistic. Further, fat-tails could also be a reflection of the changing conditional volatility, which can be modelled under suitable conditional heteroscedastic models such as exponentially weighted moving average used in RiskMetrics (J. P. Morgan/Reuters, 1996) or more advanced models such as ARCH, GARCH, etc. (Engle 1982; Bollerslev, 1986; Wong et al., 2003). Under normality of such conditional distributions, the expression of the VaR estimates is $(\mu_t + \sigma_t z_\alpha)$, where μ_t and σ_t are the time-varying/conditional mean and standard deviation of returns, respectively.

A.3 VaR Measurement Using Tail-Index

The fat tails of unconditional returns distribution can be handled through extreme value theory using tail-index (for instance), which measures the amount of tail fatness. One can, therefore, estimate the tail-index and measure the VaR based on the underlying distribution. The basic premise of this idea stems from the result that the tails of every fat-tailed distribution converge to the tails of the Pareto distribution. The upper tail of such distributions can be modelled simply as:

$$\text{Prob}[X > x] \approx C^{\alpha} |x|^{-\alpha} \ (\text{i.e. Prob}[X \le x] \approx 1 - C^{\alpha} |x|^{-\alpha});\ \ x > C \qquad \text{(A.1)}$$

where the symbol $\approx$ indicates approximately equal; C is the threshold above which the Pareto law holds; |x| denotes the absolute value of x; and the parameter α is the tail-index.

Similarly, the lower tail of a fat-tailed distribution can be modelled as:

$$\text{Prob}[X > x] \approx 1 - C^{\alpha}\ x^{-\alpha} \quad (\text{i.e. Prob}[X \le x] \approx C^{\alpha} x^{-\alpha});\ \ x < C \qquad \text{(A.2)}$$

where C is the threshold below which the Pareto law holds, and the parameter α (called the tail-index) measures the tail-fatness.

In practice, observations in the upper-tail of the return distribution are generally positive, and those in the lower-tail are negative. Thus, Eqs. (A.1) and (A.2) are important in VaR measurement. The holder of a short financial position suffers a loss when the returns on the underlying assets are positive; therefore, the method concentrates on the upper-tail of the distribution (i.e., Eq. A.1) while calculating the VaR (Tsay, 2002, p. 258). Similarly, the holder of a long financial position would model the lower-tail of return distribution (i.e., use Eq. A.2), since a negative return on the underlying assets would lead to losses. In either case, the estimation of the VaR is crucially dependent on the estimation of the tail-index α. There are several methods for estimating tail-index, such as Hill's (1975) estimator and the estimator under the ordinary least squares (OLS) framework suggested by van den Goorbergh (1999). These two methods are presented below.

A.3.1 Hill's Estimator for Tail-Index

For a given threshold C in the right-tail, Hill's (1975) maximum likelihood estimator of $\eta = 1/\alpha$ is

$$\hat{\eta} = \frac{1}{n}\sum_{i=1}^{n}\log\left(\frac{X_i}{C}\right) \tag{A.3}$$

where X_i's i = 1,2, …..,n are n observations (exceeding C) from the right-tail of the distribution.

For estimating the parameters for the left-tail, we simply multiply the observations by –1 and repeat the calculations applicable to the right-tail of the distribution.

In reality, C is usually unknown and needs to be estimated. If the sample observations come from a Pareto distribution, C would be estimated by the minimum observed value (the minimum order statistic). However, we are not modelling the complete portion of the Pareto distribution. We are dealing with only a fat-tailed distribution that has a right-tail that is approximated by the tail of a Pareto distribution. Therefore, one has to select a threshold level (say C) above which the Pareto law holds. In practice, Eq. (A.3) can be evaluated based on order statistics in the right-tail; thus, the selection of the order statistics truncation number assumes importance. In other words, one needs to select the number of extreme observations n needed to operationalise Eq.(A.3). Mills (1999, p. 186) discusses a number of available strategies for selecting n. The method used in this paper is adapted from Phillips et al. (1996). They suggest that the optimal value of n should be one, which minimises the Mean-Square-Error (MSE) of the limiting distribution of $\hat{\eta}$. To implement this strategy, we need the estimates of γ for truncation numbers $n_1 = N^{\beta}$ and $n_2=N^{\tau}$, where $0 < \beta < 2/3 < \tau < 1$. Let $\hat{\eta}_j$ be the estimate of γ for $n = n_j$, j = 1,2. The optimal choice for truncation number is $n = [a\ T^{2/3}]$, where 'a' is a constant estimated as $\hat{a} = |(\hat{\gamma}_1/\sqrt{2})(T/n_2)(\hat{\gamma}_1 - \hat{\gamma}_2)|^{2/3}$. Phillips et al. (1996) recommended setting $\beta = 0.6$ and $\tau = 0.9$ (see Mills, 1999, p. 186).

