The Pennsylvania State University

The Graduate School

Eberly College of Science

VALUE-AT-RISK FORECASTING BASED ON THE ASYMMETRIC EXPONENTIAL POWER DISTRIBUTION

A Thesis in

Statistics

by

Lu Ou

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science

August 2016

The thesis of Lu Ou was reviewed and approved* by the following:

Zhibiao Zhao

Associate Professor of Statistics

Thesis Advisor

Runze Li

Professor of Statistics

David Hunter

Professor of Statistics

Department Head

*Signatures are on file in the Graduate School.

Abstract

Value-at-risk (VaR) is a standard measure of market risk in financial markets. This paper proposes a novel, adaptive, and flexible method to forecast volatility, asymmetry, and VaR. As an extension of the existing exponential smoothing as well as GARCH formulations, the method is motivated from an Asymmetric Exponential Power distribution, which includes the Laplace, Normal, and Uniform distributions as special cases and takes into account the potentially time-varying nature in volatility and skewness of financial return series. Results from a Monte Carlo simulation study and an empirical application to the S&P 500 index illustrate that the proposed method offers more accurate and robust VaR forecasts across a range of models at different confidence levels compared to other popular parametric alternatives.

Table of Contents

List of Figures

List of Tables

Acknowledgments

Special thanks go to Dr. Zhibiao Zhao, who supports and helps me all the way towards the completion of this master thesis. I appreciate all the efforts and time that he has invested in this project.

I also feel grateful for the great advice and help from Professors Runze Li and David Hunter.

I appreciate the support of Dr. Sy-Miin Chow, my PhD advisor in Human Development and Family Studies, in I obtaining this concurrent degree.

I thank Melanie Mckinney and Christa Kreps for providing me with resources that relate to the thesis. Without their guidance and support, I would not be able to achieve this milestone smoothly.

I very much appreciate the company and encouragement of my friends as well.

Chapter 1
Introduction

In financial portfolio management, the return and its associated risk are the two most important factors of interest. A return is a price change defined relative to some initial price. While the average return by definition is the mean value of the return of a portfolio, the risk of the return concerns its uncertainty and potentially large loss. Long urged by regulators to make efforts to understand and control financial risk, financial institutions periodically monitor their risk in dynamic portfolio management.

Among many other risk measures such as standard deviation, expected shortfall, entropic risk measure, and tail conditional expectation, Value-at-risk (VaR) plays an important role in financial management. Despite criticism, VaR has been endorsed by regulatory authorities and popularized in industry by the RiskMetrics model of J. P. Morgan 1996. In its most general form, VaR measures the potential loss in value of an asset, a portfolio, or a firm over a defined period for a given confidence interval. That is, for a given time horizon Δt and confidence level α, the VaR is the threshold VaR_α (for simplicity we drop the dependence on the fixed time period

Δt of loss in market value such that

$$\mathbb{P}\{\text{Loss during } \Delta t \geq \text{VaR}_\alpha\} = \alpha, \tag{1.1}$$

where the confidence level α is often taken to be 5% or 1%. For example, if the VaR of an asset is \$1 million at a confidence level of 99% during a fixed period of one-week, the potential loss of the asset over that one-week period exceeds the threshold of \$1 million with probability 1%. In other words, the probability that the market value of the asset will drop more than \$1 million over that week is 1%. In practice, the loss is the negative value of the return. Thus, (1.1) can be equivalently defined as

$$\mathbb{P}\{\text{Return during } \Delta t \leq -\text{VaR}_\alpha\} = \alpha. \tag{1.2}$$

To estimate VaR at a given confidence level, we usually need to define the probability distribution of the loss (or return) of interest. In practice, the return X_{t+1} from time t to $t+1$ is the log price change or continuously-compounded return, defined as

$$X_{t+1} = \ln(\frac{P_{t+1}}{P_t}) = p_{t+1} - p_t, \tag{1.3}$$

where P_t and $p_t = \ln(P_t)$ are the price and log price at time t, respectively. For estimation and forecasting purposes, we need to specify the evolution and distribution of the returns over time. A widely used class of models that describes the evolution of returns over time is based on modeling financial log prices as a random walk model such that

$$X_{t+1} = \sigma_{t+1}\epsilon_{t+1}, \quad \epsilon_{t+1} \sim N(0,1), \tag{1.4}$$

where we follow the convention to drop the mean term, and the standard deviation σ_{t+1} of log price changes varies over time, mirroring the fluctuations of returns around zero and their time-dependent volatility in their time series realizations. In addition, σ_{t+1} is often successively modeled as a function of past information up to time t as in the autoregressive conditional heteroscedasticity (ARCH) model (Engle, 1982, 1983). Since the introduction of the basic ARCH model (Engle, 1982), extensions include generalized ARCH (GARCH; Bollerslev, 1986) and Integrated GARCH (IGARCH), with the latter being a special case of the former. In a 0-intercept IGARCH(1,1),

$$X_{t+1}|\mathcal{I}_t \sim N(0, \sigma_t^2), \tag{1.5}$$

$$\sigma_{t+1}^2 = \lambda\sigma_t^2 + (1-\lambda)X_t^2, \tag{1.6}$$

where $\mathcal{I}_t$ is the information set (σ-field) of all information up to time t, and the decay factor $\lambda \in (0,1)$ determines the relative weights that are applied to the returns and the effective amount of data used in estimating volatility. If we recursively substitute the previous conditional variance using the one-step ahead equation (1.6), we can obtain $\sigma_{T+1}^2 = \sum_{i=0}^{\infty}(1-\lambda)\lambda^i X_{T-i}^2$. Since $\lambda \in (0,1)$, when the number of available time points T is sufficiently large so that $\lambda^i \approx 0, i \geq T$, can be ignored, the volatility forecast at time $T+1$ given information up to time T can be estimated as

$$\hat{\sigma}_{T+1,standard}^2 = \sum_{i=0}^{T-1}(1-\lambda)\lambda^i X_{T-i}^2 = (1-\lambda)\sum_{i=1}^{T}\lambda^{T-i}X_i^2, \tag{1.7}$$

which puts geometrically declining weights on past squared returns and yields standard exponentially weighted moving average (EWMA; J. P. Morgan, 1996) estimator of the σ^2_{T+1}. If the average of the past returns is 0, this estimator approximates the sample variance $\hat{\sigma}^2 = \frac{1}{T}\sum_{i=1}^{T} X_i^2$ as $\lambda \to 1$, given a sufficiently large T. With the standard EWMA estimator $\hat{\sigma}^2_{T+1,standard}$, we can then use (1.2) to forecast the VaR at time $T+1$ as

$$\widehat{\text{VaR}}_{T+1,standard} = -\hat{\sigma}_{T+1,standard}\Phi^{-1}(\alpha), \tag{1.8}$$

where $\Phi^{-1}(\cdot)$ is the inverse cumulative distribution function (cdf) of the standard Normal distribution.

