

Volatility filtering in estimation of kurtosis (and variance)

Stanislav Anatolyev*

CERGE-EI and New Economic School

January 2019

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.

MSC: 62F10, 62F12, 62F25, 62H12, 62M10, 62P20, 91B84.

*Addresses: CERGE-EI, Politických vězňů 7, 11121 Prague 1, Czech Republic; New Economic School, 45 Skolkovskoe shosse, 121353, Moscow, Russia. E-mail: stanislav.anatolyev@cerge-ei.cz.

“The use of kurtosis should be banished from use with fat-tailed distributions.”
Gabaix, Gopikrishnan, Plerou, and Stanley ([17], p.491)

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 highly unreliable (e.g., [3], [28]). 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 ([28]).

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 ARCH(1) process ([10]) with news impact parameter α , unit variance and standard normal standardized innovations. We set the sample size to 10000 which is larger than typical samples. Table 1 shows 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 θ is the corresponding theoretical value (equaling $3(1 - \alpha^2)/(1 - 3\alpha^2)$ or 1, respectively). One can see that when α is relatively small, so is the median bias, but as α 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.

It follows, in particular, that sample kurtoses documented in applied literature are very likely to be underestimating their true values. For example, in [5] the sample estimates for

Table 1: Median relative biases of sample kurtosis and sample variance for different values of α .

Sample kurtosis					
α	0.30	0.40	0.50	0.56	0.57
	-1.8%	-7.8%	-33.1%	-78.1%	-90.0%
Sample variance					
α	0.50	0.80	0.90	0.95	0.99
	-0.4%	-11.4%	-33.4%	-57.5%	-89.5%

the kurtosis coefficient (left side of Table 1, p.544) are way below the values implied by the estimated GARCH-t model parameters.¹

The empirical finance literature documents stylized facts that real-world financial returns exhibit heavy tails, sometimes to the extent that fourth or even second moments may be infinite (see, among others, [31] and [7]). For example, when power laws of return distributions are estimated, the tail index estimates often fall into the interval $(2, 4)$ implying that no finite kurtosis exists ([31], [17]) or may even turn out smaller than 2 implying that no finite variance exists ([32], [24]).²

While the population moments may not exist, the sample analogs are always finite,³ 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 [14] (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 [7] are also suspect, and, according to the experimental evidence above, the actual kurtosis may be seriously larger, even if it is finite.

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 (see formula (3) ahead)

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

²See the detailed discussion of heavy-tailed power law distributions, their estimation and application to financial data in [25], and of their connection with GARCH models in [37].

³See, for example, [9] on the limiting behavior of sample variances and autocovariances in such cases.

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 [1] 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.⁴

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 using gaussian quasi-maximum likelihood (QML) estimation, 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 the errors in asymptotic variance estimation in time series (see [2]). 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) is often statistically sufficient for that purpose ([20]).

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, up to almost always for sufficiently big samples.

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 (to be more exact, 8th) moments of the return series. The asymptotic CI for VF in contrast require more relaxed

⁴The ‘variance targeting’ detour for GARCH models ([12], [16]) helps reduce the dimensionality of the problem, 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.

moment conditions (existence of 8th moments as well, but of the innovation series instead of the return series) and hence are much better motivated. Even though they involve computations of certain derivatives, they are still easy to construct numerically. We also utilize the robust CI by [26] based on group t statistics. We compare coverage rates in simulations and find out severe distortions in the case of MM 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 MM.

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 analyzes construction of confidence intervals, while Section 5 compares their coverage properties. Section 6 outlines an extension to the multivariate case. Section 7 shows an illustration with real data. Section 8 concludes. The Appendix contains auxiliary technical details and proofs of propositions.

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)

$$\sigma^2 = Er_t^2$$

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

$$\kappa = \frac{Er_t^4}{\sigma^4}$$

are simply

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

and

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

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 ([4]) 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

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \alpha r_{t-1}^2. \quad (1)$$

For further use, denote $\theta = (\omega, \alpha, \beta)'$. The variance of r_t can be expressed as (see [4])

$$\sigma^2 = \frac{\omega}{1 - \beta - \alpha}, \quad (2)$$

provided that it is finite, which happens when

$$\beta + \alpha < 1.$$

The kurtosis coefficient of r_t can be expressed as (see [21])

$$\kappa = \nu_4 \frac{1 - (\beta + \alpha)^2}{1 - (\beta^2 + 2\alpha\beta + \alpha^2\nu_4)}, \quad (3)$$

provided that is finite, which happens when

$$\beta^2 + 2\alpha\beta + \alpha^2\nu_4 < 1.$$

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

Step 1. Run⁵ GARCH(1,1) QML 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

$$\hat{\sigma}_{VF}^2 = \frac{\hat{\omega}}{1 - \hat{\beta} - \hat{\alpha}}.$$

⁵For raw returns, a conditional mean (e.g., a constant, or AR(1) as appropriate) should be added.

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,⁶ compute

$$\hat{\kappa}_{VF} = \hat{\nu}_4 \frac{1 - (\hat{\beta} + \hat{\alpha})^2}{1 - (\hat{\beta}^2 + 2\hat{\alpha}\hat{\beta} + \hat{\alpha}^2\hat{\nu}_4)}.$$

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

For the GARCH(2,2) process

$$\sigma_t^2 = \omega + \beta_1\sigma_{t-1}^2 + \beta_2\sigma_{t-2}^2 + \alpha_1r_{t-1}^2 + \alpha_2r_{t-2}^2,$$

denote

$$\begin{aligned}\gamma_{11} &= \beta_1 + \alpha_1, \\ \gamma_{21} &= \beta_2 + \alpha_2, \\ \gamma_{12} &= \beta_1^2 + 2\alpha_1\beta_1 + \alpha_1^2\nu_4, \\ \gamma_{22} &= \beta_2^2 + 2\alpha_2\beta_2 + \alpha_2^2\nu_4, \\ \tilde{\gamma}_{12} &= \beta_1\beta_2 + \alpha_2\beta_1 + \alpha_1\beta_2 + \alpha_1\alpha_2\nu_4.\end{aligned}$$

