Statistica Sinica **13**(2003), 1015-1044

ON CONDITIONAL MOMENTS OF GARCH MODELS, WITH APPLICATIONS TO MULTIPLE PERIOD VALUE AT RISK ESTIMATION

Chi-Ming Wong and Mike K. P. So

The Hong Kong University of Science & Technology

Abstract: In this article, the exact conditional second, third and fourth moments of returns and their temporal aggregates are derived under Quadratic GARCH models. Three multiple period Value at Risk estimation methods are proposed. Two methods are based on the exact second to fourth moments and the other adopts a Monte Carlo approach. Some simulations show that the multiple period Value at Risk calculated from an asymmetric t-distribution with the variance, skewness parameter and the degrees of freedom chosen to match the second to fourth moments of the aggregate returns is close to the one obtained by Monte Carlo simulations. Using some market indices for illustration, the proposed Value at Risk estimation methods are found to be superior to some standard approaches such as RiskMetrics.

Key words and phrases: Aggregate returns, heteroskedastic models, kurtosis, Monte Carlo methods, skewness, square root of time rule, volatility.

1. Introduction

This article studies the conditional moments of temporal aggregate returns under some GARCH specifications. Let r_t be the return at time t and Ω_t be the information up to time t. The aggregate return $R_{t,h}$ at time t for a horizon h is given by

$$R_{t,h} = r_{t+1} + \cdots + r_{t+h}.$$

Denote the conditional variance of r_t given Ω_{t-1} by σ_t^2. It is well-known that if the variances are constant, $\sigma_t^2 = \sigma^2$, and the returns r_t are uncorrelated, the variance of the aggregate returns $R_{t,h}$ is simply $h\sigma^2$. In other words, under the random walk hypothesis, the standard deviation or volatility of $R_{t,h}$ is obtained by scaling σ with $\sqrt{h}$. This simple scaling method is called the square root of time rule, or $\sqrt{h}$ rule. For example, if σ is the constant standard deviation of daily returns, the annual standard deviation is usually referred to $\sqrt{252}\sigma$, under the assumption that we have approximately 252 trading days per year. Although this square root of time rule is widely accepted by practitioners to do annualization and to measure the risk in different horizons, its restrictions and problems are well known. For example, J. P. Morgan (1996, p.87) stated that '*Typically, the square root of time rule results from the assumption that variances are constant*'. Also, Diebold, Hickman, Inoue and Schuermann (1998) stated that '*The common practice of converting* 1*-day volatility estimates to* h*-day estimates by scaling by* $\sqrt{h}$ *is inappropriate and produces overestimates of the variability of long-horizon volatility*'. In light of the restrictions and problems of the $\sqrt{h}$ rule pointed out in the literature, it is important to further examine the rule in various scenarios.

We focus our study on the GARCH framework. The appropriateness of the rule depends on the conditional second moment properties of the aggregate returns. Recently, many studies have investigated the moment properties of GARCH processes. See, for example, He and Teräsvirta (1999a, 1999b) and Duan, Gauthier and Simonato (1999). While existing results on GARCH moments involve mainly the unconditional moments, Chapters 3 and 7 of Tsay (2002) studied the multiple period volatility forecasts under GARCH models. To examine the $\sqrt{h}$ rule and to study the tail properties of the aggregate returns, we derive the exact conditional variance, skewness and kurtosis of $R_{t,h}$ given Ω_t for some GARCH processes. Through this variance, we provide theoretical justification for the adaptation of the square root of time rule in some cases such as the RiskMetrics model of J. P. Morgan. More importantly, the variance, skewness and kurtosis enable us to construct two new methods for estimating multiple period Value at Risk (VaR).

VaR is a common measure of risk. It is the loss of a portfolio that will be exceeded with a predetermined probability over a time period. In general, if C is the current market value of a portfolio and h is the holding period, the h-period VaR of that portfolio is given by

$$\text{VaR} = -C \times V_h, \tag{1}$$

where V_h is the cutoff value which is exceeded by h-period returns with probability $1 - p$. Therefore, estimating the VaR amounts to computing a percentile of the h-period portfolio return distribution. Several approaches have been developed, including the historical simulation method, variance-covariance method and Monte Carlo simulation method. Danielsson and de Vires (1997) discussed a newly-developed method which is based on extreme value theory. Ho, Burridge, Cadle and Theobald (2000) applied extreme value theory to some Asian market indices. Lucas (2000) considered the misspecification of tail properties in the return distribution and its effect on VaR estimation. For comprehensive reviews of the VaR, one can refer to Duffie and Pan (1997), Jorion (1997), Dowd (1998) and Tsay (2002).

Most of the existing researches focus on the one-period VaR estimation, that is, the time horizon is one unit. For the calculation of the VaR in long horizons, we need to know the distribution of $R_{t,h}$ given Ω_t, which is generally not feasible. A traditional method which applies the $\sqrt{h}$ rule treats $R_{t,h}$ as a normal variable with mean zero and variance calculated by scaling σ^2_{t+1} with h. RiskMetrics adopted this $\sqrt{h}$ method in the multiple period VaR estimation. Beltratti and Morana (1999) applied this method with GARCH models to daily and half-hourly data. Rather than following the square root of time rule, we make use of the tail behavior of the aggregate return distribution. We developed two new VaR estimation methods based on the exact conditional variance, skewness and kurtosis and a Monte Carlo method. Simulation and empirical results demonstrate that our proposed methods outperform the $\sqrt{h}$ method in many cases.

The rest of the article is organized as follows. Section 2 gives the derivation of the exact conditional variance. Section 3 derives the exact conditional third and fourth moments of the aggregate returns in some GARCH processes. Section 4 discusses problems of the multiple period VaR estimation. Three new methods for estimating long horizon VaR are also introduced in this section. One approach uses the exact conditional variance derived in Section 2 while regarding $R_{t,h}$ as normal variables. Another approach uses the exact conditional variance but assumes $R_{t,h}$ has a skewed t-distribution with the skewness and kurtosis matching that of $R_{t,h}$. The last approach uses some Monte Carlo simulation methods. Section 5 studies the distribution of the aggregate returns $R_{t,h}$ in various scenarios. Section 6 presents results for comparing our proposed multiple period VaR estimation methods with the commonly used $\sqrt{h}$ method. Section 7 contains empirical applications using daily returns of seven market indices.

2. Exact Conditional Variance of Aggregates

In the general heteroskedastic models considered in Engle (1982) and Bollerslev (1986), the conditional variance of r_t is independent of the sign of r_t. However, as it is commonly observed in the literature that the variance of returns responds asymmetrically to the rise and drop in the stock markets, we adopt the Quadratic GARCH model in this paper (Engle (1990), Sentana (1991) and Campbell and Hentschel (1992)). Specifically, the return generating process follows the QGARCH(p,q) model is

$$r_t = \mu + \bar{r}_t, \qquad \bar{r}_t = \sigma_t \epsilon_t, \qquad \epsilon_t \sim D(0,1), \tag{2}$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i (\bar{r}_{t-i} - b_i)^2 + \sum_{j=1}^{p} \beta_j \sigma^2_{t-j}, \tag{3}$$

where $D(0,1)$ denotes a distribution with mean 0 and variance 1. As usual, the random errors ϵ_t are uncorrelated. In the model μ is the unconditional mean, $\bar{r}_t = r_t - \mu$ is the 'centered return' and the b_i's are the asymmetric variance parameters whose values equal to zero gives the traditional GARCH(p, q) model. An interesting particular case is the IGARCH(1,1) model as adopted in RiskMetrics, where $\mu = \alpha_0 = b_1 = 0$, $\alpha_1 = 1 - \lambda$, $\beta_1 = \lambda$ and $D(0,1)$ is the standard normal distribution. Writing (3) as $\sigma_t^2 = \alpha_0' + \sum_{i=1}^q \alpha_i \bar{r}_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2 - 2\sum_{s=1}^q \alpha_s b_s \bar{r}_{t-s}$, $\alpha_0' = \alpha_0 + \sum_{s=1}^q \alpha_s b_s^2$, we have the following results.

Proposition 1. $\mathrm{Var}\,(R_{t,h} \mid \Omega_t) = \sum_{k=1}^h E[\bar{r}_{t+k}^2 \mid \Omega_t]$.

Proposition 2. *Let $\gamma_{t,s}$ be the conditional expectation $E[\bar{r}_t^2 \mid \Omega_s]$. Define $m = \max\{p, q\}$ and $\phi_i = \alpha_i + \beta_i$, $i = 1, \ldots, m$ where $\alpha_i = 0$ for $i > q$ and $\beta_i = 0$ for $i > p$. Then, for $k \geq m + 1$, $\gamma_{t+k,t} = \alpha_0' + \sum_{i=1}^m \phi_i \gamma_{t+k-i,t}$.*

Proofs are given in Appendix A.1 and A.2. Similar forecasting results under GARCH are also discussed in Sections 3.4 and 7.3 of Tsay (2002). Using Propositions 1 and 2, we can get the aggregate conditional variance $\mathrm{Var}\,(R_{t,h} \mid \Omega_t)$ recursively. In particular, if $p = q = 1$,

$$\mathrm{Var}\,(R_{t,h} \mid \Omega_t) = \begin{cases} \dfrac{\alpha_0'}{1-\phi_1}[h - \dfrac{1-\phi_1^h}{1-\phi_1}] + \dfrac{1-\phi_1^h}{1-\phi_1}\sigma_{t+1}^2 & \text{if } \phi_1 < 1 \\ \dfrac{(h-1)h}{2}\alpha_0' + h\sigma_{t+1}^2 & \text{if } \phi_1 = 1 \end{cases}. \tag{4}$$

For the RiskMetrics model that has $\phi_1 = 1$ and $\alpha_0' = 0$, the volatility of $R_{t,h}$ is given by the volatility at time $t + 1$, that is σ_{t+1} multiplied by $\sqrt{h}$. Therefore, the square root of time rule adopted by many practitioners is obeyed in the RiskMetrics set-up. This is also mentioned in Section 7.2 of Tsay (2002). In the stationary case $\phi_1 < 1$, $\phi_1^h \to 0$ as $h \to \infty$ and so

$$\mathrm{Var}\,(R_{t,h} \mid \Omega_t) \approx \frac{\alpha_0'}{1-\phi_1}h - \frac{\alpha_0'}{(1-\phi_1)^2} + \frac{\sigma_{t+1}^2}{1-\phi_1}, \tag{5}$$

when h is large. The long horizon forecast variance is roughly h times $\alpha_0'/(1-\phi_1)$, *independent of* t, contradicting the $\sqrt{h}$ rule that $\mathrm{Var}\,(R_{t,h} \mid \Omega_t)$ is h multiples of σ_{t+1}^2. Needless to say, in the unconditional context, the $\sqrt{h}$ rule holds in the stationary case because $\mathrm{Var}\,(R_{t,h})$ is $h\,\mathrm{Var}\,(r_{t+1})$.

The above conclusions are in accord with the argument in Diebold, Hickman, Inoue and Schuermann (1998) that in stationary GARCH(1,1) models, applying the $\sqrt{h}$ rule produces wrong fluctuation in the long horizon volatility forecasts. Using the results in Drost and Nijman (1993), Diebold, Hickman, Inoue and

Schuermann (1998) demonstrated that aggregation diminishes the volatility fluctuation as h increases. For instance, if $\{r_t\}$ is GARCH(1,1) with $\mu = b_1 = 0$, the aggregates $R_{th,h} = r_{th+1} + \cdots + r_{th+h}$ will follow an implied GARCH(1,1) process with the conditional variance $\sigma_t^{(h)2} = \mathrm{Var}\,(R_{th,h} \mid \Omega_t^{(h)})$ given by

$$\sigma_t^{(h)2} = \alpha_0^{(h)} + \alpha_1^{(h)} R^2_{(t-1)h,h} + \beta_1^{(h)} \sigma_{t-1}^{(h)2}, \tag{6}$$

where $\Omega_t^{(h)}$ is the set of aggregate returns $R_{0,h}, \ldots, R_{(t-1)h,h}$. As h tends to ∞, $\alpha_0^{(h)}$ tends to $h\alpha_0/(1-\phi_1)$ and both $\alpha_1^{(h)}$ and $\beta_1^{(h)}$ tend to zero (see Drost and Nijman (1993)). Therefore, the volatility fluctuation disappears and the conditional variance $\sigma_t^{(h)2}$ converges to $h\alpha_0/(1-\phi_1)$. This finding is consistent with (5) in that $\mathrm{Var}\,(R_{th,h} \mid \Omega_{th})$ converges to h times the unconditional variance of r_t as h tends to ∞. Although we have the coherent limit result for the variance of $R_{th,h}$ from the implied h-period volatility model in (6) and the associated 1-period model, the variance forecast of $R_{th,h}$ given in (4), that is $\mathrm{Var}\,(R_{th,h} \mid \Omega_{th})$, is different from $\mathrm{Var}\,(R_{th,h} \mid \Omega_t^{(h)})$. The former incorporates information of all 1-period returns up to time th whereas the latter uses the h-period returns $R_{0,h}, \ldots, R_{(t-1)h,h}$.

The advice put forward in Diebold, Hickman, Inoue and Schuermann (1998) is that a h-period volatility model should be used if we are interested in the h-period volatilities. For example, if we have 2500 daily observations and we want monthly volatility forecasts or a holding period of $h = 20$ days, we can only use 125 monthly observations instead of the 2500 daily observations to construct a monthly return model. As far as the parameter accuracy is concerned, this reduction in the number of observations is certainly not desirable. As Ω_{th} contains more information than $\Omega_t^{(h)}$, if $\mathrm{Var}\,(R_{th,h} \mid \Omega_{th})$ can be worked out numerically or analytically, which is feasible for the QGARCH processes in (2) and (3), it is more natural and appropriate to use $\mathrm{Var}\,(R_{th,h} \mid \Omega_{th})$ rather than $\mathrm{Var}\,(R_{th,h} \mid \Omega_t^{(h)})$ to forecast the variance of $R_{th,h}$. Hence, we suggest fitting models of 1-period returns rather than models of aggregate returns for multiple period volatility forecasting.

