Volatility filtering in estimation of kurtosis
(and variance)

Stanislav Anatolyev*
New Economic School

Abstract

The kurtosis of the distribution of financial returns characterized by high volatility persistence and thick tails is notoriously difficult to estimate precisely. We propose a simple but effective procedure of estimating the kurtosis coefficient (and variance) based on volatility filtering that uses a simple GARCH model. In addition to an estimate, the proposed algorithm issues a signal of whether the kurtosis (or variance) is finite or infinite. We also show how to construct confidence intervals around the proposed estimates. Simulations indicate that the proposed estimates are much less median biased than the usual method-of-moments estimates, their confidence intervals having much more precise coverage probabilities. The procedure also works well when the underlying volatility process is not the one the filtering technique is based on. We illustrate how the algorithm works using several actual series of returns.

Key words: returns, persistence, thick tails, kurtosis, variance, GARCH.

*Address: Stanislav Anatolyev, New Economic School, 100A Novaya Street, Skolkovo, Moscow, 143026, Russia. Phone: +7 (495) 956-9508. E-mail: sanatoly@nes.ru.
1 Introduction

Every empirical paper reports summary statistics for the data used in the study. Many textbooks do so as well for educational purposes. Among this output one can always see the sample variance and often sample kurtosis (raw or more frequently in the form of a kurtosis coefficient). When the data are financial returns, these statistics, especially the sample kurtosis, are unreliable. Indeed, when the data are highly persistent and thick-tailed, as are typical financial returns, the sample moments typically underestimate the true moments thus giving a misleading picture. The fact that higher-order moments of financial returns are notoriously hard to precisely estimate has even prompted some researchers to look for alternative measures of those data features they are meant to capture (see, e.g., Kim & White, 2004).

Technically, even though the mathematical expectation of (non-centered) sample moments matches the population moments exactly, their distributions are highly skewed so that the major probability mass is concentrated away from the true parameter. The following experiment gives an illustration. We generate a series following simple zero mean \textit{ARCH}(1) process (Engle, 1982) with news impact parameter $\alpha$, unit variance and standard normal standardized innovations. We set the sample size to 10000 which is larger than typical samples. The following tables show median biases of the sample kurtosis and variance expressed in percentage points, i.e. medians of

$$\frac{\hat{\theta} - \theta}{\theta} \times 100\%$$

across 1,000,000 simulations, where $\hat{\theta}$ is sample kurtosis or sample variance, and $\theta$ is the corresponding theoretical value (equaling $3(1 - \alpha^2)/(1 - 3\alpha^2)$ or 1, respectively). One can see that when $\alpha$ is relatively small, so is the median bias, but as $\alpha$ approaches the boundary of existence of the population kurtosis and variance ($1/\sqrt{3}$ and 1, respectively), the median bias becomes enormous. Needless to say, often financial returns operate in a close proximity of these boundaries.

| $\alpha$ | 0.30 | 0.40 | 0.50 | 0.56 | 0.57 |
|----------|------|------|------|------|------|
| Median bias of sample kurtosis | -1.8% | -7.8% | -33.1% | -78.1% | -90.0% |
| Median bias of sample variance |
|-------------------------------|
| $\alpha$ | 0.50 | 0.80 | 0.90 | 0.95 | 0.99 |
|        | $-0.4\%$ | $-11.4\%$ | $-33.4\%$ | $-57.5\%$ | $-89.5\%$ |

It follows, in particular, that sample kurtoses documented in applied literature are very likely to be underestimating their true values. For example, in Bollerslev (1987) the sample estimates for the kurtosis coefficient (left side of Table 1, p.544) are way below the values implied by the estimated GARCH-t model parameters.¹

A related and equally important issue is that the population moments may not exist at all. At the same time the sample analogs are always finite, and only their relatively large value may give a hint that the moments may not in fact exist. The figures for the sample kurtosis as 99.7 and 119.2 for some daily stock return indices documented in Frances & van Dijk (2000, Table 1.1) probably reflect the fact that the population kurtosis is infinite. The kurtosis figures like 63 or 77 for exchange rate futures documented in Cont (2001) are also suspect, and, according to the experimental evidence above, the actual kurtosis may be seriously larger, even if it is finite.

Given the degree of persistence, the problem of biased sample moments is much more severe for the fourth moment (i.e. kurtosis) than for the second moment (i.e. variance). However, even for the latter it is practically relevant. The simulation evidence in Anatolyev & Khrapov (2015) shows that for a GARCH model the ‘variance targeted’ estimate of the variance (i.e. sample variance) is (up to) twice as median biased as its estimate obtained within the GARCH framework.²

¹For example, for the British pound (left side of first line in Table 1 on p.544) the reported sample estimate is 4.81 while the GARCH-t parameters (right side of first line in Table 1 on p.544, and first line in Table 2 on p.545) imply the unconditional kurtosis coefficient of

$$\kappa = 4.63 \frac{1 - (0.921 + 0.057)^2}{1 - (0.921^2 + 2 \cdot 0.057 \cdot 0.921 + 0.057^2 \cdot 4.63)} \approx 6.35$$

(see formula (2) ahead).

²The ‘variance targeting’ detour for GARCH models (Engle & Mezrich, 1996; Francq, Horvath & Zakoian, 2011) helps reduce the problem dimensionality which is useful in a multivariate context. The method boils down to estimating the variance from the sample at the first step, subsequently plugging this estimate into the GARCH equation.
In this article we propose a method that in a way is ‘opposite’ to variance targeting, which is also (and primarily) applicable to computing the kurtosis coefficient. In a nutshell, the proposed volatility filtering (VF) technique implies passing the return series through a GARCH filter, computing kurtosis and variance of the filtered series, and restoring the implied kurtosis and variance of the original series by ‘reverting’ the estimated filter. The resulting estimates are going to be more precise because computation of moments for highly persistent series (that are subject to big estimation biases) is thus avoided. The trick is reminiscent of prewhitening errors in asymptotic variance estimation in time series (see Andrews & Monahan, 1992). Note that the underlying volatility process need not be the same GARCH as long as the filtering significantly reduces the degree of return persistence. The filtering can be based on the simple GARCH(1,1) process that often happens to be suitable for description of the conditional variance of financial returns, and, if the out-of-sample performance can be regarded as a signal of a ‘true’ model, GARCH(1,1) if often statistically sufficient for that purpose (Hansen & Lunde, 2005).

We verify how the VF technique works in simulations relative to the method of moments (MM). The VF estimates are much less median biased than the MM estimates, not only when the true underlying process is GARCH of the same order, but also when it is nonlinear GARCH from another class or when its order is higher. For practical purposes, filtering using orders (1,1) appears to be sufficient indeed. When the true kurtosis is infinite, the algorithm issues such signal often enough, and this frequency approaches unity for sufficiently big sample sizes.

We also show how to construct asymptotic confidence intervals (CI) for MM and VF estimators and verify their actual coverage probabilities. The MM asymptotic CI are poorly motivated because of problems with existence of higher order moments. The asymptotic CI for VF in contrast require more relaxed moment conditions and hence are much better motivated. Even though they involve computations of certain derivatives, they are still easy to construct numerically. We compare coverage rates in simulations and find out severe distortions in case MM is used, and much smaller distortions in the case of VF.

Finally, we illustrate discrepancies among different estimates and their standard errors using several actual series of various nature – stock index returns, individual stock returns,
exchange rate returns, – and frequency – monthly, weekly, daily. These discrepancies turn out to be quite large for estimates of the kurtosis coefficient, while a comparison of standard errors confirms inadequacy of those for the MM method.

The article is organized as follows. Section 2 describes the technique. Section 3 contains simulation evidence on the performance of the VF method in comparison with the method of moments. Section 4 analyses construction of confidence intervals, while Section 5 compares their coverage properties. Section 6 outlines an extension of the technique to the multivariate case. Section 7 shows an illustration with real data. Section 8 concludes. The Appendix contains auxiliary technical details.

2 Point estimation: How does it work?

For observable zero mean strictly stationary observable series of returns $r_t$, the method-of-moment (MM) estimates for the second moment (i.e. variance)

$$\hat{\sigma}^2 = E r_t^2$$

and fourth moment normalized by squared variance (i.e. kurtosis coefficient)

$$\kappa = \frac{E r_t^4}{\sigma^4}$$

are simply

$$\hat{\sigma}^2_{MM} = \frac{1}{T} \sum_{t=1}^{T} r_t^2$$

and

$$\hat{\kappa}_{MM} = \frac{1}{\hat{\sigma}^4_{MM}} \frac{1}{T} \sum_{t=1}^{T} r_t^4.$$  

Now we show how estimates of the variance and kurtosis coefficient can be constructed using the GARCH(1,1) and GARCH(2,2) volatility filtering.