Then the moments are:

$$\sigma^2 = \frac{\omega}{1 - \gamma_{11} - \gamma_{21}}, \quad (4)$$

provided that

$$\gamma_{11} + \gamma_{21} < 1,$$

and (see [21])

$$\kappa = \nu_4 \frac{(1 - \gamma_{11} - \gamma_{21})((1 + \gamma_{11} + \gamma_{21})(1 - \gamma_{21}) + 2\gamma_{11}\gamma_{21})}{(1 - \gamma_{21})(1 - \gamma_{12} - \gamma_{22}) - 2\gamma_{11}\tilde{\gamma}_{12}}, \quad (5)$$

provided that

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

The estimation algorithm is adjusted accordingly:

⁶Alternatively, one could also set $\hat{\kappa}_{VF} = \infty$ before checking for finiteness of kurtosis if $\hat{\sigma}_{VF}^2 = \infty$ is issued at step 2, because the kurtosis is infinite whenever variance is infinite. However, we stick to the algorithm in the text because $\hat{\sigma}_{VF}^2 = \infty$ is sometimes issued spuriously even when both variance and kurtosis are finite.

Step 1. Run⁷ GARCH(2,2) QML 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

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

Also, compute

$$\begin{aligned} \hat{\gamma}_{11} &= \hat{\beta}_1 + \hat{\alpha}_1, \\ \hat{\gamma}_{21} &= \hat{\beta}_2 + \hat{\alpha}_2, \\ \hat{\gamma}_{12} &= \hat{\beta}_1^2 + 2\hat{\alpha}_1\hat{\beta}_1 + \hat{\alpha}_1^2\hat{\nu}_4, \\ \hat{\gamma}_{22} &= \hat{\beta}_2^2 + 2\hat{\alpha}_2\hat{\beta}_2 + \hat{\alpha}_2^2\hat{\nu}_4, \\ \hat{\tilde{\gamma}}_{12} &= \hat{\beta}_1\hat{\beta}_2 + \hat{\alpha}_2\hat{\beta}_1 + \hat{\alpha}_1\hat{\beta}_2 + \hat{\alpha}_1\hat{\alpha}_2\hat{\nu}_4. \end{aligned}$$

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

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

Again, the inequality conditions can be alternatively verified by first computing the estimate, and then checking whether 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 1000 runs of artificial data of length $T = 500, 2000$ and 10000 , and read off 25%, 50% and 75% quantiles of distributions of estimates of κ and σ^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 κ , because it is for the kurtosis that the problem under consideration is most acute; still, we look at the results for the variance σ^2 as well.⁸ The parameter ω is set to unity unless otherwise noted.

⁷Again, for raw returns, a conditional mean (e.g., a constant, or AR(1) as appropriate) should be added.

⁸The performance of GARCH parameter estimates, which is not of interest here, is described in [1].

3.1 Estimation of kurtosis

Table 2a shows the results for kurtosis estimates when the standardized innovations ε_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.

Table 2a: Estimates of kurtosis coefficient from MM and GARCH(1,1) VF procedures, true process is GARCH(1,1), innovations are normal.

Quantiles of $\hat{\kappa}$									
	25%	50%	75%	25%	50%	75%	25%	50%	75%
	$T = 500$			$T = 2000$			$T = 10000$		
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
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
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

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 2b shows the results for the former two parameter combinations when the standardized innovations ε_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.

Table 2b: Estimates of kurtosis coefficient from MM and GARCH(1,1) VF procedures, true process is GARCH(1,1), innovations are Student with 10 degrees of freedom.

		Quantiles of $\hat{\kappa}$								
		25%	50%	75%	25%	50%	75%	25%	50%	75%
		$T = 500$			$T = 2000$			$T = 10000$		
DGP		$\alpha = 0.15, \beta = 0.80, \nu_4 = 4$								
True		$\kappa = 13.00$								
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
DGP		$\alpha = 0.20, \beta = 0.70, \nu_4 = 4$								
True		$\kappa = 10.86$								
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

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 2c shows the results with standard normal standardized innovations for the first of previous parameter combinations with normal innovations but inflated news impact parameter α 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.

Table 2c: Estimates of kurtosis coefficient from MM and GARCH(1,1) VF procedures, true process is GARCH(1,1), innovations are normal.

Quantiles of $\hat{\kappa}$									
	25%	50%	75%	25%	50%	75%	25%	50%	75%
	$T = 500$			$T = 2000$			$T = 10000$		
DGP	$\alpha = 0.19, \beta = 0.80, \nu_4 = 3$								
True	$\kappa = \infty$								
MM	4.13	5.12	6.93	6.03	8.09	12.61	10.42	16.13	28.23
VF	4.75	∞	∞	∞	∞	∞	∞	∞	∞
VF % of ∞	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	∞	9.12	39.27	∞	20.57	126.64	∞
VF % of ∞	36			43			45		

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 is 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 is 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 3a shows the results for kurtosis estimates when the underlying process is GARCH(2,2) with standardized normal innovations. In the second parameter combination, the weights placed on second lags are much larger than in the first parameter combination, otherwise the sum of α 's and sum of β '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 3a: Estimates of kurtosis coefficient from MM and GARCH(1,1) VF procedures, true process is GARCH(2,2), innovations are normal.

		Quantiles of $\hat{\kappa}$								
		25%	50%	75%	25%	50%	75%	25%	50%	75%
		$T = 500$			$T = 2000$			$T = 10000$		
DGP	GARCH(2,2): $\alpha_1 = 0.10, \beta_1 = 0.50, \alpha_2 = 0.05, \beta_2 = 0.30$									
True	$\kappa = 4.51$									
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): $\alpha_1 = 0.05, \beta_1 = 0.30, \alpha_2 = 0.10, \beta_2 = 0.50$									
True	$\kappa = 4.19$									
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 3b shows the results for kurtosis estimates when the underlying process is one of three nonlinear GARCH(1,1) models with standardized normal innovations, and a stochastic

volatility (SV) model. The GARCH models are GJR ([18]), TGARCH ([38]), and EGARCH ([33]), and the SV model is SARV ([36]).

Table 3b: Estimates of kurtosis coefficient from MM and GARCH(1,1) VF procedures, true process is nonlinear GARCH or SV, innovations are normal.