Although the conditional normal distribution assumed in the 0-intercept IGARCH(1,1) (Equation 1.5) yields an unconditional distribution that is leptokurtic as desired for financial returns, the standard EWMA estimator is asymptotically inefficient, putting too much weight on extreme observations (Guermat & Harris, 2001, 2002). Guermat & Harris (2001) assumed the conditional distribution of the returns follows a Laplace or double-exponential distribution, whose probability density function is given by

$$f(r) = \frac{1}{\sqrt{2}\sigma} \exp\left\{-\frac{\sqrt{2}|r|}{\sigma}\right\}, \tag{1.9}$$

and proposed a more robust EWMA-based volatility estimator, which is an EWMA version of the maximum likelihood estimator (MLE) of the standard deviation, $\hat{\sigma} = \frac{1}{T}\sum_{i=1}^{T}\sqrt{2}|X_i|$, of the Laplace distribution. By introducing the decay factor $\lambda \in (0,1)$, we obtain the robust-EWMA estimator as an exponentially weighted average of past absolute returns, given by

$$\hat{\sigma}_{T+1,robust} = (1-\lambda)\sum_{i=1}^{T}\lambda^{T-i}|X_i|, \tag{1.10}$$

where the approximation $\sum_{t=1}^{T}\lambda^{T-i} \approx \frac{1}{(1-\lambda)}$ has been applied, given a sufficiently large T. Using absolute returns rather than squared returns allows the estimator to be less sensitive to extreme values and thus more efficient when the conditional distribution of returns is leptokurtic. Accordingly, we forecast the VaR at time $T+1$ as

$$\widehat{\text{VaR}}_{T+1,robust} = -\hat{\sigma}_{T+1,robust}\frac{\log(2(1-\alpha))}{\sqrt{2}}. \tag{1.11}$$

While the Laplace distribution can account for the heavy tails of the conditional distributions of returns, it does not take skewness or time-varying higher moments into consideration. With respect to this regard, Gerlach et al. (2013) employed the asymmetric Laplace distribution to model the conditional distribution of returns, yielding a skewed EWMA estimator of the conditional volatility and a skewed-EWMA-based forecast of VaR, for which the equations will be presented as a special case of the estimators and forecast we will discuss later, in that we are motivated to model the conditional distribution of returns using a more generalized distribution to allow for more flexibility in choosing different probability distributions.

Recently, Wang & Zhao (2016) studied conditional VaR using a semiparametric approach. In their approach, the noises are independent and identically distributed and independent of the historical information. In the current approach, we follow Gerlach et al. (2013) and allow the conditional distribution of the noises to be dependent on the historical information through some time-varying parameters. If we view the noises as the random shocks to the market, this time-varying framework allows the distribution of these random shocks to change over time.

The rest of the thesis is organized as follows. Chapter 2 studies some properties of the asymmetric exponential power (AEP) distribution, which forms the basis of our approach. In Chapter 3, we discuss the new VaR forecasting method based on the AEP distribution. Chapter 4 presents some numerical studies, including both simulation studies and a real application.

Chapter 2

Asymmetric Exponential Power (AEP) Distribution

Throughout, we write the gamma function $\Gamma(t) = \int_0^\infty y^{t-1}e^{-y}dy, t > 0$. Introduce

$$f(x|\beta, \sigma, p) = \frac{1}{\sigma\Gamma(1+1/\beta)} \exp\left\{-\left[\frac{\mathbf{1}_{x>0}}{p^\beta} + \frac{\mathbf{1}_{x\leq 0}}{(1-p)^\beta}\right]\left|\frac{x}{\sigma}\right|^\beta\right\}, \quad x \in \mathbb{R}, \quad (2.1)$$

for parameters $\beta > 0, \sigma > 0, p \in (0,1)$, where $\mathbf{1}_{o(\cdot)} : \mathbb{R} \to \{0,1\}$ is an indicator function such that $\mathbf{1}_{o(x)} = 1$ when the logical function $o(\cdot) : \mathbb{R} \to \{true, false\}$ is true at x and $\mathbf{1}_{o(x)} = 0$ when $o(x)$ is evaluated as false. It is easy to show that, for $\beta > 0$ and $c > 0$,

$$\int_0^\infty \exp\left\{-\left|\frac{x}{c}\right|^\beta\right\} dx = \int_{-\infty}^0 \exp\left\{-\left|\frac{x}{c}\right|^\beta\right\} dx = c\Gamma(1+1/\beta). \quad (2.2)$$

Thus, $\int_{-\infty}^\infty f(x|\beta,\sigma,p)dx = (\int_0^\infty + \int_{-\infty}^0) f(x|\beta,\sigma,p)dx = p + (1-p) = 1$, and consequently $f(x|\beta,\sigma,p)$ is a density function. This distribution is often termed as asymmetric exponential power (hereafter, AEP) distribution [Ayebo and Kozubowski (2003), under different parametrizations].

Example 1. If $p = 1/2$, $f(x|\beta, \sigma, p) \propto \exp(-c|x|^\beta)$ with $c = (2/\sigma)^\beta$, which is the generalized error distribution. Furthermore, $\beta = 2$ corresponds to the normal distribution, and $\beta = 1$ corresponds to the Laplace distribution, respectively.

Example 2. If $p \neq 1/2$, $f(x|\beta, \sigma, p) \propto \exp\{-(c_1|x|^\beta \mathbf{1}_{x>0} + c_2|x|^\beta \mathbf{1}_{x\leq 0})\}$ for some $c_1 \neq c_2$. That is, the coefficients c_1 and c_2 for the power form $|x|^\beta$ are different for the two cases $x > 0$ and $x \leq 0$, thus creating asymmetry. Furthermore, if $\beta = 1$, then it reduces to the skewed Laplace distribution in Gerlach et al. (2013).

Proposition 1 presents some properties of the AEP distribution.

Proposition 1. *Assume* $X \overset{D}{=} f(x|\beta, \sigma, p)$. *Then*

(i) $\mathbb{P}(X > 0) = p$ *and* $\mathbb{P}(X \leq 0) = 1 - p$.

(ii) $X \overset{D}{=} \sigma\varepsilon$ *with* $\varepsilon \overset{D}{=} f(x|\beta, 1, p)$.

(iii) for any $\nu \geq 0$,

$$\mathbb{E}(|X|^\nu \mathbf{1}_{X>0}) = \frac{\Gamma(\frac{\nu+1}{\beta})}{\Gamma(\frac{1}{\beta})}\sigma^\nu p^{\nu+1} \tag{2.3}$$

and

$$\mathbb{E}(|X|^\nu \mathbf{1}_{X\leq 0}) = \frac{\Gamma(\frac{\nu+1}{\beta})}{\Gamma(\frac{1}{\beta})}\sigma^\nu (1-p)^{\nu+1}. \tag{2.4}$$

From Proposition 1(i), the parameter p is the probability of $X > 0$. When $p < 1/2$, the probability of $X \leq 0$ is larger than that of $X > 0$, so the density function is skewed to the left, whereas the density function is skewed to the right when $p > 1/2$. From Proposition 1(ii), the parameter σ can be interpreted as the scale parameter. We can also add a location parameter μ by considering $f(x-\mu|\beta,\sigma,p)$, but for simplicity we consider only $\mu = 0$.

The parameter β determines the main shape, in particular the tail behavior, of the AEP distribution. The smaller β, the relatively heavier tail. To measure the heaviedness of the tail of a random variable, we use the kurtosis defined as

$$\kappa_X = \frac{\mathbb{E}[(X-\mu_1)^4]}{\{\mathbb{E}[(X-\mu_1)^2]\}^2}, \quad \text{where} \quad \mu_1 = \mathbb{E}(X).$$

For normal distribution, the kurtosis is 3. A distribution with kurtosis larger than 3 is said to have a heavy tail.