Consider the GARCH(1,1) model (Bollerslev, 1986) with possibly non-normal standardized innovations:

$$r_t = \sigma_t \varepsilon_t$$
where \( \varepsilon_t \sim i.i.d. \mathcal{D}(0, 1) \) with finite fourth moment (kurtosis) \( \nu_4 = E \varepsilon_t^4 \), and volatility (conditional variance) follows
\[
\sigma_t^2 = \omega + \beta \sigma_{t-1}^2 + \alpha \varepsilon_{t-1}^2.
\]
For further use, denote \( \theta = (\omega, \alpha, \beta)' \). The variance of \( r_t \) can be expressed as (see Bollerslev, 1986)
\[
\sigma^2 = \frac{\omega}{1 - \beta - \alpha},
\]
provided that it is finite which happens when
\[\beta + \alpha < 1.\]

The kurtosis coefficient of \( r_t \) can be expressed as (see He and Teräsvirta, 1999a)
\[
\kappa = \nu_4 \frac{1 - (\beta + \alpha)^2}{1 - (\beta^2 + 2\alpha\beta + \alpha^2\nu_4)},
\]
provided that is finite which happens when
\[\beta^2 + 2\alpha\beta + \alpha^2\nu_4 < 1.\]

The algorithm of obtaining volatility filtered (VF) estimates goes as follows:

Step 1. Run \(^3\) GARCH(1,1) on \( r_t \). Obtain GARCH parameters \( \hat{\omega}, \hat{\beta}, \) and \( \hat{\alpha} \) and standardized errors \( \hat{\varepsilon}_t \). Compute an estimates of kurtosis of standardized innovations as
\[
\hat{\nu}_4 = \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_t^4.
\]

Step 2. If \( \hat{\beta} + \hat{\alpha} \geq 1 \), set \( \hat{\sigma}_{VF}^2 = \infty \). Otherwise, compute
\[
\hat{\sigma}_{VF}^2 = \frac{\hat{\omega}}{1 - \hat{\beta} - \hat{\alpha}}.
\]

Step 3. If \( \hat{\beta}^2 + 2\hat{\alpha}\hat{\beta} + \hat{\alpha}^2 \hat{\nu}_4 \geq 1 \), set \( \hat{\kappa}_{VF} = \infty \). Otherwise, compute
\[
\hat{\kappa}_{VF} = \hat{\nu}_4 \frac{1 - (\hat{\beta} + \hat{\alpha})^2}{1 - (\hat{\beta}^2 + 2\hat{\alpha}\hat{\beta} + \hat{\alpha}^2 \hat{\nu}_4)}.
\]

\(^3\)For raw returns, a conditional mean (e.g., a constant, or AR(1) as appropriate) should be added.
The inequality conditions can be alternatively verified by first computing the estimate and then checking if it is positive.

For the GARCH(2,2) process

\[ \sigma_t^2 = \omega + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2 + \alpha_1 r_{t-1}^2 + \alpha_2 r_{t-2}^2, \]

denote

\[ \gamma_{11} = \beta_1 + \alpha_1, \]
\[ \gamma_{21} = \beta_2 + \alpha_2, \]
\[ \gamma_{12} = \beta_1^2 + 2\alpha_1 \beta_1 + \alpha_1^2 \nu_4, \]
\[ \gamma_{22} = \beta_2^2 + 2\alpha_2 \beta_2 + \alpha_2^2 \nu_4, \]
\[ \tilde{\gamma}_{12} = \beta_1 \beta_2 + \alpha_2 \beta_1 + \alpha_1 \beta_2 + \alpha_1 \alpha_2 \nu_4. \]

Then the moments are:

\[ \sigma^2 = \frac{\omega}{1 - \gamma_{11} - \gamma_{21}}, \quad (3) \]

provided that

\[ \gamma_{11} + \gamma_{21} < 1, \]

and (see He and Teräsvirta, 1999a)

\[ \kappa = \nu_4 \frac{(1 - \gamma_{11} - \gamma_{21}) (1 + \gamma_{11} + \gamma_{21})(1 - \gamma_{21}) + 2\gamma_{11} \gamma_{21})}{(1 - \gamma_{21}) (1 - \gamma_{12} - \gamma_{22}) - 2\gamma_{11} \tilde{\gamma}_{12}}, \quad (4) \]

provided that

\[ 2\gamma_{11} \tilde{\gamma}_{12} < (1 - \gamma_{21}) (1 - \gamma_{12} - \gamma_{22}). \]

The estimation algorithm is adjusted accordingly:

Step 1. Run\(^4\) GARCH(2,2) on \(r_t\). Obtain GARCH parameters \(\hat{\omega}, \hat{\beta}_1, \hat{\beta}_2, \hat{\alpha}_1\) and \(\hat{\alpha}_2\) and standardized errors \(\hat{\varepsilon}_t\). Compute an estimate of variance and kurtosis of standardized innovations as

\[ \hat{\nu}_4 = \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_t^4. \]

\(^4\)Again, for raw returns, a conditional mean (e.g., a constant, or AR(1) as appropriate) should be added.
Also, compute

\[
\hat{\gamma}_{11} = \hat{\beta}_1 + \hat{\alpha}_1, \\
\hat{\gamma}_{21} = \hat{\beta}_2 + \hat{\alpha}_2, \\
\hat{\gamma}_{12} = \hat{\beta}_1^2 + 2\hat{\alpha}_1\hat{\beta}_1 + \hat{\alpha}_1^2\hat{\nu}_4, \\
\hat{\gamma}_{22} = \hat{\beta}_2^2 + 2\hat{\alpha}_2\hat{\beta}_2 + \hat{\alpha}_2^2\hat{\nu}_4, \\
\hat{\gamma}_{12} = \hat{\beta}_2\hat{\beta}_1 + \hat{\alpha}_2\hat{\beta}_1 + \hat{\alpha}_1\hat{\beta}_2 + \hat{\alpha}_1\hat{\alpha}_2\hat{\nu}_4.
\]

Step 2. If \(\hat{\gamma}_{11} + \hat{\gamma}_{21} \geq 1\), set \(\hat{\sigma}_{VF}^2 = \infty\). Otherwise, compute

\[
\hat{\sigma}_{VF}^2 = \frac{\hat{\omega}}{1 - \hat{\gamma}_{11} - \hat{\gamma}_{21}}.
\]

Step 3. If \(2\hat{\gamma}_{11}\hat{\gamma}_{12} \geq (1 - \hat{\gamma}_{21})(1 - \hat{\gamma}_{12} - \hat{\gamma}_{22})\), set \(\hat{\kappa}_{VF} = \infty\). Otherwise, compute

\[
\hat{\kappa}_{VF} = \hat{\nu}_4 \frac{(1 - \hat{\gamma}_{11} - \hat{\gamma}_{21})((1 + \hat{\gamma}_{11} + \hat{\gamma}_{21})(1 - \hat{\gamma}_{21}) + 2\hat{\gamma}_{11}\hat{\gamma}_{21})}{(1 - \hat{\gamma}_{21})(1 - \hat{\gamma}_{12} - \hat{\gamma}_{22}) - 2\hat{\gamma}_{11}\hat{\gamma}_{12}}.
\]

Again, the inequality conditions can be alternatively verified by first computing the estimate and then checking if it is positive.

### 3 Point estimation: How well does it work?

To verify the performance of the VF technique and compare it with that of MM we run simulations with 1,000 runs of artificial data of length \(T = 500, 2000\) and 10000, and read off 25%, 50% and 75% quantiles of distributions of estimates of \(\kappa\) and \(\sigma^2\) from both procedures.

We concentrate on quantiles because of likely problems with existence of moments. Most of the time we focus on the kurtosis coefficient \(\kappa\) because it is for the kurtosis that the problem under consideration is most acute; still, we look at the results for the variance \(\sigma^2\) as well.

The parameter \(\omega\) is set to unity unless otherwise noted.

#### 3.1 Estimation of kurtosis

Table 1a shows the results for kurtosis estimates when the standardized innovations \(\varepsilon_t\) are standard normal implying \(\nu_4 = 3\), for several combinations of variance equation parameters.
The first two combinations are empirically plausible and imply a moderately big kurtosis coefficient between 5 and 6. The third combination implies reduction to a simpler ARCH(1) true process. The fourth combination implies very weak volatility persistence and almost mesokurtic tails.

One can see that in the case of strongly persistent GARCH process (the first two panels) the VF estimates are much less biased than the MM estimates. When the sample size is large, the VF estimates are nearly median unbiased, while for the MM estimator the [25%,75%] interquartile range does not even contain the true value of kurtosis! For the less persistent ARCH process (the third panel) the VF estimates are quite precise even for the smallest sample size, while the true parameter is barely contained in the MM [25%,75%] interquartile range for the largest sample size. When the volatility persistence is very weak (the fourth panel), both procedures yield very similar results.