Quantiles of $\hat{\kappa}$									
	25%	50%	75%	25%	50%	75%	25%	50%	75%
	$T = 500$			$T = 2000$			$T = 10000$		
DGP	GJR(1,1): $\alpha = 0.10, \gamma = 0.10, \beta = 0.80$								
True	$\kappa = 6.50$								
MM	3.37	3.78	4.50	3.98	4.45	5.29	4.52	5.03	5.91
VF	3.35	4.17	5.94	4.69	5.61	7.63	5.29	5.86	6.63
DGP	TGARCH(1,1): $\alpha^+ = 0.10, \alpha^- = 0.20, \beta = 0.80$								
True	$\kappa = 4.52$								
MM	3.33	3.65	4.13	3.68	3.95	4.32	3.94	4.11	4.34
VF	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	3.86	4.29	4.84	4.20	4.47	4.87	4.45	4.62	4.81
VF	4.03	4.54	5.43	4.29	4.60	5.00	4.52	4.68	4.88
DGP	SARV(1): $\omega = -0.25, \beta = 0.80, \sigma_v^2 = 0.16$								
True	$\kappa = 4.68$								
MM	3.90	4.31	4.90	4.27	4.51	4.91	4.48	4.64	4.81
VF	4.17	4.78	5.64	4.35	4.81	5.28	4.76	4.98	5.23

The formulas for kurtosis coefficients for these models can be found in the Appendix A.1. The GJR equation is

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + (\alpha + \gamma\mathbb{I}_{\{r_{t-1} < 0\}}) r_{t-1}^2, \quad (6)$$

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 2a, the kurtosis coefficient is larger because of this asymmetry. The TGARCH equation is formulated for the conditional standard deviation:

$$\sigma_t = \omega + \beta\sigma_{t-1} + (\alpha^+\mathbb{I}_{\{r_{t-1}>0\}} + \alpha^-\mathbb{I}_{\{r_{t-1}<0\}}) |r_{t-1}|. \quad (7)$$

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

$$\log \sigma_t^2 = \omega + \beta \log \sigma_{t-1}^2 + \alpha \frac{r_{t-1}}{\sigma_{t-1}}. \quad (8)$$

The SARV equation is also formulated for logs but with an unobservable volatility shock in place of the observable driving process:

$$\log \sigma_t^2 = \omega + \beta \log \sigma_{t-1}^2 + v_t, \quad (9)$$

where $v_t \sim i.i.d. \mathcal{N}(0, \sigma_v^2)$ independent of ε_t .

As expected, the MM estimates are severely biased downward, and in the three nonlinear GARCH cases the VF estimator still fares better in median terms. For the SV process, GARCH volatility filtering has a harder time whitening the squared returns, due to the unobservable component being the driving process of volatility. Interestingly, for the TGARCH process the sign of the bias for the VF estimator is positive rather than negative, and so is it for the SARV process. This may be a consequence of a combination of high asymmetry and additional highly nonlinear transformation of the conditional variance in the TGARCH case and unobservable driver of volatility in the SARV case. These are the cases 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 higher order GARCH 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 4 contains the results when instead the GARCH(1,2) and GARCH(2,2) filters are employed, and the actual processes are those from Tables 3a and 3b. Here we set the sample size to $T = 2000$.

Table 4: Estimates of kurtosis coefficient from MM and GARCH(1,1) VF procedures, true process is nonlinear GARCH, SV or GARCH(2,2), sample size is 2000.

Quantiles of $\hat{\kappa}$									
	25%	50%	75%	25%	50%	75%	25%	50%	75%
	GARCH(1,1)			GARCH(1,2)			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	3.89	4.36	5.24	3.97	4.45	5.15
VF	4.70	5.59	7.45	4.50	5.37	7.12	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	3.68	3.95	4.32	3.70	3.95	4.31
VF	4.36	4.94	5.93	4.38	5.05	6.02	4.42	5.11	6.48
DGP	EGARCH(1,1): $\omega = -0.25, \alpha = 0.40, \beta = 0.80$								
True	$\kappa = 4.68$								
MM	4.20	4.47	4.87	4.18	4.48	4.85	4.24	4.54	4.90
VF	4.29	4.60	5.00	4.37	4.73	5.33	4.38	4.71	5.14
DGP	SARV(1): $\omega = -0.25, \beta = 0.80, \sigma_v^2 = 0.16$								
True	$\kappa = 4.68$								
MM	4.27	4.51	4.91	4.24	4.53	4.88	4.25	4.52	4.87
VF	4.35	4.81	5.28	4.62	5.02	5.70	4.58	4.97	5.46
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	3.54	3.85	4.34	3.57	3.85	4.31
VF	3.86	4.37	5.15	3.86	4.32	5.16	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	3.69	4.01	3.46	3.70	4.08
VF	3.69	4.05	4.62	3.66	4.04	4.62	3.70	4.03	4.59

Interestingly and unexpectedly, for nonlinear/SV 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 a higher order GARCH filter is used in place of GARCH(1,1). In the EGARCH and SV cases there is some improvement, but it is not large. When the true process has orders (2,2) but only (1,1) (or (1,2), for that matter) 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.

This phenomenon can be explained as follows. First, even though the GARCH(2,2) model better accounts for the true persistence structure, most of it can be well captured by the GARCH(1,1) structure. Indeed, the GARCH literature suggests that the GARCH(1,1) model is able to provide volatility forecasts that are no worse than those provided by more sophisticated models ([20]). Secondly, an addition of one or two more parameters is a sharp increase in model complexity, and it seems that estimation thereof reduces the precision of ingredients of the kurtosis estimate to a greater degree than the deviation from the true persistence structure.

3.4 Estimation of variance

Finally, we compare MM and VF variance estimates. Table 5 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.

Table 5: Estimates of variance from MM and GARCH(1,1) VF procedures, true process is GARCH(1,1), innovations are normal.

