Article

Right on Target, or Is it? The Role of Distributional Shape in Variance Targeting

Stanislav Anatolyev * and Stanislav Khrapov

New Economic School, 100A Novaya Street, Skolkovo, Moscow, 143026 Russia;
E-Mail: skhrapov@nes.ru

* Author to whom correspondence should be addressed; E-Mail: sanatoly@nes.ru;
  Tel.: +7-495-956-9508.

Abstract: Estimation of GARCH models can be simplified by augmenting quasi-maximum likelihood (QML) estimation with variance targeting, which reduces the degree of parameterization and facilitates estimation. We compare the two approaches and investigate, via simulations, how non-normality features of the return distribution affect the quality of estimation of the volatility equation and corresponding value-at-risk predictions. We find that most GARCH coefficients and associated predictions are more precisely estimated when no variance targeting is employed. Bias properties are exacerbated for a heavier-tailed distribution of standardized returns, while the distributional asymmetry has little or moderate impact, these phenomena tending to be more pronounced under variance targeting. Some effects further intensify if one uses ML based on a leptokurtic distribution in place of normal QML. The sample size has also a more favorable effect on estimation precision when no variance targeting is used. Thus, if computational costs are not prohibitive, variance targeting should probably be avoided.

Keywords: GARCH; variance targeting; non-normality; heavy tails; skewness; quasi-maximum likelihood

JEL classifications: C22; C53; C58

1. Introduction

The technique of variance targeting for GARCH models has been proposed by Engle and Mezrich [1] in order to reduce the degree of parameterization, especially in multivariate contexts (see Pedersen and Rahbek [2] and Francq *et al.* [3]), and to facilitate the computation of maximum likelihood (ML) and quasi-maximum likelihood (QML) estimates. Variance targeting has become popular in empirical work, is included in popular textbooks and monographs and gets incorporated into some econometric software.[1]

Besides serving the main purpose, variance targeting has other merits, such as the superiority of long-term volatility and value-at-risk (VaR) predictions when the variance model is misspecified (Francq *et al.* [4]) because the unconditional variance is guaranteed to be consistently estimated. However, variance targeting does not come at no cost. One of its disadvantages over 'regular' QML estimation is the two-step nature, which results in efficiency losses (Francq *et al.* [4]) and a need for standard error corrections (Kristensen and Linton [5]). Francq *et al.* [4] quantify efficiency losses in a stylized ARCH model and find out that they are negligible when conditional heteroskedasticity is weak, but for some parameters, they may be dramatic when it is large enough to drive the unconditional kurtosis sufficiently high.

More generally, the heaviness of tails plays an important role in variance targeting. Francq *et al.* [4] prove the asymptotic normality of QML estimates under variance targeting when the unconditional kurtosis is finite. Francq *et al.* [4] establish, in particular, that asymptotic variances of both estimators, as well as their difference, are positively related to the fourth moment of returns. Vaynman and Beare [6] develop an asymptotic theory for the case of infinite unconditional kurtosis when the rate of convergence and asymptotic distribution cease to be standard. A quick look at estimated parameters when the distribution is heavy tailed reveals the drastic differences in values that they can take depending on whether variance targeting is applied or not. For example, five arbitrary runs of our simulation procedure described below yield for one of the parameters whose true value is unity the following estimates: {0.50, 0.54, 0.34, 0.32, 0.43} with variance targeting and {4.74, 0.67, 1.70, 2.02, 0.63} without variance targeting.[2] One can see that while the estimates are severely biased downward, though not highly dispersed in the first case, they tend to take very different values, sometimes having little to do with the true value, in the second case. These differences, of course, pass over to the other output from the GARCH analysis.

The literature on econometric modeling of time-varying volatility has been expanding ever since the seminal contribution of Engle [7] and Bollerslev [8]. One is welcome to explore just the tip of this iceberg surveyed by Bollerslev [9] and Bauwens *et al.* [10] for univariate and multivariate models, respectively. In spite of all of the advances in financial econometrics literature, we analyze the issues at hand on a relatively simple univariate GARCH model of Bollerslev [8] and its simple extension, GJR-GARCH, of Glosten *et al.* [11]. These choices are justified from several perspectives. First, they allow one to focus exclusively on the consequences of applying the two-step variance targeting estimation procedure and

---

[1] For example, the `rugarch` package in R.

[2] Using the notation that follows, this experiment corresponds to the case $\rho = 0.99$, $\beta = 0.8$, $\eta = 5$, $\lambda = 0$, $T = 2000$. The parameter under consideration is σ^2 with the true value of one.

the impact of the distributional features of the data. We believe that the effects described in this paper are inherited by more complicated nonlinear and multivariate models. Second, thanks to their simplicity and the fact that the properties are well studied, GARCH and GJR are still widely used among practitioners of financial markets and regulators for risk management purposes. Finally, the GJR model is the most straightforward extension of GARCH allowing for leverage effects.