Table 1b shows the results for the former two parameter combinations when the standardized innovations $\varepsilon_t$ are (standardized) Student with $\eta = 10$ degrees of freedom implying $\nu_4 = 4$. This fattens the tails significantly resulting in a kurtosis coefficient larger than 10.

A larger kurtosis makes both estimators more variable, hence wider interquartile ranges. The VF estimator is especially prone to this, especially the right tail. However, in all cases its [25%,75%] interquartile range covers the true value of kurtosis, which does not fall far from the median, except for the smallest sample size. The MM estimates, in contrast, are very biased to the left, and their [25%,75%] interquartile ranges never cover the true value, even for the largest sample size.

Table 1c shows the results with standard normal standardized innovations for the first of previous parameter combinations with normal innovations but inflated news impact parameter $\alpha$ so that the kurtosis is either infinite or huge. The table additionally reports percentages of cases when the VF procedure classifies the kurtosis to be infinite.

In these cases the MM estimation gives ridiculously small estimates of kurtosis, the downward bias is huge. The VF estimates are also median biased downward, but to a much lesser degree. When the actual kurtosis in infinite, so is the median VF estimate for all sample sizes; in fact, the whole [25%,75%] interquartile range is infinite when the sample size is big enough, and the score in classification is large for the medium sample size and almost perfect.
for the largest sample size. When the actual kurtosis in finite but very large, it takes quite a big sample to reach the true value of kurtosis in median terms, and the proportion of infinite classifications, although moderately large, does not exceed 50%.

### 3.2 Kurtosis estimation under volatility misspecification

The volatility process followed by real data may not necessarily be GARCH(1,1). Now we verify how well the GARCH(1,1) volatility filtering works when the actual process is in fact more complex. We consider two cases: when the actual process is linear GARCH but of higher orders, (2,2), and when the actual process is some nonlinear GARCH, but of the same orders, (1,1).

Table 2a shows the results for kurtosis estimates when the underlying process is GARCH(2,2) with standardized normal innovations. In the second parameter combination the weight placed on the second lags are much larger than in the first parameter combination, otherwise the sum of $\alpha$’s and sum of $\beta$’s are kept the same. The true kurtosis coefficients are also similar. One can see that GARCH(1,1) filtering when the true process is of higher order still works very well, with the largest sample size yielding nearly median unbiased VF estimates, while the MM estimates are still far from the true parameter both in median and interquartile terms.

Table 2b shows the results for kurtosis estimates when the underlying process is one of three nonlinear GARCH(1,1) models with standardized normal innovations. These models are GJR (Glosten, Jagannathan & Runkle, 1993), TGARCH (Zakoïan, 1994) and EGARCH (Nelson, 1991). The formulas for kurtosis coefficients for these models can be found in the Appendix A.1. The GJR equation is

$$\sigma_t^2 = \omega + \beta_2 \sigma_{t-1}^2 + (\alpha + \gamma \mathbb{I}_{(r_{t-1}<0)}) r_{t-1}^2,$$

where $\mathbb{I}_{(r_{t-1}<0)}$ is indicator function. In contrast to the linear GARCH, it incorporates the leverage effect, i.e. asymmetry with respect to the sign of the past innovation. We set the news impact coefficients pretty different in size for the innovations of different sign. Although on average the news impact is the same as in the first GARCH(1,1) process in Table 1a, the kurtosis coefficient is larger because of this asymmetry. The TGARCH equation is
formulated for the conditional standard deviation:

$$\sigma_t = \omega + \beta \sigma_{t-1} + \left( \alpha^+ \mathbb{I}_{\{r_{t-1} > 0\}} + \alpha^- \mathbb{I}_{\{r_{t-1} < 0\}} \right) |r_{t-1}|.$$  \hfill (6)

We incorporate a distinguished leverage effect like that for the GJR model. The (restricted) EGARCH equation is formulated for logs of conditional variances:

$$\log \sigma_t^2 = \omega + \beta \log \sigma_{t-1}^2 + \frac{\eta_{t-1}}{\sigma_{t-1}}.$$  \hfill (7)

As expected, the MM estimates are severely biases downward, and in all three cases the VF estimator still fares better in median terms. Interestingly, for the TGARCH model the sign of the bias for the VF estimator is positive rather than negative. This is the only case when the VF estimator does not attain the goal for large sample sizes; for smaller ones the positive and negative biases partially cancel each other.

### 3.3 Kurtosis estimation with GARCH(2,2) filtering

The evidence presented in the previous subsection poses a question: if the actual volatility dynamics is nonlinear relative to GARCH, or its order is higher, would increasing the orders of the GARCH process used in filtering help better capture the deviations from the actual process and make a positive impact on the precision of VF estimates? Table 3 contains the results when instead the GARCH(2,2) filtering is relied on, and the actual processes are those from Tables 2a and 2b. Here we set the sample size to $T = 2000$.

Interestingly and unexpectedly, for nonlinear deviations from GARCH the use of higher orders does not bring improvements and in fact sometimes the performance of VF estimates deteriorates. In particular, in the GJR and TGARCH cases the median bias is larger when the GARCH(2,2) filter is used in place of GARCH(1,1). In the EGARCH case there is some improvement, but it is not large. Similarly, when the true process has orders (2,2) but only (1,1) are assumed, the performance of the VF technique changes very little, if at all. This is true even when the coefficients on second lags are bigger than those on first lags in the GARCH(2,2) specification.

A plausible explanation of this phenomenon is that even though the GARCH(2,2) model better accounts for the true persistence structure, estimation of two additional parameters reduces the precision of ingredients of the kurtosis estimate to a greater degree.
3.4 Estimation of variance

Finally, we compare MM and VF variance estimates. Table 4 shows the results with two processes with high persistence. The first parameter combination implies IGARCH structure and hence infinite variance. In the second parameter combination the news impact parameter is a bit smaller resulting in a covariance stationary process implying big but finite variance.

When the actual variance is finite, the VF [25%,75%] interquartile range is wider than that of the MM estimator. However, it is centered much closer to the true parameter, and the discrepancy diminishes quickly with the sample size. The percentages of incorrect classification diminishes very fast with the sample size. When the actual variance is infinite, typical VF estimates increase very quickly with the sample size, with the percentage of correct classification also growing although more slowly.

4 Confidence intervals: How does it work?

One may want to surround the point estimates, if they turn out to be finite, by confidence intervals. In the case of MM estimation the confidence intervals can be constructed in a straightforward way using the large sample theory and the Delta method. The asymptotic variance estimate for the MM estimator \((\hat{\sigma}_2^2, \hat{\kappa}_MM)^{'}\) is

\[
\hat{V}_{MM} = \hat{G}_{MM} \hat{W}_{MM} \hat{G}_{MM}^{'};
\]

where

\[
\hat{G}_{MM} = \begin{pmatrix}
1 & 0 \\
-2\hat{\sigma}_MM^2\hat{\kappa}_MM & \hat{\sigma}_MM^4
\end{pmatrix}
\]

is a consistent estimator of the matrix \(G\) of analytical first derivatives of the mapping from the vector of sample second and fourth moments to the vector \((\sigma^2, \kappa)^{'}\), and \(\hat{W}_{MM}\) is a HAC (e.g., Newey and West, 1987) estimator of the (long-run) variance of the vector of sample second and fourth moments

\[
W_{MM} = \sum_{j=-\infty}^{\infty} \text{cov} \left( \begin{pmatrix} \hat{r}_t^2 \\ \hat{r}_t^4 \end{pmatrix} , \begin{pmatrix} \hat{r}_{t-j}^2 \\ \hat{r}_{t-j}^4 \end{pmatrix} \right).
\]

Note that the MM asymptotic CI are poorly motivated as these computations presume existence of as high as eighth moments of returns. In the situations under consideration
such moments are unlikely to exist, and even when they do, their method-of-moment estimates may yield very poor approximations. This factor, together with biasedness of the estimates, is expected to make the MM confidence intervals rather spurious. In contrast, normal asymptotics holds for the GARCH parameters hold without a requirement of finite eighth moments; in fact, asymptotic normality may not fail even when the unconditional kurtosis is infinite provided that the conditional kurtosis is finite (e.g., Lee & Hansen, 1994; Francq & Zakoïan, 2004).