Quantiles of $\hat{\sigma}^2$									
	25%	50%	75%	25%	50%	75%	25%	50%	75%
	$T = 500$			$T = 2000$			$T = 10000$		
DGP	$\alpha_1 = 0.20, \beta_1 = 0.80, \varepsilon_t \sim N(0, 1)$								
True	$\sigma^2 = \infty$								
MM	51.6	80.8	143.7	74.4	111.3	195.1	116.1	160.6	251.4
VF	64.1	172.9	∞	134.9	434.8	∞	331.0	4113.5	∞
VF % of ∞	35			39			47		
DGP	$\alpha_1 = 0.19, \beta_1 = 0.80, \varepsilon_t \sim N(0, 1)$								
True	$\sigma^2 = 100$								
MM	36.5	51.6	79.6	49.0	61.9	86.0	62.4	74.1	94.6
VF	41.7	72.4	322.4	60.4	87.8	188.9	76.9	98.7	135.3
VF % of ∞	18			12			1		

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 (CI). We consider two ways of their construction: one is based on the Delta method supported by the conventional large sample theory; the other is correlation and heterogeneity robust inference by [26] based on finite sample inference for group t statistics.

4.1 Asymptotic confidence intervals

Consider first the inference based on conventional large sample theory. In a nutshell, it amounts to applying the Delta method to time series asymptotics for basic estimators obtained from the sample – the second and fourth sample moments for MM, and the GARCH parameter estimates and fourth sample moments of standardized returns for VF.

For the MM estimator, we have

Proposition 1. Suppose the return series r_t is stationary and ergodic with $E[r_t^8] < \infty$.

The MM estimator $(\hat{\sigma}_{MM}^2, \hat{\kappa}_{MM})'$ is consistent and asymptotically normal with the asymptotic variance

$$V_{MM} = G_{MM}W_{MM}G'_{MM},$$

where

$$G_{MM} = \begin{pmatrix} 1 & 0 \\ -2\sigma^{-2}\kappa & \sigma^{-4} \end{pmatrix}$$

is a matrix of analytical first derivatives of the mapping from the vector of sample second and fourth moments to the vector $(\sigma^2, \kappa)'$, and

$$W_{MM} = \sum_{j=-\infty}^{\infty} \text{cov} \left(\begin{pmatrix} r_t^2 \\ r_t^4 \end{pmatrix}, \begin{pmatrix} r_{t-j}^2 \\ r_{t-j}^4 \end{pmatrix} \right)$$

is the long-run variance of the vector of sample second and fourth moments.

The result of Proposition 1 follows by the Delta method applied to the transformation from the second and fourth moments of returns to the moments figuring in the definitions of σ^2 and κ . Because the former exhibit serial correlation, HAC variance estimation is needed.

The asymptotic variance V_{MM} can be consistently estimated by

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

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_{MM} , and \hat{W}_{MM} is heteroskedasticity and autocorrelation consistent (HAC, e.g., [34]) estimator of W_{MM} .

To construct a VF confidence interval one needs the joint asymptotics of the QML estimator $\hat{\theta} = (\hat{\omega}, \hat{\alpha}, \hat{\beta})'$ of the GARCH parameters and sample fourth moments of standardized returns $\hat{\nu}_4$. Denote $v_r = E[\varepsilon_t^r]$ for $r = 6, 8$. We have

Proposition 2. Suppose the return series r_t follows a strong GARCH(1,1) process (1) with $\beta^2 + 2\alpha\beta + \alpha^2\nu_4 < 1$, $v_8 < \infty$, and θ lying in the interior of a compact parameter set. The VF estimator $(\hat{\sigma}_{VF}^2, \hat{\kappa}_{VF})'$ is consistent and asymptotically normal with the asymptotic variance

$$V_{VF} = G_{VF}W_{VF}G'_{VF},$$

where G_{VF} is the matrix of first derivatives of the mapping from the vector of GARCH parameters and fourth moments of standardized returns to the vector $(\sigma^2, \kappa)'$ embedded in the VF algorithm, and W_{VF} is a covariance matrix of GARCH parameters and fourth moment of standardized returns

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

where

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

and

$$\ell_t = \frac{r_t^2}{\sigma_t^2} + \log \sigma_t^2.$$

The result of Proposition 2 follows by the Delta method applied to the transformation from the GARCH parameters and fourth moments of standardized returns to the moments figuring in the expressions (2) and (3) or (4) and (5). Because the score function and standardized returns are martingale differences, HAC variance estimation is not needed.

The matrix G_{VF} is analytically rather convoluted, but it can be easily computed using numerical methods once the mapping is programmed, and estimated by plugging in the available estimates of components. The matrix W_{VF} can be computed using the asymptotics in the Proposition that extends the asymptotics for GARCH coefficients in [15]. It can be easily estimated during the numerical minimization of the average quasi-likelihood. Then, an asymptotic variance estimate for the VF estimator $(\hat{\sigma}_{VF}^2, \hat{\kappa}_{VF})'$ 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} .

Note that the asymptotics for the VF estimator is well motivated when $v_8 < \infty$. This moment condition is much weaker for persistent data than the condition $E[r_t^8] < \infty$ needed for asymptotics of the MM estimator to hold. 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 yield very poor approximations (the effects described in

the Introduction for second and fourth moments are, of course, further exacerbated for eight moments). This factor, together with biasedness of the estimates, is expected to make the MM confidence intervals rather spurious. In contrast, the normal asymptotics holds for the GARCH parameters 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., [29], [15]). Also note that MM inference uses HAC variance estimates, which often have poor finite sample properties, while VF inference does not. The decisive factor, however, in different precision of CI is expected to be the different biasedness properties of the two point estimators.