3. Exact Conditional Third and Fourth Moments of Aggregates

Common conditional heteroskedastic models, such as GARCH models, are defined by the predictive distribution of r_{t+1} conditional on Ω_t. Although the conditional distribution $f(r_{t+1} \mid \Omega_t)$ is fixed in the model formulation, $f(r_{t+h} \mid \Omega_t)$ or even $f(R_{t,h} \mid \Omega_t)$ are usually very complicated and unknown if $h > 1$. As the construction of the h-period VaR is based on the percentiles of $f(R_{t,h} \mid \Omega_t)$, some properties of $f(R_{t,h} \mid \Omega_t)$ are likely to be helpful in improving the VaR estimation. In this section, we focus on the conditional third and fourth moments of $R_{t,h}$ under the QGARCH(p,q) model in (2) and (3) with symmetric ϵ_t. Take the kurtosis of ϵ_t to be $K = E\left[\epsilon_t^4\right] > E\left[\epsilon_t^2\right]^2 = 1$. Define the aggregate centered return as $\bar{R}_{t,h} = \bar{r}_{t+1} + \cdots + \bar{r}_{t+h}$, which relates to the aggregate return through $R_{t,h} = h\mu + \bar{R}_{t,h}$. Under the symmetry of ϵ_t, we have

$$E\left[\bar{R}_{t,h}^3 \mid \Omega_t\right] = 3\sum_{i=2}^{h} L_{t,i}, \qquad h \geq 2, \tag{7}$$

and the conditional fourth moment $A_{t,h} = E\left[\bar{R}_{t,h}^4 \mid \Omega_t\right]$ given by

$$A_{t,h} = K\sigma_{t+1}^4 + 6\sum_{j=2}^{h} E_{t,j} + \sum_{j=2}^{h} P_{t+j,t+j}, \qquad h \geq 2, \tag{8}$$

where $L_{t,h} = E\left[\bar{R}_{t,h-1}\bar{r}_{t+h}^2 \mid \Omega_t\right]$, $E_{t,h} = E\left[\bar{R}_{t,h-1}^2\bar{r}_{t+h}^2 \mid \Omega_t\right]$ and $P_{t+l,t+k} = E\left[\bar{r}_{t+l}^2\bar{r}_{t+k}^2 \mid \Omega_t\right]$ can be computed via some recursions. A proof of (7) and (8) and detailed procedures in calculating $A_{t,h}$ are given in Appendix A.3 and A.4. If there is no variance asymmetry, that is, $b_i = 0$, the third moment $E\left[\bar{R}_{t,h}^3 \mid \Omega_t\right]$ will vanish and so the skewness of $R_{t,h}$ is zero. Therefore the conditional skewness of the aggregate returns is induced by the presence of a variance asymmetry effect. Furthermore, $E\left[R_{t,h}^3 \mid \Omega_t\right] = h^3\mu^3 + 3h\mu\, E\left[\bar{R}_{t,h}^2 \mid \Omega_t\right] + E\left[\bar{R}_{t,h}^3 \mid \Omega_t\right]$ and $E\left[R_{t,h}^4 \mid \Omega_t\right] = h^4\mu^4 + 6h^2\mu^2 E\left[\bar{R}_{t,h}^2 \mid \Omega_t\right] + 4h\mu\, E\left[\bar{R}_{t,h}^3 \mid \Omega_t\right] + E\left[\bar{R}_{t,h}^4 \mid \Omega_t\right]$, for $h \geq 1$. The availability of $E\left[R_{t,h}^3 \mid \Omega_t\right]$ and $E\left[R_{t,h}^4 \mid \Omega_t\right]$ helps us understand the tail behavior of $f(R_{t,h} \mid \Omega_t)$, important for working out the percentiles of $R_{t,h}$ accurately given the information up to time t. Given the exact conditional variance $\mathrm{Var}\left(R_{t,h} \mid \Omega_t\right)$ we introduce, in the next section, a new multiple-period VaR estimation method that is likely to outperform other methods that do not make use of the tail properties of $f(R_{t,h} \mid \Omega_t)$.

An important special case of (3) is the RiskMetrics model:

$$r_t = \sigma_t\epsilon_t, \qquad \sigma_t^2 = (1-\lambda)\, r_{t-1}^2 + \lambda\, \sigma_{t-1}^2. \tag{9}$$

In this case, $p = q = 1$, $\mu = \alpha_0 = b_1 = 0$, $\alpha_1 = 1 - \lambda$ and $\beta_1 = \lambda$. Following (8), the conditional kurtosis of the 1-period return r_{t+h} and the aggregate return $R_{t,h}$ given Ω_t, denoted by $K_{r_{t+h}|\Omega_t}$ and $K_{R_{t,h}|\Omega_t}$ respectively, can be written down in closed form as

$$K_{r_{t+h}|\Omega_t} = KG^{h-1}, \tag{10}$$

$$K_{R_{t,h}|\Omega_t} = \frac{K}{h}\left[1 + \left(\frac{G^h - 1}{h(G-1)} - 1\right)\left(\frac{6H}{G-1} + 1\right)\right], \tag{11}$$

where $G = (K-1)(1-\lambda)^2 + 1$ and $H = 1 - \lambda + \frac{\lambda}{K}$. The derivations of (10) and (11) are presented in Appendix A.5. It is interesting to see that both $K_{r_{t+h}|\Omega_t}$ and $K_{R_{t,h}|\Omega_t}$ are independent of t. Since G is greater than one (as $K > 1$), the conditional kurtosis of r_{t+h} increases exponentially with h, while the conditional kurtosis of $R_{t,h}$ tends to infinity as h tends to infinity. This long-horizon behavior of $K_{R_{t,h}|\Omega_t}$ indicates that the distribution of $R_{t,h}$ becomes more heavy tailed as the forecast horizon or the holding period h get longer. Therefore, we cannot be surprised if a small percentile of $R_{t,h}$ is poorly estimated under the normality assumption of $f(R_{t,h} \mid \Omega_t)$, especially when h is large.

4. Multiple Period VaR Estimation

Value at Risk is a measure of the maximum loss of a portfolio over a predetermined horizon. More precisely, it is the loss that will be exceeded with probability p over a time horizon of h periods. According to this definition, the Value at Risk can be formulated as in (1), where C is the current market value of the portfolio and V_h is the h-period return pth percentile. Obviously, an VaR estimate depends very much on the parameters p and h. The choices of p and h can be subjective. For example, Jorion (1997, p.20) stated that p can range from 1% to 5% according to the individual preference of different commercial banks. Moreover, the time horizon or the holding period can vary quite a lot in different applications (see Christoffersen, Diebold and Schuermann (1998) and Jorion (1997)). In 1996, the Bank for International Settlements (BIS) put forward an amendment to the Capital Accord to Incorporate Market Risks. According to the guidelines of the amendment, the VaR associated with $p = 1\%$ and $h = 10$ days should be calculated for the determination of the market risk capital. In practice, the selection of $h \neq 1$ leads to much complication in the estimation of VaR. In the actual calculation of VaR, we usually assume a time series model for 1-period returns, such as the one given in (2) and (3). The time unit for a single period depends on the frequency of the available related financial data. For example, for equity indices data, daily or even hourly returns can be collected and so the time unit can be set at 1 day. To implement the BIS regulation based on a model of daily return data, we set $h = 10$.

Suppose that a model for 1-period returns is formulated as in (2) and (3). Using the notations set out in Section 2, given the information up to time t, we can determine V_h as $V_h = F_{t,h}^{-1}(p)$, where $F_{t,h}(\cdot)$ is the probability distribution of the h-period return $R_{t,h}$ given Ω_t, i.e., $F_{t,h}(x) = \Pr(R_{t,h} \leq x \mid \Omega_t)$. If we want to obtain VaR as in (1), we need the inverse of $F_{t,h}(\cdot)$ evaluated at p. In particular, if $h = 1$, V_h is $\sigma_{t+1} D^{-1}(p)$, where $D^{-1}(\cdot)$ is the inverse of the error distribution

$D(0,1)$. However, $F_{t,h}(\cdot)$ is generally analytically intractable, especially when h is large. Even though for $h > 1$, the conditional distribution of $R_{t,h}$ given Ω_t can be written as

$$F_{t,h}(x) = \int_{R_{t,h} \le x} f(R_{t,h} \mid \Omega_{t+h-1}) \prod_{i=1}^{h-1} f(r_{t+i} \mid \Omega_{t+i-1})\, d(r_{t+1}, \cdots, r_{t+h}),$$

the evaluation of V_h has to involve high-dimension integration. Therefore, the exact value of V_h is usually unavailable when h is greater than one. A commonly used estimator for V_h is $\hat{V}_h^{[1]} = h\mu + \sqrt{h}\sigma_{t+1}\Phi^{-1}(p)$, where $\Phi(\cdot)$ is the standard normal distribution function. RiskMetrics adopts $\hat{V}_h^{[1]}$ with $\mu = 0$ for the h-period VaR estimation. The rationale is based on a normality assumption and the square root of time rule. In the model assumed by RiskMetrics, $D(0,1)$ is the standard normal and so $\hat{V}_1^{[1]} = \sigma_{t+1}\Phi^{-1}(p)$ gives the exact value of V_1. Borrowing the idea of the $\sqrt{h}$ rule, $\hat{V}_h^{[1]}$ is constructed by scaling $\hat{V}_1^{[1]}$ with $\sqrt{h}$. This scaling method has been widely accepted by practitioners, for example, the BIS suggested using the $\sqrt{h}$ rule to convert a 1-day VaR estimate to a 10-day VaR estimate for calculating capital requirement. If the distribution $F_{t,h}(\cdot)$ is normal and $\mathrm{Var}\,(R_{t,h} \mid \Omega_t) = h\sigma_{t+1}^2$, $\hat{V}_h^{[1]}$ is equivalent to V_h. According to (4), $\mathrm{Var}\,(R_{t,h} \mid \Omega_t) = h\sigma_{t+1}^2$ holds in the QGARCH(1,1) model with $\mu = b_1 = \alpha_0 = 0$ and $\alpha_1 + \beta_1 = 1$. Therefore, using $\hat{V}_h^{[1]}$ under the RiskMetrics model setting can provide a good estimate of V_h if $F_{t,h}(\cdot)$ is reasonably close to normal. To investigate whether $\hat{V}_h^{[1]}$ is an appropriate estimator for V_h, we examine the discrepancy between $F_{t,h}(\cdot)$ and a normal distribution having the same variance in the next section.

Since in most cases, such as modeling the return r_t with a stationary QGARCH(1,1) model, neither $\mathrm{Var}\,(R_{t,h} \mid \Omega_t) = h\sigma_{t+1}^2$ nor $F_{t,h}(\cdot)$ is normal, using $\hat{V}_h^{[1]}$ to provide a good estimate of V_h is questionable. In this paper, we propose a natural alternative to $\hat{V}_h^{[1]}$ as $\hat{V}_h^{[2]} = h\mu + \sqrt{\mathrm{Var}\,(R_{t,h} \mid \Omega_t)}\,\Phi^{-1}(p)$. This estimator is constructed by treating $F_{t,h}(\cdot)$ as normal with variance $\mathrm{Var}\,(R_{t,h} \mid \Omega_t)$. An advantage of $\hat{V}_h^{[2]}$ over $\hat{V}_h^{[1]}$ is that using the exact variance of $F_{t,h}(\cdot)$ in $\hat{V}_h^{[2]}$ bypasses the potential bias of $\hat{V}_h^{[1]}$ due to 'mis-scaling'. For example, under a stationary QGARCH(1,1) model, using $\hat{V}_h^{[1]}$ for the long horizon VaR estimation can be problematic because, when h is large, $\mathrm{Var}\,(R_{t,h} \mid \Omega_t) \approx h\alpha_0'/(1-\phi_1)$. This can be very different from $h\sigma_{t+1}^2$. The new estimator $\hat{V}_h^{[2]}$ is expected to be superior to $\hat{V}_h^{[1]}$ in many cases. Obviously, when $\mathrm{Var}\,(R_{t,h} \mid \Omega_t) = h\sigma_{t+1}^2$, $\hat{V}_h^{[2]} = \hat{V}_h^{[1]}$.

Although $\hat{V}_h^{[2]}$ can overcome the mis-scaling problem in using $\hat{V}_h^{[1]}$, the error in the VaR estimation due to the departure of $F_{t,h}(\cdot)$ from normal can be very significant. We propose another estimator for V_h which is more general than $\hat{V}_h^{[2]}$ by incorporating also the skewness and tail properties of $F_{t,h}(\cdot)$. A new estimator is constructed using the skewed t-distribution introduced in Theodossiou (1998). Its probability density function is

$$f(x) = \begin{cases} C\left[1+\frac{2}{\nu-2}\left(\dfrac{x+a}{\theta(1-\tau)}\right)^2\right]^{-\frac{(\nu+1)}{2}} & \text{if } x < -a, \\ C\left[1+\frac{2}{\nu-2}\left(\dfrac{x+a}{\theta(1+\tau)}\right)^2\right]^{-\frac{(\nu+1)}{2}} & \text{if } x \geq -a, \end{cases} \tag{12}$$

where τ and ν are parameters of the distribution,

$$C = \frac{B(\frac{3}{2},\frac{\nu-2}{2})^{\frac{1}{2}}S(\tau)}{B(\frac{1}{2},\frac{\nu}{2})^{\frac{3}{2}}}, \qquad \theta = \frac{\sqrt{2}}{S(\tau)},$$

$$a = \frac{2\tau B(1,\frac{\nu-1}{2})}{S(\tau)B(\frac{1}{2},\frac{\nu}{2})^{\frac{1}{2}}B(\frac{3}{2},\frac{\nu-2}{2})^{\frac{1}{2}}}, \qquad S(\tau) = \left[1+3\tau^2 - \frac{4\tau^2 B(1,\frac{\nu-1}{2})^2}{B(\frac{1}{2},\frac{\nu}{2})B(\frac{3}{2},\frac{\nu-2}{2})}\right]^{\frac{1}{2}}$$

and $B(\cdot)$ is the beta function. The above distribution has mean 0, variance 1,

$$E[x^3] = \frac{4\tau(1+\tau^3)B(2,\frac{\nu-3}{2})B(\frac{1}{2},\frac{\nu}{2})^{\frac{1}{2}}}{B(\frac{3}{2},\frac{\nu-2}{2})^{\frac{3}{2}}S(\tau)^3} - 3a - a^3 \qquad \text{and}$$

$$E[x^4] = \frac{3(\nu-2)(1+10\tau^2+5\tau^4)}{(\nu-4)S(\tau)^4} - 4aE[x^3] - 6a^2 - a^4.$$

If $\tau = 0$, the skewed t is the usual symmetric t-distribution.