A.3.2 Ordinary Least Squares (OLS) for Estimating Tail-Index

The tail-index can be estimated via alternative approaches. An OLS-based tail-index estimation was suggested by van den Goorbergh (1999). The same approach was discussed and implemented in van den Goorbergh and Vlaar (1999).[7]

As the fat-tails declined by the Pareto-distribution type power, the OLS-based method for lower tail works as follows (using Eq.A.3):

[7] For an OLS estimation of tail-index, we followed van den Goorbergh and Vlaar (1999).

$$\text{Prob}[X \le x] \approx C^{\alpha}\,|x|^{-\alpha}); \text{ i.e., Log Prob}[X \le x] \approx \alpha \log C - \alpha \log |x| = a - \alpha \log |x| \quad \text{(A.4)}$$

where C is the lower-tail threshold for the Pareto distribution to hold good, and a = α log C is a constant.

The tail-index α in Eq. (A.4) can be estimated through OLS. The tail-index for the right-tail can be estimated similarly via a regression equation corresponding to Eq. (A.2). Alternatively, one could multiply the underlying variable with -1 and carry out the OLS analysis for the lower-tail of the transformed–variable (i.e., negative of the variable) to obtain the tail-index.

A.3.3 Estimating VaR Using Tail-Index

We follow van den Goorbergh and Vlaar (1999) for the measurement of VaR using the estimated tail-index. Let p and q (p < q) be the two tail probabilities; x_p and x_q are the corresponding percentiles. One gets $p \approx C^{\alpha} (x_p)^{-\alpha}$ and $q \approx C^{\alpha} (x_q)^{-\alpha}$, indicating that $x_p \approx x_q (q/p)^{1/\alpha}$. Assuming that the threshold in the left-tail of the return (in percentage) distribution corresponds to the m^th^ order statistics (in ascending order), the estimate of x_p would be

$$\hat{x}_p = R_{(m)}\left(\frac{m}{np}\right)^{\hat{\gamma}} \quad \text{(A.5)}$$

where $R_{(m)}$ is the m^{th} order statistics in the ascending order of n observations chosen from the tail of the underlying distribution; *p* is the given probability level for which VaR is being estimated; $\hat{\gamma}$ is the estimate of γ. Knowing the estimated percentile $\hat{x}_p$, one can easily calculate the VaR.

The methodology described here estimates the tail-index and VaR for the right-tail of a distribution. To estimate the parameters for the left-tail, we simply multiply the observations by –1 and repeat the calculations.

Bibliography

Acharya, Viral, Robert Engle and Matthew Richardson (2012). “Capital Shortfall: A New Approach to Ranking and Regulating Systemic Risks." American Economic Review, Vol. 102, No. 3, pp 59-64.

Acharya, Viral, Lasse Pedersen, Thomas Philippon and Matthew Richardson (2010). “Measuring Systemic Risk”, Federal Reserve Bank of Cleveland, Working Paper 10-02, March.

Adams, Zeno, Ronald Fuss and Reint Gropp (2010). “Modelling Spillover Effects Among Financial Institutions: A State-Dependent Sensitivity Value-at-Risk (SDSVaR) Approach”, EBS Working Paper.

Adrian, Tobias and Brunnermeier, Marcus (2008). “CoVaR”. FRB of New York. Staff Report No. 348, September (Revised September, 2011), accessed online.

Artzner, Philippe, Freddy Delbaen, Jean-Marc Eber and David Heath (1999), Coherent Measures of Risk, *Mathematical Finance*, Vol. 9, No. 3 (July), pp. 203-28.

Basel Committee (1996a). *Amendment to the Capital Accord to Incorporate Market Risks*, Bank for International Settlements.

Basel Committee (1996b). *Supervisory Framework for the Use of ‘Backtesting’ in Conjunction with Internal Models Approach to Market Risk*, Bank for International Settlements.

Bauer, Christian (2000). “Value at Risk Using Hyperbolic Distributions”, *Journal of Economics and Business*, Vol. 52, pp. 455-67.

Bera, Anil K. and Mathew L. Higgins (1993), ARCH Models: Properties, Estimation and Testing, *Journal of Economic Surveys*, Vol. 7, No. 4, pp. 305-62.

Berkowitz, Jeremy and James O’Brien (2002). “How Accurate are Value-at-Risk Models at Commercial Banks?”, *Journal of Finance*, Vol. LVII, No. 3, June, pp. 1093-111.

Bollerslev, T.(1986). “Generalized Autoregressive Conditional Heteroskedasticity”, *Journal of Econometrics*, Vol. 31, pp. 307-27.

Box, G.E.P. and D.R. Cox (1964). “An Analysis of Transformations” (with Discussion), *Journal of Royal Statistical Society,* Vol. 26, No. 2, pp. 211-252.