4.2 Robust confidence intervals

Now consider the correlation and heterogeneity robust inference by [26] (IM henceforth). This way of constructing confidence intervals avoids tedious, especially for VF, estimation of derivatives and asymptotic variances; on the other hand, it requires repeated estimation of the GARCH model.⁹ Suppose we divide the sample into a fixed number q of equal non-overlapping groups of size τ , where $T = \tau q$, and compute estimates $(\hat{\sigma}_j^2, \hat{\kappa}_j)'$, $j = 1, \dots, q$, on all groups. Suppose that $2 \leq q \leq 14$ and $\phi \leq 0.10$, where ϕ is the CI level or test size. Then, we have

Proposition 3. Suppose that $\{(\hat{\sigma}_j^2, \hat{\kappa}_j)'\}_{j=1}^q$ are consistent, asymptotically normal and asymptotically independent as $\tau \rightarrow \infty$, with asymptotic variances uniformly bounded from below. Then, the t statistics over the groups,

$$t_{\sigma^2} = \sqrt{q} \frac{\overline{\hat{\sigma}^2} - \sigma^2}{s_{\hat{\sigma}^2}},$$

$$t_{\kappa} = \sqrt{q} \frac{\overline{\hat{\kappa}} - \kappa}{s_{\hat{\kappa}}},$$

where $\overline{\hat{\sigma}^2} = q^{-1} \sum_{j=1}^q \hat{\sigma}_j^2$, $s_{\hat{\sigma}^2}^2 = (q-1)^{-1} \sum_{j=1}^q (\hat{\sigma}_j^2 - \overline{\hat{\sigma}^2})^2$, $\overline{\hat{\kappa}} = q^{-1} \sum_{j=1}^q \hat{\kappa}_j$, and $s_{\hat{\kappa}}^2 = (q-1)^{-1} \sum_{j=1}^q (\hat{\kappa}_j - \overline{\hat{\kappa}})^2$, are both asymptotically distributed as Student's t with $q-1$ degrees of freedom under the null that σ^2 and κ are true values of the variance and kurtosis coefficient.

⁹[35] uses robust inference in the context of GARCH models, but the inference concerns the parameters of the conditional mean equation.

These ‘robust’ t tests are finite sample in the sense that a finite sample Student’s distribution is used for a sample of fixed size q , and asymptotic in the sense that an asymptotically infinite sample is needed to ensure uncorrelatedness of these q ‘observations’. By inverting the IM tests, one can construct IM confidence intervals for σ^2 and κ simply as $\overline{\hat{\sigma}^2} \mp cv_\phi s_{\hat{\sigma}^2} / \sqrt{q}$ and $\overline{\hat{\kappa}} \mp cv_\phi s_{\hat{\kappa}} / \sqrt{q}$, where cv_ϕ is the $(1 - \frac{1}{2}\phi)$ -quantile of the Student’s t distribution with $q - 1$ degrees of freedom.

As mentioned above, the IM confidence intervals require no tedious estimation of asymptotic variance ingredients, but require estimation of the GARCH model multiple times (which is, however, not a big cost). The latter fact in the current context imposes a pressure on the number of groups from above: each group has to be long enough for the GARCH model to be estimable and these estimates to be reliable. There is yet another peculiarity of IM confidence intervals: because they are centered around an average over groups rather than around the whole sample point estimate, the point estimate may be severely shifted away from the CI center or may not belong to the CI at all. Such outcomes may leave an empirical researcher who is used to symmetric confidence intervals a bit dissatisfied.

5 Confidence intervals: How well does it work?

Table 6 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 $\nu_4 = 3$, the other has Student’s tails implying $\nu_4 = 4$. 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 estimators, MM and VF, and two inference tools, Δ M (for ‘Delta method’) and IM. In the case of asymptotic CI, we use the HAC variance estimator by [24] for the MM, and the two-sided difference scheme for obtaining partial derivatives with the step $h = 10^{-5}$ times the parameter value, for VF.¹⁰ In the case of IM inference, we use $q = 2$ when $T = 500$, $q = 3$ when $T = 2000$, and $q = 5$ when $T = 10000$; these choices are mostly driven by the need to allocate a sufficient number of observations into each group for the volatility filter to be computationally feasible, and to set q to be slowly increasing

¹⁰This guarantees the precision of order $O(h^2)$ (see [27], section 7.7).

Table 6: Coverage probabilities of 95% asymptotic confidence intervals from MM and GARCH(1,1) VF procedures.

	κ	σ^2	κ	σ^2	κ	σ^2
	$T = 500$		$T = 2000$		$T = 10000$	
DGP	GARCH(1,1): $\alpha = 0.15, \beta = 0.80, \nu_4 = 3$					
True	$\kappa = 5.57, \sigma^2 = 20$					
MM with ΔM	24%	57%	32%	68%	44%	76%
MM with IM	71%	94%	52%	92%	50%	93%
VF with ΔM	69%	84%	89%	92%	95%	95%
VF with IM	85%	94%	89%	94%	94%	95%
DGP	GARCH(1,1): $\alpha = 0.15, \beta = 0.80, \nu_4 = 4$					
True	$\kappa = 13, \sigma^2 = 20$					
MM with ΔM	7%	55%	13%	67%	24%	76%
MM with IM	43%	92%	23%	92%	21%	92%
VF with ΔM	41%	82%	59%	90%	81%	95%
VF with IM	78%	94%	83%	95%	93%	94%
DGP	GARCH(2,2): $\alpha_1 = 0.10, \beta_1 = 0.50, \alpha_2 = 0.05, \beta_2 = 0.30, \nu_4 = 3$					
True	$\kappa = 4.51, \sigma^2 = 20$					
MM with ΔM	32%	54%	43%	64%	55%	72%
MM with IM	79%	94%	64%	93%	62%	94%
VF with ΔM	80%	86%	92%	91%	97%	95%
VF with IM	90%	95%	88%	90%	94%	95%
DGP	GJR(1,1): $\alpha = 0.10, \gamma = 0.10, \beta = 0.80, \nu_4 = 3$					
True	$\kappa = 6.50, \sigma^2 = 20$					
MM with ΔM	17%	56%	26%	67%	34%	75%
MM with IM	63%	93%	43%	92%	40%	93%
VF with ΔM	60%	83%	80%	92%	85%	94%
VF with IM	83%	94%	88%	92%	87%	95%

Note: For IM, $q = 2$ when $T = 500$, $q = 3$ when $T = 2000$, and $q = 5$ when $T = 10000$.

with T .

First consider the asymptotic inference based on the Delta method. 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 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 almost 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 do 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 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 for 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 uncertainty in estimation of a fourth moment of standardized returns.

Now consider the inference based on the IM method by. For both estimators and across all simulation designs, the IM confidence intervals have much better coverage properties than those based on the Delta method. For the largest sample size and estimation of variance, the VF confidence intervals based both on ΔM and on IM provide ideal coverage, though the coverage distortions for kurtosis kick in even for the largest sample size when the Delta method is used. For medium and small samples, the IM method clearly dominates in terms of coverage, be it for variance or for kurtosis. Having such good coverage is accompanied though with the CI being non-symmetric around the possibly biased VF point estimates and not even including them from time to time.