Proposition 2 studies the kurtosis for the AEP distribution. Since Kurtosis is invariant under linear transformation, we consider the special case $\sigma = 1$.

Proposition 2. *Assume* $X \stackrel{D}{=} f(x|\beta,1,p)$. *Then*

$$\kappa_X = \frac{\mu_4 - 4\mu_3\mu_1 + 6\mu_2\mu_1^2 - 3\mu_1^4}{(\mu_2-\mu_1^2)^2},$$

where

$$\mu_k \ := \ \mathbb{E}(X^k) = \frac{\Gamma(\frac{k+1}{\beta})}{\Gamma(\frac{1}{\beta})}[p^{k+1} + (-1)^k(1-p)^{k+1}]. \tag{2.5}$$

To visualize how kurtosis changes with (β,p), Figure 2.1 plots the kurtosis for different choice of (β,p). We observe that: (i) the kurtosis decreases quickly with $\beta \in [1,2]$; (ii) the kurtosis changes slowly for $\beta \geq 4$; and (iii) the kurtosis is largely affected by β but relatively not sensitive to p. Thus, by using different choice of (β,p), the AEP distribution can be used to model data with the kurtosis in a wide range.

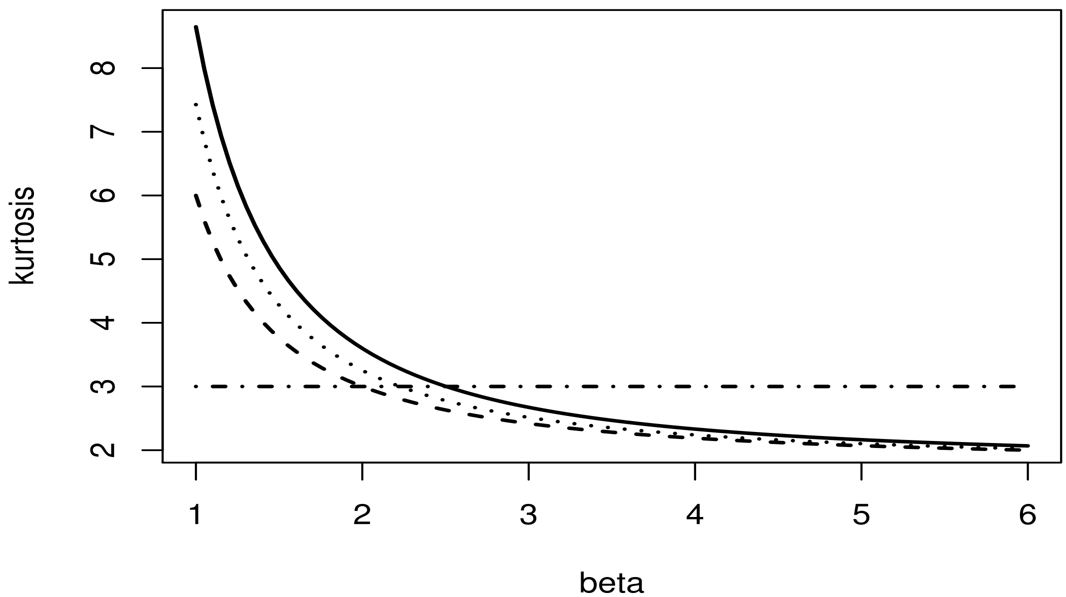

Figure 2.1: Kurtosis for $X \stackrel{D}{=} f(x|\beta, 1, p)$. Solid curve is the kurtosis for $p = 0.15$ and varying $\beta \in [1, 6]$. Similarly, dotted and dash curves are kurtosis curves corresponding to $p = 0.3$ and $p = 0.5$, respectively, and $\beta \in [1, 6]$. The horizontal dot-dashed line is the kurtosis $\kappa_X = 3$ for normal distribution.

The behavior of AEP distribution when $\beta \approx 0$: When $\beta \approx 0$, using the approximation $|x|^\beta \approx 1 + \beta \log(|x|)$, we have for any fixed $c > 0$ and $x \neq 0$,

$$\exp(-c|x|^\beta) \approx \exp\{-c[1 + \beta \log(|x|)]\} = \exp(-c)|x|^{-c\beta}.$$

Thus, the exponential power distribution with a small shape parameter β behaves like the power law. For the AEP distribution (2.1), when $\beta \approx 0$, we have

$$f(x|\beta, \sigma, p) \propto \exp(-c_1/\beta)|x|^{-c_1}\mathbf{1}_{x>0} + \exp(-c_2/\beta)|x|^{-c_2}\mathbf{1}_{x\leq 0},$$

where $c_1 = \beta(p\sigma)^{-\beta}$ and $c_2 = \beta[(1-p)\sigma]^{-\beta}$. For Student-t distribution with $\nu > 0$ degrees of freedom, its density is proportional to $(1+x^2/\nu)^{-(\nu+1)/2} \propto x^{-(\nu+1)}$ for large x. For VaR analysis, we are interested in the extreme tail. Thus, the AEP distribution with a small shape parameter β can capture the tail behavior of student-t distribution; moreover, when $p \neq 1/2$, it has the flexibility of modeling different tail behavior of the two sides $x > 0$ and $x \leq 0$ by allowing $c_1 \neq c_2$.

Proposition 3 studies the cumulative distribution function (cdf) of the AEP distribution.

Proposition 3. *Consider the AEP distribution (2.1) with $\sigma = 1$. The cdf is given by*

$$F(x|\beta,1,p) = \int_{-\infty}^{x} f(x|\beta,1,p)dx = \begin{cases} 1-p+\frac{p}{\Gamma(1/\beta)}\underline{\Gamma}((x/p)^{\beta},1/\beta), & x>0; \\ 1-p-\frac{1-p}{\Gamma(1/\beta)}\underline{\Gamma}([-x/(1-p)]^{\beta},1/\beta), & x \leq 0. \end{cases}$$

Here, $\underline{\Gamma}(z,t) = \int_0^z y^{t-1}e^{-y}dy$ is the lower incomplete gamma function. Consequently, the $\alpha \in (0,1)$ quantile is given by the inverse of $F(x|\beta,1,p)$:

$$F^{-1}(\alpha|\beta,1,p) = \begin{cases} p[\underline{\Gamma}^{-1}(z_1,1/\beta)]^{1/\beta}, & \alpha > 1-p; \\ -(1-p)[\underline{\Gamma}^{-1}(z_2,1/\beta)]^{1/\beta}, & \alpha \leq 1-p. \end{cases}$$

Here, $\underline{\Gamma}^{-1}(z,t)$ is the inverse function of $\underline{\Gamma}(z,t)$ with respect to z, $z_1 = [1-(1-\alpha)/p]\Gamma(1/\beta)$, and $z_2 = [1-\alpha/(1-p)]\Gamma(1/\beta)$.

