ESTIMATING VALUE AT RISK: From JP Morgan’s Standard-EWMA to Skewed-EWMA Forecasting*

Zudi Lu
School of Mathematical Sciences
The University of Adelaide

Hai Huang
Industrial and Commercial Bank of China
China

Richard Gerlach
Faculty of Economics and Business
The University of Sydney

Abstract

Significantly driven by JP Morgan’s RiskMetrics system with EWMA (exponentially weighted moving average) forecasting technique, value-at-risk (VaR) has turned to be a popular measure of the degree of various risks in financial risk management. In this paper we propose a new approach termed skewed-EWMA to forecast the changing volatility and formulate an adaptively efficient procedure to estimate the VaR. Differently from the JP Morgan’s standard-EWMA, which is derived from a Gaussian distribution, and the Guermat and Harris (2001)’s robust-EWMA, from a Laplace distribution, we motivate and derive our skewed-EWMA procedure from an asymmetric Laplace distribution, where both skewness and heavy tails in return distribution and the time-varying nature of them in practice are taken into account. An EWMA-based procedure that adaptively adjusts the shape parameter controlling the skewness and kurtosis in the distribution is suggested. Backtesting results show that our proposed skewed-EWMA method offers a viable improvement in forecasting VaR.

January 2010

OME Working Paper No: 01/2010
http://www.econ.usyd.edu.au/ome/research/working_papers

* Partially supported by an Internal Research Grant of Curtin University of Technology and a Discovery grant from the Australian Research Council.
ESTIMATING VALUE AT RISK: From JP Morgan’s Standard-EWMA to Skewed-EWMA Forecasting*

ZUDI LU a and HAI HUANG b and RICHARD GERLACH c

a School of Mathematical Sciences, The University of Adelaide, Australia
b Industrial and Commercial Bank of China, Beijing, China
c Discipline of Operations Management and Econometrics, University of Sydney, Australia.

Abstract: Significantly driven by JP Morgan’s RiskMetrics system with EWMA (exponentially weighted moving average) forecasting technique, value-at-risk (VaR) has turned to be a popular measure of the degree of various risks in financial risk management. In this paper we propose a new approach termed skewed-EWMA to forecast the changing volatility and formulate an adaptively efficient procedure to estimate the VaR. Differently from the JP Morgan’s standard-EWMA, which is derived from a Gaussian distribution, and the Guermat and Harris (2001)’s robust-EWMA, from a Laplace distribution, we motivate and derive our skewed-EWMA procedure from an asymmetric Laplace distribution, where both skewness and heavy tails in return distribution and the time-varying nature of them in practice are taken into account. An EWMA-based procedure that adaptively adjusts the shape parameter controlling the skewness and kurtosis in the distribution is suggested. Backtesting results show that our proposed skewed-EWMA method offers a viable improvement in forecasting VaR.

Key words: Asymmetric Laplace distribution, Exponentially weighted moving average (EWMA), forecasting, Skewed EWMA, Skewness and heavy tails, Time-varying shape parameter, Value-at-risk (VaR).

JEL: C51, C52, C53, G15

1. Introduction

Prompted by the globalization of the world economy, financial innovation and the growth of the world’s financial centers, value-at-risk (VaR) has become a popular measurement of the degree of various risks that a financial asset/portfolio is exposed to in the risky financial markets. For example, regulators have urged market participants to make major efforts to understand and control financial risk with a benchmark for measuring market risk by VaR, in order to underpin the solvency and stability of the world banking system. Early in 1993 the Group of Thirty (G-30) advised to value positions using market prices and to assess financial risks with VaR. The Basel Committee on Banking Supervision in 1996 endorsed the use of

*Partially supported by an Internal Research Grant of Curtin University of Technology and a Discovery grant from the Australian Research Council.
VaR model, contingent on important qualitative and quantitative standards. Just as Jorion (2001, preface) pointed out, risk management had truly experienced a revolution in the last few years, and this was started by value at risk (VaR), a new method to measure financial market risk that was developed in response to the financial/derivative disasters of the early 1990s. We particularly note that the recent Great Financial Crisis (GFC) has once again called into the question of financial risk management (FRM) methods and practice. For background reviews and developments on VaR, the reader is referred to Duffie and Pan (1997), Dowd (1998), Hull and White (1998a, b), Jorion (2001), Dempster (2002), Allen (2003), Holton (2003), Dupacova and Polivka (2007) and Kaut et al. (2007), among others.

One important driving force behind this popularity of VaR is no doubt due to the release to the public of the JP Morgan’s (1996) RiskMetrics™ Technical Document, where an exponentially weighted moving average (EWMA) estimator is suggested to forecast the conditional volatility of short horizon asset returns in terms of conditional variance. It is generally well-known that EWMA methods have been around since the 1950s, and are still the most popular forecasting methods used in business and industry in the forecasting literature (see Hyndman et al. (2008) for a recent review). We will term this JP Morgan’s method of forecasting volatility a standard-EWMA for convenience in the following. This standard-EWMA estimator is appropriate for financial assets if the return series are really or approximately from a conditional Gaussian distribution. In practice, the conditional distribution of financial returns is, however, usually skewed and fat-tailed with time varying nature, which deviates from the Gaussian assumption. Therefore the standard EWMA estimator is inefficient in the sense that it will attach too much weight to extreme returns. To solve this problem suffered due to heavy tails in returns, Guermat and Harris (2001) put forward a robust-EWMA procedure that is derived from Laplace distribution. Their empirical applications showed that a robust-EWMA estimator in terms of the absolute return values, rather than squared ones, can offer an important improvement over the standard-EWMA estimator. However, we note that the Laplace distribution is symmetric with a constant kurtosis, which may not be well consistent with the real return series.