To construct a VF confidence interval based on the asymptotics of the QML estimator of the GARCH parameters and the Delta method, one needs an estimate of the covariance matrix $W_{VF}$ of parameters and average fourth moment of standardized returns and the matrix $G_{VF}$ of first derivatives of the mapping from the vector of GARCH parameters and average fourth moment of standardized returns to the vector $(\sigma^2, \kappa)'$ embedded in the VF algorithm. Then the asymptotic variance estimate for the VF estimator $(\hat{\sigma}^2_{VF}, \hat{\kappa}_{VF})'$ is

$$\hat{V}_{VF} = \hat{G}_{VF}\hat{W}_{VF}\hat{G}'_{VF},$$

where $\hat{G}_{VF}$ consistently estimates $G_{VF}$, and $\hat{W}_{VF}$ consistently estimates $W_{VF}$. The latter matrix is rather analytically convoluted, but it can be easily computed using numerical methods once the mapping is programmed. The former matrix can be computed using the following asymptotics (see derivation in Appendix A.2) that extends asymptotics of GARCH coefficients in Francq & Zakoïan (2004). We have

$$\sqrt{T} \left( \hat{\theta} - \theta \right) \to^d N \left( 0, W_{VF} \right),$$

where $\hat{\theta} = (\hat{\omega}, \hat{\alpha}, \hat{\beta})'$,

$$W_{VF} = \begin{bmatrix}
(v_4 - 1) J^{-1} & (v_6 - v_4) J^{-1} F_v \\
(v_6 - v_4) J^{-1} F_v' & v_8 - v_4^2
\end{bmatrix},$$

where

$$J = E \left[ \frac{1}{\sigma_i^2} \frac{\partial \sigma_i^2}{\partial \theta} \frac{\partial \sigma_i^2}{\partial \theta'} \right], \quad F = E \left[ \frac{1}{\sigma_i^2} \frac{\partial \sigma_i^2}{\partial \theta} \right],$$

$v_r = E [\varepsilon_t^r], r = 4, 6, 8$, and

$$\ell_t = \frac{r^2_t}{\sigma_i^2} + \log \sigma_i^2$$. 

13
being the (minus) true normal quasi-likelihood for observation $t$. The asymptotic variance $W_{VF}$ may be easily estimated during numerical minimization of the average quasi-likelihood. Note that this asymptotics is well motivated only if $v_8 < \infty$. This moment condition is much weaker for persistent data than the condition $E[r_t^3] < \infty$ needed for asymptotics of the MM estimator to hold.

5 Confidence intervals: How well does it work?

Table 5 contains results of simulation experiments with several DGPs used before. Two are correctly specified GARCH(1,1) models with basic parameter values except the tail thickness of standardized returns – one has normal tails implying $v_4 = 3$, the other has Student’s tails implying $v_4 = 4$. Two other DGPs are misspecified – one follows GARCH(2,2), and the other follows GJR(1,1). The true variance is $\sigma^2 = 20$ in all DGPs, the true value of kurtosis varies across DGPs. We report actual coverage probabilities of nominally 95% asymptotic confidence intervals from both procedures. We use the Newey and West (1987) HAC variance estimator for the MM, and the two-sided difference scheme for obtaining partial derivatives with the step $h = 10^{-5}$ times the parameter value, for the VF.\(^5\)

One can notice immediately that the MM confidence intervals have a dramatic undercoverage problem, especially those for the kurtosis coefficient. Even though the MM coverage probabilities increase with the sample size, even for with $T = 10000$ the actual coverage is far from nominal: for variance, the maximal coverage is about 75%, for kurtosis – mere 55% (and often much smaller). The VF confidence intervals are undercovered to a much lesser degree, uniformly. Still pretty serious undercoverage for relatively small samples quickly diminishes when the sample gets larger, and with $T = 10000$ the actual coverage for variance gets about nominal, while for kurtosis it rises to relatively tolerable levels.

The VF coverage probabilities for variance do not vary or vary very little with DGP and does not depend much whether it is correctly or misspecified. Comparing the VF coverage probabilities in two top panels and two bottom panels one can conclude that the coverage properties do not vary much with whether the true DGP is misspecified or not in the VF

\(^5\)This guarantees the precision of order $O(h^2)$ (see Judd, 1998, section 7.7).
procedure. It is the value of kurtosis that affects coverage: the higher the kurtosis coefficient, the higher are coverage distortions. Highest size distortions for kurtosis appear in the second DGP where the kurtosis is highest because the tails of standardized returns are thickest. Indeed, the problems of estimation of fourth moments partially carry over to standardized returns even though they are not persistent. All remaining VF coverage distortions evidently originate from two sources: one is a bias in estimates of GARCH coefficients, and the other is uncertainly in estimation of a fourth moment of standardized returns.

6 Extension to multivariate case

Extension to a multivariate context is straightforward, albeit technically tedious. Here we write out the VF algorithm when series of returns $r_t$ is $\ell$-dimensional, $\ell > 1$, when the returns are conditionally spherically distributed (cf. Hafner, 2003).

The MM estimate for the covariance matrix

$$\Sigma = E r_t r_t'$$

is

$$\hat{\Sigma}_{MM} = \frac{1}{T} \sum_{t=1}^{T} r_t r_t'.$$

For the collection of coefficients of kurtoses and co-kurtoses

$$\kappa_{ii} = \frac{E r_{i,t}^4}{\Sigma_{ii}^2}, \quad \kappa_{ij} = \frac{E r_{i,t}^2 r_{j,t}^2}{\Sigma_{ii} \Sigma_{jj}},$$

$i, j = 1, ..., \ell$, $i \neq j$, the collection of MM estimates is

$$\hat{\kappa}_{iiMM} = \frac{1}{\Sigma_{MMii}^2} \frac{1}{T} \sum_{t=1}^{T} r_{i,t}^4, \quad \hat{\kappa}_{ijMM} = \frac{1}{\Sigma_{MMii} \Sigma_{MMjj}} \frac{1}{T} \sum_{t=1}^{T} r_{i,t}^2 r_{j,t}^2.$$

Consider the general multivariate GARCH(1,1) model (Bollerslev, Engle & Wooldridge, 1988) with possibly non-normal standardized innovations:

$$r_t = H_t^{1/2} \varepsilon_t$$

where $\varepsilon_t \sim i.i.d. \mathcal{D}(0, I_\ell)$, and vectorized conditional covariance matrix follows

$$\text{vech}(H_t) = \text{vech}(\Omega) + B\text{vech}(H_{t-1}) + A\text{vech}(r_{t-1} r_{t-1}').$$
The distribution \( D \) of standardized innovations is assumed to be spherical with fourth cross-moments \( c = E \varepsilon_{i,t}^2 \varepsilon_{j,t}^2 = 3E \varepsilon_{i,t}^4 \) for all \( i, j = 1, ..., \ell, i \neq j \).

Denote \( n = \ell(\ell + 1)/2 \). The vectorized variance of \( r_t \) can be expressed as (see Engle & Kroner, 1995)

\[
\text{vech}(\Sigma) = (I_n - B - A)^{-1} \text{vech}(\Omega),
\]

provided that it is finite which happens when all eigenvalues of \( B + A \) lie inside the unit circle. From Hafner (2003), the kurtoses and co-kurtoses are elements of the following vector:

\[
K = \text{vec} \left( E \left[ \text{vech}(r_t r_t') \text{vech}(r_t r_t')' \right] \right)
\]

\[
= G_\ell (I_n^2 - Z)^{-1} \text{vec} \left( \begin{array}{c}
\text{vech}(\Omega)\text{vech}(\Omega)' \\
+(B + A)\text{vech}(\Sigma)\text{vech}(\Omega)' \\
+\text{vech}(\Omega)\text{vech}(\Sigma)'(B + A)'
\end{array} \right),
\]

where

\[
Z = (A \otimes A)G_\ell + B \otimes A + A \otimes B + B \otimes B,
\]

\[
G_\ell = c \left( 2(L_\ell \otimes D_\ell^+) (I_\ell \otimes C_\ell \otimes I_\ell) (D_\ell \otimes D_\ell) + I_n^2 \right),
\]

and \( L_\ell, D_\ell, D_\ell^+, C_\ell \) are certain matrix manipulation matrices (see Hafner, 2003), provided that this quantity is finite which happens when all eigenvalues of \( Z \) lie inside the unit circle.

The collection of coefficients of kurtoses and co-kurtoses can be extracted from \( K \) and \( \Sigma \):

\[
\kappa_{ii} = \frac{K_{kk_i}}{\Sigma_{ii}^2}, \quad \kappa_{ij} = \frac{K_{kk_i k_k j}}{\Sigma_{ii} \Sigma_{jj}},
\]

where \( k_i = (i - 1)(\ell - i/2) + i, \ i, j = 1, ..., \ell, i \neq j \).

The algorithm of obtaining VF estimates then goes as follows:

Step 1. Run\(^6\) unrestricted or restricted multivariate GARCH(1,1) on \( r_t \). Obtain MGARCH parameters \( \hat{\Omega}, \hat{B} \) and \( \hat{A} \) and standardized innovations \( \hat{\varepsilon}_t \). Compute an estimate of fourth cross-moments as, for example,

\[
\hat{c} = \frac{1}{3\ell} \sum_{i=1}^{\ell} \frac{1}{T} \sum_{t=1}^{T} \hat{\varepsilon}_{i,t}^4.
\]

Compute estimate \( \hat{Z} \) of \( Z \) from estimates \( \hat{\Omega}, \hat{B}, \hat{A} \) and \( \hat{c} \).