Chapter 3

VaR Forecasting Using AEP Distribution

3.1 A general framework

Let $\mathcal{I}_t$ be the information set up to time t. Due to the nice properties of the AEP distribution in Section 2, we propose to model the conditional distribution $X_{t+1}|\mathcal{I}_t$ using the AEP distribution $f(x|\beta, \sigma, p)$. Thus, similar to (1.5), we propose

$$X_{t+1}|\mathcal{I}_t \sim f(x|\beta, \sigma_{t+1}, p_{t+1}), \tag{3.1}$$

where $\beta > 0$ is a time-invariant parameter, σ_{t+1} and p_{t+1} are time-varying parameters. That is, we allow the scale parameter σ_{t+1} and the probability parameter p_{t+1} in the AEP distribution (2.1) to be time-varying and dependent on the historical information $\mathcal{I}_t$. By Proposition 1(i), we have

$$\mathbb{P}\{X_{t+1} > 0|\mathcal{I}_t\} = p_{t+1}. \tag{3.2}$$

That is, p_{t+1} can be interpreted as the conditional probability of a positive return at time $t+1$ given the historical information $\mathcal{I}_t$. For classical ARCH, GARCH, and their variants, $\mathbb{P}\{X_{t+1} > 0|\mathcal{I}_t\}$ is constant (in fact, 1/2). In practice, it is more reasonable to use a time-varying probability p_{t+1} instead of a constant probability 1/2 as in the classical approach.

By Proposition 1(ii), $X_{t+1}|\mathcal{I}_t \sim \sigma_{t+1}\varepsilon_{t+1}$ with $\varepsilon_{t+1} \sim f(x|\beta, 1, p_{t+1})$. Let $F^{-1}(u|\beta, 1, p)$ be the inverse function of the cdf $F(x|\beta, 1, p)$ in Proposition 3. Then at level α the theoretical VaR at time $T+1$ is

$$\mathrm{VaR}_{T+1} = -\sigma_{T+1}F^{-1}(\alpha|\beta, 1, p_{T+1}). \tag{3.3}$$

This theoretical VaR is infeasible as it involves the unknown time-varying quantities (σ_{T+1}, p_{T+1}). Below we impose some evolution dynamics for σ_{t+1} and p_{t+1}.

Different specification of the dynamics for σ_{t+1} and p_{t+1} results in different approach. However, a reasonable specification should be such that, at each time point, the time-varying model quantities should optimally (in some sense) use the information available prior to the current time. Thus, it is reasonable to adopt the maximum likelihood approach.

3.2 Motivation from maximum likelihood estimation (MLE)

Now we discuss the Maximum Likelihood Estimator (MLE) for the AEP distribution (2.1). To motivate the idea, in (3.1) we assume that $\sigma_{t+1} = \sigma$ and $p_{t+1} = p$ are time-invariant constants and we will study how the MLE of (σ, p) relates to samples, which will then motivate the model dynamics for (σ_{t+1}, p_{t+1}).

Given i.i.d. samples $X_1, \ldots, X_n$ from $f(x|\beta, \sigma, p)$, the likelihood function is

$$\begin{aligned} L(\beta, \sigma, p) &= \prod_{t=1}^{n} f(X_t|\beta, \sigma, p) \\ &= \frac{1}{[\sigma\Gamma(1+1/\beta)]^n} \prod_{t=1}^{n} \exp\left\{ -\left[\frac{\mathbf{1}_{X_t>0}}{p^\beta} + \frac{\mathbf{1}_{X_t\le 0}}{(1-p)^\beta} \right] \left| \frac{X_t}{\sigma} \right|^\beta \right\}, \end{aligned}$$

and the log likelihood function is

$$\begin{aligned} \ell(\beta, \sigma, p) &= \log L(\beta, \sigma, p) \\ &= -n\log(\sigma) - n\log\Gamma(1+1/\beta) - \frac{nA}{(p\sigma)^\beta} - \frac{nB}{[(1-p)\sigma]^\beta}, \end{aligned} \quad (3.4)$$

where A and B are given by

$$A = \frac{1}{n}\sum_{t=1}^{n} |X_t|^\beta \mathbf{1}_{X_t>0}, \quad (3.5)$$

$$B = \frac{1}{n}\sum_{t=1}^{n} |X_t|^\beta \mathbf{1}_{X_t\le 0}. \quad (3.6)$$

The MLE of $(\beta_1, \beta_2, \sigma, p)$ is obtained by maximizing $\ell(\beta, \sigma, p)$. Due to the complicated dependence on the parameter (β, σ, p), we will proceed in two steps.

(Step 1): We first assume that β is known and (σ, p) is unknown.

Theorem 1. *Assume β is known. Then the MLE of (σ, p) is*

$$p = \frac{A^{1/(\beta+1)}}{A^{1/(\beta+1)} + B^{1/(\beta+1)}}, \quad (3.7)$$

$$\sigma^\beta = \frac{\beta A}{p^\beta} + \frac{\beta B}{(1-p)^\beta}. \quad (3.8)$$

In Section 3.3 below, we use Theorem 1 to derive some dynamic models with some time-varying versions of (σ, p), leading to the proposed new VaR approach.

(Step 2): When β is unknown, for each given β, denote by $\sigma(\beta)$ and $p(\beta)$ the solution given in the equations (3.7) and (3.8). Plugging $\sigma(\beta)$ and $p(\beta)$ into (3.4), the MLE of β is obtained through

$$\hat{\beta} = \underset{\beta}{\text{argmin}} \left\{ -n\log(\sigma) - n\log\Gamma(1+1/\beta) - \frac{nA}{(p\sigma)^\beta} - \frac{nB}{[(1-p)\sigma]^\beta} \right\}, \qquad (3.9)$$

with $\sigma = \sigma(\beta)$ and $p = p(\beta)$. Note that A and B are also functions of β. Finally, the MLE of (σ, p) is $(\sigma(\hat{\beta}), p(\hat{\beta}))$.

Theorem 2 below is from Komunjer (2007) who established the asymptotic normality of the MLE.

Theorem 2 (Komunjer (2007))**.** *Let $\theta = (\beta, \sigma, p)$. Assume θ is in a compact set. Let $\hat{\theta}$ be the MLE. Then*

$$\sqrt{n}(\hat{\theta} - \theta) \to N(0, \mathcal{J}^{-1}),$$

where $\mathcal{J} = \mathbb{E}\{[\partial \log f(X_0|\theta)/\partial\theta][\partial \log f(X_0|\theta)/\partial\theta]'\}$ is the Fisher information matrix.

3.3 Time-varying dynamics for (σ_{t+1}, p_{t+1})

As discussed in Section 3.4, when we pretend that $\sigma_{t+1} = \sigma$ and $p_{t+1} = p$ are time-invariant constants, then by Theorem 1, for given β the MLE of (σ, p) has explicit forms depending on A and B defined in (3.5)–(3.6). Naturally, we use these explicit solutions with some time-varying versions of A and B to construct time-varying versions (σ_{t+1}, p_{t+1}).