In this paper, we will propose a new VaR forecasting model, termed skewed (robust) EWMA, or simply skewed-EWMA. Differently from the standard-EWMA in JP Morgan’s RiskMetrics that is derived from a Gaussian distribution and the robust-EWMA by Guermat and Harris (2001) from a Laplace distribution, we are motivating our skewed-EWMA procedure from an asymmetric Laplace distribution. One important advantage of an asymmetric Laplace distribution lies in its ability to capture both the skewed and heavy tailed behaviors in financial data. Our proposed skewed-EWMA method can be seen as a generalization of the EWMA estimators in the literature, nesting the robust-EWMA estimator as a special case. Most importantly, we will not only take into account both skewness and heavy tails in financial return distribution, and also suggest an EWMA-based procedure to adaptively adjust the shape parameter controlling the skewness and kurtosis, which adapts to the time-changing nature of skewness and heavy tails in financial practice. We will empirically apply our new estimator to forecast the value at risks in the foreign exchange rates and aggregate equity portfolios from the US, the UK, China and Japan financial markets to examine the out-of-sample performances of the three VaR forecasting models, which illustrates that the proposed skewed-EWMA
method outperforms both the standard-EWMA and the robust-EWMA methods in terms of the outcome of the backtesting techniques widely applied in the literature.

The structure of the paper is as follows. Section 2 first outlines the standard-EWMA method in RiskMetrics and the robust-EWMA in Guermat and Harris (2001), and then focuses on developing our proposed methodology in depth, with an introduction to asymmetric Laplace distribution and our new VaR model — skewed-EWMA estimator. Section 3 examines the forecasting and evaluation of the VaR with empirical applications of different models. From the empirical out-of-sample testing the superiority of the skewed-EWMA method over the standard- and robust-EWMA methods is obviously viable. The last section concludes.

2. Methodology

Various methodologies have been developed for VaR estimation and forecasting in the literature, basically including two categories of indirect and direct methods. The indirect methods try to first estimate the distribution of financial return series and then calculate VaR from the estimated distribution. Most of the methodologies in the VaR literature belong to this category, such as distribution based parametric approaches (e.g., JP Morgan-RiskMetrics’ standard-EWMA, Guermat and Harris(2001)’s robust-EWMA), historical simulation, Monte Carol simulation, extreme value theory. Alternatively, direct method tries to estimate the VaR value directly without need to estimate the distribution of financial return series, for example, the quantile regression based CAViaR model developed by Engle & Manganelli (2004). In this paper we are concerned with the parametric EWMA forecasting. First of all we will review the basic ideas in standard-EWMA and robust-EWMA as a motivation to our skewed-EWMA method.

2.1. Standard-EWMA and robust-EWMA

It is a common practice in financial risk modelling that the mean value of return series can be approximately taken as zero in comparison with the quantity of volatility. We follow this convention in the following of this paper.

The JP Morgan RiskMetrics’ standard-EWMA assumes conditional normality for the distribution of return series, $r_t$, with volatility modelled as an IGARCH(1,1), namely

$$ VaR_{t+1} = \sigma_{t+1} \Phi^{-1}(\alpha), $$

$$ \sigma_{t+1}^2 = \lambda \sigma_t^2 + (1 - \lambda) r_t^2, \ 0 < \lambda < 1, $$

where $\alpha$ is the confidence level of VaR, say $\alpha = 99\%$, $\Phi^{-1}(\cdot)$ is the inverse function of the cumulative distribution function of standard normal distribution, and the second equality can be equivalently expressed as

$$ \sigma_{t+1}^2 = (1 - \lambda) r_t^2 + (1 - \lambda) \lambda r_{t-1}^2 + (1 - \lambda) \lambda^2 r_{t-2}^2 + \ldots $$

$$ = (1 - \lambda) \sum_{i=0}^{\infty} \lambda^k r_{t-i}^2. $$
which is an EWMA version of the usual sample variance $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} r_i^2$. This model puts geometrically declining weights on past observations, but assigning greater importance to recent observations. As Nelson and Foster (1994) pointed out, when returns are conditionally normally distributed, standard EWMA is optimal.

In reality, however, the distribution of financial return series is often skewed and heavy-tailed, with departure from normality (c.f., Hull and White, 1998a). Guermat and Harris (2001) therefore, based on a Laplace distribution, suggested a robust-EWMA in the form:

$$\text{VaR}_{t+1} = -\frac{\sigma_{t+1}}{\sqrt{2}} \ln \{2(1 - \alpha)\},$$

$$\sigma_{t+1} = \lambda \sigma_t + (1 - \lambda) \sqrt{2} |r_t|, \quad 0 < \lambda < 1.$$  

Note that the volatility $\sigma_{t+1}$ is an EWMA version of the maximum likelihood estimator of the standard deviation, $\hat{\sigma} = \frac{1}{n} \sum_{i=1}^{n} \sqrt{2} |r_i|$, in the traditional Laplace distribution. It accounts for heavy tails in financial return series but no skewness is taken into account.

### 2.2. Skewed-EWMA

In this subsection, we propose a skewed-EWMA VaR model, that is motivated from an asymmetric Laplace distribution (ALD) to take into account both skewness and heavy tails in financial return distributions, and suggest an EWMA based procedure that adaptively adjusts the shape parameter which controls the skewness and kurtosis in the ALD. This procedure is adaptive to time-varying nature of financial systems in practice. We will see in later sections that the proposed skewed-EWMA outperforms both the standard- and robust- EWMA in VaR forecasting.

#### 2.2.1. Asymmetric Laplace distribution

Asymmetric Laplace distribution (ALD) has been defined in some different ways and found to be useful in defining quantile and and quantile regression in the literature. For example, the quantile regression minimisation is equivalent to the maximisation of a log likelihood based on the ALD. There are now many papers recognising and building on this link. See, for example, the work of Koenker and Machedo (1999), Yu et al. (2003), Yu and Zhang (2005) and Geraci and Bottai (2007), among others.

In this paper, we define an asymmetric Laplace distribution (ALD) in an alternative form of parametrisation to that of Kotz et al. (2002) and Yu and Zhang (2005).

**Definition 2-1** If a random variable $X$ has the following distribution density, we call it is asymmetric Laplace distributed (ALD), denoted $x \sim AL(\mu, \sigma, p)$,

$$f(x|\mu, \sigma, p) = \frac{k}{\sigma} \exp \left\{ - \left( \frac{1-p}{\pi} I_{[x>\mu]} + \frac{1}{p} I_{[x<\mu]} \right) \frac{k}{2} |x - \mu| \right\},$$

where $\mu$, $\sigma$ and $p$ are the location, scale and shape parameters, respectively, and $k = k(p) = \sqrt{p^2 + (1 - p)^2}$.