6 Extension to multivariate case

The extension to a multivariate context is straightforward, albeit technically tedious. Here we write out the VF algorithm when series of returns r_t is ℓ -dimensional, $\ell > 1$, and the returns are conditionally spherically distributed (cf. [19]).

The MM estimate for the covariance matrix

$$\Sigma = Er_t r_t'$$

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

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

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

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

Consider the general multivariate GARCH(1, 1) model ([6]) 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 \mathcal{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 $i, j = 1, \dots, \ell$, $i \neq j$.

Denote $n = \ell(\ell + 1)/2$. The vectorized variance of r_t can be expressed as (see [11])

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

provided that it is finite, which happens when all eigenvalues of $B + A$ lie inside the unit circle. From [19], the kurtoses and co-kurtoses are elements of the following vector:

$$\begin{aligned} \mathbf{K} &= \text{vec} \left(E[\text{vech}(r_t r_t') \text{vech}(r_t r_t)'] \right) \\ &= G_\ell (I_{n^2} - Z)^{-1} \text{vec} \left(\begin{array}{l} \text{vech}(\Omega) \text{vech}(\Omega)' \\ + (B + A) \text{vech}(\Sigma) \text{vech}(\Omega)' \\ + \text{vech}(\Omega) \text{vech}(\Sigma)' (B + A)' \end{array} \right), \end{aligned}$$

where

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

$$G_\ell = c(2(L_\ell \otimes D_\ell^+)(I_\ell \otimes C_{\ell\ell} \otimes I_\ell)(D_\ell \otimes D_\ell) + I_{n^2}),$$

and $L_\ell, D_\ell, D_\ell^+, C_{\ell\ell}$ are certain matrix manipulation matrices (see [19]), 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 Σ :

$$\kappa_{ii} = \frac{K_{k_i k_i}}{\Sigma_{ii}^2}, \quad \kappa_{ij} = \frac{K_{k_i k_j}}{\Sigma_{ii} \Sigma_{jj}},$$

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

The algorithm of obtaining VF estimates then goes as follows:

Step 1. Run¹¹ unrestricted or restricted multivariate GARCH(1,1) QML 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} .

Step 2. If the largest eigenvalue of $\hat{B} + \hat{A}$ is larger than unity in modulus, set $\hat{\Sigma}_{VF} = \infty$. Otherwise, compute

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

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 $i, j = 1, \dots, \ell$, $i \neq j$.

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.

¹¹For raw returns, a conditional mean (e.g., a constant vector, or VAR(1) as appropriate) should be added.

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 σ^2 and kurtosis coefficient κ from the MM procedure and VF procedures using GARCH(1,1) and GARCH(2,2) in addition to a constant conditional mean. Table 7a 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.

Table 7a: 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	
			σ^2	κ	σ^2	κ	σ^2	κ
daily S&P500	2000-2014	3773	1.65×10^{-4} (0.18)	11.02 (2.00)	1.40×10^{-4} (0.44)	∞	1.39×10^{-4} (0.40)	∞
weekly S&P500	1990-2014	1304	5.38×10^{-4} (0.71)	9.51 (3.20)	6.53×10^{-4} (3.10)	∞	5.26×10^{-4} (2.33)	25.76 (124.21)
monthly S&P500	1950-2014	780	1.78×10^{-3} (0.16)	4.99 (1.06)	1.99×10^{-3} (0.63)	13.99 (31.23)	1.89×10^{-3} (0.23)	15.62 (44.61)
daily IBM	2000-2014	3773	2.90×10^{-4} (0.25)	10.79 (2.12)	3.24×10^{-4} (1.25)	∞	3.17×10^{-4} (1.06)	∞
weekly IBM	1990-2014	1304	1.46×10^{-3} (0.12)	5.83 (0.46)	1.52×10^{-3} (0.41)	10.72 (8.73)	1.55×10^{-3} (0.45)	11.25 (10.51)
monthly IBM	1962-2014	636	4.83×10^{-3} (0.49)	4.70 (0.55)	4.70×10^{-3} (0.89)	5.34 (2.18)	4.82×10^{-3} (0.90)	6.02 (3.20)
daily \$/£	1990-2014	6286	3.51×10^{-5} (0.20)	7.40 (1.05)	3.28×10^{-5} (0.52)	6.48 (1.90)	3.30×10^{-5} (0.55)	6.64 (2.10)
monthly \$/£	1971-2014	527	5.65×10^{-4} (0.66)	4.79 (0.64)	6.18×10^{-4} (1.36)	9.90 (14.80)	6.16×10^{-4} (1.43)	9.02 (11.82)
daily ¥/\$	1990-2014	6286	4.71×10^{-5} (0.23)	7.43 (0.85)	4.75×10^{-5} (0.62)	8.39 (2.27)	4.69×10^{-5} (0.60)	8.00 (1.95)
monthly ¥/\$	1971-2014	528	7.26×10^{-4} (0.66)	4.14 (0.34)	7.16×10^{-4} (0.63)	4.17 (0.89)	7.27×10^{-4} (0.83)	4.33 (1.04)

Note: Asymptotic standard errors are below point estimates.