\(^6\)For raw returns, a conditional mean (e.g., a constant vector, or VAR(1) as appropriate) should be added.
Step 2. If the largest eigenvalue of $\hat{B} + \hat{A}$ is larger than unity in modulus, set $\hat{\Sigma}_{VF} = \infty$. Otherwise, compute

$$\text{vech}(\hat{\Sigma}_{VF}) = \left( I_n - \hat{B} - \hat{A} \right)^{-1} \text{vech}(\hat{\Omega}).$$

Step 3. If the largest eigenvalue of $\hat{Z}$ is larger than unity in modulus, set $\hat{K}_{VF} = \infty$. Otherwise, compute $\hat{K}_{VF}$ and extract relevant values of $\hat{\kappa}_{iiVF}$ and $\hat{\kappa}_{ijVF}$ for $i, j = 1, ..., \ell, i \neq j$.

If the whole covariance/co-kurtosis matrix is classified as infinite, it is possible that some variances/kurtoses are still finite. If they are of interest, one may reduce the dimensionality of the return vector and repeat the procedure for a subset of returns.

7 Empirical illustration

In this section we demonstrate the VF technique on typical financial return series of different frequencies (daily, weekly, monthly) and different nature (stock index, individual stock, exchange rate). We compare the estimates of variance $\sigma^2$ and kurtosis coefficient $\kappa$ from the MM procedure and VF procedures using GARCH(1,1) and GARCH(2,2) in addition to a constant conditional mean. Table 6a contains the results.

While the variance point estimates are in agreement across the methods, the point estimates of the kurtosis coefficient are not. While the former tend to be a bit higher, although not always, under volatility filtering, the latter are often two to three times larger (when finite) with filtering than without it. In some cases the VF method signals on non-existence of the fourth moment. This tendency is, of course, more clear for series that are likely to have thicker tails (possibly due to higher volatility persistence) – high frequency index and stock returns, while for exchange rates the difference is not so high.

Most dramatic are differences in standard errors provided by the two methods. Those for the VF variance estimates are typically 2-4 times larger than those for MM estimates. The latter standard errors provide too optimistic picture of actual estimation uncertainty. This tendency is exacerbated for kurtosis estimates. Even for moderately kurtic data (like monthly IBM returns and monthly dollar/yen exchange rate) the VF standard errors exceed those provided by MM by a fewfold. For more kurtic returns the ratio may be greater than
ten or even twenty. For very kurtic data with a probably finite kurtosis (like monthly S&P500 returns), the difference may be a few dozenfold. In cases of likely non-existent kurtosis, the MM shows moderate figures for the kurtosis coefficient, with inadequately small standard errors. Often, the MM estimate plus two MM standard deviations falls way short of the VF point estimates for the kurtosis coefficient.

Interestingly, while the figures across different orders of GARCH filtering are pretty similar, there are discrepancies for (most persistent and most thick-tailed) index returns. For the weekly S&P500 returns, the GARCH(1,1) filtering detects infinite kurtosis while the GARCH(2,2) filtering still regards it finite but very large (almost three times larger than the MM estimate computed from the raw data), and the huge VF standard errors reflect high uncertainty embedded in the estimate.

We also consider the bivariate case for daily and monthly exchange rate pairs. We compare the estimates of covariance matrix $\Sigma$ and relevant elements of the co-kurtosis coefficient matrix $K$ from the MM procedure and VF procedure that uses BEKK(1,1) (Engle & Kroner, 1995) in addition to a constant conditional mean vector. Table 6b contains the results. The MM variance and kurtosis estimates are equal or very similar to those from the univariate MM procedure, and the covariance and co-kurtosis values are pretty moderate. While the estimates of the former from the multivariate volatility filtering procedure tend to go up as expected, the estimates of the co-measures also increase in absolute value, especially the co-kurtosis coefficients (up to almost twice).

8 Concluding remarks

The volatility filtering method is an effective way to compute second and especially fourth moments of persistent series by avoiding a necessity to apply the method of moments to a persistent series. As our simulations show, the VF estimates are much less median biased compared to the MM estimates, and the asymptotic confidence intervals have actual coverage properties much closer to nominal. This holds not only when the GARCH(1,1) process used in filtering coincides with the true process driving the persistence, but also for a wide range of other GARCH-type processes with different persistence structures.
The remaining median biases and coverage distortions of the VF kurtosis estimator in a correctly specified persistence model are mainly due to the biasedness (even in mean terms) of GARCH coefficients which is intrinsic even in a correctly specified GARCH model (see, e.g., Anatolyev & Khrapov, 2015). While these fade off as the sample size grows, this happens pretty slowly. One may increase efficiency of GARCH and fourth moment estimates by using QML based on heavy-tailed distributions (Fan, Qi & Xiu, 2014), or by applying bias correction to GARCH parameters – analytical (Linton, 1997) or via bootstrap (Corradi & Iglesias, 2008). However, these refinements are likely to complicate the algorithm seriously threatening its simplicity.

One bigger development of the method may constitute an attempt to tie the form (e.g., the lag order) of the process used to filter the volatility to some computable persistence characteristics. Another possible extension is changing the volatility process to one that allows analytic computation of all moments of the series, not only of the fourth and second.
References

Anatolyev, S. & Khaplov, S. (2015). Right on target, or is it? The role of distributional shape in variance targeting. *Econometrics*, 3(3), 610–632.

Andrews, D. W. K. & Monahan, J. C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica*, 60, 953–966.

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

Bollerslev, T. (1987). Conditionally heteroskedastic time series model for speculative prices and rates of return. *Review of Economics and Statistics*, 69, 542–547.

Bollerslev, T., Engle, R. F. & Wooldridge, J. M. (1988). A Capital Asset Pricing Model with time-varying covariances. *Journal of Political Economy*, 96, 116–131.

Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2), 223–236.

Corradi, V. & Iglesias, E. M. (2008). Bootstrap refinements for QML estimators of the GARCH(1,1) parameters. *Journal of Econometrics*, 144(2), 500–510.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4), 987–1007.

Engle, R. F. & Kroner, K. F. (1995) Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11, 122–150.

Engle, R. F. & Mezrich, J. (1996). GARCH for groups. *Risk*, 9, 36–40.

Fan, J., Qi, L. & Xiu, D. (2014). Quasi-maximum likelihood estimation of GARCH models with heavy-tailed likelihoods. *Journal of Business & Economics Statistics*, 32(2), 178–191.

Frances, P. H. & van Dijk, D. (2000). *Nonlinear Time Series Models in Empirical Finance*. Cambridge University Press.

Francq, C. & Zakoïan, J.-M. (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli*, 10(4), 605–637.
Francq, C., Horvath, L., & Zakoïan, J.-M. (2011). Merits and drawbacks of variance targeting in GARCH models. *Journal of Financial Econometrics*, 9(4), 619–656.

Glosten, L. R., Jagannathan, R. and 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.

Hafner, C. (2003). Fourth moment structure of multivariate GARCH models. *Journal of Financial Econometrics*, 1(1), 26–54.

Hansen, P. R. & Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a GARCH (1,1)? *Journal of Applied Econometrics*, 20, 873–889.

He, C. and Teräsvirta, T. (1999a). Fourth moment structure of the GARCH\((p,q)\) process. *Econometric Theory*, 15, 824–846.

He, C. and Teräsvirta, T. (1999b). Properties of moments of a family of GARCH processes. *Journal of Econometrics*, 92, 173–192.

He, C., Teräsvirta, T. and Malmsten, H. (2002). Fourth moment structure of a family of first-order exponential GARCH models. *Econometric Theory*, 18, 868–885.

Judd, K. (1998). *Numerical Methods in Economics*, MIT Press.

Kim, T.-H. and White, H. (2004). On more robust estimation of skewness and kurtosis. *Finance Research Letters*, 1, 56–73.

Lee, S.W. and Hansen, B.E. (1994). Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory*, 10, 29–52.

Linton, O. (1997) An asymptotic expansion in the GARCH(1,1) model. *Econometric Theory*, 13, 558–581.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 59(2), 347–370.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55, 703–708.

Zakoïan, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, 18, 931–955.
Appendix

8.1 Kurtosis and variance formulas for non-GARCH processes

The following formulas for the kurtosis coefficients can be found in He and Teräsvirta (1999b) and He, Teräsvirta and Malmsten (2002).