Note that the three parameters in (2.6) are different from the parameterization of ALD in Yu and Zhang (2005, page 1867). Our parameters in (2.6) are of particular meanings. The shape parameter $p$ is the
probability that \( X < \mu \), where \( \mu \) is the mode of the distribution; i.e. \( p = Pr(X < \mu) \) and thus \( p \in [0, 1] \), and in particular, the variance is \( \text{Var}(X) = \sigma^2 \). As pointed out at the beginning of Subsection 2.1, for financial return data, \( \mu \) is often close to 0 in practice, and as we are only concerned with forecasting of the risk (volatility) of the return, rather than the return itself, in this paper, we will take \( \mu = 0 \) as in Guermat and Harris (2001) for simplicity, and simply denote \( X \sim ALD(\sigma, p) \), in the following.

The shape parameter \( p \) controls the skewness and kurtosis of the asymmetric Laplace distribution \( X \sim ALD(\sigma, p) \), which can be calculated in the formula

\[
S_k = \frac{2[(1-p)^3 - p^3]}{k^3}, \quad K_u = \frac{9(1-p)^4 + 6(1-p)^2p^2 + 9p^4}{k^4}.
\]

(2.7)

Different \( p \) value leads to positive or negative skewness. It follows from (2.6) with \( \mu = 0 \) that if \( p < 0.5 \), for \( x > 0, |x| \) has a larger weight, causing the density function skewed to the right with a positive skewness; if \( p = 0.5 \), ALD reduces to traditional Laplace distribution which is symmetric; and if \( p > 0.5 \), similarly the density function is skewed to the left with a negative skewness.


2.2.2. Maximum likelihood estimate

In order to motivate our skewed-EWMA models, as done in Guermat and Harris (2001), we first consider the maximum likelihood estimate of the important unknown parameters \( \sigma \) and \( p \) in the ALD (2.6) (with \( \mu = 0 \)) by assuming \( \sigma \) and \( p \) constant in an unconditional setting. Suppose the observations are \( r_1, \cdots, r_n \). Then the likelihood

\[
L(\sigma, p) = \prod_{i=1}^{n} \left\{ \frac{k}{\sigma} \exp \left[ - \left( \frac{1}{1-p} I_{r_i, >0} + \frac{1}{p} I_{r_i, <0} \right) \frac{k}{\sigma} |r_i| \right] \right\},
\]

(2.8)

\[
= \frac{k^n}{\sigma^n} \exp \left\{ - \sum_{i=1}^{n} \left( \frac{1}{1-p} I_{r_i, >0} + \frac{1}{p} I_{r_i, <0} \right) \frac{k}{\sigma} |r_i| \right\}.
\]

(2.9)

The log likelihood

\[
\ell(\sigma, p) \equiv \ln L(\sigma, p)
\]

\[
= n \ln k - n \ln \sigma - \sum_{i=1}^{n} \left\{ \left( \frac{1}{1-p} I_{r_i, >0} + \frac{1}{p} I_{r_i, <0} \right) \frac{k}{\sigma} |r_i| \right\}.
\]

(2.10)

Here \( \ell(\cdot) \) is differentiable with respect to \( \sigma, p \). To maximise \( \ell(\cdot) \), we let the first order partial derivatives of \( \ell(\cdot) \) with respect to \( \sigma \) and \( p \) equal zero, respectively, leading to

\[
\hat{\sigma} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{k}{p} I_{r_i, <0} + \frac{k}{1-p} I_{r_i, >0} \right) |r_i|,
\]

(2.11)

For simplicity we write the likelihood as if the observations are i.i.d.; otherwise we could understand the likelihood as the conditional likelihood given the initial information with one-period conditional probability density function in the form of (2.6) with \( \mu = 0 \).
\[ p = \frac{1}{1 + \sqrt{u/v}}, \quad (2.12) \]

where
\[ u = \frac{1}{n} \sum_{i=1}^{n} |r_i| I[r_i > 0], \quad v = \frac{1}{n} \sum_{i=1}^{n} |r_i| I[r_i < 0]. \quad (2.13) \]

Obviously, \( u \) is the averaged positive return while \( v \) is the absolute value of averaged negative return. The larger \( u \), the better; but the larger \( v \) the worse for investment. These formulae are very important to develop our skewed-EWMA forecasting below.

### 2.2.3. Skewed-EWMA based VaR modelling

Based on asymmetric Laplace distribution (2.6) with \( \mu = 0 \), if the parameters \( \sigma \) and \( p \) are known, then the VaR value at the confidence level \( \alpha \) can be easily derived, which is

\[ \text{VaR} = -\sigma \left[ 1 + \left( \frac{1-p}{p} \right)^2 \right]^{-1/2} \ln \frac{1-\alpha}{p}. \quad (2.14) \]

Theoretically, the appeal of ALD lies in its simplicity as that of Gaussian distribution but with skewness and kurtosis taken into account. Practically, as a referee noted, it may be difficult to understand the intuitive appeal of assuming the returns distribution is an ALD. For example, Yu and Zhang (2005) plot the ALD density function for various values of \( p \). It has a sharp point at the mode and slopes in a concave fashion towards the extreme tails. It seems to bear little resemblance to plots of historical returns. However, as well-known, any parametric density function is only an approximation to the real distribution function. Furthermore, note that, as done in Guermat and Harris (2001), the above derivation in Subsection 2.2.2 is only a motivation for our time-varying models developed below, and we do not apply the ALD density function for the returns distribution directly in practice. For example, it is well documented in financial literature that practically, the volatility \( \sigma^2 \) is often not constant, but clustered and time-varying. We will also show that the shape parameter \( p \) may change with time. Therefore we will develop our parameters \( \sigma \) and \( p \) in a conditional setting, that is they are time-varying and stochastic. This means that the unconditional distribution of the returns developed in our models below is itself non-ALD in general.