In (3.5), A is a simple sample average of $|X_t|^\beta \mathbf{1}_{X_t>0}$ with uniform weights for all returns. To create a time-varying version of A at time t, it is intuitively reasonable to use larger weights for returns that are closer to the current time t. One popular approach is the exponentially weighted moving average (EWMA; the RiskMetrics by J. P. Morgan, 1996):

$$A_t = \sum_{i=0}^{\infty}(1-\lambda_1)\lambda_1^i |X_{t-i}|^{\beta}\mathbf{1}_{X_{t-i}>0}, \tag{3.10}$$

where $\lambda_1 \in (0,1)$ is some decay factor. Essentially, A_t uses the exponential weights $\omega_i = (1-\lambda_1)\lambda_1^i$, satisfying $\sum_{i=0}^{\infty}\omega_i = 1$, to replace the uniform weights in the simple sample average version A, so that distant returns have exponentially decaying weights. Note that A_t has the equivalent recursive representation

$$A_t = \lambda_1 A_{t-1} + (1-\lambda_1)|X_t|^{\beta}\mathbf{1}_{X_t>0}. \tag{3.11}$$

Similarly, we create the time-varying version of B as

$$B_t = \sum_{i=0}^{\infty}(1-\lambda_2)\lambda_2^i |X_{t-i}|^{\beta}\mathbf{1}_{X_{t-i}\le 0}, \tag{3.12}$$

for another decay factor $\lambda_2 \in (0,1)$, and B_t has the equivalent recursive representation

$$B_t = \lambda_2 B_{t-1} + (1-\lambda_2)|X_t|^{\beta}\mathbf{1}_{X_t\le 0}. \tag{3.13}$$

Substituting (3.11)–(3.13) into (3.7)–(3.8), we obtain the time-varying version (p_{t+1}, σ_{t+1}):

$$p_{t+1} = \frac{A_t^{1/(\beta+1)}}{A_t^{1/(\beta+1)} + B_t^{1/(\beta+1)}}, \tag{3.14}$$

$$\sigma_{t+1}^{\beta} = \frac{\beta A_t}{p_{t+1}^{\beta}} + \frac{\beta B_t}{(1-p_{t+1})^{\beta}}. \tag{3.15}$$

In the recursive equation for σ_{t+1} in (3.15), both the power and the coefficients of σ_t are different. The coefficients of A_t and B_t impose different effects from positive return $X_t > 0$ and negative return $X_t \leq 0$. Also, the coefficients are time-varying, while the coefficients are usually deterministic constants in classical GARCH type models.

Example 3. If we use constant probability $p_{t+1} = p$, then (3.15) becomes

$$\sigma_{t+1}^{\beta} = \beta \left[\frac{A_t}{p^{\beta}} + \frac{B_t}{(1-p)^{\beta}} \right]. \tag{3.16}$$

Furthermore, if we use $\lambda_1 = \lambda_2 = \lambda$ in (3.11) and (3.13), then (3.16) has the recursive representation

$$\sigma_{t+1}^{\beta} = \lambda \sigma_t^{\beta} + \rho_t |X_t|^{\beta}, \quad \rho_t = \beta(1-\lambda) \left[\frac{\mathbf{1}_{X_t>0}}{p^{\beta}} + \frac{\mathbf{1}_{X_t \leq 0}}{(1-p)^{\beta}} \right], \tag{3.17}$$

which has a GARCH(1,1) structure. The coefficient ρ_t is different for the two regions $X_t > 0$ and $X_t \leq 0$, and thus this model can model some asymmetry effects due to positive ($X_t > 0$) and negative ($X_t \leq 0$) returns. In particular, if $\beta = 2$, (3.17) becomes the GJR-GARCH specification which has the form $\sigma_{t+1}^2 = a_0 + a_1 X_t^2 + a_2 \sigma_t^2 + a_3 X_t^2 \mathbf{1}_{X_t<0}$. Thus, (3.17) is a generalization of the GJR-GARCH (Glosten et al., 1993) specification to a general power other than the square. In the symmetric case $p = 1/2$, (3.17) further reduces to

$$\sigma_{t+1}^{\beta} = \lambda \sigma_t^{\beta} + (1-\lambda)\beta 2^{\beta} |X_t|^{\beta}. \tag{3.18}$$

which includes Bollerslev (1986)'s GARCH(1,1) as a special case with $\beta = 2$ (see also Bollerslev et al., 1992, 1994).

From Example 3, the proposed evolution dynamics (3.15) is flexible enough to include several existing popular GARCH models as special cases. It is worth mentioning that, in (3.14) and (3.15), since (A_t, B_t) and so σ_{t+1} depends on $(X_t, X_{t-1}, \ldots)$, σ_{t+1} has the ARCH(∞) structure.

Next, we argue that the proposed time-varying probability p_{t+1} framework can model the phenomenon of clustering of positive returns in a bull market, clustering of negative returns in financial crisis, and a roughly equal mixture of positive and negative returns in a range market. In (3.14), $\{p_{t+1}\}$ is a sequence of random variables as it depends on (A_t, B_t). We consider its "progressively averaged" version

$$p_{t+1} = \frac{\overline{A_t}^{1/(\beta+1)}}{\overline{A_t}^{1/(\beta+1)} + \overline{B_t}^{1/(\beta+1)}}, \tag{3.19}$$

by replacing (A_t, B_t) therein with their "progressively" conditional expectations $(\overline{A_t}, \overline{B_t})$ defined as

$$\begin{aligned}
\overline{A_t} &= \sum_{i=0}^{\infty}(1-\lambda_1)\lambda_1^i \mathbb{E}(|X_{t-i}|^{\beta_1}\mathbf{1}_{X_{t-i}>0}|\mathcal{I}_{t-i-1}), \\
\overline{B_t} &= \sum_{i=0}^{\infty}(1-\lambda_2)\lambda_2^i \mathbb{E}(|X_{t-i}|^{\beta_1}\mathbf{1}_{X_{t-i}\le 0}|\mathcal{I}_{t-i-1}).
\end{aligned}$$

The expectations are taken conditioning on the immediately past information. Below we argue that if $(\sigma_s, p_s) \approx (\sigma, p)$, for some (σ, p), has persisted for a period of time $s = t, t-1, \ldots, t-J$, for some reasonably large J, then $(\sigma_{t+1}, p_{t+1}) \approx (\sigma, p)$. That is, the same pattern will likely continue to persist, given it has already persisted for some time period. In the infinite order representation (3.10), due to

$\lambda \in (0,1)$ and the exponentially decaying factor λ_1^i, when J is large, the terms beyond J are relatively very small. Applying this truncation and (2.3)-(2.4) with $(\sigma_s, p_s) \approx (\sigma, p)$ for $t - J \le s \le t$, we have

$$\begin{aligned}
\overline{A_t} &\approx \sum_{i=0}^{J} (1-\lambda_1)\lambda_1^i \mathbb{E}(|X_{t-i}|^\beta \mathbf{1}_{X_{t-i}>0} | \mathcal{I}_{t-i-1}) \\
&= \sum_{i=0}^{J} (1-\lambda_1)\lambda_1^i \frac{\Gamma(1+1/\beta)}{\Gamma(1/\beta)} \sigma_{t-i}^\beta p_{t-i}^{\beta+1} \\
&\approx \frac{\sigma^\beta p^{\beta+1}}{\beta}.
\end{aligned}$$

Here the last expressions comes from $\Gamma(1+x) = x\Gamma(x)$ and $\sum_{i=0}^{J}(1-\lambda_1)\lambda_1^i \approx 1$ for large J, and $(\sigma_s, p_s) \approx (\sigma, p)$ for $s = t, t-1, \ldots, t-J$. Similarly, we can show $\overline{B_t} \approx \sigma^\beta (1-p)^{\beta+1}/\beta$. Substituting these approximations to (3.19), we have $p_{t+1} \approx p$.

Similarly, if we further consider the "progressively averaged" version of σ_{t+1} by replacing (A_t, B_t) in (3.15 with their "progressively" conditional expectations $(\overline{A_t}, \overline{B_t})$ and use the approximation $p_{t+1} \approx p$ that we just obtained, similar calculations show that $\sigma_{t+1} \approx \sigma$.