Christoffersen, P.F. (1998). “Evaluating Interval Forecasts”, *International Economic Review*, 39, pp. 841-62.

Christoffersen, P., Jinyong Hahn and Atsushi Inoue (2001). “Testing and Comparing Value-at-Risk Measures”, *Journal of Empirical Finance*, Vol. 8, No. 3, July, pp. 325-42.

Engle, R. F. (1982). “Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation”, *Econometrica*, Vol. 50, No. 4, July, pp. 987-1007.

Hill, B.M. (1975). “A Simple General Approach to Inference About the Tail of a Distribution”, *Annals of Statistics,* 35, pp. 1163-73.

Jarque, C.M. and A.K. Bera (1987). “A Test for Normality of Observations and Regression Residuals”, *International Statistical Review*, Vol. 55, pp. 163-72.

J.P.Morgan/Reuters (1996). *RiskMetrics: Technical Document*, Fourth Edition, New York, USA.

John, J.A. and N.R. Draper (1980). “An Alternative Family of Transformations”, *Appl. Statist.*, Vol. 29, pp. 190-97.

Jorian, P. (2001). *Value-at-Risk – The New Benchmark for Managing Financial Risk, Second Edition*, McGraw Hill.

J.P.Morgan Bank (1996). *RiskMetrics – Technical Document*, New York, J. P. Morgan Bank.

Kupiec, P.(1995), “Techniques for Verifying the Accuracy of Risk Measurement Models”, *Journal of Derivatives*, Vol. 2, pp. 73-84.

Kupiec, P. (1999), “Risk Capital and VaR”, Journal of Derivatives, Vol. 7, No. 2, Winter, pp. 41-52.

Linden, Mikael (2001). “A Model for Stock Return Distribution”, *International Journal of Finance and Economics*, April, Vol. 6, No. 2, pp. 159-69.

Lopez, Jose A. (1998), “Methods for Evaluating Value-at-Risk Estimates”, *Research and Market Analysis Group, Federal Reserve Bank of New York.*

Majumder, Neeta and Debasish Majumder (2002), Measuring Income Risk to Promote Macro Markets, *Journal of Policy Modeling*, Vol. 24, pp. 607-19.

Mills, Terence C. (1993). *The Econometric Modelling of Financial Time Series*, 2nd Edition, Cambridge University Press.

Phillips, P.C.B., J.W. McFarland and P.C. McMahon (1996). “Robust Tests of Forward Exchange Market Efficiency with Empirical Evidence from the 1920s”, *Journal of Applied Econometrics*, Issue 1 (Jan-Feb), pp.1-22.

Puig, Pedro and Michael A. Stephens (2007), Goodness of Fit Tests for the Skew-Laplace Distribution, SORT, Vol. 31, No. 1, January-June, pp. 45-54.

Samanta, G.P. (2008), Value-at-Risk Using Transformations to Normality, in Jayaram, N. and R.S. Deshpande [Eds.] (2008), *Footprints of Development and Change – Essays in Memory of Professor V.K.R.V.Rao Commemorating His Birth Centenary*, Academic Foundation, New Delhi.

Sarma, Mandira, Susan Thomas and Ajay Shah (2003). “Selection of Value-at-Risk Models”, *Journal of Forecasting*, 22(4), pp. 337-358.

Shiller, R.J. (1993a), Macro markets: Creating institutions for managing society’s largest economic risk, New York: Clarendon Press.

Shiller, R.J. (1993b), Aggregate income risk and hedging mechanism, NBER Working Paper.

Shiller, R.J. and A. Wincoop (1999), Macro markets and financial security, Economic Policy Review, Federal Reserve Bank of New York, 5, pp. 21-38.

Tsay, Ruey S. (2002). *Analysis of Financial Time Series*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc.

van den Goorbergh, R.W.J. and P.J.G. Vlaar (1999). “Value-at-Risk Analysis of Stock Returns Historical Simulation, Variance Techniques or Tail Index Estimation?”, *DNB Staff Reports*, No. 40, De Nederlandsche Bank.

Wilson, Thomas C. (1998). “Value at Risk”, in Alexander, Carol [Ed.], *Risk Management and Analysis*, Vol. 1, Measuring and Modelling Financial Risk, 1998, John Wiley & Sons.

Wong, Michael Chak Sham, Wai Yan Cheng and Clement Yuk Pang Wong (2003). “Market Risk Management of Banks: Implications from the Accuracy of Value-at-Risk Forecasts”, *Journal of Forecasting*, 22, pp. 23-33.

Yeo, In-Kwon and Richard A. Johnson (2000). “A New Family of Power Transformations to Improve Normality or Symmetry”, *Biometrika*, Vol. 87, No. 4, pp. 954-59.