We study the differences in properties of parameter estimates, as well as VaR predictions, which result from either using or abstaining from the variance targeting technique. We primarily focus on the influence of the shape of the conditional distribution of returns on discrepancies between the outcomes of the 'regular' QML procedure and the one augmented by variance targeting. Particularly, we simulate data from a GARCH model with standardized returns following the skewed Student distribution (Hansen [12]) allowing thick-tailedness and asymmetry, compute biases for parameter estimates and VaR predictions and track their dependence on these features. As the leading estimation method, we use the normal QML, but also experiment with regular ML based on the symmetric Student t and skewed Student distributions.

Francq *et al.* [4] obtain in their simulations that "...the finite-sample performance of the VTE seems quite satisfactory" and that the "experiments on daily stock returns do not show sensible differences between the estimated parameters of the two methods." The results of our simulation exercises do not conform to these conclusions. For most parameters and associated predictions, with a notable exception of unconditional variance and its functions, the bias under variance targeting is larger, at least in median terms, sometimes by a few-fold. This tendency is typically exacerbated for a heavier-tailed distribution of standardized returns, while the distributional asymmetry has little or a moderate impact. A larger sample size also has a more favorable effect on estimation precision when no variance targeting is used. For the unconditional variance mentioned above as an exception, the estimator dispersion may be quite high because of a long right tail when no variance targeting is employed (see the numerical example above). However, the median bias can be so much larger under variance targeting that statistics explicitly containing (in particular, long-run VaR predictions) this parameter may exhibit a very big degree of bias. Some effects further intensify if one uses ML based on a leptokurtic distribution in place of normal QML; in particular, the median bias for long-run VaR predictions shrinks significantly with the tail heaviness of the return distribution when no variance targeting is used, but does not under variance targeting. To summarize, we conclude that, especially when estimates of unconditional variance are involved in the statistic of interest, one should better avoid variance targeting provided that its computational benefits are not overwhelming.

Finite sample properties of GARCH parameter estimates have been previously documented in a number of studies. For example, Lumsdaine [13] concludes in particular that small sample distributions of parameter estimates of the GARCH(1,1) and IGARCH(1,1) models are skewed and leptokurtic. Linton [14] derives second-order asymptotic bias and skewness for QML parameter estimates in the GARCH(1,1) model. Some components of the second-order bias[3] and skewness are positively related (in absolute value) to the fourth cumulant, which explains some of the dependencies on return thick

---

[3]   Thus, of the median bias, as well; see Footnote 12 in Linton [14].

tailedness obtained here. Moreover, it is known that two-step estimators tend to possess more asymptotic bias components than one-step estimators (e.g., Anatolyev and Gospodinov [15], Chapter 5) which may explain why these tendencies are more severe when variance targeting is exploited.

The paper is organized as follows. In Section 2, we describe the simulation design, including the model, data generating process, estimation methods and statistics of interest. In Section 3, we show and analyze the results of simulation experiments. Section 4 concludes. Appendix A "Illustrations of DGP" contains some clarifying plots, while Appendix B "Selected Bias Distributions" shows noteworthy graphs representative of the results discussed in the article. An online Appendix containing a more complete collection of graphs and tables is available on the web at is.gd/vartarget.

2. Simulation Design

2.1. Models

The conditional variance model starts from the decomposition of returns:

$$r_t = \sqrt{h_t}\varepsilon_t, \tag{1}$$

where the conditional mean is absent, h_t is conditional variance, and ε_t is IID standardized return having distribution $\mathcal{D}$ with zero mean and unit variance. The leading process for the conditional variance is the GARCH(1,1) model of Bollerslev [8]:

$$h_t = \omega + \alpha r_{t-1}^2 + \beta h_{t-1}. \tag{2}$$

Let us introduce the parameter:

$$\rho \equiv \alpha + \beta, \tag{3}$$

which can be interpreted as a persistence of conditional variance. The GARCH equation is assumed to be stable in the sense that $\rho < 1$. The unconditional variance equals:

$$\sigma^2 = \frac{\omega}{1 - \rho}. \tag{4}$$

We also experiment with the GJR-GARCH conditional variance specification of Glosten *et al.* [11] containing leverage effects (see Rodríguez and Ruiz [16] and McAleer [17]). The GJR(1,1,1) equation is:

$$h_t = \omega + \alpha r_{t-1}^2 + \gamma r_{t-1}^2 \mathbf{1}\left[r_{t-1} < 0\right] + \beta h_{t-1}. \tag{5}$$

The persistence for this process is defined as (see Rodríguez and Ruiz [16]):

$$\rho \equiv \alpha + \gamma/2 + \beta, \tag{6}$$

and the unconditional variance is still given by (4) with ρ changed accordingly.[4]

As a matter of normalization, in all DGPs, we set $\sigma^2 = 1$.