Therefore, the derived model evolution dynamics (3.14)–(3.15) implies that if the market enters into a bull, bear, or range market for a period of time, it will have the tendency to continue the same pattern. However, the above analysis is based on the "progressively averaged" version and there are always uncertainties and outliers in the market. When substantial outliers kick in to break the pattern, new pattern will emerge and persist until new outliers kick in.

3.4 MLE of $(\beta, \lambda_1, \lambda_2)$ and the VaR forecasting algorithm

The VaR forecasting for X_{T+1} is obtained by substituting the model dynamics (3.14)–(3.15) into (3.3). However, the unknown parameters $(\beta, \lambda_1, \lambda_2)$ need to be estimated.

Now we discuss the MLE of $(\beta, \lambda_1, \lambda_2)$ in model (3.1), (3.14) and (3.15). Given $X_1, \ldots, X_n$, denote by $f(X_{t+1}|\mathcal{I}_t)$ the conditional density of X_{t+1} given $\mathcal{I}_t$. From (3.1), $f(X_{t+1}|\mathcal{I}_t) = f(X_{t+1}|\beta, \sigma_{p+1}, p_{t+1})$. Then, given X_1, the conditional log likelihood of $(X_2, \ldots, X_n)$ is

$$\begin{aligned}
& \sum_{t=1}^{n-1} \log f(X_{t+1}|\mathcal{I}_t) \\
= & \sum_{t=1}^{n-1} \log f(X_{t+1}|\beta, \sigma_{p+1}, p_{t+1}) \\
= & -n \log \Gamma(1+1/\beta) - \sum_{t=1}^{n-1} \left\{ \log(\sigma_{t+1}) + \left[\frac{\mathbf{1}_{X_{t+1}>0}}{p_{t+1}^{\beta}} + \frac{\mathbf{1}_{X_{t+1}\le 0}}{(1-p_{t+1})^{\beta}} \right] \left| \frac{X_{t+1}}{\sigma_{t+1}} \right|^{\beta} \right\},
\end{aligned}$$

where p_{t+1} and σ_{t+1} are given by (3.14) and (3.15). The MLE of $(\beta, \lambda_1, \lambda_2)$ is obtained by maximizing the above function with respect to $(\beta, \lambda_1, \lambda_2)$. Denote the MLE by $(\hat{\beta}, \hat{\lambda}_1, \hat{\lambda}_2)$. Plugging these estimates into (3.10) and (3.12) and using their truncations (in practice, we only have obtains $X_1, \ldots, X_n$), we obtain the following estimates of A_t and B_t:

$$\begin{aligned}
\hat{A}_t &= \sum_{i=0}^{t-1} (1-\hat{\lambda}_1)\hat{\lambda}_1^i |X_{t-i}|^{\beta} \mathbf{1}_{X_{t-i}>0}, \\
\hat{B}_t &= \sum_{i=0}^{t-1} (1-\hat{\lambda}_2)\hat{\lambda}_2^i |X_{t-i}|^{\beta} \mathbf{1}_{X_{t-i}\le 0}.
\end{aligned}$$

Plugging $\hat{A}_t$ and $\hat{B}_t$ into (3.14) and (3.15) to obtain

$$\begin{aligned}\hat{p}_{t+1} &= \frac{\hat{A}_t^{1/(\hat{\beta}+1)}}{\hat{A}_t^{1/(\hat{\beta}+1)} + \hat{B}_t^{1/(\hat{\beta}+1)}}, \\ \hat{\sigma}_{t+1}^{\hat{\beta}} &= \frac{\hat{\beta}\hat{A}_t}{\hat{p}_{t+1}^{\hat{\beta}}} + \frac{\hat{\beta}\hat{B}_t}{(1-\hat{p}_{t+1})^{\hat{\beta}}}.\end{aligned}$$

Finally, from (3.3), the forecasted VaR for X_{T+1} is

$$\widehat{\mathrm{VaR}}_{T+1} = \begin{cases} -\hat{\sigma}_{T+1}\hat{p}_{T+1}[\underline{\Gamma}^{-1}(\hat{z}_1, 1/\hat{\beta})]^{1/\hat{\beta}}, & \alpha > 1-\hat{p}_{T+1}; \\ \hat{\sigma}_{T+1}(1-\hat{p}_{T+1})[\underline{\Gamma}^{-1}(\hat{z}_2, 1/\hat{\beta})]^{1/\hat{\beta}}, & \alpha \le 1-\hat{p}_{T+1}. \end{cases}$$

Here, $\underline{\Gamma}^{-1}(z,t)$ is defined as in Proposition 3,

$$\begin{aligned}\hat{z}_1 &= \left(1 - \frac{1-\alpha}{\hat{p}_{T+1}}\right)\Gamma(1/\hat{\beta}), \\ \hat{z}_2 &= \left(1 - \frac{\alpha}{1-\hat{p}_{T+1}}\right)\Gamma(1/\hat{\beta}).\end{aligned}$$

3.5 Proofs

Proof of Proposition 1. (i) follows from (2.2). (ii) follows from the definition (2.1). To prove (iii), consider $\mathbb{E}(|X|^\nu \mathbf{1}_{X>0})$ since the other case follows similarly. Note that

$$\begin{aligned}\mathbb{E}(|X|^\nu \mathbf{1}_{X>0}) &= \frac{1}{\sigma\Gamma(1+1/\beta)}\int_0^\infty x^\nu \exp\left\{-\left(\frac{x}{p\sigma}\right)^\beta\right\}dx \\ \left(\text{let } \left(\frac{x}{p\sigma}\right)^\beta = y\right) \quad &= \frac{(p\sigma)^{\nu+1}}{\sigma\beta\Gamma(1+1/\beta)}\int_0^\infty y^{(\nu+1)/\beta-1}\exp(-y)dy \\ &= \frac{(p\sigma)^{\nu+1}}{\sigma\beta\Gamma(1+1/\beta)}\Gamma((\nu+1)/\beta).\end{aligned}$$

The proof is completed in view of $\Gamma(t+1) = t\Gamma(t)$. $\diamondsuit$

Proof of Proposition 2. For integer $k \geq 1$, $\mathbb{E}(X^k) = \mathbb{E}(|X|^k \mathbf{1}_{X>0}) + (-1)^k \mathbb{E}(|X|^k \mathbf{1}_{X \leq 0}$. Thus, by Proposition 1(iii), it can be easily seen that $\mathbb{E}(X^k) = \mu_k$ with μ_k given in 2.5). $\diamondsuit$

Proof of Proposition 3. For cdf, we consider $x > 0$ only since the other case follows similarly. For $x > 0$, $F(x|\beta, \sigma, p) = \mathbb{P}(X \leq 0) + \mathbb{P}(0 < X \leq x) = 1 - p + \int_0^x f(x|\beta, \sigma, p)dx$. Using the transformation $y = [x/(p\sigma)]^\beta$ and by the definition of the lower incomplete gamma function, we can immediately obtain the result. The quantile function immediately follows from inverting the cdf. $\diamondsuit$