By encompassing the third and fourth moments structure of the aggregate returns, the new estimator is

$$\hat{V}_h^{[3]} = h\mu + \sqrt{\mathrm{Var}\left(R_{t,h} \mid \Omega_t\right)}\, f^{-1}(p), \tag{13}$$

where $f^{-1}(p)$ is the pth percentile of (12). We choose τ and ν to match the skewness and kurtosis of the skewed t-distribution and that of the aggregate returns. In other words, the two parameters are found by solving the two equations:

$$\frac{E\left[\bar{R}_{t,h}^3 \mid \Omega_t\right]}{\mathrm{Var}\left(R_{t,h} \mid \Omega_t\right)^{\frac{3}{2}}} = E[x^3], \tag{14}$$

$$\frac{E\left[\bar{R}_{t,h}^4 \mid \Omega_t\right]}{\mathrm{Var}\left(R_{t,h} \mid \Omega_t\right)^2} = E[x^4]. \tag{15}$$

In particular, when the volatility responds symmetrically to good and bad news, that is $b_i = 0$, we have $E\left[\bar{R}^3_{t,h} \mid \Omega_t\right] = 0$ and thus solving (14) and (15) gives $\tau = 0$ and

$$\nu = \frac{6 - 4\,K_{R_{t,h}|\Omega_t}}{3 - K_{R_{t,h}|\Omega_t}} \quad \text{or} \quad 4 + \frac{6}{K_{R_{t,h}|\Omega_t} - 3}, \tag{16}$$

where $K_{R_{t,h}|\Omega_t} = E[x^4]$. Then, the estimator in (13) is simplified to $\hat{V}_h^{[3]} = h\mu + \sqrt{\mathrm{Var}\left(R_{t,h} \mid \Omega_t\right)}\, t_\nu^{-1}(p)$, where $t_\nu(\cdot)$ is the standardized t-distribution with variance 1 and degrees of freedom ν. Under the RiskMetrics model specification in (9), ν in (16) depends on λ, K and h only because, according to (11), $K_{R_{t,h}|\Omega_t}$ is time-independent under the RiskMetrics model. If ϵ_t is standard normally distributed ($K = 3$), we have the following values of $K_{R_{t,h}|\Omega_t}$, ν and $t_\nu^{-1}(p)$ for $p = 1\%$ and 5%, $h = 5$, 10 and 50 and $\lambda = 0.94$ and 0.97.

Table 1.

h	$K_{R_{t,h}\|\Omega_t}$	ν	$t_\nu^{-1}(0.01)$	$d_\nu(0.01)$	$t_\nu^{-1}(0.05)$	$d_\nu(0.05)$
$\lambda = 0.94$						
5	3.31613	22.98	-2.389	-0.063	-1.638	0.007
10	3.39271	19.28	-2.401	-0.075	-1.636	0.009
50	3.77838	11.71	-2.450	-0.124	-1.626	0.019
$\lambda = 0.97$						
5	3.15075	43.80	-2.359	-0.033	-1.642	0.003
10	3.17822	37.67	-2.364	-0.038	-1.641	0.004
50	3.27081	26.16	-2.381	-0.055	-1.639	0.006

Knowing the standard normal percentiles, we also present $d_\nu(p) = t_\nu^{-1}(p) - \Phi^{-1}(p)$ in various scenarios. We observe from the table that $t_\nu^{-1}(0.01)$ decreases with h whereas $t_\nu^{-1}(0.05)$ increases with h. The magnitude of $d_\nu(p)$ grows with h for both $p = 1\%$ and 5%, implying that there is greater discrepancy between the standard normal and the t-distribution used to match $K_{R_{t,h}|\Omega_t}$. Although all $d_\nu(0.05)$ reported are positive, their magnitude is so small that $\hat{V}_h^{[2]}$ and $\hat{V}_h^{[3]}$ are likely to be close in data implementation. Therefore, it is not surprising to see satisfactory performance in using the RiskMetrics method for $p = 5\%$ even though the fat-tailed characteristics of $f(R_{t,h} \mid \Omega_t)$ have not been accounted for. On the other hand, all $d_\nu(0.01)$ are large and negative, implying that $\hat{V}_h^{[2]}$ is substantially greater than $\hat{V}_h^{[3]}$ for $p = 1\%$. Hence, replacing -2.326 with $t_\nu^{-1}(0.01)$, our third estimator offers a simple way to reduce the usual upward bias in the RiskMetrics method for estimating the 1% V_h.

The last estimator we propose is based on some Monte Carlo samples of $R_{t,h}$ from $F_{t,h}(\cdot)$. This method avoids making any assumptions on the distribution of $F_{t,h}(\cdot)$. If the number of Monte Carlo samples obtained is large enough, this method is likely to produce a good estimate of V_h. Because of the decomposition $f(r_{t+1},\cdots,r_{t+h} \mid \Omega_t) = \prod_{i=1}^{h} f(r_{t+i} \mid \Omega_{t+i-1})$, i.i.d. samples from the joint density can be simulated by the method of composition (Tanner (1993, pp.30-33)). Given Ω_t, σ^2_{t+1} is known. For $i = 1,\ldots,N$ where N is the number of replications, we

1. simulate $r^{(i)}_{t+1} \sim \mu + \sigma_{t+1}D(0,1)$ and set $j = 2$,
2. calculate $\sigma^{(i)}_{t+j}$ from (3) using $r^{(i)}_{t+j-1},\ldots,r^{(i)}_{t+1}$ and Ω_t,
3. simulate $r^{(i)}_{t+j} \sim \mu + \sigma^{(i)}_{t+j}D(0,1)$,
4. repeat steps 2 and 3 for $j = 3,\ldots,h$.

Then $(r^{(i)}_{t+1},\ldots,r^{(i)}_{t+h})$ is a draw from the joint density $f(r_{t+1},\cdots,r_{t+h} \mid \Omega_t)$ and $R^{(i)}_{t,h} = r^{(i)}_{t+1} + \cdots + r^{(i)}_{t+h}$, $i = 1,\ldots,N$, forms an independent sample from $f(R_{t,h} \mid \Omega_t)$. Finally, we propose a Monte Carlo estimator for V_h given by $\hat{V}^{[4]}_h$ = sample p percentile of $R_{t,h}$. It was shown in Serfling (1980, pp.74-75) that $\hat{V}^{[4]}_h$, constructed by the i.i.d. sample $R^{(i)}_{t,h}$, $i = 1,\ldots,N$, converges to V_h with probability one. So this fourth estimator converges to V_h as N increases. When N is sufficiently large, the empirical distribution of the Monte Carlo sample can well approximate the target distribution $F_{t,h}(\cdot)$ and the sample percentile $\hat{V}^{[4]}_h$ can give us a good estimate of the desired VaR.

5. Distribution of the Aggregates $R_{t,h}$

In Sections 2 and 3, we have shown how to calculate the exact conditional variance, skewness and kurtosis of the aggregate return $R_{t,h}$ given Ω_t for QGARCH(p,q) models. In this section, we study in detail the fourth moment properties of the aggregates distribution. We also examine by simulations how close the distribution of the aggregates is to the normal and t-distributions used to construct $\hat{V}^{[2]}_h$ and $\hat{V}^{[3]}_h$ respectively, for different horizons h. The following two sub-sections describe the design of the simulation study and report the results.

5.1. Simulation design

We have considered the QGARCH(1,1) model defined in (2) and (3) with $\mu = 0$ and $b_1 = 0$. The focus is on the symmetric GARCH model as it is commonly adopted in financial research. The parameters $\alpha_0 = 1$, $\alpha_1 = 0.1$ and three values of β_1 ($\beta_1 = 0.8, 0.85$ and 0.895) were chosen in the simulations. The parameters μ and α_0 of the GARCH(1,1) model are only location and scaling factors which would not affect the shape of the aggregates distribution. The parameters α_1 and β_1 were set to comport with common results from data analyses that β_1 is large and $\alpha_1 + \beta_1$ is close to one. The choice of $\alpha_1 + \beta_1$ close to one is to capture the stylized fact of high persistent volatility. We focus on the forecast horizons $h = 1, \ldots, 150$ and two distributions of errors ϵ_t in (2), namely the standard normal distribution and the t-distribution with 5 degrees of freedom. For each model considered, a series of sample size $t = 3528$ 1-period returns, together with their conditional variances up to σ_{t+1}^2, was generated. Starting from time $t+1$ with σ_{t+1}^2 being fixed, $N = 200,000$ replications were formed. In the ith replication, a sample path consisting of $r_{t+1}^{(i)}, r_{t+2}^{(i)}, \ldots, r_{t+h}^{(i)}$ was generated, and aggregates $R_{t,1}^{(i)}, R_{t,2}^{(i)}, \ldots, R_{t,h}^{(i)}$ were calculated.

We computed the Kolmogorov-Smirnov (K-S) one-sample goodness-of-fit test statistic $\sup_x |S_N(x) - F(x)|$, where $S_N(x)$ is the empirical distribution of $R_{t,h}^{(1)}$, $\ldots, R_{t,h}^{(N)}$ and $F(x)$ is the null distribution. Here, $F(x)$ stands for either the normal distribution with mean 0 and variance $\mathrm{Var}\,(R_{t,h} \mid \Omega_t)$ for constructing $\hat{V}_h^{[2]}$, or the t-distribution with mean 0, variance $\mathrm{Var}\,(R_{t,h} \mid \Omega_t)$ and kurtosis $K_{R_{t,h}|\Omega_t}$ for defining $\hat{V}_h^{[3]}$. The K-S test statistic was used to measure the maximum distance between the empirical distribution of the aggregate return $R_{t,h}$ and the null distribution $F(x)$.

5.2. Simulation results

Figures 1 and 2 summarize the simulation results of the distributions of the 1-period return r_{t+h} and the aggregate return $R_{t,h}$ for horizons $h = 1, \ldots, 150$. In Figure 1, the excess kurtosis (kurtosis $-$ 3) was plotted against the horizon h. Excess kurtosis measures the tail thickness of the distribution and a positive excess kurtosis indicates a leptokurtic distribution. The horizontal line in each plot locates the zero excess kurtosis which corresponds to normality. We can see from Figure 1 that all excess kurtoses are positive. In parts (a) to (d), the excess kurtosis of the 1-period return r_{t+h} (dotted line) converges to some value as the horizon h increases. The larger the value of β_1, the further that value is above zero and the longer it takes to converge. The excess kurtosis of aggregate return $R_{t,h}$ (solid line) tends to decrease over time horizons where the 1-period return r_{t+h} has similar kurtosis. In parts (e) and (f), corresponding to the near nonstationary case of $\beta_1 = 0.895$, both the excess kurtoses of r_{t+h} and $R_{t,h}$ seem to increase exponentially with h. This particular finding agrees with the characteristics of the RiskMetrics model documented in (10) and (11). The simulation results in Figure 1 indicate that the predictive density $f(R_{t,h} \mid \Omega_t)$ deviates substantially from normality, especially when $\alpha_1 + \beta_1 \approx 1$. Therefore, assuming $f(R_{t,h} \mid \Omega_t)$ to be normal in constructing $V_h^{[1]}$ and $V_h^{[2]}$ is arguable.

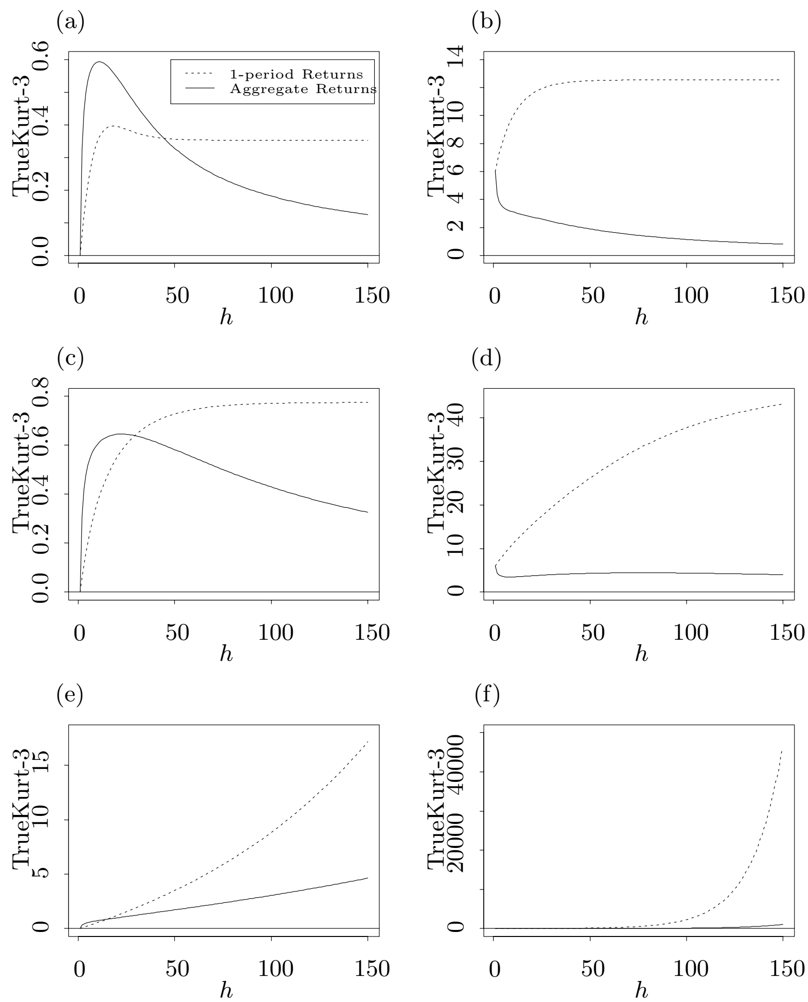

Figure 1. Plots of the true excess kurtosis (kurtosis − 3) as a function of horizon h for both 1-period return r_{t+h} (dotted line) and aggregate return $R_{t,h}$ (solid line) generated from a GARCH(1,1) process. Parts (a) and (b) are for $\beta_1 = 0.80$; parts (c) and (d) are for $\beta_1 = 0.85$, and parts (e) and (f) are for $\beta_1 = 0.895$. Parts (a), (c) and (e) are for normal distributed ϵ_t; parts (b), (d) and (f) are for t-distributed ϵ_t with 5 degrees of freedom.