For the GJR(1,1) model (5),

\[
\sigma^2 = \frac{\omega}{1 - \beta - (\alpha + \frac{1}{2}\gamma)}
\]

and

\[
\kappa = \nu_4 \frac{1 - (\beta + (\alpha + \frac{1}{2}\gamma))^2}{1 - (\beta^2 + \frac{1}{2} \beta + (\alpha^2 + \alpha\gamma + \frac{1}{2}\gamma^2) \nu_4)}.\]

For the TGARCH(1,1) model (6),

\[
\kappa = \nu_4 \frac{\Gamma_4 (1 - \gamma_C) (1 - \gamma_{C2})}{(1 + \gamma_C)^2 (1 - \gamma_{C3}) (1 - \gamma_{C4})},
\]

where

\[
\gamma_C = \beta + \nu_1 \frac{\alpha^+ + \alpha^-}{2},
\]

\[
\gamma_{C2} = \beta^2 + \nu_1 \beta (\alpha^+ + \alpha^-) + \nu_2 \frac{(\alpha^+)^2 + (\alpha^-)^2}{2},
\]

\[
\gamma_{C3} = \beta^3 + 3\nu_1 \beta^2 \frac{\alpha^+ + \alpha^-}{2} + 3\nu_2 \beta (\alpha^+)^2 + (\alpha^-)^2 + \nu_3 \frac{(\alpha^+)^3 + (\alpha^-)^3}{2},
\]

\[
\gamma_{C4} = \beta^4 + 2\nu_1 \beta^3 (\alpha^+ + \alpha^-) + 3\nu_2 \beta^2 \left((\alpha^+)^2 + (\alpha^-)^2\right) + 2\nu_3 \beta \left((\alpha^+)^3 + (\alpha^-)^3\right) + \nu_4 \frac{(\alpha^+)^4 + (\alpha^-)^4}{2},
\]

\[
\Gamma_4 = 1 + 3\gamma_C + 5\gamma_{C2} + 3\gamma_{C3} + 3\gamma_C \gamma_{C2} + 5\gamma_C \gamma_{C3} + 3\gamma_{C2} \gamma_{C3} + \gamma_C \gamma_{C2} \gamma_{C3},
\]

and for standard normal standardized innovations, \(\nu_1 = \sqrt{2/\pi}, \nu_2 = 1, \nu_3 = 2\sqrt{2/\pi}\) and \(\nu_4 = 3\).

For the EGARCH(1,1) model (7) with normal innovations,

\[
\kappa = 3 \exp \left( \frac{\alpha^2}{1 - \beta^2} \right).
\]

8.2 Derivation of asymptotic variance for VF estimator

Following Francq & Zakoïan (2004), formula (2.10), denote

\[
J = E \left[ \frac{\partial^2 \ell_t}{\partial \theta \partial \theta'} \right] = E \left[ \frac{1}{\sigma_t^4} \frac{\partial \sigma_t^2}{\partial \theta} \frac{\partial \sigma_t^2}{\partial \theta'} \right],
\]

22
where \( \ell_t \) is the true likelihood for observation \( t \). The joint asymptotics for average \( \partial \ell_t / \partial \theta \) and average \( \varepsilon_t^4 \) is

\[
\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left( \frac{\partial \ell_t}{\partial \theta} \right) \rightarrow^d N(0, W_0),
\]

where

\[
W_0 = \begin{bmatrix}
E \left[ \frac{\partial \ell_t}{\partial \theta} \frac{\partial \ell_t}{\partial \theta} \right] & E \left[ (\varepsilon_t^4 - v_4) \frac{\partial \ell_t}{\partial \theta} \right] \\
E \left[ (\varepsilon_t^4 - v_4) \frac{\partial \ell_t}{\partial \theta} \right] & E \left[ (\varepsilon_t^4 - v_4)^2 \right] \\
\end{bmatrix}
\]

using the representation (4.12) from Francq & Zakoïan (2004), where

\[
F = E \left[ \frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2}{\partial \theta} \right].
\]

Further,

\[
\sqrt{T} \left( \frac{\hat{\theta} - \theta}{\hat{v}_4 - v_4} \right) =^d \text{diag} \left\{- \left( \frac{1}{T} \sum_{t=1}^{T} \frac{\partial^2 \ell_t}{\partial \theta \partial \theta} \right)^{-1}, 1 \right\} \left( \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \left( \frac{\partial \ell_t}{\partial \theta} \right) \right) \rightarrow^d N(0, W_{VF}),
\]

where

\[
W_{VF} = \text{diag}\{ -J^{-1}, 1 \} W_0 \text{diag}\{ -J^{-1}, 1 \}
\]

\[
= \begin{bmatrix}
(v_4 - 1) J^{-1} & (v_6 - v_4) J^{-1} F \\
(v_6 - v_4) J^{-1} F' & v_8 - v_4^2
\end{bmatrix}.
\]

Finally, it follows that the asymptotic variance of VF estimator can be computed via the Delta method:

\[
V_{VF} = G_{VF} W_{VF} G_{VF}',
\]

where \( G_{VF} \) is the matrix of first derivatives of the mapping from the vector of parameters and average fourth moments of standardized returns to the vector \((\sigma^2, \kappa)'\).
### Tables and Figures

Table 1a: Estimates of kurtosis coefficient from MM and GARCH(1,1) VF procedures, true process is GARCH(1,1), standardized innovations are standard normal.

| Quantiles of $\hat{\kappa}$ | 25% | 50% | 75% | 25% | 50% | 75% | 25% | 50% | 75% |
|------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| T = 500                      |     |     |     |     |     |     |     |     |     |
| DGP $\alpha = 0.15$, $\beta = 0.80$, $\nu_4 = 3$ |     |     |     |     |     |     |     |     |     |
| True                         | $\kappa = 5.57$ |     |     |     |     |     |     |     |     |
| MM                           | 3.32 | 3.74 | 4.34 | 3.82 | 4.25 | 4.90 | 4.31 | 4.68 | 5.30 |
| VF                           | 3.74 | 4.77 | 7.88 | 4.52 | 5.26 | 7.03 | 5.07 | 5.54 | 6.18 |

| T = 2000                     |     |     |     |     |     |     |     |     |     |
| DGP $\alpha = 0.20$, $\beta = 0.70$, $\nu_4 = 3$ |     |     |     |     |     |     |     |     |     |
| True                         | $\kappa = 5.18$ |     |     |     |     |     |     |     |     |
| MM                           | 3.37 | 3.76 | 4.39 | 3.82 | 4.17 | 4.77 | 4.27 | 4.64 | 5.17 |
| VF                           | 3.86 | 4.69 | 6.84 | 4.41 | 5.08 | 6.13 | 4.78 | 5.19 | 5.64 |

| T = 10000                    |     |     |     |     |     |     |     |     |     |
| DGP $\alpha = 0.40$, $\beta = 0.00$, $\nu_4 = 3$ |     |     |     |     |     |     |     |     |     |
| True                         | $\kappa = 4.85$ |     |     |     |     |     |     |     |     |
| MM                           | 3.47 | 3.86 | 4.52 | 3.82 | 4.19 | 4.77 | 4.19 | 4.50 | 4.88 |
| VF                           | 3.84 | 4.62 | 5.97 | 4.36 | 4.91 | 5.80 | 4.60 | 4.84 | 5.13 |

| DGP $\alpha = 0.05$, $\beta = 0.05$, $\nu_4 = 3$ |     |     |     |     |     |     |     |     |     |
| True                         | $\kappa = 3.02$ |     |     |     |     |     |     |     |     |
| MM                           | 2.83 | 2.97 | 3.13 | 2.92 | 3.00 | 3.08 | 2.98 | 3.01 | 3.05 |
| VF                           | 2.84 | 3.00 | 3.16 | 2.93 | 3.01 | 3.12 | 2.98 | 3.02 | 3.05 |
Table 1b: Estimates of kurtosis coefficient from MM and GARCH(1,1) VF procedures, true process is GARCH(1,1), standardized innovations are standardized Student with 10 degrees of freedom.

|                | Quantiles of $\hat{\kappa}$ |                |                |                |                |                |
|----------------|------------------------------|----------------|----------------|----------------|----------------|----------------|
|                | 25% | 50% | 75%         | 25% | 50% | 75% | 25% | 50% | 75% |
| $T = 500$      |     |     |             |     |     |     |     |     |     |
| DGP $\alpha = 0.15, \beta = 0.80, \nu_4 = 4$ |     |     |             |     |     |     |     |     |     |
| True           | $\kappa = 13.00$             | $\kappa = 13.00$ | $\kappa = 13.00$ | $\kappa = 13.00$ | $\kappa = 13.00$ | $\kappa = 13.00$ |
| MM             | 4.06 | 4.80 | 6.06         | 5.01 | 5.91 | 7.56 | 6.13 | 7.01 | 8.95 |
| VF             | 4.87 | 7.32 | 59.15        | 7.06 | 10.71 | 25.78 | 9.37 | 12.48 | 20.74 |
| $T = 2000$     |     |     |             |     |     |     |     |     |     |
| DGP $\alpha = 0.20, \beta = 0.70, \nu_4 = 4$ |     |     |             |     |     |     |     |     |     |
| True           | $\kappa = 10.86$             | $\kappa = 10.86$ | $\kappa = 10.86$ | $\kappa = 10.86$ | $\kappa = 10.86$ | $\kappa = 10.86$ |
| MM             | 4.14 | 4.81 | 6.04         | 4.40 | 5.83 | 5.02 | 5.98 | 6.82 | 8.32 |
| VF             | 5.02 | 7.50 | 30.79        | 6.87 | 9.57 | 18.67 | 8.54 | 10.32 | 13.92 |
Table 1c: Estimates of kurtosis coefficient from MM and GARCH(1,1) VF procedures, true process is GARCH(1,1), standardized innovations are standard normal.