Proof of Theorem 1. For the log likelihood in (3.4), setting the partial derivatives, with respect to σ and p, equal to zero, we have

$$\begin{aligned} \frac{\partial \ell(\beta, \sigma, p)}{\partial \sigma} = 0 \quad &\Longrightarrow \quad -\frac{n}{\sigma} + \frac{nA\beta p}{(p\sigma)^{\beta+1}} + \frac{nB\beta(1-p)}{[(1-p)\sigma]^{\beta+1}} = 0, \\ \frac{\partial \ell(\beta, \sigma, p)}{\partial p} = 0 \quad &\Longrightarrow \quad \frac{nA\beta\sigma}{(p\sigma)^{\beta+1}} - \frac{nB\beta\sigma}{[(1-p)\sigma]^{\beta+1}} = 0. \end{aligned}$$

The first equation immediately yields (3.8). From the second equation, we can get $B/A = (1/p - 1)^{\beta+1}$, which yields (3.7). $\diamondsuit$

Chapter 4

Forecasting and evaluation

To examine the performance of the proposed generalized-EWMA-based VaR forecasting approach, we compared it to some of the other existing parametric competitors in a Monte Carlo simulation study as well as an empirical application to the daily loss of the S&P 500 index. We compared five parametric VaR forecasting methods: 1) the standard-EWMA approach in the RiskMetrics model of J. P. Morgan (1996), 2) the robust-EWMA approach (Guermat & Harris, 2001), 3) the skewed-EWMA approach (i.e., the proposed generalized-EWMA approach with $\hat{\beta} = 1$) with decay factors estimated (Gerlach et al., 2013), 4) the generalized-EWMA approach with $\hat{\beta} = 2$ and decay factors estimated, and 5) the generalized-EWMA approach with β and decay factors all estimated. The decay factor was set to 0.94 in both the standard-EWMA and robust-EWMA approaches, as the value is recommended in the technical document of RiskMetrics (J. P. Morgan, 1996) and is within the range of $[0.92, 0.95]$ suggested by Guermat & Harris (2001). In the latter three approaches, the freely estimated parameters, including the decay factors and β (if applicable), were estimated as those that minimize the negative loglikelihood function, found by the **nlm** function in R (R Core Team, 2015). Following the procedure in Gerlach et al. (2013), we eliminated one parameter by constraining

λ_u and λ in our analysis.

4.1 Monte Carlo simulation study

We simulated returns from three time-series models with three noise distributions. Below are the three models that we considered.

$$\begin{aligned}
&\text{Model 1:} \quad Y_i = \theta_0 + \theta_1 Y_{i-1} + \sigma\varepsilon_i, \quad (\theta_0, \theta_1, \sigma) = (0.3, 0.4, 0.5);\\
&\text{Model 2:} \quad Y_i = \theta_0 + \theta_1 Y_{i-1} + \theta_2 Y_{i-2} + \sigma_i\varepsilon_i, \quad (\theta_0, \theta_1, \theta_2, \sigma) = (0.3, 0.4, -0.5, 0.5);\\
&\text{Model 3:} \quad Y_i = \theta_0 + \theta_1 Y_{i-1} + \varepsilon_i\sqrt{\theta_2^2 + \theta_3^2 Y_{i-1}^2}, \quad (\theta_0, \theta_1, \theta_2, \theta_3) = (0.3, 0.4, 0.3, 0.5);
\end{aligned}$$

Model 1 is an AR(1) model, Model 2 is an AR(2) model, and Model 3 is an AR(1)-ARCH(1) model. The noise ε_i are from three distributions: a) the standard normal distribution, $N(0,1)$, b) Student-t distribution with 3 degrees of freedom and a normalizer of $\sqrt{3}$ making the variance one, $t_3/\sqrt{3}$, or 3) a mixture normal distribution, $.5N(-.6, .75) + .25N(.4, .75)$, which is skewed with mean -.1 and variance 1. In all settings we use sample size $T = 1000$ (not including the initial states in simulation). That is, for each time series, we use the past 999 samples to predict the VaR at the 1000th time point. For each setting, there are $n = 1000$ Monte Carlo realizations.

To numerically examine the different estimators, we consider the empirical violation rates (i.e., the empirical proportion out of n realizations that the observed loss at time T exceeds the forecasted VaR using data prior to T). Since in the Monte Carlo simulation, the realizations are independent and under the null hypothesis that $P = P\{\text{Loss} \geq VaR_\alpha\} = \alpha$ is true, whether the observed loss at time T exceeds the predicted VaR follows a binomial distribution with parameters n and

α. Therefore, we can formally test the null hypothesis that the empirical violation rate, P, is equal to the confidence level, α, by using a test statistic $\frac{\sqrt{n}(P-\alpha)}{\sqrt{P(1-P)}}$, which should follow $N(0,1)$ under the null hypothesis.

Table 4.1 summarizes the empirical violation rates for the five VaR forecasting methods. With more underlined numbers in the lower part of the table, we failed to reject the null hypothesis at significance level 5% more often when we used the generalized-EWMA approaches, regardless of whether we set $\hat{\beta}$ to 1 or 2 or allow $\hat{\beta}$ to be estimated. (the underlined numbers in Table 4.1). That is, generally, for the models we considered, the generalized-EWMA approaches had comparable and better performance than the the standard-EWMA and robust-EWMA methods, which were likely to overestimate the VaR such that the violation proportions were low. The generalized-EWMA approach with $\hat{\beta}$ estimated had top performance across different model settings.

Table 4.1: Empirical violation rates for the five VaR forecasting methods from the Monte Carlo simulation study

True Model	Model 1			Model 2			Model 3		
	Noise								
Method	$N(0,1)$	$\frac{t_3}{\sqrt{3}}$	Mixture	$N(0,1)$	$\frac{t_3}{\sqrt{3}}$	Mixture	$N(0,1)$	$\frac{t_3}{\sqrt{3}}$	Mixture
Standard-EWMA									
$\alpha = .01$	.000	.001	.001	.001	$\underline{.007}$	.000	.000	.001	.000
$\alpha = .05$	.002	.004	.003	.009	.014	.010	.000	.005	.001
$\alpha = .10$	.004	.009	.012	.032	.023	.038	.001	.006	.002
Robust-EWMA									
$\alpha = .01$	.000	.000	.000	.000	.003	.000	.000	.001	.000
$\alpha = .05$	.001	.003	.003	.003	.012	.002	.000	.002	.001
$\alpha = .10$	.004	.008	.012	.030	.024	.038	.001	.005	.002
Skewed-EWMA									
$\alpha = .01$	$\underline{.009}$	$\underline{.016}$	$\underline{.016}$	$\underline{.006}$	$\underline{.014}$	.002	.020	$\underline{.014}$	.022
$\alpha = .05$	$\underline{.054}$	$\underline{.049}$	.071	$\underline{.057}$	$\underline{.064}$	$\underline{.050}$	$\underline{.053}$	.031	$\underline{.061}$
$\alpha = .10$	.081	.075	$\underline{.109}$	$\underline{.118}$	$\underline{.111}$	$\underline{.114}$	$\underline{.083}$	.048	$\underline{.092}$
Generalized-EWMA $\hat{\beta} = 2$									
$\alpha = .01$	$\underline{.012}$	$\underline{.006}$	.020	$\underline{.017}$	$\underline{.015}$	$\underline{.013}$	$\underline{.010}$	.004	.024
$\alpha = .05$	.038	.013	$\underline{.063}$	$\underline{.055}$	.033	$\underline{.056}$	.026	.011	$\underline{.049}$
$\alpha = .10$	.066	.021	$\underline{.087}$	$\underline{.090}$	.054	$\underline{.103}$	.054	.013	.070
Generalized-EWMA $\hat{\beta}$ estimated									
$\alpha = .01$	$\underline{.012}$	$\underline{.017}$	.021	$\underline{.015}$	$\underline{.015}$	$\underline{.012}$	$\underline{.010}$	.020	.022
$\alpha = .05$	.035	$\underline{.045}$	$\underline{.063}$	$\underline{.055}$	$\underline{.050}$	$\underline{.056}$	.033	.031	$\underline{.051}$
$\alpha = .10$	.068	.066	$\underline{.085}$	$\underline{.092}$	$\underline{.105}$	$\underline{.103}$	.066	.042	.082

The underlined numbers mean that the empirical violation rate is not significantly (at significance level 5%) different from the confidence level, α, which are desired.