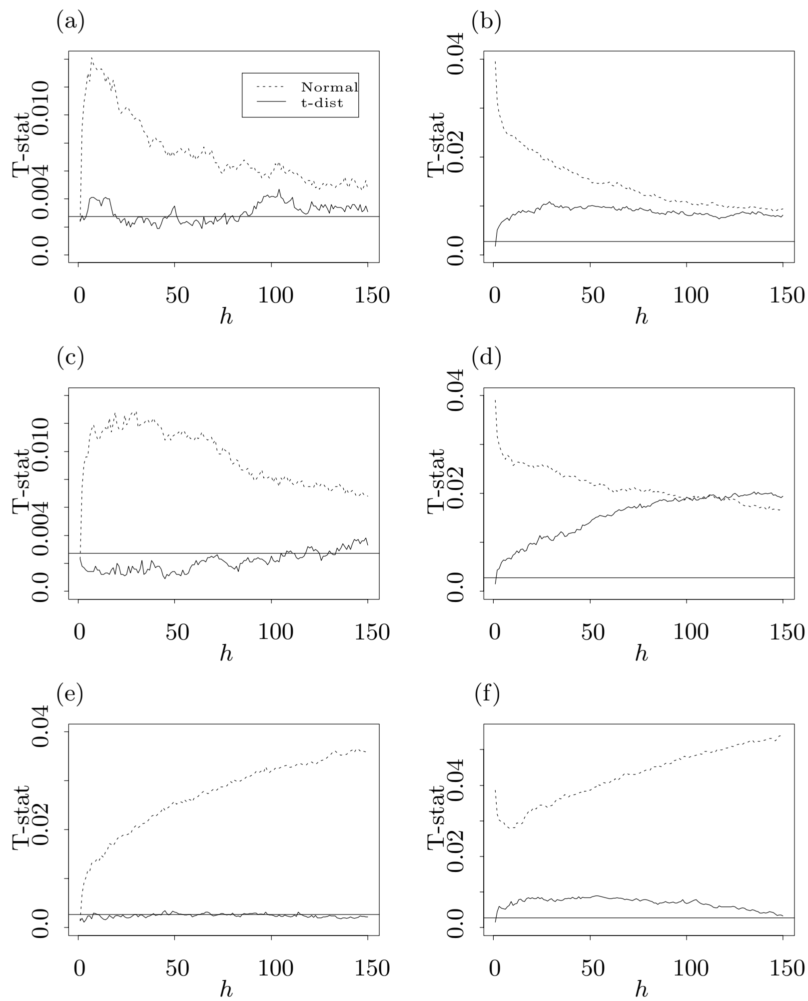

Figure 2. Plots of the K-S test statistic (T-stat) as a function of horizon h for the aggregate return $R_{t,h}$ generated from a GARCH(1,1) process. The horizontal line is the critical value of the K-S test at 1% significance level. The dotted line represents T-stat of the null normal distribution with variance $\mathrm{Var}\,(R_{t,h}|\Omega_t)$ and the solid line represents T-stat of the null t-distribution with variance $\mathrm{Var}\,(R_{t,h}|\Omega_t)$ and kurtosis $K_{R_{t,h}|\Omega_t}$. Parts (a) and (b) are for $\beta_1 = 0.80$; parts (c) and (d) are for $\beta_1 = 0.85$; parts (e) and (f) are for $\beta_1 = 0.895$. Parts (a), (c) and (e) are for normal distributed ϵ_t; parts (b), (d) and (f) are for t-distributed ϵ_t with 5 degrees of freedom.

In Figure 2, we want to see how close the conditional distribution of the aggregate return $f(R_{t,h} \mid \Omega_t)$ is to the normal distribution with the same variance, and to the t-distribution with the same variance and kurtosis. In other words, these normal and t-distributions are the null distributions for computing the K-S test statistic. The horizontal line in each plot marks the 1% critical value of the K-S test for reference. Parts (a), (c) and (e) of Figure 2 correspond to GARCH(1,1) models with standard normally distributed ϵ_t. The K-S test statistic associated with the null t-distribution lies very close to the 1% critical value while the K-S test statistic associated with the null normal distribution is well above the 1% critical value. This indicates that the t-distribution that matches the true conditional variance and kurtosis of $f(R_{t,h} \mid \Omega_t)$ is a good approximation to the desired conditional distribution. For GARCH(1,1) models with ϵ_t distributed as standardized t with 5 degrees of freedom, parts (b), (d) and (f) of Figure 2 show that the null t-distribution is still 'closer' to $f(R_{t,h} \mid \Omega_t)$ than the normal. Since the case with $\beta_1 = 0.895$ resembles the RiskMetrics model, it is anticipated that the pattern of the K-S test statistics for the RiskMetrics model is very similar to Figures 2(e) and (f). Hence, we should not be surprised if $V_h^{[3]}$, based on the null t-distribution, outperforms $V_h^{[1]}$ and $V_h^{[2]}$ in VaR estimation under the GARCH and RiskMetrics framework.

6. Comparing the Four VaR Estimation Methods

Since the Monte Carlo estimator $\hat{V}_h^{[4]}$ approaches V_h as the number of replications N tends to infinity, it can be regarded as the benchmark among the four estimators discussed in Section 4 if N is large enough. In this section, we set $N = 200,000$ and use the same simulation setup in Section 5 to compare the four VaR estimation methods. To facilitate the comparison of $\hat{V}_h^{[1]}$, $\hat{V}_h^{[2]}$ and $\hat{V}_h^{[3]}$ with the chosen benchmark, we compute the percentage difference between each of the first three methods and the Monte Carlo method, $(\hat{V}_h^{[i]}/\hat{V}_h^{[4]} - 1) \times 100\%$ for $i = 1, 2, 3$. We expect that good estimation methods are able to produce VaR estimates that are close to that generated by the Monte Carlo method, so small absolute percentage differences are desirable.

In Figure 3, the percentage differences of the three estimation methods are plotted against the horizon h for the GARCH(1,1) model where ϵ_t is t-distributed with 5 degrees of freedom, $p = 1\%$ and 5%, and $\beta_1 = 0.8, 0.85$ and 0.895. From parts (a) to (f) of Figure 3, $\hat{V}_h^{[1]}$ (dashed line) has the largest magnitude in percentage difference among the three methods. The large deviation of $\hat{V}_h^{[1]}$ from $\hat{V}_h^{[4]}$ is due to the mis-scaling problem of using the $\sqrt{h}$ rule. By incorporating the exact variance, $\hat{V}_h^{[2]}$ (dotted line) shows great improvement over $\hat{V}_h^{[1]}$. However, systematic bias is recorded in $\hat{V}_h^{[2]}$ by having negative and positive percentage differences when $p = 1\%$ and 5% respectively. This is due to the fact that the distribution of $R_{t,h}$ is leptokurtic (see Figure 1) and $R_{t,h}$ is assumed to be normal

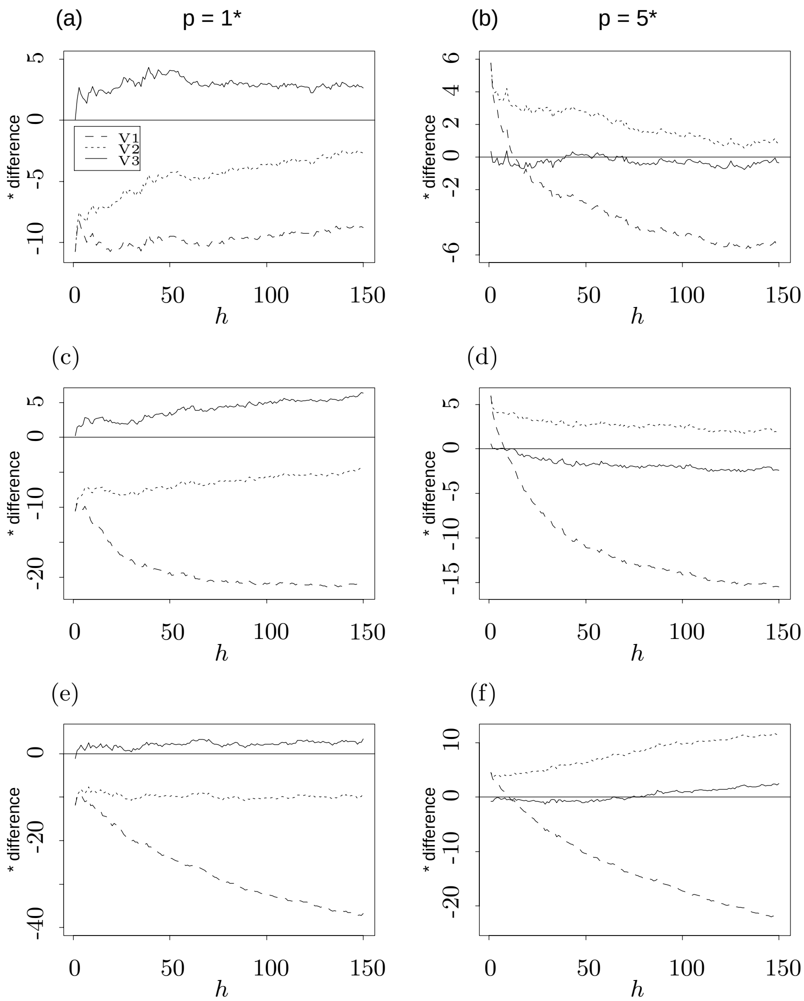

Figure 3. Plots of the percentage difference between $\hat{V}_h^{[1]}$ and $\hat{V}_h^{[4]}$ (dashed line), $\hat{V}_h^{[2]}$ and $\hat{V}_h^{[4]}$ (dotted line), and $\hat{V}_h^{[3]}$ and $\hat{V}_h^{[4]}$ (solid line) as a function of the horizon h for GARCH(1,1) model, ϵ_t is t-distributed with 5 degrees of freedom. Parts (a) and (b) are for $\beta_1 = 0.80$; parts (c) and (d) are for $\beta_1 = 0.85$; parts (e) and (f) are for $\beta_1 = 0.895$. Parts (a), (c) and (e) are for $p = 1\%$; parts (b), (d) and (f) are for $p = 5\%$.

when deriving $\hat{V}_h^{[2]}$. In terms of the magnitude of the percentage differences, $\hat{V}_h^{[3]}$ (solid line) generally performs better than $\hat{V}_h^{[2]}$. In the simulations using normal distributed ϵ_t, we observe similar results as above, that $\hat{V}_h^{[3]}$ is able to produce estimates that are closest to $\hat{V}_h^{[4]}$ among the three estimators in most horizons. For the RiskMetrics model, the estimator $\hat{V}_h^{[2]}$ is identical to $\hat{V}_h^{[1]}$ as

$\mathrm{Var}(R_{t,h} \mid \Omega_t) = h\sigma_{t+1}^2$, so we present only two curves in each plot of Figure 4. Again, we can observe that $\hat{V}_h^{[3]}$ (solid line) is much better than $\hat{V}_h^{[2]}$ (dotted line) if $p = 1\%$ or 5%, in the sense that $\hat{V}_h^{[3]}$ is closer to the benchmark in most cases. We can also see from the percentage difference of $\hat{V}_h^{[2]}$ that $\hat{V}_h^{[4]} < \hat{V}_h^{[2]}$ when $p = 1\%$, and the opposite is true when $p = 5\%$. The large discrepancy between $\hat{V}_h^{[2]}$ and $\hat{V}_h^{[4]}$ for $p = 1\%$ explains the usual upward bias observed when applying the RiskMetrics VaR estimator to real data.

To conclude, the VaR estimation method $\hat{V}_h^{[3]}$, which uses t-distribution to match the conditional variance and kurtosis, is the best among the three estimation methods. It has performance similar to that of the Monte Carlo method $\hat{V}_h^{[4]}$, but can be calculated instantly. In practice, we can use $\hat{V}_h^{[3]}$ as a substitute of $\hat{V}_h^{[4]}$, to avoid long execution time for large N.

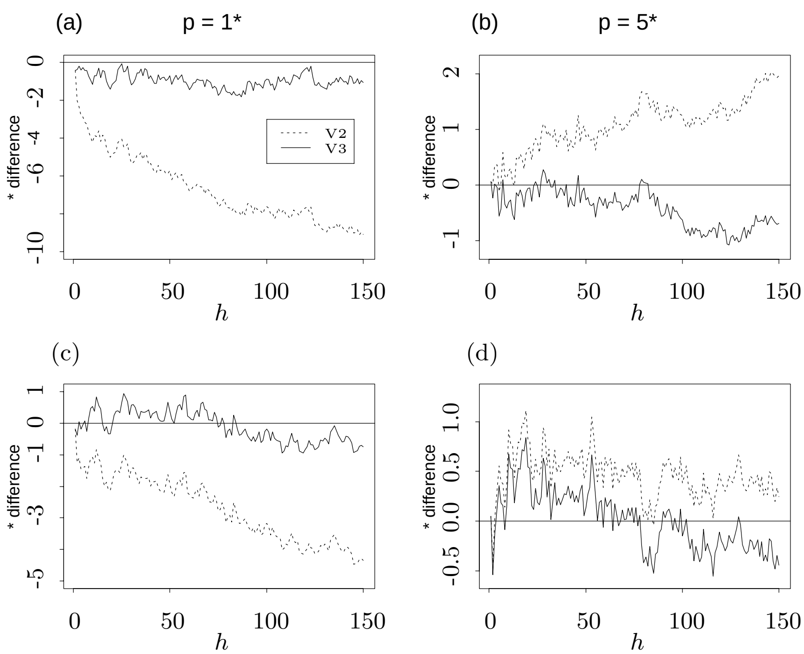

Figure 4. Plots of the percentage difference between $\hat{V}_h^{[2]}$ and $\hat{V}_h^{[4]}$ (dotted line), and $\hat{V}_h^{[3]}$ and $\hat{V}_h^{[4]}$ (solid line) as a function of the horizon h for the RiskMetrics model, ϵ_t is normal. Parts (a) and (b) are for $\lambda = 0.94$; parts (c) and (d) are for $\lambda = 0.97$; parts (a) and (c) are for $p = 1\%$; parts (b) and (d) are for $p = 5\%$.

7. Empirical Applications

In this section, we apply the four VaR estimation methods with two

QGARCH(1,1) models and the RiskMetrics model to the daily returns of seven market indices. The indices we have used are the AOI (Australia) from 1990 to 1998; the CAC 40 (France) from 1991 to 1998; the DAX (Germany) from 1991 to 1998; the FTSE 100 (UK) from 1990 to 1998; the HSI (Hong Kong) from 1990 to 1998; the Nikkei 225 (Japan) from 1990 to 1998; the S & P 500 (USA) from 1990 to 1998. For each market index, we have its daily returns for the period of 1990 to 1998 (1991 to 1998 for the France CAC 40 and Germany DAX). The models we considered here are: (a) *QGARCH*, QGARCH(1,1) model with t-distributed ϵ_t; (b) *GARCH*, QGARCH(1,1) model with $\mu = b_1 = 0$ and t-distributed ϵ_t; (c) *RiskMetrics* model with normally distributed ϵ_t. For the *QGARCH* and *GARCH* models, the parameters were obtained by maximum likelihood estimation using the initial five years daily data r_j where $j = 1, \ldots, t$, and $t \approx 1,250$ (initial four years for CAC 40 and DAX: $t \approx 1,000$). The number of trading days in each year is slightly different from market to market but is roughly equal to 252 days. For the *RiskMetrics* model, the decay factor was set to $\lambda = 0.94$, as suggested by J. P. Morgan (1996) for daily data.