---

[4] This correspondence is exact for symmetric standardized error distributions and approximate for asymmetric ones.

2.2. Data Generation

In data generation, we use two values for the feedback parameter β: 0.8 and 0.9, and one value for the persistence parameter ρ: 0.99.[5] These values approximately match GARCH parameters of daily stock and exchange rate returns (e.g., Bollerslev [18]). The two combinations imply two values for the news impact parameter α: 0.09 and 0.19. Note that in the case of normal standardized returns, the condition for the existence of fourth moments $2\alpha^2 + \rho^2 < 1$ (Bollerslev [8]) is satisfied when $\beta = 0.9$, but fails when $\beta = 0.8$. However, even in the former case under conditional non-normality, the corresponding condition $(E\left[\varepsilon_t^4\right] - 1)\alpha^2 + (\alpha + \beta)^2 < 1$ may fail if ε_t is leptokurtic enough. This failure, however, need not necessarily lead to a failure of asymptotic normality as long as $E[\varepsilon_t^4]$ is finite (Francq *et al.* [4], Lee and Hansen [19]). Still, the moment inequality when satisfied is quite close to being binding, so the loss in asymptotic precision as a result of variance targeting may be pretty large (Francq *et al.* [4]). In the GJR specification, the value of the additional (leverage) parameter equals $\gamma = 0.1$, and persistence is maintained at $\rho = 0.99$. For such parameter combinations, the fourth moment does not exist even if ε_t is mesokurtic (see Rodríguez and Ruiz [16]).

For the distribution $\mathcal{D}$, we use the skewed Student distribution of Hansen [12], which serves as a workhorse for estimating skewed conditional distributions in empirical finance; see, for example, Jondeau and Rockinger [20]. The skewed Student distribution with zero mean and unit variance has the following probability density function:

$$f\left(x|\eta, \lambda\right) = bc \left[1 + \frac{1}{\eta - 2}\left(\frac{a + bx}{1 + \mathrm{sgn}(x + a/b)\lambda}\right)^2\right]^{-(\eta+1)/2}, \tag{7}$$

where $2 < \eta < \infty$ and $-1 < \lambda < 1$. The constants a, b and c are given by:

$$a = 4\lambda c\frac{\eta - 2}{\eta - 1}, \quad b^2 = 1 + 3\lambda^2 - a^2, \quad c = \frac{\Gamma\left(\frac{\eta+1}{2}\right)}{\sqrt{\pi\left(\eta - 2\right)}\Gamma\left(\frac{\eta}{2}\right)}. \tag{8}$$

The shape parameters are η (degrees of freedom) and λ (degree of asymmetry). We vary the value of the kurtosis parameter η within the set $\{30, 10, 5, 4, 3\}$, implying a wide range of tail thickness. The value of 30 makes $\mathcal{D}$ indistinguishable (when $\lambda = 0$) from the normal distribution,[6] the value of three implies very heavy tails when third and fourth moments fail to exist,[7] while the intermediate values match degrees of freedom of daily stock and exchange rate returns (e.g., Bollerslev [18]). While we do most of experiments for zero skewness to focus on the effect of heavy tails, we also vary the asymmetry parameter λ within the set $\{-0.1, -0.3, -0.5, -0.8\}$, where -0.1 approximately matches the asymmetry coefficient for the S&P 500 index returns, while the other values imply several-fold exaggerated skewness. Figure 1 on page 624 presents plots of the skewed Student density corresponding to (some of the) combinations

---

[5]   We have also experimented with $\rho = 0.95$ with no qualitative changes in the conclusions.

[6]   We have also run experiments with standardized returns simulated from the standard normal distribution. The results are very similar to those corresponding to the case $\eta = 30$.

[7]   For data with infinite higher-order moments, Hill and Renault [21] propose robust variance targeting based on tail-trimming techniques.

of shape parameters that we employ. The density corresponding to the combination $\eta = 30$, $\lambda = 0$ (the darkest curves) is indistinguishable from the standard normal density shown in red.

We report most results for the sample length $T = 2000$, which is typical in GARCH estimation on daily data. We also analyze an effect of a sample size setting T to 500, 1000, 5000 and 10,000.