4.2 Empirical application to the S&P 500 Index

We also considered and examined VaR forecasting for US S&P500 index daily loss. The full dataset runs for 10 years from January, 3, 2005 to December, 31, 2014. The stock indices were obtained from the Yahoo Finance website. Days with no trading in a market were removed. There are $N = 2518$ observations. We obtained the daily loss series by calculating $Y_i = -\ln(\frac{S_i}{S_{i-1}})$, where S_i is the index value at day i. Figure 4.1 shows the over-time plot of the loss series $\{Y_i\}$ over the ten-year time period. As there is clear evidence of volatility clustering in the plot, we naturally decided to use the five parametric forecasting methods that model the conditional variance as time-varying.

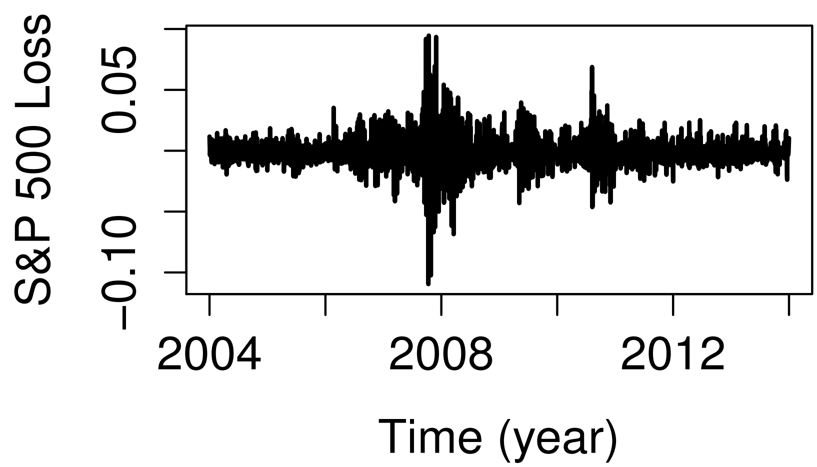

Figure 4.1: Time series plot of daily S&P 500 index loss (i.e., negative logarithm return) during the ten-year period January 2005–December, 2015.

For the last $n = 1000$ daily losses in our data, which roughly corresponds to the daily losses during the last four years 2011–2014, we sequentially predicted the VaR at time t based on the daily losses prior to time t and compared it to the observed loss. Table 4.2 summarizes the empirical violation rates for each method at confidence levels 1%, 5%, and 10%. Overall, the empirical violation rates are close to the confidence level, but most of them are larger than α, which indicated that these parametric methods tend to underestimate the VaR for the S&P 500 dailay loss. The robust-EWMA method had top performance among the five, with the empirical violation rates generally very close to the confidence levels. The generalized-EWMA approaches with $\hat{\beta} = 1$ (the skewed-EWMA approach) and $\hat{\beta}$ estimated shared the second, with almost identical performance. The standard-EWMA and generalized-EWMA with $\hat{\beta} = 2$ did not predict VaR well at confidence levels 1% and 5%, but had best predictions at confidence level 10%. Their performances may be affected by the extreme values in real data, while those of the other three approaches are more robust.

Table 4.2: Empirical violation rates for the five VaR forecasting methods from the empirical application to S&P 500 daily loss

	Confidence level		
Method	$\alpha = .01$	$\alpha = .05$	$\alpha = .10$
Standard-EWMA	.026	.060	.096
Robust-EWMA	.010	.052	.105
Skewed-EWMA	.014	.067	.114
Generalized-EWMA $\hat{\beta} = 2$	.032	.076	.101
Generalized-EWMA $\hat{\beta}$ estimated	.014	.067	.114

Chapter 5

Discussion

Based on the Monte Carlo simulation study and the empirical analysis, we concluded that the proposed generalized-EWMA approach performs well in forecasting VaR, with the empirical violation rate in both studies close to the confidence. Allowing the power parameter β freely estimated in the generalized-EWMA approach can lead to robust VaR estimates across various models, which was a major disadvantage of parametric methods, compared to non-parametric ones. However, the iterative procedure of the generalized-EWMA approach requires more computational resources than its parametric competitors.

Another contribution of this study is that we have presented the definition and properties of an easy-to-interpret Asymmetric Exponential Power distribution, which includes the Laplace, Normal, and Uniform distributions, as well as their asymmetric variations. In our proposed generalized-EWMA approach, the model dynamics of the time-varying shape parameter and positive return probability parameter are motivated from the maximum likelihood estimates of the AEP distribution. The derived evolution dynamics is flexible enough to include several existing popular GARCH models as special cases. Furthermore, the proposed time-varying framework can model the phenomenon of clustering of positive returns in a bull market, clustering of negative returns in financial crisis, and a roughly equal mixture of positive and negative returns in a range market.

One weakness of this study is that we only focused on parametric methods, not taking into consideration the semiparametric and nonparametric approaches. An extension of the current project to compare a broader range of methods in terms of both statistical and computational efficiency could be a future direction.

References

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.

Bollerslev, T., Chou, R. Y., & Kroner, K. F. (1992). Arch modeling in finance: A review of the theory and empirical evidence. *Journal of econometrics*, *52*(1-2), 5–59.

Bollerslev, T., Engle, R. F., & Nelson, D. B. (1994). Arch models. In R. F. Engle & D. McFadden (Eds.), *Handbook of econometrics* (Vol. 4, pp. 2959–3038). Amsterdam: Elsevier.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, *50*, 987–1007.

Engle, R. F. (1983). Estimates of the variance of U.S. inflation based upon the arch model. *Journal of Money, Credit and Banking*, 286–301.

Gerlach, R., Lu, Z., & Huang, H. (2013). Exponentially smoothing the skewed laplace distribution for value-at-risk forecasting. *Journal of Forecasting*, *32*(6), 534–550.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, *48*(5), 1779–1801.

Guermat, C., & Harris, R. D. (2001). Robust conditional variance estimation and Value-at-Risk. *Journal of Risk*, *4*, 25–41.

Guermat, C., & Harris, R. D. (2002). Forecasting value at risk allowing for time variation in the variance and kurtosis of portfolio returns. *International Journal of Forecasting*, *18*(3), 409–419.

J. P. Morgan. (1996, 12). *RiskMetrics—Technical Document* (4th ed.; Tech. Rep.).

Komunjer, I. (2007). Asymmetric power distribution: Theory and applications to risk measurement. *Journal of Applied Econometrics*, *22*(5), 891–921.

R Core Team. (2015). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/`

Wang, C., & Zhao, Z. (2016). Conditional value-at-risk: Semiparametric estimation and inference. *Manuscript*.