The four types of VaR estimates $\hat{V}_h^{[i]}$, $i = 1, \ldots, 4$, were computed based on the models (a) to (c) for $h = 5, 10$ and 50 and probabilities $p = 1\%$, 2.5% and 5% at the time point t. The actual h-period returns $R_{t,h}$ for h=5, 10 and 50 were also computed from the daily returns of the market indices. Then the estimation window was shifted forward by one day and the *QGARCH* and *GARCH* parameters were re-estimated using the daily returns r_j, $j = 2, \ldots, t+1$. The computation of VaR estimates and actual multiple period returns were performed again at the time point $t+1$. This rolling sample analysis was repeated until the whole validation period (1995-1998) was covered. At the end, the VaR estimates $\hat{V}_h^{[i]}$, $i = 1, \ldots, 4$, together with the actual multiple period returns $R_{t,h}$ for h = 5, 10 and 50 were obtained at the time points $t, \ldots, t+n$ where $n \approx 1,008$ (four year validation period: 1995 to 1998). For each combination of values of p, type of VaR estimates i and horizon h, the proportion of $R_{t,h}$ that falls below its VaR estimates $\hat{V}_h^{[i]}$ denoted by $\hat{p}$ was calculated. If the assumed model for the 1-period returns is correct, we expect that a good VaR estimation method will have $\hat{p}$ close to p or the ratio $\hat{p}/p$ close to 1.

Table 2 lists the ratio $\hat{p}/p$ of the seven market indices for $h = 10$ in the four year validation period (1995 to 1998). For each market index and given p, the ratios closest to 1 were put in boxes. For $p = 2.5\%$ and 5%, the ratios $\hat{p}/p$ do not vary much and are similar. The major factor that determines the difference in the ratios seems to be the underlying dynamic model we assumed for the 1-period returns. For these two moderately small p, it is evident that *GARCH* produces more reliable VaR estimates than *QGARCH* and *RiskMetrics*. The differences among the four VaR estimation methods are small within each model, except for some cases of $QGARCH$. It is also interesting to note the extraordinary large variation in the ratios of the Nikkei 225.

Table 2. Ratio of the proportion $\hat{p}$ of 10-day returns less than the estimated V_h to the actual probability p ($h = 10$).

	QGARCH				*GARCH*				*RiskMetrics*		
	$\hat{V}_h^{[1]}$	$\hat{V}_h^{[2]}$	$\hat{V}_h^{[3]}$	$\hat{V}_h^{[4]}$	$\hat{V}_h^{[1]}$	$\hat{V}_h^{[2]}$	$\hat{V}_h^{[3]}$	$\hat{V}_h^{[4]}$	$\hat{V}_h^{[2]}$	$\hat{V}_h^{[3]}$	$\hat{V}_h^{[4]}$
					$p = 1\%$						
HSI	2.76	2.86	2.02	$\boxed{1.84}$	2.35	2.35	$\boxed{1.84}$	1.94	2.25	1.94	1.94
Nikkei	0.82	0.51	0.10	0.20	$\boxed{1.02}$	0.92	0.82	0.82	1.12	$\boxed{1.02}$	$\boxed{1.02}$
SP500	2.10	2.10	1.50	1.50	1.50	1.50	$\boxed{1.20}$	1.30	1.50	1.50	1.50
AOI	1.89	1.79	0.70	0.70	1.20	$\boxed{1.10}$	$\boxed{0.90}$	$\boxed{0.90}$	1.40	1.20	1.20
FTSE	2.40	2.40	1.60	1.50	1.40	1.40	1.20	1.20	1.20	$\boxed{0.90}$	$\boxed{0.90}$
CAC	1.11	1.11	$\boxed{0.91}$	$\boxed{0.91}$	0.71	0.71	0.61	0.71	1.22	1.12	$\boxed{0.91}$
DAX	2.92	2.62	2.12	1.92	2.12	2.02	1.51	$\boxed{1.41}$	2.62	2.22	2.12
					$p = 2.5\%$						
HSI	1.96	1.88	1.61	1.63	1.51	1.51	1.47	$\boxed{1.43}$	1.51	1.47	1.47
Nikkei	0.57	0.45	0.21	0.29	$\boxed{0.86}$	0.78	0.74	0.78	1.18	$\boxed{1.14}$	$\boxed{1.14}$
SP500	1.36	1.48	1.08	1.24	$\boxed{0.96}$	$\boxed{0.96}$	$\boxed{0.96}$	$\boxed{0.96}$	0.92	0.92	0.88
AOI	1.47	1.43	1.08	1.16	$\boxed{1.04}$	$\boxed{1.04}$	$\boxed{1.04}$	$\boxed{1.04}$	1.20	1.16	1.12
FTSE	1.48	1.40	1.32	1.32	0.96	0.96	0.92	0.96	$\boxed{1.00}$	0.96	0.96
CAC	0.73	0.77	0.69	0.69	0.69	0.69	0.69	0.69	0.89	0.89	$\boxed{0.93}$
DAX	1.81	1.73	1.49	1.53	1.53	$\boxed{1.41}$	$\boxed{1.41}$	$\boxed{1.41}$	1.53	1.49	1.53
					$p = 5\%$						
HSI	1.96	1.78	1.72	1.80	1.41	$\boxed{1.27}$	1.39	1.37	1.39	1.41	1.41
Nikkei	0.73	0.53	0.44	0.55	0.90	0.82	$\boxed{0.92}$	$\boxed{0.92}$	1.20	1.25	1.22
SP500	$\boxed{1.02}$	0.96	0.94	0.96	0.74	0.74	0.76	0.76	0.72	0.74	0.74
AOI	1.31	1.27	1.18	1.24	$\boxed{0.96}$	$\boxed{0.96}$	$\boxed{0.96}$	$\boxed{0.96}$	1.14	1.18	1.16
FTSE	1.06	1.04	0.92	$\boxed{1.00}$	0.72	0.72	0.72	0.70	0.82	0.82	0.82
CAC	0.95	$\boxed{1.01}$	0.95	0.95	0.75	0.75	0.75	0.75	0.95	0.97	0.95
DAX	1.25	1.23	1.23	1.23	0.97	$\boxed{0.99}$	1.03	$\boxed{1.01}$	1.03	1.05	1.03

Figures in the boxes are the ratios $\hat{p}/p$ closest to 1.

For $p = 1\%$, the differences in $\hat{p}/p$ among the four estimation methods can be substantial within each model. For example, the ratios of $QGARCH$ vary from 1.84 to 2.86 for HSI, and from 0.70 to 1.89 for AOI. In the estimation of this extreme percentile, the boxes cluster in $GARCH$ and locate mostly in $\hat{V}_h^{[3]}$ and $\hat{V}_h^{[4]}$. This indicates that they are superior to $\hat{V}_h^{[1]}$ and $\hat{V}_h^{[2]}$. Incorporating also the skewness and kurtosis of $R_{t,h}$ in $\hat{V}_h^{[3]}$ significantly improves the VaR estimation results. While $\hat{V}_h^{[3]}$ and $\hat{V}_h^{[4]}$ work equally well in this competition, $\hat{V}_h^{[3]}$ costs much less in computational time and so is recommended for applications. In Table 3, the holding period is shortened to 5 days ($h = 5$). The estimation method $\hat{V}_h^{[3]}$ associated with *GARCH* is consistently the best for $p = 1\%$ and 2.5% (except for Nikkei and AOI with $p = 1\%$). In Table 4, the holding period is increased to 50 days ($h = 50$). In this case, all VaR estimation methods perform equally poorly when $p = 1\%$. Allowing the mean and asymmetric parameters in *QGARCH* seems to have some advantages in estimating the fifth percentile, but it does not lead to any noticeable improvement for $p = 1\%$ and 2.5%.

Table 3. Ratio of the proportion $\hat{p}$ of 5-day returns less than the estimated V_h to the actual probability p ($h = 5$).

	QGARCH				*GARCH*				*RiskMetrics*		
	$\hat{V}_h^{[1]}$	$\hat{V}_h^{[2]}$	$\hat{V}_h^{[3]}$	$\hat{V}_h^{[4]}$	$\hat{V}_h^{[1]}$	$\hat{V}_h^{[2]}$	$\hat{V}_h^{[3]}$	$\hat{V}_h^{[4]}$	$\hat{V}_h^{[2]}$	$\hat{V}_h^{[3]}$	$\hat{V}_h^{[4]}$
						$p = 1\%$					
HSI	3.35	3.25	2.57	2.64	2.64	2.64	$\boxed{2.03}$	2.24	3.05	2.64	2.54
Nikkei	$\boxed{1.02}$	0.81	0.43	0.51	1.32	1.22	0.81	0.81	1.52	1.32	1.22
SP500	2.98	3.08	1.99	2.09	1.69	1.69	$\boxed{0.99}$	1.29	1.79	1.69	1.69
AOI	2.08	1.88	1.59	1.59	1.59	1.49	1.49	$\boxed{1.39}$	2.28	2.18	2.08
FTSE	2.38	2.38	2.18	2.09	2.09	1.99	$\boxed{1.79}$	1.89	2.28	2.28	2.28
CAC	1.41	1.41	1.11	1.11	1.11	1.11	$\boxed{1.01}$	0.91	1.41	1.31	1.31
DAX	2.41	2.41	2.01	2.11	2.01	2.01	$\boxed{1.71}$	1.81	2.01	1.81	1.81
						$p = 2.5\%$					
HSI	2.07	2.03	1.73	1.59	1.63	$\boxed{1.50}$	$\boxed{1.50}$	$\boxed{1.50}$	1.79	1.79	1.79
Nikkei	0.89	0.89	0.43	0.53	$\boxed{1.02}$	0.98	0.94	0.94	1.30	1.30	1.30
SP500	1.95	1.99	1.51	1.71	1.39	1.39	$\boxed{1.31}$	1.35	$\boxed{1.31}$	$\boxed{1.31}$	$\boxed{1.31}$
AOI	1.43	1.43	1.35	1.39	1.23	1.31	$\boxed{1.19}$	1.23	1.47	1.47	1.47
FTSE	1.47	1.47	1.43	1.47	1.27	$\boxed{1.23}$	$\boxed{1.23}$	$\boxed{1.23}$	1.35	1.31	1.35
CAC	1.41	1.45	1.33	1.25	$\boxed{1.13}$	$\boxed{1.13}$	$\boxed{1.13}$	1.17	1.33	1.17	1.21
DAX	1.44	1.44	1.32	1.32	1.28	$\boxed{1.24}$	$\boxed{1.24}$	$\boxed{1.24}$	1.44	1.40	1.36
						$p = 5\%$					
HSI	1.77	1.71	1.81	1.81	1.38	$\boxed{1.30}$	1.54	1.46	1.54	1.57	1.59
Nikkei	0.83	0.81	0.81	0.81	1.08	$\boxed{1.02}$	1.08	1.10	1.16	1.18	1.16
SP500	1.45	1.45	1.43	1.45	$\boxed{1.01}$	$\boxed{1.01}$	1.03	1.03	$\boxed{1.01}$	$\boxed{1.01}$	$\boxed{0.99}$
AOI	1.15	1.15	1.11	1.11	1.03	$\boxed{1.01}$	1.05	1.07	1.09	1.11	1.11
FTSE	1.25	1.25	1.15	1.17	0.89	0.89	0.89	0.89	$\boxed{1.03}$	1.09	1.05
CAC	1.19	1.17	1.19	1.19	$\boxed{1.09}$	$\boxed{1.09}$	$\boxed{1.09}$	1.11	1.21	1.21	1.21
DAX	1.30	1.28	1.22	1.28	1.06	$\boxed{1.04}$	1.08	1.12	1.12	1.14	1.12

Figures in the boxes are the ratios $\hat{p}/p$ closest to 1.

Table 4. Ratio of the proportion $\hat{p}$ of 50-day returns less than the estimated V_h to the actual probability p ($h = 50$).