For \( \sigma \), an EWMA version estimator corresponding to (2.11) can be suggested as follows:

\[ \sigma_{t+1} = (1 - \lambda) \sum_{i=0}^{\infty} \lambda^i \left( \frac{k}{1-p} I[r_{t-i} > 0] + \frac{k}{p} I[r_{t-i} < 0] \right) |r_{t-i}|, \quad (2.15) \]

where \( k \) is defined as in (2.6), and \( \lambda \) is a decaying factor, the choice of which will be discussed later on. Through iteration, (2.15) can be re-expressed as

\[ \sigma_{t+1} = \lambda \sigma_t + (1 - \lambda) \left( \frac{k}{1-p} I[r_t > 0] + \frac{k}{p} I[r_t < 0] \right) |r_t|. \quad (2.16) \]

Clearly, if \( p = 0.5 \), that is corresponding to symmetric Laplace distribution, then (2.14) and (2.16) reduce to (2.4) and (2.5), respectively, which was proposed by Guermat and Harris (2001). However, if \( p \neq 0.5 \), then the contribution of the positive/negative value of \( r_t \) to \( \sigma_{t+1} \) is quite different — this feature is particularly interesting for the effects of good news \((r_t > 0)\) and bad news \((r_t < 0)\) are well characterized in (2.16). In
fact, this skewed-EWMA estimate of (2.16) is a special first order threshold GARCH (TGARCH) model of Zakoian (1994) in the form

$$\sigma_{t+1} = \sigma_t + \alpha_1^+ r_t^+ + \alpha_1^- r_t^- + \beta_1 \sigma_t,$$  \hspace{1cm} (2.17)

where \( r_t^+ = |r_t|I_{|r_t|>0} \), \( r^- = |r_t|I_{|r_t|<0} \), and \( \alpha_0, \alpha_1^+, \alpha_1^- \) and \( \beta_1 \) are parameters, satisfying \( \alpha_0 \geq 0, \alpha_1^+ \geq 0, \alpha_1^- \geq 0 \) and \( \beta_1 \geq 0 \). Obviously, if \( \alpha_0 = 0, \alpha_1^+ = \frac{k(1-\lambda)}{1+k}, \alpha_1^- = \frac{k(1-\lambda)}{p} \) and \( \beta_1 = \lambda \), then the TGARCH(1,1) model (2.17) reduces to the skewed-EWMA estimate of volatility in (2.16).

Up to now, the parameter \( p \) in skewed-EWMA of (2.16) is dealt with as a constant, which can be estimated by (2.12) combined with (2.13). According to our experience, however, though this estimated \( p \) could work better than the fixed constant of 0.5 in robust-EWMA, the improvement is still limited. In fact, with a constant \( p \), the skewness and kurtosis defined in (2.7) keeps constant and does not well adapt to the time-changing nature of financial systems in practice. Therefore we suggest an alternative EWMA based estimate for the parameter \( p \).

First, we define two EWMA estimates for \( u \) and \( v \) in view of (2.13) in the following forms:

$$u_{t+1} = (1 - \beta_u) \sum_{i=0}^{\infty} \beta_u^i |r_{t-i}| I_{|r_{t-i}|>0},$$  \hspace{1cm} (2.18)

$$v_{t+1} = (1 - \beta_v) \sum_{i=0}^{\infty} \beta_v^i |r_{t-i}| I_{|r_{t-i}|<0},$$  \hspace{1cm} (2.19)

where \( \beta_u \) and \( \beta_v \) are two decaying factors in these two EWMA estimates, respectively, which may be different from the decaying factor \( \lambda \) used in (2.16) and will be specified later on in Section 3.2. Once \( \beta \)'s are determined in (2.18) and (2.19), the iterative forms for \( u_{t+1} \) and \( v_{t+1} \) can be written as

$$u_{t+1} = \beta_u u_t + (1 - \beta_u) |r_t| I_{|r_t|>0},$$  \hspace{1cm} (2.20)

$$v_{t+1} = \beta_v v_t + (1 - \beta_v) |r_t| I_{|r_t|<0},$$  \hspace{1cm} (2.21)

Then using (2.12) we can suggest an EWMA based estimate for \( p_{t+1} \) as follows

$$p_{t+1} = \frac{1}{1 + \sqrt{u_{t+1}/v_{t+1}}}. $$  \hspace{1cm} (2.22)

Note that with \( p \) taken in this form, the skewness and kurtosis defined in (2.7) will be time-varying and automatically adapt to the changing nature of financial practice (see also Guermat and Harris (2002) for a time-varying kurtosis, but no skewness).

It would be useful to have some intuition regarding the \( u_t \) and \( v_t \), which are smoothed in expressions (2.20) and (2.21). Clearly \( u_{t+1} \) and \( v_{t+1} \) are the absolute values of EWMA of the positive and negative historical returns up to time \( t \), respectively. Differently from Guermat and Harris (2001) which assume \( u_t \equiv v_t \) at all times and take \( p(t) \equiv 0.5 \), we let the data speak for whether or not \( u_t \) and \( v_t \) will be similar in value and so \( p_t \) will be close to 0.5. This would be more flexible and sensible, and the fitting of the ALD is not analogous to the fitting of the median quantile in general.

Now our proposed skewed-EWMA based estimate of VaR can be finally defined on the basis of (2.22) as follows: with reference to (2.16),

$$\sigma_{t+1} = \lambda \sigma_t + (1 - \lambda) \left( \frac{k_{t+1}}{1 - p_{t+1}} I_{|r_t|>0} + \frac{k_{t+1}}{p_{t+1}} |r_t| I_{|r_t|<0} \right) |r_t|,$$  \hspace{1cm} (2.23)
where \( k_{t+1} = k(p_{t+1}) = \sqrt{p_{t+1}^2 + (1 - p_{t+1})^2} \), and in view of (2.14),
\[
VaR_{t+1} = -\sigma_{t+1} \left[ 1 + \left( \frac{1 - p_{t+1}}{P_{t+1}} \right)^2 \right]^{-1/2} \ln \frac{1 - \alpha}{P_{t+1}}.
\]
(2.24)

Applying (2.20)–(2.24), we can calculate the skewed-EWMA forecasting of the VaR value.