Table 7b: 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	
			Σ	K	Σ	K
daily \$/£-¥/\$	1990-2014	6286	$\begin{pmatrix} 3.51 \\ [1.68, 5.33] \\ -0.89 \\ [-2.26, 0.47] \\ 4.70 \\ [3.25, 6.15] \end{pmatrix}$	$\begin{pmatrix} 7.40 \\ [1.35, 10.45] \\ 2.53 \\ [-0.02, 4.62] \\ 7.42 \\ [3.84, 9.83] \end{pmatrix}$	$\begin{pmatrix} 3.44 \\ [1.98, 4.67] \\ -1.03 \\ [-2.03, -0.04] \\ 4.76 \\ [3.36, 5.98] \end{pmatrix}$	$\begin{pmatrix} 9.95 \\ [1.87, 17.68] \\ 3.19 \\ [1.71, 4.22] \\ 7.18 \\ [5.30, 8.60] \end{pmatrix}$
			$\times 10^{-5}$	$\times 10^{-5}$	$\times 10^{-5}$	$\times 10^{-5}$
monthly \$/£-¥/\$	1971-2014	527	$\begin{pmatrix} 5.65 \\ [-13.67, 24.95] \\ -2.37 \\ [-19.87, 15.10] \\ 7.21 \\ [5.32, 9.04] \end{pmatrix}$	$\begin{pmatrix} 4.75 \\ [-1.33, 10.68] \\ 1.32 \\ [1.19, 1.47] \\ 4.01 \\ [2.71, 5.29] \end{pmatrix}$	$\begin{pmatrix} 6.06 \\ [-33.63, 47.07] \\ -2.47 \\ [-29.69, 23.50] \\ 7.45 \\ [-4.87, 21.21] \end{pmatrix}$	$\begin{pmatrix} 9.73 \\ [-25.56, 30.42] \\ 2.54 \\ [-1.55, 55.62] \\ 4.45 \\ [-5.10, 15.53] \end{pmatrix}$
			$\times 10^{-4}$	$\times 10^{-4}$	$\times 10^{-4}$	$\times 10^{-4}$

Note: Robust IM 95% confidence intervals are below point estimates. With daily data, $q = 4$; with monthly data, $q = 2$.

We also consider the bivariate case for daily and monthly exchange rate pairs. We compare the estimates of covariance matrix Σ and relevant elements of the co-kurtosis coefficient matrix K from the MM procedure and VF procedure that uses BEKK(1,1) ([11]) in addition to a constant conditional mean vector. Here, because the use of the Delta method is even more cumbersome with the multivariate model, we construct robust IM 95% level confidence intervals instead of asymptotic ones. Table 7b 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). The 95% confidence intervals are quite narrow for the daily data, but are little informative for the monthly data, with the confidence intervals for positive parameters often covering large negative areas.

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 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., [1]). 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 ([13]), or by applying bias correction to GARCH parameters – analytical ([30]) or via bootstrap ([8]). 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.

9 Acknowledgements

I am grateful to the editor, associate editor and two anonymous referees for useful comments and suggestions. I also thank audiences at the EEA Congress 2016 at the Université de Genève, 10th CFE Conference at the University of Seville, 9th Biennial CES Conference in Prague, and Financial Econometrics Workshop 2018 at Örebro University.

References

- [1] Anatolyev, S. and S. Khrapov (2015). Right on target, or is it? The role of distributional shape in variance targeting. *Econometrics* 3(3), 610-632.
- [2] Andrews, D.W.K. and J.C. Monahan (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60, 953-966.
- [3] Bai, J. and S. Ng (2005). Tests of skewness, kurtosis, and normality for time series data. *J. Bus. Econom. Statist.* 32(2), 178-191.
- [4] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* 31(3), 307-327.
- [5] Bollerslev, T. (1987). Conditionally heteroskedastic time series model for speculative prices and rates of return. *Rev. Econ. Stat.* 69, 542-547.
- [6] Bollerslev, T., R.F. Engle, and J.M. Wooldridge (1988). A Capital Asset Pricing Model with time-varying covariances. *J. Polit. Econ.* 96, 116-131.
- [7] Cont, R. (2001). Empirical properties of asset returns: stylized facts and statistical issues. *Quant. Finance* 1(2), 223-236.
- [8] Corradi, V. and E.M. Iglesias (2008). Bootstrap refinements for QML estimators of the GARCH(1,1) parameters. *J. Econometrics* 144(2), 500-510.
- [9] Davis, R.A. and T. Mikosch (1998). The sample autocorrelations of heavy-tailed processes with applications to ARCH. *Ann. Statist.* 26(5), 2049-2080.
- [10] Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50(4), 987-1007.
- [11] Engle, R. F. and K.F. Kroner (1995). Multivariate simultaneous generalized ARCH. *Econometric Theory* 11, 122-150.
- [12] Engle, R. F. and J. Mezrich (1996). GARCH for groups. *Risk* 9, 36-40.
- [13] Fan, J., Qi, L. and D. Xiu (2014). Quasi-maximum likelihood estimation of GARCH models with heavy-tailed likelihoods. *J. Bus. Econom. Statist.* 32(2), 178-191.

- [14] Frances, P.H. and D. van Dijk (2000). *Nonlinear Time Series Models in Empirical Finance*. Cambridge: Cambridge University Press.
- [15] Francq, C. and J.-M. Zakoïan (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10(4), 605-637.
- [16] Francq, C., L. Horvath, and J.-M. Zakoïan (2011). Merits and drawbacks of variance targeting in GARCH models. *J. Financial Econometrics* 9(4), 619-656.
- [17] Gabaix, X., P. Gopikrishnan, V. Plerou, and H. Stanley (2006). Institutional investors and stock market volatility. *Quart. J. Econ.* 121(2), 461-504.
- [18] Glosten, L.R., R. Jagannathan, and D.E. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *J. Finance* 48(5), 1779-1801.
- [19] Hafner, C. (2003). Fourth moment structure of multivariate GARCH models. *J. Financial Econometrics* 1(1), 26-54.
- [20] Hansen, P.R. and A. Lunde (2005). A forecast comparison of volatility models: does anything beat a GARCH (1,1)? *J. Appl. Econometrics* 20, 873-889.
- [21] He, C. and T. Teräsvirta (1999a). Fourth moment structure of the GARCH(p,q) process. *Econometric Theory* 15, 824-846.
- [22] He, C. and T. Teräsvirta (1999b). Properties of moments of a family of GARCH processes. *J. Econometrics* 92, 173-192.
- [23] He, C., T. Teräsvirta, and H. Malmsten (2002). Fourth moment structure of a family of first-order exponential GARCH models. *Econometric Theory* 18, 868-885.
- [24] Ibragimov M., R. Ibragimov, and P. Kattuman (2013). Emerging markets and heavy tails. *J. Bank. Finance* 37, 2546-2559.
- [25] Ibragimov M., R. Ibragimov, and J. Walden (2015). *Heavy-Tailed Distributions and Robustness in Economics and Finance*. New York, NY: Springer.
- [26] Ibragimov, R. and U.K. Müller (2010). t-Statistic based correlation and heterogeneity robust inference. *J. Bus. Econom. Statist.* 28(4), 453-468.