	QGARCH				*GARCH*				*RiskMetrics*		
	$\hat{V}_h^{[1]}$	$\hat{V}_h^{[2]}$	$\hat{V}_h^{[3]}$	$\hat{V}_h^{[4]}$	$\hat{V}_h^{[1]}$	$\hat{V}_h^{[2]}$	$\hat{V}_h^{[3]}$	$\hat{V}_h^{[4]}$	$\hat{V}_h^{[2]}$	$\hat{V}_h^{[3]}$	$\hat{V}_h^{[4]}$
$p = 1\%$											
HSI	6.50	6.50	4.31	4.37	5.11	4.69	3.51	3.62	3.09	$\boxed{2.88}$	$\boxed{2.88}$
Nikkei	1.28	0.00	0.00	0.00	2.23	$\boxed{0.85}$	0.21	0.32	3.51	2.34	2.34
SP500	2.70	2.81	$\boxed{1.36}$	1.66	0.31	0.31	0.00	0.00	0.31	0.10	0.10
AOI	2.28	$\boxed{1.04}$	0.10	0.10	0.10	0.10	0.10	0.10	0.00	0.00	0.00
FTSE	2.49	2.49	2.09	2.08	1.77	1.77	$\boxed{1.66}$	$\boxed{1.66}$	$\boxed{1.66}$	$\boxed{1.66}$	$\boxed{1.66}$
CAC	2.43	2.43	2.11	2.11	$\boxed{2.01}$	$\boxed{2.01}$	$\boxed{2.01}$	$\boxed{2.01}$	$\boxed{2.01}$	$\boxed{2.01}$	$\boxed{2.01}$
DAX	3.68	3.89	3.67	3.36	2.84	2.84	2.84	2.84	2.94	$\boxed{2.73}$	$\boxed{2.73}$
$p = 2.5\%$											
HSI	3.28	3.66	3.17	3.28	2.64	2.51	2.47	2.47	2.34	$\boxed{2.22}$	$\boxed{2.22}$
Nikkei	1.74	0.47	0.00	0.13	2.13	1.58	$\boxed{1.40}$	1.45	2.51	2.51	2.47
SP500	1.25	1.25	$\boxed{1.17}$	1.21	$\boxed{0.83}$	$\boxed{0.83}$	$\boxed{0.83}$	$\boxed{0.83}$	$\boxed{0.83}$	0.79	0.79
AOI	1.74	1.41	$\boxed{0.87}$	$\boxed{0.87}$	0.12	0.12	0.12	0.12	0.46	0.33	0.33
FTSE	1.16	1.16	1.13	1.16	$\boxed{0.92}$	$\boxed{0.92}$	$\boxed{0.92}$	$\boxed{0.92}$	$\boxed{0.92}$	$\boxed{0.92}$	$\boxed{0.92}$
CAC	1.14	1.27	1.10	1.18	$\boxed{0.97}$	$\boxed{0.97}$	0.89	0.93	1.06	0.93	0.93
DAX	1.64	1.76	1.79	1.68	1.34	1.43	1.43	1.39	$\boxed{1.30}$	$\boxed{1.30}$	$\boxed{1.30}$
$p = 5\%$											
HSI	2.26	2.32	2.22	2.32	1.68	1.81	1.94	1.94	$\boxed{1.66}$	1.70	1.70
Nikkei	1.38	$\boxed{0.66}$	0.38	$\boxed{0.66}$	2.02	1.70	1.85	1.83	2.21	2.26	2.26
SP500	$\boxed{0.89}$	0.79	0.67	0.67	0.54	0.52	0.54	0.54	0.54	0.54	0.54
AOI	1.29	1.27	$\boxed{1.08}$	1.12	0.44	0.37	0.39	0.37	0.44	0.48	0.46
FTSE	0.85	$\boxed{0.89}$	0.84	0.85	0.56	0.56	0.56	0.56	0.56	0.56	0.56
CAC	$\boxed{0.87}$	0.84	0.76	0.78	0.68	0.61	0.61	0.65	0.70	0.70	0.72
DAX	$\boxed{1.01}$	1.03	1.10	$\boxed{1.01}$	0.82	0.86	0.86	0.86	0.82	0.84	0.84

Figures in the boxes are the ratios $\hat{p}/p$ closest to 1.

The overall picture we get from the tables is as follows. First, the data generating model of the return is important in the estimation of VaR. Broadly speaking, suitable choices are either the RiskMetrics model or the symmetric GARCH model. For the horizons $h = 5$ and 10, the GARCH model is likely to be a promising alternative to the RiskMetrics model. While the QGARCH model is able to capture the volatility asymmetry in financial markets, it seems to be too complicated for predicting the return percentiles and yields poorer performance than the GARCH model. The additional conditional skewness of $R_{t,h}$ induced by the parameters b_i does not evidently help forecast the VaR. Second, when p

is small, the VaR estimation method becomes important and $\hat{V}_h^{[3]}$ is usually the best or at par with other methods. So even if we follow the RiskMetrics model, our proposed third estimator is likely to outperform the classical $\hat{V}_h^{[1]}$ based on the $\sqrt{h}$ rule.

Acknowledgements

The authors would like to thank Professor Ruey S. Tsay and two anonymous referees for their valuable suggestions and comments. Financial support by the Hong Kong RGC Direct Allocation Grants 00/01.BM25 and 00/01.BM28 is gratefully acknowledged.

Appendix

A.1 Proof of Proposition 1

For $i > j > 0$, $E[\,\bar{r}_{t+i}\bar{r}_{t+j}\mid\Omega_t\,] = E[\,E[\bar{r}_{t+i}\bar{r}_{t+j}\mid\Omega_{t+i-1}]\mid\Omega_t\,] = E[\,\bar{r}_{t+j}\,E[\bar{r}_{t+i}\mid\Omega_{t+i-1}]\mid\Omega_t\,] = 0$, as $E[\,\bar{r}_{t+i}\mid\Omega_{t+i-1}\,] = 0$. Obviously, the above result implies that $E[\,\bar{r}_{t+i}\bar{r}_{t+j}\mid\Omega_t\,] = 0$, for $i, j > 0$ and $i \neq j$. The proposition follows as $E[\,\bar{r}_{t+i}\mid\Omega_t\,] = 0$, for $i > 0$.

A.2. Proof of Proposition 2

Recall that we have the general model $\bar{r}_t = \sigma_t\,\epsilon_t$, $\epsilon_t \sim D(0,1)$, $\sigma_t^2 = \alpha_0' + \sum_{i=1}^q \alpha_i\,\bar{r}_{t-i}^2 + \sum_{j=1}^p \beta_j\,\sigma_{t-j}^2 - 2\sum_{s=1}^q \alpha_s b_s \bar{r}_{t-s}$. Using the notation $\gamma_{t,s} = E[\,\bar{r}_t^2\mid\Omega_s\,]$, we have for $k \geq m+1$,

$$\begin{aligned}\gamma_{t+k,t} &= E\left[\,\bar{r}_{t+k}^2\mid\Omega_t\,\right] = E\left[\,E\left[\,\bar{r}_{t+k}^2\mid\Omega_{t+k-1}\,\right]\mid\Omega_t\,\right]\\ &= E\Big[\,\alpha_0' + \sum_{i=1}^q \alpha_i\,\bar{r}_{t+k-i}^2 + \sum_{j=1}^p \beta_j\,\sigma_{t+k-j}^2 - 2\sum_{s=1}^q \alpha_s b_s \bar{r}_{t+k-s}\mid\Omega_t\,\Big]\\ &= \alpha_0' + \sum_{i=1}^q \alpha_i\,\gamma_{t+k-i,t} + \sum_{j=1}^p \beta_j\,\gamma_{t+k-j,t} = \alpha_0' + \sum_{i=1}^m \phi_i\,\gamma_{t+k-i,t}.\end{aligned}$$

The second last equality is valid because $E[\sigma_{t+k-j}^2\mid\Omega_t] = E[E[\bar{r}_{t+k-j}^2\mid\Omega_{t+k-j-1}]\mid\Omega_t] = E[\bar{r}_{t+k-j}^2\mid\Omega_t]$ and $E[\bar{r}_{t+k-j}\mid\Omega_t] = 0$ when $k-j \geq 1$.

A.3. Derivation of the exact conditional third moment of aggregates

Define $T_{t+k,t+h} = E\left[\bar{r}_{t+k}\bar{r}_{t+h}^2\mid\Omega_t\right]$ and $L_{t,h} = E\left[\bar{R}_{t,h-1}\bar{r}_{t+h}^2\mid\Omega_t\right]$. For $h \geq 2$, $E\left[\bar{R}_{t,h}^3\mid\Omega_t\right] = E\left[(\bar{R}_{t,h-1}+\bar{r}_{t+h})^3\mid\Omega_t\right] = E[\bar{R}_{t,h-1}^3 + 3\bar{R}_{t,h-1}^2\bar{r}_{t+h} + 3\bar{R}_{t,h-1}\bar{r}_{t+h}^2 + \bar{r}_{t+h}^3\mid\Omega_t] = E\left[\bar{R}_{t,h-1}^3\mid\Omega_t\right] + 3L_{t,h}$. From this recursion, the conditional third moment of aggregates is $E\left[\bar{R}^3_{t,h} \mid \Omega_t\right] = 3\sum_{i=2}^{h} L_{t,i},\ h \geq 2$, as $E\left[\bar{R}^3_{t,1} \mid \Omega_t\right] = E\left[\bar{r}^3_{t+1} \mid \Omega_t\right] = 0$, where

$$L_{t,h} = E\Big[\bar{R}_{t,h-1}\bar{r}^2_{t+h} \mid \Omega_t\Big] = E\Big[\sum_{i=1}^{h-1}\bar{r}_{t+i}\bar{r}^2_{t+h} \mid \Omega_t\Big] = \sum_{i=1}^{h-1} T_{t+i,t+h}.$$

Therefore, to find the conditional third moments, it suffices to compute $T_{t+k,t+h}$. When $h=k$, $T_{t+h,t+h}=E\left[\bar{r}_{t+h}\bar{r}^2_{t+h} \mid \Omega_t\right]=0$. If $h<k$, $T_{t+k,t+h}=E\left[\bar{r}_{t+k}\bar{r}^2_{t+h} \mid \Omega_t\right] = E\left[\bar{r}^2_{t+h}E[\bar{r}_{t+k} \mid \Omega_{t+k-1}] \mid \Omega_t\right] = 0$. For $h > k$ and $h \geq m+1$,

$$\begin{aligned}
T_{t+k,t+h} &= E\left[\bar{r}_{t+k}\bar{r}^2_{t+h} \mid \Omega_t\right] \\
&= E\left[E[\bar{r}_{t+k}\bar{r}^2_{t+h} \mid \Omega_{t+h-1}] \mid \Omega_t\right] \\
&= E\left[\bar{r}_{t+k}\sigma^2_{t+h} \mid \Omega_t\right] \qquad \text{(as } h > k\text{)} \\
&= E\Big[\bar{r}_{t+k}\Big(\alpha_0' + \sum_{i=1}^{q}\alpha_i\bar{r}^2_{t+h-i} + \sum_{j=1}^{p}\beta_j\sigma^2_{t+h-j} - 2\sum_{s=1}^{q}\alpha_s b_s\bar{r}_{t+h-s}\Big) \mid \Omega_t\Big] \\
&= \alpha_0' E\left[\bar{r}_{t+k} \mid \Omega_t\right] + \sum_{i=1}^{q}\alpha_i E\left[\bar{r}_{t+k}\bar{r}^2_{t+h-i} \mid \Omega_t\right] + \sum_{j=1}^{p}\beta_j E\left[\bar{r}_{t+k}\sigma^2_{t+h-j} \mid \Omega_t\right] \\
&\quad -2\sum_{s=1}^{q}\alpha_s b_s E\left[\bar{r}_{t+k}\bar{r}_{t+h-s} \mid \Omega_t\right] \qquad (17) \\
&= \sum_{i=1}^{q}\alpha_i T_{t+k,t+h-i} + \sum_{j=1}^{p}\beta_j T_{t+k,t+h-j} - 2\alpha_{h-k}b_{h-k}\gamma_{t+k,t} I(1 \leq h-k \leq q).
\end{aligned}$$

Using (17), $T_{t+k,t+h}$ can be computed recursively. In the particular case of no variance asymmetry, i.e., $b_i = 0$, $T_{t+k,t+h} = L_{t,h} = 0$ and so the conditional third moment $E\left[\bar{R}^3_{t,h} \mid \Omega_t\right]$ vanishes.

A.4. Derivation of the Exact Conditional Fourth Moment of Aggregates

Recall that $\gamma_{t+h,t} = E\left[\bar{r}^2_{t+h} \mid \Omega_t\right]$, $K = E\left[\epsilon^4_t\right]$, $m = \max\{p,q\}$, $A_{t,h} = E\left[\bar{R}^4_{t,h} \mid \Omega_t\right]$, $E_{t,h} = E\left[\bar{R}^2_{t,h-1}\bar{r}^2_{t+h} \mid \Omega_t\right]$ and $P_{t+l,t+k} = E\left[\bar{r}^2_{t+l}\bar{r}^2_{t+k} \mid \Omega_t\right]$. In addition, we define $Q_{t+l,t+k} = E\left[\bar{r}^2_{t+l}\sigma^2_{t+k} \mid \Omega_t\right]$. For $h \geq 2$,

$$\begin{aligned}
A_{t,h} &= E\left[\bar{R}^4_{t,h} \mid \Omega_t\right] \\
&= E\left[\left(\bar{R}_{t,h-1} + \bar{r}_{t+h}\right)^4 \mid \Omega_t\right]
\end{aligned}$$

$$
\begin{aligned}
&= E\left[\bar{R}^4_{t,h-1} + 4\bar{R}^3_{t,h-1}\bar{r}_{t+h} + 6\bar{R}^2_{t,h-1}\bar{r}^2_{t+h} + 4\bar{R}_{t,h-1}\bar{r}^3_{t+h} + \bar{r}^4_{t+h} \mid \Omega_t\right] \\
&= A_{t,h-1} + 6E_{t,h} + P_{t+h,t+h}. \qquad (18)
\end{aligned}
$$

The last equality in (18) follows because $E\Big[\bar{R}^3_{t,h-1}\bar{r}_{t+h} \mid \Omega_t\Big] = E\Big[E\Big[\bar{R}^3_{t,h-1}\bar{r}_{t+h} \mid \Omega_{t+h-1}\Big] \mid \Omega_t\Big] = E\left[\bar{R}^3_{t,h-1}E\left[\bar{r}_{t+h} \mid \Omega_{t+h-1}\right] \mid \Omega_t\right] = 0$ as $E\left[\bar{r}_{t+h} \mid \Omega_{t+h-1}\right] = 0$, and $E\left[\bar{R}_{t,h-1}\bar{r}^3_{t+h} \mid \Omega_t\right] = E\left[E\left[\bar{R}_{t,h-1}\bar{r}^3_{t+h} \mid \Omega_{t+h-1}\right] \mid \Omega_t\right] = E\Big[\bar{R}_{t,h-1}E\Big[\bar{r}^3_{t+h} \mid \Omega_{t+h-1}\Big] \mid \Omega_t\Big] = 0$, as ϵ_{t+h} is symmetric about 0. From (18), it is not difficult to see that

$$
A_{t,h} = A_{t,1} + 6\sum_{j=2}^{h} E_{t,j} + \sum_{j=2}^{h} P_{t+j,t+j}, \qquad h \geq 2, \tag{19}
$$

where $A_{t,1} = K\sigma^4_{t+1}$. Therefore, it suffices to calculate $E_{t,j}$ and $P_{t+j,t+j}$, $j = 2, \ldots, h$, for evaluating the conditional fourth moment $E[\bar{R}^4_{t+h}|\Omega_t]$. Since $P_{t+l,t+k} = P_{t+k,t+l}$, we only need to consider the two cases (i) $k = l$, and (ii) $k < l$ for $P_{t+l,t+k}$. Assume that $k \geq m + 1$.