Just as the JP Morgan’s standard-EWMA (2.2) is a special GARCH(1,1) model of Bollerslev (1986), this skewed-EWMA estimate, when \( p \) is assumed constant, is a special TGARCH(1,1) model of Zakoian (1994), which in general reduces the number of parameters of TGARCH(1,1) model (which has 4 parameters in (2.17)). In fact, in our empirical application below, we find we can further take \( \beta_u = \beta_v = \beta \) for simplicity to reduce the number of the decaying factors in the skewed-EWMA forecasting (which has only two decaying factors, \( \beta \) in (2.20) and (2.21) and \( \lambda \) in (2.23)). We update the value of the parameter \( p \) by an EWMA based procedure, so our proposed skewed-EWMA in (2.23) is a varying-coefficient TGARCH model, which can adjust the skewness and kurtosis conveniently in calculating the VaR by the EWMA based procedure and also is more efficient and automatically adaptive in modelling the real time-varying financial system.

3. Forecasting and evaluation: empirical performance

We now turn to examine and compare the empirical performance of the proposed skewed-EWMA with those of the standard-EWMA and the robust-EWMA using real data sets.

3.1. Data

We will consider and examine the VaR forecasting for 6 financial series, including 3 exchange rates (British Pound/US Dollar, Canadian Dollar/US Dollar and Japanese Yen/US Dollar, denoted by BU, CU and JU, respectively) and 3 stock exchange indexes (Hong Kong Hang Seng index, Shanghai Composite Index and Shenzhen Component Index, denoted by HK, SH and SZ, respectively) for length of 10 years from 1 January 1992 to 31 December 2001. The data can be obtained from the Yahoo Finance website. We are concerned with the daily return series, defined by \( \ln P_t - \ln P_{t-1} \), where \( P_t \) is the close price of day \( t \).

The histograms of the 6 return series for 4 years from 1 January 1998 to 31 December 2001 are plotted in Fig 1. Clearly the changes of the 3 exchange rates are relatively small, basically within the interval of \( \pm 5\% \), in particular for CU within \( \pm 2\% \) and for JU its largest negative return is a little bit greater than 5%. The changes of the 3 stock indexes are quite large, in particular for HK the largest positive daily return is greater than 10%, while for SH and SZ their daily rerun series are within \( \pm 10\% \) due to the up/down limits of 10%. The distributions of the 6 return series all are quite highly peaked, obviously deviating from the Gaussian distribution, and also skewed, with the 3 exchange rates series skewed to the left and the index series to the right.

In the following we consider the daily return series in unit of \( \% \), that is multiplied by 100, i.e. \( r_t = 100(\ln P_t - \ln P_{t-1}) \). Table 1 lists the preliminary values of the sample statistics of \( r_t \) including mean, standard deviation (S.D.), skewness (Sk) and kurtosis (Ku) based on the daily return series multiplied by
Fig. 1: Histogram of the 6 return series: The first three are for the exchange rates of British Pond, Canadian Dollar and Japanese Yen with respect to US Dollar; the second three are for the stock exchange indexes of Hong Kong Hang Seng Composite Index, Shanghai Composite Index and Shenzhen Component Index.
Table 1. Preliminary Statistics

Based on the daily return series multiplied by 100 from January 1998 to December 2001. BU, CU and JU stand for the 3 exchange rates of British Pound/US Dollar, Canadian Dollar/US Dollar and Japanese Yen/US Dollar, respectively, and HK, SH and SZ for the 3 stock exchange indexes of Hong Kong Hang Seng Composite Index, Shanghai Composite Index and Shenzhen Component Index, respectively.

<table>
<thead>
<tr>
<th>Name</th>
<th>Mean</th>
<th>S.D.</th>
<th>Sk</th>
<th>Ku</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>BU</td>
<td>0.0122</td>
<td>0.4767</td>
<td>-0.2308</td>
<td>3.6455</td>
<td>1001</td>
</tr>
<tr>
<td>CU</td>
<td>0.0108</td>
<td>0.3612</td>
<td>-0.1727</td>
<td>4.6227</td>
<td>1001</td>
</tr>
<tr>
<td>JU</td>
<td>-0.0011</td>
<td>0.8354</td>
<td>-0.7383</td>
<td>7.5545</td>
<td>1001</td>
</tr>
<tr>
<td>HK</td>
<td>0.0062</td>
<td>2.0953</td>
<td>0.2041</td>
<td>6.3751</td>
<td>986</td>
</tr>
<tr>
<td>SH</td>
<td>0.0333</td>
<td>1.4723</td>
<td>0.2177</td>
<td>9.1200</td>
<td>964</td>
</tr>
<tr>
<td>SZ</td>
<td>-0.0238</td>
<td>1.6362</td>
<td>0.5061</td>
<td>7.9279</td>
<td>964</td>
</tr>
</tbody>
</table>

100 from January 1998 to December 2001. Clearly the coefficients of skewness and kurtosis of the 6 return series are all quite large, indicating that the distributions of the 6 return series are non-Gaussian.

3.2. Determination of decaying factors

In the skewed-EWMA estimate of VaR proposed in the above, there are important decaying factors $\lambda$ and $\beta$'s that need to be specified appropriately in application. We suggest applying likelihood principle (c.f., Hyndmana, et al., 2002, Page 444) to select these decaying factors as our skewed-EWMA estimate is proposed on the basis of maximum likelihood estimate (MLE) of the standard deviation in the asymmetric Laplace distribution. The procedure is described as follows.