| DGP        | $\alpha = 0.19$, $\beta = 0.80$, $\nu_4 = 3$ | $\kappa = \infty$ |
|------------|---------------------------------------------|---------------------|
| True       |                                             |                     |
| MM         | 4.13 5.12 6.93 6.03 8.09 12.61 10.42 16.13 28.23 |                     |
| VF         | 4.75 $\infty$ $\infty$ $\infty$ $\infty$ $\infty$ $\infty$ $\infty$ $\infty$ |                     |
| VF % of $\infty$ | 62 88 99 |                     |
| DGP        | $\alpha = 0.17$, $\beta = 0.80$, $\nu_4 = 3$ |                     |
| True       |                                             | $\kappa = 136.38$   |
| MM         | 3.63 4.26 5.54 4.47 5.28 6.71 5.93 7.16 9.59 |                     |
| VF         | 4.84 11.14 $\infty$ 9.12 39.27 $\infty$ 20.57 126.64 $\infty$ |                     |
| VF % of $\infty$ | 36 43 45 |                     |
Table 2a: Estimates of kurtosis coefficient from MM and GARCH(1,1) VF procedures, true process is GARCH(2,2), standardized innovations are standard normal.

|                | 25%   | 50%   | 75%   | 25%   | 50%   | 75%   | 25%   | 50%   | 75%   |
|----------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|                | $T = 500$ | $T = 2000$ | $T = 10000$ |
| **DGP**        | GARCH(2,2): $\alpha_1 = 0.10$, $\beta_1 = 0.50$, $\alpha_2 = 0.05$, $\beta_2 = 0.30$ | $\kappa = 4.51$ |
| **True**       |       |       |       |       |       |       |       |       |       |
| MM             | 3.19  | 3.51  | 3.95  | 3.57  | 3.89  | 4.37  | 3.90  | 4.14  | 4.56  |
| VF             | 3.39  | 3.99  | 5.15  | 3.92  | 4.33  | 5.13  | 4.22  | 4.47  | 4.77  |
| **DGP**        | GARCH(2,2): $\alpha_1 = 0.05$, $\beta_1 = 0.30$, $\alpha_2 = 0.10$, $\beta_2 = 0.50$ | $\kappa = 4.19$ |
| **True**       |       |       |       |       |       |       |       |       |       |
| MM             | 3.12  | 3.38  | 3.74  | 3.46  | 3.71  | 4.09  | 3.74  | 3.97  | 4.28  |
| VF             | 3.11  | 3.56  | 4.40  | 3.69  | 4.05  | 4.62  | 3.95  | 4.14  | 4.37  |
Table 2b: Estimates of kurtosis coefficient from MM and GARCH(1,1) VF procedures, true process is various nonlinear GARCH of orders (1,1), standardized innovations are standard normal.

| DGP          | GJR(1,1): $\alpha = 0.10$, $\gamma = 0.10$, $\beta = 0.80$ | True $\kappa = 6.50$ | MM   | VF   |
|--------------|-------------------------------------------------------------|-----------------------|------|------|
|              |                                                             |                       | 25%  | 50%  | 75%  | 25%  | 50%  | 75%  | 25%  | 50%  | 75%  | 25%  | 50%  | 75%  |
| $T = 500$    |                                                             |                       | 3.37 | 3.78 | 4.50 | 3.98 | 4.45 | 5.29 | 4.52 | 5.03 | 5.91 |
| $T = 2000$   |                                                             |                       | 3.35 | 4.17 | 5.94 | 4.69 | 5.61 | 7.63 | 5.29 | 5.86 | 6.63 |
| $T = 10000$  |                                                             |                       | 3.33 | 3.65 | 4.13 | 3.68 | 3.95 | 4.32 | 3.94 | 4.11 | 4.34 |
|              |                                                             |                       | 3.60 | 4.32 | 5.75 | 4.36 | 4.94 | 5.93 | 4.76 | 5.11 | 5.55 |

| DGP          | TGARCH(1,1): $\alpha^+ = 0.10$, $\alpha^- = 0.20$, $\beta = 0.80$ | True $\kappa = 4.52$ | MM   | VF   |
|--------------|---------------------------------------------------------------------|-----------------------|------|------|
|              |                                                                      |                       | 3.33 | 3.65 | 4.13 | 3.68 | 3.95 | 4.32 | 3.94 | 4.11 | 4.34 |
|              |                                                                      |                       | 3.60 | 4.32 | 5.75 | 4.36 | 4.94 | 5.93 | 4.76 | 5.11 | 5.55 |

| DGP          | EGARCH(1,1): $\omega = -0.25$, $\alpha = 0.40$, $\beta = 0.80$    | True $\kappa = 4.68$ | MM   | VF   |
|--------------|---------------------------------------------------------------------|-----------------------|------|------|
|              |                                                                      |                       | 3.86 | 4.29 | 4.84 | 4.20 | 4.47 | 4.87 | 4.45 | 4.62 | 4.81 |
|              |                                                                      |                       | 4.03 | 4.54 | 5.43 | 4.29 | 4.60 | 5.00 | 4.52 | 4.68 | 4.88 |
Table 3: Estimates of kurtosis coefficient from MM and GARCH(1,1) VF procedures, true process is GARCH(2,2), standardized innovations are standard normal. The sample size is $T = 2000$.

| Quantiles of $\hat{\kappa}$ | 25% | 50% | 75% | 25% | 50% | 75% |
|-----------------------------|-----|-----|-----|-----|-----|-----|
|                            | GARCH(1,1) | GARCH(2,2) |
| DGP GJR(1,1): $\alpha = 0.10$, $\gamma = 0.10$, $\beta = 0.80$ |                   |                   |
| True $\kappa = 6.50$ |                   |                   |
| MM 3.98 | 4.45 | 5.29 | 3.97 | 4.45 | 5.15 |
| VF 4.70 | 5.59 | 7.45 | 4.43 | 5.37 | 7.15 |
| DGP TGARCH(1,1): $\alpha^+ = 0.10$, $\alpha^- = 0.20$, $\beta = 0.80$ |                   |                   |
| True $\kappa = 4.52$ |                   |                   |
| MM 3.68 | 3.95 | 4.32 | 3.70 | 3.95 | 4.31 |
| VF 4.36 | 4.94 | 5.93 | 4.42 | 5.11 | 6.48 |
| DGP EGARCH(1,1): $\omega = -0.25$, $\alpha = 0.40$, $\gamma = 0$, $\beta = 0.80$ |                   |                   |
| True $\kappa = 4.68$ |                   |                   |
| MM 4.20 | 4.47 | 4.87 | 4.24 | 4.54 | 4.90 |
| VF 4.29 | 4.60 | 5.00 | 4.38 | 4.71 | 5.14 |
| DGP GARCH(2,2): $\alpha_1 = 0.10$, $\beta_1 = 0.50$, $\alpha_2 = 0.05$, $\beta_2 = 0.30$ |                   |                   |
| True $\kappa = 4.51$ |                   |                   |
| MM 3.57 | 3.89 | 4.37 | 3.57 | 3.85 | 4.31 |
| VF 3.86 | 4.37 | 5.15 | 3.92 | 4.38 | 5.23 |
| DGP GARCH(2,2): $\alpha_1 = 0.05$, $\beta_1 = 0.30$, $\alpha_2 = 0.10$, $\beta_2 = 0.50$ |                   |                   |
| True $\kappa = 4.19$ |                   |                   |
| MM 3.46 | 3.71 | 4.09 | 3.46 | 3.70 | 4.08 |
| VF 3.69 | 4.05 | 4.62 | 3.70 | 4.03 | 4.59 |
Table 4: Estimates of variance from MM and GARCH(1,1) VF procedures, true process is GARCH(1,1), standardized innovations are standard normal.