Case 1. $k = l$

$$
\begin{aligned}
P_{t+l,t+k} &= E\left[\bar{r}^2_{t+l}\bar{r}^2_{t+k} \mid \Omega_t\right] = E\left[\bar{r}^4_{t+k} \mid \Omega_t\right] \\
&= E\left[E\left[\bar{r}^4_{t+k} \mid \Omega_{t+k-1}\right] \mid \Omega_t\right] = E\left[K\sigma^4_{t+k} \mid \Omega_t\right] \\
&= KE\Big[\Big(\alpha_0' + \sum_{i=1}^{q}\alpha_i\bar{r}^2_{t+k-i} + \sum_{j=1}^{p}\beta_j\sigma^2_{t+k-j} - 2\sum_{s=1}^{q}\alpha_s b_s\bar{r}_{t+k-s}\Big)^2 \mid \Omega_t\Big] \\
&= KE\Big[\alpha_0'^2 + \Big(\sum_{i=1}^{q}\alpha_i\bar{r}^2_{t+k-i}\Big)^2 + \Big(\sum_{j=1}^{p}\beta_j\sigma^2_{t+k-j}\Big)^2 + 2\alpha_0'\sum_{i=1}^{q}\alpha_i\bar{r}^2_{t+k-i} \\
&\quad +2\alpha_0'\sum_{j=1}^{p}\beta_j\sigma^2_{t+k-j} + 2\Big(\sum_{i=1}^{q}\alpha_i\bar{r}^2_{t+k-i}\Big)\Big(\sum_{j=1}^{p}\beta_j\sigma^2_{t+k-j}\Big) + 4\Big(\sum_{s=1}^{q}\alpha_s b_s\bar{r}_{t+k-s}\Big)^2 \\
&\quad -4\alpha_0'\Big(\sum_{s=1}^{q}\alpha_s b_s\bar{r}_{t+k-s}\Big) - 4\Big(\sum_{s=1}^{q}\alpha_s b_s\bar{r}_{t+k-s}\Big)\Big(\sum_{i=1}^{q}\alpha_i\bar{r}^2_{t+k-i}\Big) \\
&\quad -4\Big(\sum_{s=1}^{q}\alpha_s b_s\bar{r}_{t+k-s}\Big)\Big(\sum_{j=1}^{p}\beta_j\sigma^2_{t+k-j}\Big) \mid \Omega_t\Big] \\
&= KE\Big[\alpha_0'^2 + \sum_{i=1}^{q}\alpha_i^2\bar{r}^4_{t+k-i} + 2\sum\sum_{i>i'}\alpha_i\alpha_{i'}\bar{r}^2_{t+k-i}\bar{r}^2_{t+k-i'} + \sum_{j=1}^{p}\beta_j^2\sigma^4_{t+k-j} \\
&\quad +2\sum\sum_{j>j'}\beta_j\beta_{j'}\sigma^2_{t+k-j}\sigma^2_{t+k-j'} + 2\alpha_0'\sum_{i=1}^{q}\alpha_i\bar{r}^2_{t+k-i} + 2\alpha_0'\sum_{j=1}^{p}\beta_j\sigma^2_{t+k-j}
\end{aligned}
$$

$$
\begin{aligned}
&+2\sum_{i=1}^{q}\sum_{j=1}^{p}\alpha_i\beta_j\bar r^2_{t+k-i}\sigma^2_{t+k-j}+4\sum_{s=1}^{q}\alpha_s^2 b_s^2\bar r^2_{t+k-s}\\
&+8\sum\sum_{s>s'}\alpha_s\alpha_{s'}b_s b_{s'}\bar r_{t+k-s}\bar r_{t+k-s'}-4\alpha_0'\sum_{s=1}^{q}\alpha_s b_s\bar r_{t+k-s}\\
&-4\sum_{s=1}^{q}\sum_{i=1}^{q}\alpha_s b_s\bar r_{t+k-s}\alpha_i\bar r^2_{t+k-i}-4\sum_{s=1}^{q}\sum_{j=1}^{p}\alpha_s b_s\bar r_{t+k-s}\beta_j\sigma^2_{t+k-j}\mid\Omega_t\Big]\\
=&\,K\Big[\alpha_0'^2+\sum_{i=1}^{q}\alpha_i^2 P_{t+k-i,t+k-i}+2\sum\sum_{i>i'}\alpha_i\alpha_{i'}P_{t+k-i,t+k-i'}\\
&+\frac{1}{K}\sum_{j=1}^{p}\beta_j^2 P_{t+k-j,t+k-j}+2\sum\sum_{j>j'}\beta_j\beta_{j'}Q_{t+k-j',t+k-j}\\
&+2\alpha_0'\sum_{i=1}^{q}\alpha_i\gamma_{t+k-i,t}+2\alpha_0'\sum_{j=1}^{p}\beta_j\gamma_{t+k-j,t}+2\sum_{i=1}^{q}\sum_{j=1}^{p}\alpha_i\beta_j Q_{t+k-i,t+k-j}\\
&+4\sum_{s=1}^{q}\alpha_s^2 b_s^2\gamma_{t+k-s,t}-4\sum_{s=1}^{q}\sum_{i=1}^{q}\alpha_s b_s\alpha_i T_{t+k-s,t+k-i}\\
&-4\sum_{j=1}^{p}\sum_{s>j}^{q}\alpha_s b_s\beta_j T_{t+k-s,t+k-j}\Big].
\end{aligned}
\tag{20}
$$

The last equality in (20) is valid because of (a)−(e) below.

(a) For $k\geq m+1$, $E\left[\sigma^4_{t+k-j}\mid\Omega_t\right]=E\left[(1/K)E\left[\bar r^4_{t+k-j}\mid\Omega_{t+k-j-1}\right]\mid\Omega_t\right]=(1/K)E\left[\bar r^4_{t+k-j}\mid\Omega_t\right]=(1/K)P_{t+k-j,t+k-j}$, as $t+k-j-1\geq t+m-j\geq t$.

(b) For $k\geq m+1$, $j>j'$, $E\left[\sigma^2_{t+k-j}\sigma^2_{t+k-j'}\mid\Omega_t\right]=E\Big[\sigma^2_{t+k-j}E\left[\bar r^2_{t+k-j'}\mid\Omega_{t+k-j'-1}\right]\mid\Omega_t\Big]=E\left[E\left[\bar r^2_{t+k-j'}\sigma^2_{t+k-j}\mid\Omega_{t+k-j'-1}\right]\mid\Omega_t\right]=E\left[\bar r^2_{t+k-j'}\sigma^2_{t+k-j}\mid\Omega_t\right]=Q_{t+k-j',t+k-j}$, as $j>j'$ and $t+k-j'-1\geq t+m-j'\geq t$.

(c) $E\left[\bar r^2_{t+k-i}\mid\Omega_t\right]=\gamma_{t+k-i,t}$, as $t+k-i\geq t+m+1-i\geq t+1$.

(d) $E\left[\sigma^2_{t+k-j}\mid\Omega_t\right]=\gamma_{t+k-j,t}$, as $t+k-j\geq t+m+1-j\geq t+1$.

(e) For $s>j$, $E\left[\bar r_{t+k-s}\sigma^2_{t+k-j}\mid\Omega_t\right]=E\left[\bar r_{t+k-s}E\left[\bar r^2_{t+k-j}\mid\Omega_{t+k-j-1}\right]\mid\Omega_t\right]=E\left[E\left[\bar r_{t+k-s}\bar r^2_{t+k-j}\mid\Omega_{t+k-j-1}\right]\mid\Omega_t\right]=T_{t+k-s,t+k-j}$, as $t+k-s<t+k-j$. For $s\leq j$, $E\left[\bar r_{t+k-s}\sigma^2_{t+k-j}\mid\Omega_t\right]=E\left[E\left[\bar r_{t+k-s}\sigma^2_{t+k-j}\mid\Omega_{t+k-s-1}\right]\mid\Omega_t\right]=E\left[\sigma^2_{t+k-j}E\left[\bar r_{t+k-s}\mid\Omega_{t+k-s-1}\right]\mid\Omega_t\right]=0$, as $t+k-j<t+k-s$.

Case 2. $k < l$

$$P_{t+k,t+l} = E\left[\bar{r}^2_{t+k}\bar{r}^2_{t+l} \mid \Omega_t\right] = E\left[E\left[\bar{r}^2_{t+k}\bar{r}^2_{t+l} \mid \Omega_{t+l-1}\right] \mid \Omega_t\right] = E\left[\bar{r}^2_{t+k}\sigma^2_{t+l} \mid \Omega_t\right]$$
$$= E\Big[\bar{r}^2_{t+k}\Big(\alpha_0' + \sum_{i=1}^{q}\alpha_i\bar{r}^2_{t+l-i} + \sum_{j=1}^{p}\beta_j\sigma^2_{t+l-j} - 2\sum_{s=1}^{q}\alpha_s b_s\bar{r}_{t+l-s}\Big) \mid \Omega_t\Big]$$
$$= \alpha_0'\gamma_{t+k,t} + \sum_{i=1}^{q}\alpha_i E\left[\bar{r}^2_{t+k}\bar{r}^2_{t+l-i} \mid \Omega_t\right] + \sum_{j=1}^{p}\beta_j E\left[\bar{r}^2_{t+k}\sigma^2_{t+l-j} \mid \Omega_t\right]$$
$$-2\sum_{s=1}^{q}\alpha_s b_s E\left[\bar{r}^2_{t+k}\bar{r}_{t+l-s} \mid \Omega_t\right]$$
$$= \alpha_0'\gamma_{t+k,t} + \sum_{i=1}^{q}\alpha_i P_{t+k,t+l-i} + \sum_{j=1}^{p}\beta_j Q_{t+k,t+l-j} - 2\sum_{s=1}^{q}\alpha_s b_s T_{t+l-s,t+k}. \quad (21)$$

Since $t + l - 1 \geq t$ and $l > k$. Therefore, recursive formulas for $P_{t+k,t+l}$ are established in (20) and (21). For $k, l \geq m+1$, the following equation is used to evaluate $Q_{t+l,t+k}$:

$$Q_{t+l,t+k} = E\left[\bar{r}^2_{t+l}\sigma^2_{t+k} \mid \Omega_t\right]$$
$$= E\Big[\bar{r}^2_{t+l}\Big(\alpha_0' + \sum_{i=1}^{q}\alpha_i\bar{r}^2_{t+k-i} + \sum_{j=1}^{p}\beta_j\sigma^2_{t+k-j} - 2\sum_{s=1}^{q}\alpha_s b_s\bar{r}_{t+k-s}\Big) \mid \Omega_t\Big]$$
$$= \alpha_0' E\left[\bar{r}^2_{t+l} \mid \Omega_t\right] + \sum_{i=1}^{q}\alpha_i E\left[\bar{r}^2_{t+l}\bar{r}^2_{t+k-i} \mid \Omega_t\right] + \sum_{j=1}^{p}\beta_j E\left[\bar{r}^2_{t+l}\sigma^2_{t+k-j} \mid \Omega_t\right]$$
$$-2\sum_{s=1}^{q}\alpha_s b_s E\left[\bar{r}^2_{t+l}\bar{r}_{t+k-s} \mid \Omega_t\right]$$
$$= \alpha_0'\gamma_{t+l,t} + \sum_{i=1}^{q}\alpha_i P_{t+l,t+k-i} + \sum_{j=1}^{p}\beta_j Q_{t+l,t+k-j} - 2\sum_{s=1}^{q}\alpha_s b_s T_{t+k-s,t+l}. \quad (22)$$

Given initial values $P_{t+l,t+k}$ and $Q_{t+l,t+k}$, $l, k = 1, \ldots, m$, we can obtain $P_{t+j,t+j}$, $j = 2, \ldots, h$ in (19) via the recursions in (20), (21) and (22) by calculating $P_{t+1,t+m+1}, \ldots, P_{t+m+1,t+m+1}, Q_{t+m+1,t+1}, \ldots, Q_{t+m+1,t+m+1}, Q_{t+1,t+m+1}, \ldots, Q_{t+m,t+m+1}, P_{t+1,t+m+2}, \ldots, P_{t+m+2,t+m+2}, Q_{t+m+2,t+1}, \ldots, Q_{t+m+2,t+m+2}, Q_{t+1,t+m+2}, \ldots, Q_{t+m+1,t+m+2}, \ldots$. The above calculation can be further simplified by noting that for $k > l \geq 1$, $Q_{t+l,t+k} = E\left[\bar{r}^2_{t+l}\sigma^2_{t+k} \mid \Omega_t\right] = E\Big[\bar{r}^2_{t+l}\,E\Big[\bar{r}^2_{t+k} \mid \Omega_{t+k-1}\Big] \mid \Omega_t\Big] = E\left[E\left[\bar{r}^2_{t+l}\bar{r}^2_{t+k} \mid \Omega_{t+k-1}\right] \mid \Omega_t\right] = E\left[\bar{r}^2_{t+l}\bar{r}^2_{t+k} \mid \Omega_t\right] = P_{t+l,t+k}$, and for $l \geq 1$,

$$Q_{t+l,t+l} = E\left[\bar{r}^2_{t+l}\sigma^2_{t+l} \mid \Omega_t\right] = E\left[E\left[\bar{r}^2_{t+l}\sigma^2_{t+l} \mid \Omega_{t+l-1}\right] \mid \Omega_t\right]$$
$$= E\left[\sigma^2_{t+l}E\left[\bar{r}^2_{t+l} \mid \Omega_{t+l-1}\right] \mid \Omega_t\right] = E\left[\sigma^4_{t+l} \mid \Omega_t\right] = (1/K)P_{t+l,t+l}. \quad (23)$$

According to (19), it suffices to calculate $E_{t,j}$, $j = 2,\ldots, h$, to find $A_{t,h}$. For $h \geq m+1$,

$$
\begin{aligned}
E_{t,h} &= E\left[\bar{R}^2_{t,h-1}\bar{r}^2_{t+h} \mid \Omega_t\right] = E\left[E\left[\bar{R}^2_{t,h-1}\bar{r}^2_{t+h} \mid \Omega_{t+h-1}\right] \mid \Omega_t\right] \\
&= E\left[\bar{R}^2_{t,h-1}E\left[\bar{r}^2_{t+h} \mid \Omega_{t+h-1}\right] \mid \Omega_t\right] = E\left[\bar{R}^2_{t,h-1}\sigma^2_{t+h} \mid \Omega_t\right] \\
&= E\Big[\bar{R}^2_{t,h-1}\Big(\alpha_0' + \sum_{i=1}^{q}\alpha_i\bar{r}^2_{t+h-i} + \sum_{j=1}^{p}\beta_j\sigma^2_{t+h-j} - 2\sum_{s=1}^{q}\alpha_s b_s\bar{r}_{t+h-s}\Big) \mid \Omega_t\Big] \\
&= \alpha_0' E\left[\bar{R}^2_{t,h-1} \mid \Omega_t\right] + \sum_{i=1}^{q}\alpha_i E\left[\bar{R}^2_{t,h-1}\bar{r}^2_{t+h-i} \mid \Omega_t\right] \\
&\quad + \sum_{j=1}^{p}\beta_j E\left[\bar{R}^2_{t,h-1}\sigma^2_{t+h-j} \mid \Omega_t\right] - 2\sum_{s=1}^{q}\alpha_s b_s E\left[\bar{R}^2_{t,h-1}\bar{r}_{t+h-s} \mid \Omega_t\right]. \qquad (24)
\end{aligned}
$$