Based on the return series of size $T$, $r_1, \ldots, r_T$, with asymmetric Laplace distribution, the likelihood, given $r_0$, can be expressed as

$$f(r_1, \ldots, r_T) = \prod_{t=1}^{T} \frac{k_t}{\sigma_t} \exp \left\{ -\left( \frac{1}{1 - p_t} I_{\{r_t>0\}} + \frac{1}{p_t} I_{\{r_t<0\}} \right) \frac{k_t}{\sigma_t}|r_t| \right\}. \ldots (3.1)$$

where $p_t$ is the function of $\beta = (\beta_u, \beta_v)$ and $\sigma_t$ is the function of $\lambda, \beta$, that is $p_t = p_t(\beta)$ and $\sigma_t = \sigma_t(\lambda, \beta)$, and $k_t = \sqrt{(1 - p_t)^2 + p_t^2}$. For simplicity, we take $r_0 = 0$ below. Therefore the likelihood, based on asymmetric Laplace distribution, in (3.1) depends upon both $\lambda$ and $\beta$. As usual, we consider the log likelihood

$$LKHD(\lambda, \beta) = \sum_{t=1}^{T} \ln k_t - \ln \sigma_t - \left( \frac{1}{1 - p_t} I_{\{r_t>0\}} + \frac{1}{p_t} I_{\{r_t<0\}} \right) \frac{k_t}{\sigma_t}|r_t|. \ldots (3.2)$$

We will take the maximisers, $\lambda_0$ and $\beta_0$ of (3.2) as the decaying factors, satisfying

$$LKHD(\lambda_0, \beta_0) \geq LKHD(\lambda, \beta), \forall (\lambda, \beta) \in \Theta, \ldots (3.3)$$

where $\Theta$ is the parameter space of $(\lambda, \beta)$. As suggested in Hyndmana, et al. (2002, Page 444), the model
In this table, the decaying factors $\lambda$ and $\beta_u = \beta_v = \beta$ are determined by likelihood function principle based on the first 500 observations starting from 1 January 1992. The first row is the value for $\beta$ from 0.990 to 0.999. For each $\beta$, the value of likelihood scaled down by 1000, corresponding to the value for $\lambda$ listed in the parenthesis that maximizes the likelihood function, is reported. The symbol $^*$ indicates the maximum likelihood.

<table>
<thead>
<tr>
<th>Name</th>
<th>$\beta=0.990$</th>
<th>$\beta=0.995$</th>
<th>$\beta=0.996$</th>
<th>$\beta=0.997$</th>
<th>$\beta=0.998$</th>
<th>$\beta=0.999$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BU</td>
<td>$-1.3181$</td>
<td>$-1.3145$</td>
<td>$-1.3139$</td>
<td>$-1.3132$</td>
<td>$-1.3127$</td>
<td>$-1.3124^*$</td>
</tr>
<tr>
<td></td>
<td>(0.97)</td>
<td>(0.97)</td>
<td>(0.97)</td>
<td>(0.97)</td>
<td>(0.97)</td>
<td>(0.97)</td>
</tr>
<tr>
<td>CU</td>
<td>$-0.4498$</td>
<td>$-0.4471$</td>
<td>$-0.4466$</td>
<td>$-0.4461$</td>
<td>$-0.4457$</td>
<td>$-0.4455^*$</td>
</tr>
<tr>
<td></td>
<td>(0.96)</td>
<td>(0.96)</td>
<td>(0.96)</td>
<td>(0.96)</td>
<td>(0.96)</td>
<td>(0.96)</td>
</tr>
<tr>
<td></td>
<td>(0.94)</td>
<td>(0.94)</td>
<td>(0.94)</td>
<td>(0.94)</td>
<td>(0.94)</td>
<td>(0.94)</td>
</tr>
<tr>
<td></td>
<td>(0.94)</td>
<td>(0.94)</td>
<td>(0.94)</td>
<td>(0.94)</td>
<td>(0.94)</td>
<td>(0.94)</td>
</tr>
<tr>
<td></td>
<td>(0.89)</td>
<td>(0.89)</td>
<td>(0.89)</td>
<td>(0.89)</td>
<td>(0.89)</td>
<td>(0.89)</td>
</tr>
<tr>
<td></td>
<td>(0.89)</td>
<td>(0.89)</td>
<td>(0.89)</td>
<td>(0.89)</td>
<td>(0.89)</td>
<td>(0.89)</td>
</tr>
</tbody>
</table>

In our empirical application for the data sets under study, according to $AIC$, we find that we can take $\beta_u = \beta_v$ to reduce the number of the decaying factors. For notational simplicity, we denote $\beta_u = \beta_v = \beta$ below. In Table 2, the outcome of selecting $\lambda$ and $\beta$ by likelihood function principle is reported based on the first 500 observations starting from 1 January 1992, with precision for $\lambda$ up to 0.01 and for $\beta$ up to 0.001. The first row is the value for $\beta$, taking from 0.990 to 0.999. For each value of $\beta$, we find that we can take $\beta_u = \beta_v$ to reduce the number of the decaying factors. For notational simplicity, we denote $\beta_u = \beta_v = \beta$ below. In Table 2, the outcome of selecting $\lambda$ and $\beta$ by likelihood function principle is reported based on the first 500 observations starting from 1 January 1992, with precision for $\lambda$ up to 0.01 and for $\beta$ up to 0.001. The first row is the value for $\lambda$, taking from 0.990 to 0.999. For each value of $\beta$, the $\lambda$ value that maximises the likelihood LKHD is listed inside the parenthesis while outside the corresponding likelihood value scaled down by 1000, with symbol $^*$ indicated for the maximum likelihood. Clearly the optimal $\lambda$ takes values from 0.89 to 0.97, and the optimal $\beta$ is from 0.997, 0.998 to 0.999, in particular $\beta$ equals 0.999 for the GU, CU and SH return series, 0.998 for JU and HS, and 0.997 for SZ.

Based on the analysis in the above, it looks reasonable to use $\beta$ equal to 0.998 for calculation of VaR in the following, which is the optimal or close to the optimal, while $\lambda$ takes 15 different values from 0.85 to 0.99, in order to compare with the performance of the standard-EWMA and robust-EWMA, as done in
Guermat and Harris (2001).

3.3. Evaluation of VaR forecasting

Like volatility, the actual VaR value is unobservable, which requires to be evaluated indirectly. In the literature of evaluating the VaR forecasting, the likelihood ratio test based backtesting techniques due to Kupiec (1995) and Christoffersen (1998) are no doubt the most widely applied ones, which systematically compare the history of VaR forecasts with their associated subsequent returns. Backtesting is also central to the Basel Committee’s ground-breaking decision to allow internal VaR models for capital requirements (c.f., Jorion, 2001, Page 129).

All empirical outcomes reported in this subsection are based on the data from 1st January 1992 to 31st December 2001, with the first 500 observations starting from 1 January 1992 used to determine the decaying factors (see Subsection 3.2 in the above) while the remaining observations after the first 500 ones used to evaluate the VaR forecasting by the three EWMA methods. From these tests we will obviously see that our skewed-EWMA forecasting of VaR uniformly outperforms the forecasted VaR by JP Morgan’s standard-EWMA and by Guermat and Harris (2001)’s robust-EWMA.