- [27] Judd, K. (1998). *Numerical Methods in Economics*. Cambridge, MA: MIT Press.
- [28] Kim, T.-H. and H. White (2004). On more robust estimation of skewness and kurtosis. *Finance Research Letters* 1, 56-73.
- [29] Lee, S.W. and B.E. Hansen (1994). Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory* 10, 29-52.
- [30] Linton, O. (1997). An asymptotic expansion in the GARCH(1,1) model. *Econometric Theory* 13, 558-581.
- [31] Loretan, M. and P.C.B. Phillips (1994). Testing the covariance stationarity of heavy-tailed time series: An overview of the theory with applications to several financial datasets. *J. Empirical Finance* 1, 211-248.
- [32] Mittnik, S., M.S. Paolella, and S.T. Rachev (2000). Diagnosing and treating the fat tails in financial returns data. *J. Empirical Finance* 7(3-4), 389-416.
- [33] Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59(2), 347-370.
- [34] Newey, W.K. and K.D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703-708.
- [35] Pedersen, R.S. (2018). Robust inference in conditionally heteroskedastic autoregressions. Manuscript, University of Copenhagen, mpra.ub.uni-muenchen.de/81979.
- [36] Taylor, S.J. (1986). *Modeling Financial Time Series*. New York, NY: Wiley.
- [37] Xie, X. (2017). *Analysis of Heavy-Tailed Time Series*. Ph.D. dissertation, University of Copenhagen.
- [38] Zakoïan, J.-M. (1994). Threshold heteroskedastic models. *J. Econom. Dynam. Control* 18, 931-955.

A Appendix

A.1 Kurtosis and variance formulas for non-GARCH processes

The following formulas for the kurtosis coefficients can be found in [22], [23] and [36].

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

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

and

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

For the TGARCH(1,1) model (7),

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

where

$$\begin{aligned} \gamma_C &= \beta + \nu_1 \frac{\alpha^+ + \alpha^-}{2}, \\ \gamma_{C2} &= \beta^2 + \nu_1 \beta (\alpha^+ + \alpha^-) + \nu_2 \frac{(\alpha^+)^2 + (\alpha^-)^2}{2}, \\ \gamma_{C3} &= \beta^3 + 3\nu_1 \beta^2 \frac{\alpha^+ + \alpha^-}{2} + 3\nu_2 \beta \frac{(\alpha^+)^2 + (\alpha^-)^2}{2} + \nu_3 \frac{(\alpha^+)^3 + (\alpha^-)^3}{2}, \\ \gamma_{C4} &= \beta^4 + 2\nu_1 \beta^3 (\alpha^+ + \alpha^-) + 3\nu_2 \beta^2 ((\alpha^+)^2 + (\alpha^-)^2) \\ &\quad + 2\nu_3 \beta ((\alpha^+)^3 + (\alpha^-)^3) + \nu_4 \frac{(\alpha^+)^4 + (\alpha^-)^4}{2}, \\ \Gamma_4 &= 1 + 3\gamma_C + 5\gamma_{C2} + 3\gamma_{C3} + 3\gamma_C \gamma_{C2} + 5\gamma_C \gamma_{C3} + 3\gamma_{C2} \gamma_{C3} + \gamma_C \gamma_{C2} \gamma_{C3}, \end{aligned}$$

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 (8) with normal innovations,

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

For the CARV(1) model (9),

$$\kappa = 3 \exp\left(\frac{\sigma_v^2}{1 - \beta^2}\right).$$

A.2 Proofs of Propositions

Proof of Proposition 1: Straightforward, using the Ergodic Theorem, Central Limit Theorem for general dependent observations of a stationary series, and the Delta method.

□

Proof of Proposition 2: The (minus) (twice) normal quasi-likelihood for observation t is

$$\ell_t = \frac{r_t^2}{\sigma_t^2} + \log \sigma_t^2.$$

Following [15], formula (2.10), denote

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

where ℓ_t is the true likelihood for observation t . Following the proof of Theorem 2.2 in [15], the joint asymptotics for average $\partial \ell_t / \partial \theta$ and average ε_t^4 is

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T \begin{pmatrix} \frac{\partial \ell_t}{\partial \theta} \\ \varepsilon_t^4 - v_4 \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, W_0),$$

where

$$\begin{aligned} W_0 &= \begin{bmatrix} E \left[\frac{\partial \ell_t}{\partial \theta} \frac{\partial \ell_t}{\partial \theta'} \right] & E \left[(\varepsilon_t^4 - v_4) \frac{\partial \ell_t}{\partial \theta} \right] \\ E \left[(\varepsilon_t^4 - v_4) \frac{\partial \ell_t}{\partial \theta'} \right] & E \left[(\varepsilon_t^4 - v_4)^2 \right] \end{bmatrix} \\ &= \begin{bmatrix} (v_4 - 1) J & (v_6 - v_4) F \\ (v_6 - v_4) F' & v_8 - v_4^2 \end{bmatrix} \end{aligned}$$

using the representation (4.12) from [15], where

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

Further,

$$\sqrt{T} \begin{pmatrix} \hat{\theta} - \theta \\ \hat{v}_4 - v_4 \end{pmatrix} \stackrel{A}{=} \text{diag} \left\{ - \left(\frac{1}{T} \sum_{t=1}^T \frac{\partial^2 \ell_t}{\partial \theta \partial \theta'} \right)^{-1}, 1 \right\} \frac{1}{\sqrt{T}} \sum_{t=1}^T \begin{pmatrix} \frac{\partial \ell_t}{\partial \theta} \\ \varepsilon_t^4 - v_4 \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, W_{VF}),$$

where

$$\begin{aligned} W_{VF} &= \text{diag}\{-J^{-1}, 1\} W_0 \text{diag}\{-J^{-1}, 1\} \\ &= \begin{bmatrix} (v_4 - 1) J^{-1} & (v_6 - v_4) J^{-1} F \\ (v_6 - v_4) J^{-1} F' & v_8 - v_4^2 \end{bmatrix}. \end{aligned}$$

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 $(\sigma^2, \kappa)'$. \square

Proof of Proposition 3: Follows directly from [26]. \square