| DGP          | $\alpha_1 = 0.20, \beta_1 = 0.80, \varepsilon_t \sim N(0,1)$ | $\alpha_1 = 0.19, \beta_1 = 0.80, \varepsilon_t \sim N(0,1)$ |
|--------------|-------------------------------------------------------------|-------------------------------------------------------------|
|               | $\sigma^2 = \infty$                                       | $\sigma^2 = 100$                                           |
| True         | $\sigma^2 = \infty$                                       |                                                           |
| MM           | $51.6 \quad 80.8 \quad 143.7 \quad 74.4 \quad 111.3 \quad 195.1 \quad 116.1 \quad 160.6 \quad 251.4$ | $36.5 \quad 51.6 \quad 79.6 \quad 49.0 \quad 61.9 \quad 86.0 \quad 62.4 \quad 74.1 \quad 94.6$ |
| VF           | $64.1 \quad 172.9 \quad \infty \quad 134.9 \quad 434.8 \quad \infty \quad 331.0 \quad 4113.5 \quad \infty$ | $41.7 \quad 72.4 \quad 322.4 \quad 60.4 \quad 87.8 \quad 188.9 \quad 76.9 \quad 98.7 \quad 135.3$ |
| VF % of $\infty$ | $35 \quad 39 \quad 47$                                       |                                                           |

Quantiles of $\hat{\sigma}^2$

|          | $T = 500$ | $T = 2000$ | $T = 10000$ |
|----------|-----------|------------|-------------|
| $25\%$  | $50\%$    | $75\%$    | $25\%$      | $50\%$      | $75\%$      | $25\%$      | $50\%$      | $75\%$      |
| $T = 500$ | $51.6$    | $80.8$     | $143.7$     | $74.4$    | $111.3$    | $195.1$    | $116.1$    | $160.6$    | $251.4$    |
| $T = 2000$ | $64.1$    | $172.9$    | $\infty$   | $134.9$  | $434.8$   | $\infty$   | $331.0$   | $4113.5$   | $\infty$   |
| $T = 10000$ | $36.5$    | $51.6$     | $79.6$      | $49.0$   | $61.9$  | $86.0$  | $62.4$  | $74.1$  | $94.6$  |
Table 5: Coverage probabilities of 95% asymptotic confidence intervals from MM and GARCH(1,1) VF procedures.

| DGP                | True          | $T = 500$ | $T = 2000$ | $T = 10000$ |
|--------------------|---------------|-----------|-------------|-------------|
|                    | $\kappa$ | $\sigma^2$ | $\kappa$ | $\sigma^2$ | $\kappa$ | $\sigma^2$ |
| GARCH(1,1): $\alpha = 0.15, \beta = 0.80, \nu_4 = 3$ | $\kappa = 5.57, \sigma^2 = 20$ |                      |             |             |
| MM                 | 24%          | 57%       | 32%         | 68%         | 44%       | 76%         |
| VF                 | 69%          | 84%       | 89%         | 92%         | 95%       | 95%         |
| GARCH(1,1): $\alpha = 0.15, \beta = 0.80, \nu_4 = 4$ | $\kappa = 13, \sigma^2 = 20$ |                      |             |             |
| MM                 | 7%           | 55%       | 13%         | 67%         | 24%       | 76%         |
| VF                 | 41%          | 82%       | 59%         | 90%         | 81%       | 95%         |
| GARCH(2,2): $\alpha_1 = 0.10, \beta_1 = 0.50, \alpha_2 = 0.05, \beta_2 = 0.30, \nu_4 = 3$ | $\kappa = 4.51, \sigma^2 = 20$ |                      |             |             |
| MM                 | 32%          | 54%       | 43%         | 64%         | 55%       | 72%         |
| VF                 | 80%          | 86%       | 92%         | 91%         | 97%       | 95%         |
| GJR(1,1): $\alpha = 0.10, \gamma = 0.10, \beta = 0.80, \nu_4 = 3$ | $\kappa = 6.50, \sigma^2 = 20$ |                      |             |             |
| MM                 | 17%          | 56%       | 26%         | 67%         | 34%       | 75%         |
| VF                 | 60%          | 83%       | 80%         | 92%         | 85%       | 94%         |
Table 6a: Estimates of variance and kurtosis coefficient from the MM and VF procedures for real data, univariate case.

| Series    | Period   | $T$  | MM    | GARCH(1,1) VF | GARCH(2,2) VF |
|-----------|----------|------|-------|---------------|---------------|
|           |          | $\sigma^2$ | $\kappa$ | $\sigma^2$ | $\kappa$ | $\sigma^2$ | $\kappa$ |
| daily S&P500 | 2000-2014 | 3773 | $1.65 \times 10^{-4}$ | $1.40 \times 10^{-4}$ | $\infty$ | $1.39 \times 10^{-4}$ | $\infty$ |
|           |          |      | $11.02$ | $1.40 \times 10^{-4}$ | $\infty$ | $1.39 \times 10^{-4}$ | $\infty$ |
| weekly S&P500 | 1990-2014 | 1304 | $5.38 \times 10^{-4}$ | $6.53 \times 10^{-4}$ | $\infty$ | $5.26 \times 10^{-4}$ | $25.76$ |
|           |          |      | $9.51$ | $6.53 \times 10^{-4}$ | $\infty$ | $5.26 \times 10^{-4}$ | $25.76$ |
| monthly S&P500 | 1950-2014 | 780 | $1.78 \times 10^{-3}$ | $1.99 \times 10^{-3}$ | $13.99$ | $1.89 \times 10^{-3}$ | $15.62$ |
|           |          |      | $4.99$ | $1.99 \times 10^{-3}$ | $13.99$ | $1.89 \times 10^{-3}$ | $15.62$ |
| daily IBM   | 2000-2014 | 3773 | $2.90 \times 10^{-4}$ | $3.24 \times 10^{-4}$ | $\infty$ | $3.17 \times 10^{-4}$ | $\infty$ |
|           |          |      | $10.79$ | $3.24 \times 10^{-4}$ | $\infty$ | $3.17 \times 10^{-4}$ | $\infty$ |
| weekly IBM  | 1990-2014 | 1304 | $1.46 \times 10^{-3}$ | $1.52 \times 10^{-3}$ | $10.72$ | $1.55 \times 10^{-3}$ | $11.25$ |
|           |          |      | $5.83$ | $1.52 \times 10^{-3}$ | $10.72$ | $1.55 \times 10^{-3}$ | $11.25$ |
| monthly IBM | 1962-2014 | 636 | $4.83 \times 10^{-3}$ | $4.70 \times 10^{-3}$ | $5.34$ | $4.82 \times 10^{-3}$ | $6.02$ |
|           |          |      | $4.70$ | $4.70 \times 10^{-3}$ | $5.34$ | $4.82 \times 10^{-3}$ | $6.02$ |
| daily $/£$  | 1990-2014 | 6286 | $3.51 \times 10^{-5}$ | $3.28 \times 10^{-5}$ | $6.48$ | $3.30 \times 10^{-5}$ | $6.64$ |
|           |          |      | $7.40$ | $3.28 \times 10^{-5}$ | $6.48$ | $3.30 \times 10^{-5}$ | $6.64$ |
| monthly $/£$ | 1971-2014 | 527 | $5.65 \times 10^{-4}$ | $6.18 \times 10^{-4}$ | $9.90$ | $6.16 \times 10^{-4}$ | $9.02$ |
|           |          |      | $4.79$ | $6.18 \times 10^{-4}$ | $9.90$ | $6.16 \times 10^{-4}$ | $9.02$ |
| daily ¥/$   | 1990-2014 | 6286 | $4.71 \times 10^{-5}$ | $4.75 \times 10^{-5}$ | $8.39$ | $4.69 \times 10^{-5}$ | $8.00$ |
|           |          |      | $7.43$ | $4.75 \times 10^{-5}$ | $8.39$ | $4.69 \times 10^{-5}$ | $8.00$ |
| monthly ¥/$ | 1971-2014 | 528 | $7.26 \times 10^{-4}$ | $7.16 \times 10^{-4}$ | $4.17$ | $7.27 \times 10^{-4}$ | $4.33$ |
|           |          |      | $4.14$ | $7.16 \times 10^{-4}$ | $4.17$ | $7.27 \times 10^{-4}$ | $4.33$ |
Table 6b: Estimates of variance and kurtosis coefficient from the MM and VF procedures for real data, bivariate case, BEKK(1,1).

| Series          | Period    | $T$  | MM          | BEKK(1,1) VF |
|-----------------|-----------|------|-------------|--------------|
|                 |           |      | $\Sigma$   | $K$          | $\Sigma$     | $K$          |
| daily $\$/£-¥/$ | 1990-2014 | 6286 | 3.51        | 7.40         | 3.44         | 9.95         |
|                 |           |      | $-0.89 \times 10^{-5}$ | 2.53         | $-1.03 \times 10^{-5}$ | 3.19         |
|                 |           |      | 4.70        | 7.42         | 4.76         | 7.18         |
|                 |           |      | 5.65        | 4.75         | 6.06         | 9.73         |
| monthly $\$/£-¥/$ | 1971-2014 | 527  | $-2.37 \times 10^{-4}$ | 1.32         | $-2.47 \times 10^{-4}$ | 2.54         |
|                 |           |      | 7.21        | 4.01         | 7.45         | 4.45         |