In order to correspondingly compare the performance of the proposed skewed-EWMA forecasting with those of the JP Morgan’s standard-EWMA and the Guermat and Harris’s robust-EWMA forecasts, we will follow the testing as done in Guermat and Harris (2001, 2002), focusing on the commonly used methods of the unconditional coverage test, the independence test and the combined conditional coverage test; see Jorion (2001, §6.2) for details.

1. Unconditional coverage test

The most basic requirement for a good VaR model is that the proportion of the number of exceeding the estimated VaR should be close to the nominal VaR significance level $\tau = 1 - \alpha$, that is a good VaR model should facilitate a correct unconditional coverage. For example, at the significance level 1%, the proportion of exceeding the estimated VaR should be approximate to 1%. Kupiec (1995) proposed the likelihood ratio test based on the fact that the number $N$ of exceeding VaR among the sample of size $T$ is binomially distributed, with the probability proportional to $(1 - \tau)^{T-N} \tau^N$, and hence the likelihood ratio (LR) statistic

$$LR_u = -2 \ln L_\tau + 2 \ln L_N.$$  \hspace{1cm} (3.4)

where $L_\tau = (1 - \tau)^{T-N} \tau^N$, $L_N = (1 - N/T)^{T-N} (N/T)^T$. As $T$ tends to infinity, $LR_u$ is asymptotically $\chi^2(1)$ distributed under the null hypothesis.

Figures 2 and 3 display the unconditional coverage and its likelihood ratio test based on (3.4), respectively, for one-period ahead VaR forecasting by using JP Morgan’s standard-EWMA, Guermat and Harris (2001)’s robust-EWMA, and our proposed skewed-EWMA.

2. Independence test

A more strict requirement for a good VaR model is the independence requirement, that is if a VaR model really reflects the conditional distribution and the dynamic behaviours, such as time-varying volatility, of a return series, then the loss exceeding VaR should be unpredictable. The more volatile the market is, the
Fig. 2: Unconditional coverage for the three EWMA forecasts of VaR at 1% nominal level. The x-axis is the common decay factor $\lambda$ for the three EWMA forecasts, and the y-axis is the unconditional coverage. The green cross '+' line is for standard-EWMA, the blue cross 'x' line is for robust-EWMA, and the red dot line is for skewed-EWMA. The common decay factor $\lambda$ takes values from 0.85 to 0.99 for three EWMA forecasts while the decay factor $\beta = 0.998$ only in skewed EWMA. The time period of data is from 1st January 1992 to 31st December 2001 with the first 500 observations used to determine the decaying factors and the remaining observations after the first 500 ones used to evaluate the VaR forecasting by the three EWMA methods. The unconditional coverage refers to the ratio of the number of losses exceeding VaR to the whole sample size.
Fig. 3: Likelihood ratio test of unconditional coverage for the three EWMA forecasts of VaR at 1% nominal level. The $x$-axis is the common decay factor $\lambda$ for the three EWMA forecasts, and the $y$-axis is the LRT statistic. The green cross '+' line is for standard-EWMA, the blue cross 'x' line is for robust-EWMA, and the red dot line is for skewed-EWMA. The common decay factor $\lambda$ takes values from 0.85 to 0.99 for three EWMA forecasts while the decay factor $\beta$ = 0.998 only in skewed EWMA. The time period of data is from 1st January 1992 to 31st December 2001 with the first 500 observations used to determine the decaying factors and the remaining observations after the first 500 ones used to evaluate the VaR forecasting by the three EWMA methods. The null hypothesis is that the ratio of the number of losses exceeding VaR to the whole sample size is equal to the nominal level 1%. The slight green and the slight red horizontal lines are the critical values of 3.84 and 6.63 of chi square distribution of degree of freedom 1 at significance levels of 5% and 1%, respectively.
larger the VaR should be; otherwise, a smaller VaR should be obtained. Therefore the events that loss exceeds VaR should be independent, rather than clustered. In order to test this independence, Christofferson (1998) suggested the following likelihood ratio test.

Define an indicator function:

\[ I_t = \begin{cases} 
1, & \text{if } r_t < -VaR_t \\
0, & \text{otherwise} 
\end{cases} \]  

(3.5)

i.e., if the loss exceeds VaR, \( I_t \) equals 1, which we call state 1, and otherwise \( I_t \) is zero, which we call state 0. Based on \( I_1, \ldots, I_T \), Christofferson (1998) pointed out that if a VaR model is correct then the null hypothesis that the series \( \{I_t\} \) should be independent each other holds true. Christofferson (1998) gave the LR statistic as follows:

\[ LR_{in} = 2(\ln L_A - \ln L_0), \]  

(3.6)

where \( L_A = \pi_0 \frac{T_{00}}{T_{10}} \frac{T_{10}}{T_{11}} \pi_{11} \) \( L_0 = (1 - \pi) \frac{T_{00} + T_{10} + T_{11}}{T_{10} + T_{11}} \) is the times of the transforms from state \( i \) to state \( j \), \( \pi_{ij} = T_{ij} / (T_{00} + T_{10} + T_{11}) \) is the probability that the next state is \( j \) starting from state \( i \), \( \pi = (T_{01} + T_{11}) / (T_{01} + T_{11} + T_{00} + T_{10}) \) is the probability of transform to state 1. As \( T \) is large enough, \( LR_{in} \) is asymptotically \( \chi^2(1) \) distributed with degree of freedom 1 under this null hypothesis.

Figure 4 depicts the independence test based on (3.6) for one-period ahead VaR forecasting by using JP Morgan’s standard-EWMA, Guermat and Harris (2001)’s robust-EWMA, and our proposed skewed-EWMA.