Now, for $i = 1, \ldots, q$,

$$
\begin{aligned}
E\left[\bar{R}^2_{t,h-1}\bar{r}^2_{t+h-i} \mid \Omega_t\right] &= E\Big[\Big(\bar{R}_{t,h-i-1} + \sum_{l=h-i}^{h-1}\bar{r}_{t+l}\Big)^2\bar{r}^2_{t+h-i} \mid \Omega_t\Big] \\
&= E\left[\left(\bar{R}^2_{t,h-i-1} + \bar{r}^2_{t+h-i} + \cdots + \bar{r}^2_{t+h-1}\right)\bar{r}^2_{t+h-i} \mid \Omega_t\right] \\
&= E\left[\bar{R}^2_{t,h-i-1}\bar{r}^2_{t+h-i} \mid \Omega_t\right] + \sum_{l=h-i}^{h-1}E\left[\bar{r}^2_{t+l}\bar{r}^2_{t+h-i} \mid \Omega_t\right] \\
&= E_{t,h-i} + \sum_{l=h-i}^{h-1}P_{t+l,t+h-i}. \qquad (25)
\end{aligned}
$$

The second equality in (25) follows because for $l \geq h-i$, $E\Big[\bar{R}_{t,h-i-1}\bar{r}_{t+l}\bar{r}^2_{t+h-i} \mid \Omega_t\Big]$ $= E\Big[E\Big[\bar{R}_{t,h-i-1}\bar{r}^2_{t+h-i}\bar{r}_{t+l} \mid \Omega_{t+l-1}\Big] \mid \Omega_t\Big] = E\Big[\bar{R}_{t,h-i-1}E\Big[\bar{r}^2_{t+h-i}\bar{r}_{t+l} \mid \Omega_{t+l-1}\Big] \mid \Omega_t\Big]$ $= 0$, since $t+l \geq t+h-i \geq t+m-i \geq t$ and $l \geq h-i$, and $E\left[\bar{r}_{t+l}\bar{r}_{t+l'}\bar{r}^2_{t+h-i} \mid \Omega_t\right]$ $= 0$ for l, $l' \geq h-i$ and $l \neq l'$. Similarly, for $j = 1, \ldots, p$,

$$
\begin{aligned}
&E\left[\bar{R}^2_{t,h-1}\sigma^2_{t+h-j} \mid \Omega_t\right] \\
&= E\left[\left(\bar{R}_{t,h-j-1} + \bar{r}_{t+h-j} + \cdots + \bar{r}_{t+h-1}\right)^2\sigma^2_{t+h-j} \mid \Omega_t\right] \\
&= E\left[\left(\bar{R}^2_{t,h-j-1} + \bar{r}^2_{t+h-j} + \cdots + \bar{r}^2_{t+h-1}\right)\sigma^2_{t+h-j} \mid \Omega_t\right] \\
&= E\left[\bar{R}^2_{t,h-j-1}\sigma^2_{t+h-j} \mid \Omega_t\right] + E\left[\sum_{l=h-j}^{h-1}\bar{r}^2_{t+l}\sigma^2_{t+h-j} \mid \Omega_t\right]
\end{aligned}
$$

$$= E\left[\bar{R}^2_{t,h-j-1}E\left[\bar{r}^2_{t+h-j} \mid \Omega_{t+h-j-1}\right] \mid \Omega_t\right] + \sum_{l=h-j}^{h-1} E[\bar{r}^2_{t+l}\sigma^2_{t+h-j} \mid \Omega_t]$$
$$= E\left[\bar{R}^2_{t,h-j-1}\bar{r}^2_{t+h-j} \mid \Omega_t\right] + \sum_{l=h-j}^{h-1} E[\bar{r}^2_{t+l}\sigma^2_{t+h-j} \mid \Omega_t]$$
$$= E_{t,h-j} + \sum_{l=h-j}^{h-1} Q_{t+l,t+h-j}. \tag{26}$$

The second equality in (26) follows because for $l \geq h-j$, $E\Big[\bar{R}_{t,h-j-1}\bar{r}_{t+l}\sigma^2_{t+h-j} \mid \Omega_t\Big] = E\Big[E\Big[\bar{R}_{t,h-j-1}\bar{r}_{t+l}\sigma^2_{t+h-j} \mid \Omega_{t+l-1}\Big] \mid \Omega_t\Big] = E\Big[\bar{R}_{t,h-j-1}\sigma^2_{t+h-j}E\left[\bar{r}_{t+l} \mid \Omega_{t+l-1}\right] \mid \Omega_t\Big] = 0$, since $t+l-1 \geq t+h-j-1 \geq t+m-j \geq t$ and $h-j \leq l$, and $E\left[\bar{r}_{t+l}\bar{r}_{t+l'}\sigma^2_{t+h-j} \mid \Omega_t\right] = 0$ for $l,\, l' \geq h-j$ and $l \neq l'$. Next, for $h \geq m+1 > s$,

$$E\left[\bar{R}^2_{t,h-1}\bar{r}_{t+h-s} \mid \Omega_t\right]$$
$$= E\Big[\Big(\bar{R}_{t,h-s-1} + \sum_{l=h-s}^{h-1}\bar{r}_{t+l}\Big)^2\bar{r}_{t+h-s} \mid \Omega_t\Big]$$
$$= E\Big[\Big(\bar{R}^2_{t,h-s-1} + \sum_{l=h-s}^{h-1}\bar{r}^2_{t+l} + 2\bar{R}_{t,h-s-1}\sum_{l=h-s}^{h-1}\bar{r}_{t+l} + 2\sum_{l=h-s}^{h-1}\sum_{l'>l}\bar{r}_{t+l}\bar{r}_{t+l'}\Big)\bar{r}_{t+h-s} \mid \Omega_t\Big]$$
$$= E\left[\bar{r}^3_{t+h-s} \mid \Omega_t\right] + \sum_{l=h-s+1}^{h-1} E\left[\bar{r}^2_{t+l}\bar{r}_{t+h-s} \mid \Omega_t\right] + 2E\left[\bar{R}_{t,h-s-1}\bar{r}^2_{t+h-s} \mid \Omega_t\right]$$
$$= \sum_{l=h-s+1}^{h-1} T_{t+h-s,t+l} + 2L_{t,h-s}, \quad \text{as } \epsilon_{t+h-s} \text{ is symmetric about zero.} \tag{27}$$

The third equality of (27) follows because $E\left[\bar{R}^2_{t,h-s-1}\bar{r}_{t+h-s} \mid \Omega_t\right] = 0$ as $h-s-1 \geq 0$, $E\left[\bar{R}_{t,h-s-1}\bar{r}_{t+l}\bar{r}_{t+h-s} \mid \Omega_t\right] = 0$ for $l > h-s$, and $E\left[\bar{r}_{t+l}\bar{r}_{t+l'}\bar{r}_{t+h-s} \mid \Omega_t\right] = 0$ for $l,\, l' \geq h-s$, $l \neq l'$. Substituting (25), (26) and (27) into (24), we have

$$E_{t,h} = \alpha_0'\sum_{i=1}^{h-1}\gamma_{t+i,t} + \sum_{i=1}^{q}\alpha_i\Big(E_{t,h-i} + \sum_{l=h-i}^{h-1}P_{t+l,t+h-i}\Big)$$
$$+\sum_{j=1}^{p}\beta_j\Big(E_{t,h-j} + \sum_{l=h-j}^{h-1}Q_{t+l,t+h-j}\Big)$$
$$-2\sum_{s=1}^{q}\alpha_s b_s\Big(\sum_{l=h-s+1}^{h-1}T_{t+h-s,t+l} + 2L_{t,h-s}\Big), \tag{28}$$

$h \geq m+1$, which enables us to compute $E_{t,h}$ recursively.

A.5. The exact conditional kurtosis of aggregates under RiskMetrics

From (20), we can see that under RiskMetrics, $p = q = 1$, $\mu = \alpha_0 = b_1 = 0$ implies $r_t = \bar{r}_t$ and $R_{t,h} = \bar{R}_{t,h}$, $\alpha_1 = 1 - \lambda$ and $\beta_1 = \lambda$, and for $h \geq 2$,

$$\begin{aligned} P_{t+h,t+h} &= K\Big[(1-\lambda)^2 P_{t+h-1,t+h-1} + \frac{\lambda^2}{K} P_{t+h-1,t+h-1} + 2\lambda(1-\lambda) Q_{t+h-1,t+h-1}\Big] \\ &= G P_{t+h-1,t+h-1}, \end{aligned}$$

where $G = (K-1)(1-\lambda)^2 + 1$. The last equality follows because of (23). Knowing that $P_{t+1,t+1} = K\sigma^4_{t+1}$, we get $P_{t+h,t+h} = KG^{h-1}\sigma^4_{t+1}$ and $K_{r_{t+h}|\Omega_t} = KG^{h-1}$ for $h \geq 1$.

In order to obtain $A_{t,h} = E[R^4_{t,h} \mid \Omega_t]$ under RiskMetrics, it suffices to derive $E_{t,j}$. From (28), for $j \geq 2$, we have $E_{t,j} = E_{t,j-1} + (1-\lambda+\lambda/K)P_{t+j-1,t+j-1} = E_{t,j-1} + HKG^{j-2}\sigma^4_{t+1}$, where $H = 1 - \lambda + \lambda/K$. The above implies that

$$E_{t,j} = E_{t,2} + HK\frac{G(G^{j-2}-1)}{G-1}\sigma^4_{t+1} = HK\frac{G^{j-1}-1}{G-1}\sigma^4_{t+1}, \tag{29}$$

where $E_{t,2} = E\Big[r^2_{t+1} r^2_{t+2} \,|\, \Omega_t\Big] = E\Big[r^2_{t+1}\sigma^2_{t+2} \,|\, \Omega_t\Big] = E\Big[(1-\lambda) r^4_{t+1} + \lambda r^2_{t+1}\sigma^2_{t+1} \mid \Omega_t\Big] = [K(1-\lambda)+\lambda]\ \sigma^4_{t+1} = HK\,\sigma^4_{t+1}$. Substituting (29) and $P_{t+h,t+h} = KG^{h-1}\sigma^4_{t+1}$ into (19), we get

$$\begin{aligned} A_{t,h} &= K\Big[1 + \sum_{i=2}^{h}\Big\{(G^{i-1}-1)(\frac{6H}{G-1}+1)+1\Big\}\Big]\sigma^4_{t+1} \\ &= K\Big[h + \Big(\frac{G^h-1}{G-1} - h\Big)\Big(\frac{6H}{G-1}+1\Big)\Big]\sigma^4_{t+1}. \end{aligned}$$

Dividing $A_{t,h}$ by $E\left[R^2_{t,h} \mid \Omega_t\right]^2 = h^2\sigma^4_{t+1}$ gives the result for the exact conditional kurtosis of $R_{t,h}$.

References

Beltratti, A. and Morana, C. (1999). Computing value at risk with high frequency data. *J. Empirical Finance* **6**, 431-455.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *J. Econom.* **31**, 307-327.

Campbell, J.Y. and Hentschel, L. (1992). No news is good news: an asymmetric model of changing volatility in stock returns. *J. Finan. Econom.* **31**, 281-318.

Christoffersen, P. F., Diebold, F. X. and Schuermann, T. (1998). Horizon problems and extreme events in financial risk management. *Federal Reserve Bank, New York, Economic Policy Review*, 109-118.

Danielsson, J. and de Vries, C. G. (1997). *Value-at-Risk and Extreme Returns*, London School of Economics and Institute of Economic Studies, University of Iceland.

Diebold, F. X., Hickman, A., Inoue, A. and Schuermann, T. (1998). Converting 1-day volatility to h-day volatility: scaling by $\sqrt{h}$ is worse than you think. *Risk* **11**, 104-107.

Dowd, K. (1998). *Beyond Value at Risk: The New Science of Risk Management.* Wiley, New York.

Drost, F. C. and Nijman, T. E. (1993). Temporal aggregation of GARCH processes. *Econometrica* **61**, 909-927.

Duan, J. C., Gauthier, G. and Simonato, J. G. (1999). An analytical approximation for the GARCH option pricing model. *J. Comput. Finance* **2**, 75-116.

Duffie, D. and Pan, J. (1997). An overview of value at risk. *J. Derivatives*, 7-49.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987-1008.

Engle, R. F. (1990). Discussion of stock market volatility and the crash of '87. *Rev. Finan. Stud.* **3**, 103-106.

He, C. and Teräsvirta, T. (1999a). Properties of moments of a family of GARCH processes. *J. Econom.* **92**, 173-192.

He, C. and Teräsvirta, T. (1999b). Fourth moment structure of the GARCH(p,q) process. *Econom. Theory* **15**, 824-846.

Ho, L. C., Burridge, P., Cadle, J. and Theobald, M. (2000). Value-at-risk: applying the extreme value approach to Asian markets in the recent financial turmoil. *Pacific-Basin Finance J.* **8**, 249-275.

Jorion, P. (1997). *Value at Risk: The New Benchmark for Controlling Market Risk.* Irwin, Chicago.

Morgan, J. P. (1996). *RiskMetrics Technical Document.* 4th edition. New York.

Lucas, A. (2000). A note on optimal estimation from a risk-management perspective under possibly misspecified tail behavior. *J. Bus. Econom. Statist.* **18**, 31-39.

Sentana, E. (1991). Quadratic ARCH models: a potential reinterpretation of ARCH models as second-order Taylor approximations. Working paper, London School of Economics and Political Science, London.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* Wiley, New York.

Tanner, M. A. (1993). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions.* 2nd edition. Springer-Verlag, New York.

Theodossiou, P. (1998). Financial data and the skewed generalized t distribution. *Manag. Sci.* **44**, 1650-1661.

Tsay, R. S. (2002). *Analysis of Financial Time Series.* Wiley, New York.

Department of Information and Systems Management, School of Business and Management, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong.

E-mail: imdavy@ust.hk

Department of Information and Systems Management, School of Business and Management, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong.

E-mail: immkpso@ust.hk

(Received June 2001; accepted July 2003)