3. Conditional coverage test

A complete test for VaR modelling is the testing of conditional coverage, which combines both tests of unconditional coverage and independence together. This test was put forward by Christofferson (1998), with LR statistic

\[ LR_c = -2 \ln L_T + 2 \ln L_A. \]  

(3.7)

As \( T \) tends to \( \infty \), \( LR_c \) is asymptotically \( \chi^2(2) \) distributed with degree of freedom 2. Notice that if the first observation is neglected in the calculation of unconditional coverage, then \( \pi \) is equal to \( N/T \), i.e., \( L_N = L_0 \), which implies that the three LR statistics in the above have the following relationship

\[ LR_c = LR_u + LR_{in}. \]  

(3.8)

It follows that \( LR_c \) statistic can simultaneously test the unconditional coverage and independence.

Figure 5 shows the conditional coverage test based on (3.8) for one-period ahead VaR forecasting by using JP Morgan’s standard-EWMA, Guermat and Harris (2001)’s robust-EWMA, and our proposed skewed-EWMA.

3.4. Summary of testing outcomes

From the empirical testing outcomes illustrated in Figures 2–4 for the VaR forecasting of the 6 financial return series by standard-EWMA, robust-EWMA and our skewed EWMA, we can summarise as follows:

For the unconditional coverage test in Figure 2, it shows that for the proposed skewed-EWMA based VaR forecasting, the proportion that the loss exceeds the forecasted VaR is approximately equal to the
Fig. 4: Likelihood ratio test of independence for the three EWMA forecasts of VaR at 1% nominal level. The x-axis is the common decay factor $\lambda$ for the three EWMA forecasts, and the y-axis is the LRT statistic. The green cross '+' line is for standard-EWMA, the blue cross 'x' line is for robust-EWMA, and the red dot line is for skewed-EWMA. The common decay factor $\lambda$ takes values from 0.85 to 0.99 for three EWMA forecasts while the decay factor $\beta = 0.998$ only in skewed EWMA. The time period of data is from 1\textsuperscript{st} January 1992 to 31\textsuperscript{st} December 2001 with the first 500 observations used to determine the decaying factors and the remaining observations after the first 500 ones used to evaluate the VaR forecasting by the three EWMA methods. The null hypothesis is that the events of the losses exceeding VaR are independent. The slight green and the slight red horizontal lines are the critical values of 3.84 and 6.63 of chi square distribution of degree of freedom 1 at significance levels of 5% and 1%, respectively.
Fig. 5: Likelihood ratio test of conditional coverage for the three EWMA forecasts of VaR at 1% nominal level. The $x$-axis is the common decay factor $\lambda$ for the three EWMA forecasts, and the $y$-axis is the LRT statistic. The green cross '+' line is for standard-EWMA, the blue cross 'x' line is for robust-EWMA, and the red dot line is for skewed-EWMA. The common decay factor $\lambda$ takes values from 0.85 to 0.99 for three EWMA forecasts while the decay factor $\beta = 0.998$ only in skewed EWMA. The time period of data is from 1st January 1992 to 31st December 2001 with the first 500 observations used to determine the decaying factors and the remaining observations after the first 500 ones used to evaluate the VaR forecasting by the three EWMA methods. The null hypothesis is that the ratio of the number of losses exceeding VaR to the whole sample size is equal to the nominal level 1% and the events of the losses exceeding VaR are independent. The slight green and the slight red horizontal lines are the critical values of 5.99 and 9.21 of chi square distribution of degree of freedom 2 at significance levels of 5% and 1%, respectively.
required nominal significance level 1%, while this proportion for the VaR forecasting based on standard-EWMA and robust-EWMA, respectively, all is largely greater than 1%. The likelihood ratio test in Figure 3 further validates this fact. The LR statistics of the skewed-EWMA based VaR forecasting are all less than the critical value of 6.63 at the testing significance level 1% and basically less than the 5% critical value of 3.84, indicating that the skewed-EWMA forecasting outcome is stable with respect to the decay factor and is acceptable in sense of the statistical testing. However, the LR statistics for the standard-EWMA and the robust-EWMA all basically exceed the 1% critical value, showing the worse forecasting outcomes than that by the proposed skewed-EWMA, where it is obvious that the robust-EWMA is relatively better than the standard-EWMA, as indicated in Guermat and Harris (2001).

As for the independence test indicated in Figure 4, the null hypothesis that the events of losses exceeding the corresponding VaR are independent cannot be rejected, for all three EWMA forecasts at the testing significance level of 1%, with respect to almost all $\lambda$ for the British Ponds and the Canadian Dollars, and with respect to $\lambda \leq 0.91$ for other return series. Basically our skewed-EWMA performs similarly to but still a little bit better than both standard-EWMA and robust-EWMA.

The conditional coverage test in Figure 5 is a combined test covering the unconditional coverage and independence tests. Clearly, our skewed-EWMA have much better outcomes than both teh standard-EWMA and the robust-EWMA, with LR statistics basically less than the 1% critical value. The standard-EWMA and the robust-EWMA are basically rejected at the testing significance level of 1%.

4. Conclusion

VaR, since suggested in 1993, has received wide attention both in financial practice and in academic research. Various methodologies have been developed, among which the parametric approach is most popular due to its ease of application. In particular the JP Morgan’s (1996) RiskMetrics system developed the parametric method of standard-EWMA forecasting, which is extensively applied by financial practitioners.

In this paper, we have proposed a new VaR forecasting model, named skewed-EWMA. It is a generalization of the standard EWMA method in RiskMetrics and the Guermat and Harris (2001)’s robust EWMA. It takes into account both the skewness and heavy tails of financial return distribution. We have suggested an adaptive EWMA adjustment of the shape parameter in the skewed-EWMA forecasting to characterize the time-varying nature of skewness and kurtosis in practice. We have compared the performance of the proposed skewed-EWMA with those of the standard EWMA and the robust EWMA as done in Guermat and Harris (2001). Empirical applications to the forecasting of the value at risk for 6 financial return series from foreign exchange rates and stock market indexes all illustrate that the proposed method viably outperforms the standard EWMA method of the JP Morgan’s RiskMetrics and the Guermat and Harris (2001)’s robust EWMA.

We expect that the idea of the skewed-EWMA dynamic forecasting would be promising in modelling the risks of dynamic portfolios and nonlinear derivatives, extensions to which are non-trivial (c.f., Lu and Li (2008)). Other testing of the performances can also be examined. We will leave these for future research.
References