The number of simulations is 10,000. The code for error simulation is written in Python and based on Andrew Patton's MATLAB version.[8] The initial variance, h_1, is set at the unconditional variance (4) computed from the true parameters. Since we have chosen to normalize the variance, it is always initialized at one. The other variances are generated from recursion (2). The returns r_t are obtained as a product of corresponding innovation ε_t and conditional standard deviation $\sqrt{h_t}$. These returns are then retained to become observable, and other data are discarded. The burn-in sample size is equal to 100 observations, which are also discarded from further analysis.

The estimation of GARCH parameters given the observable series of returns $\{r_t\}_{t=1}^{T}$ is performed using the code also written in Python.[9] The initial parameter values for α, β (and γ for GJR) are chosen on a small grid[10] of parameter values to minimize Gaussian likelihood. The initial value for the intercept ω is computed from (4), chosen so as to match the sample variance $(\hat{\sigma}^2)^T$ given other model parameters. For each given set of parameters and returns data, we filter out time varying variances using the recursion given in (2). After that, we have all of the data and parameters necessary to compute the likelihood function for both estimators, with and without variance targeting.

As an example, we provide a plot of simulated returns, volatility and its estimates (with and without variance targeting) in Figure 2 on page 624. Persistence is set to $\rho = 0.99$, feedback to $\beta = 0.9$, degrees of freedom of skewed Student to $\eta = 10$ and asymmetry to $\lambda = -0.3$. The volatility process is highly persistent, and returns are somewhat fat-tailed and skewed to the left. There are 500 observations simulated. On both the top and bottom plots, the blue line corresponds to the true data. On the bottom panel, we also plot the filtered variances, green for variance targeting and red for the estimator without variance targeting. First, note that the volatility is clustered with periods of both high and low volatility. Several sharp increases in volatility are the result of a few large negative and positive returns. After these "jumps", the volatility tends to its long-run unconditional level smoothly, thanks to its high persistence. In this particular example, the difference between two volatility estimators is not clearly visible to the naked eye. The difference will be obvious in the following sections, where we describe the results of a full-scale Monte Carlo study.

2.3. Estimation Methods

Normally, estimation of parameters of the GARCH model in the formulation (2) is done via normal QML, with the corresponding log-likelihood for one observation equaling:

$$\ell_t^{NT}(\omega, \alpha, \beta) = -\frac{1}{2}\left(\ln(2\pi) + \ln h_t + \frac{r_t^2}{h_t}\right), \tag{9}$$

---

[8]   All links to publicly-available software are located at is.gd/vartarget.
[9]   The code along with documentation is available at is.gd/vartarget.
[10]   The grid is the Cartesian product of $\{0.01, 0.05, 0.1, 0.2\}$ for α and γ (where appropriate), $\{0.5, 0.7, 0.9, 0.98\}$ for β.

where h_t is defined recursively in (2). Because of (4), given the estimates ω^{NT}, $\hat{\alpha}^{NT}$ and $\hat{\beta}^{NT}$ of ω, α and β, the estimate of σ^2 is formed as:

$$\left(\hat{\sigma}^2\right)^{NT} = \frac{\hat{\omega}^{NT}}{1 - \hat{\rho}^{NT}}, \quad \hat{\rho}^{NT} = \hat{\alpha}^{NT} + \hat{\beta}^{NT}. \tag{10}$$

Lee and Hansen [19] prove that these estimates are consistent and asymptotically normal provided that the fourth moment of $\mathcal{D}$ is finite.

The machinery of variance targeting works as follows [1]. An alternative reparameterization of (2) is, using (4),

$$h_t = (1 - \alpha - \beta)\,\sigma^2 + \alpha r_{t-1}^2 + \beta h_{t-1}. \tag{11}$$

The parameter σ^2 can be pre-estimated directly from the sample by the method of moments:

$$\left(\hat{\sigma}^2\right)^T = \frac{1}{T} \sum_{t=1}^{T} r_t^2. \tag{12}$$

The parameters α and β are estimated given this estimate, with the corresponding log-likelihood for one observation equaling:

$$\ell_t^T\left(\alpha, \beta\right) = -\frac{1}{2}\left(\ln\left(2\pi\right) + \ln h_t^T + \frac{r_t^2}{h_t^T}\right), \tag{13}$$

with:

$$h_t^T = (1 - \alpha - \beta)\left(\hat{\sigma}^2\right)^T + \alpha r_{t-1}^2 + \beta h_{t-1}^T. \tag{14}$$

Maximization of $\sum_t \ell_t^T\left(\alpha, \beta\right)$ yields estimates $\hat{\alpha}^T$ and $\hat{\beta}^T$ of α and β. Given these, the estimate of ω is formed as:

$$\hat{\omega}^T = \left(1 - \hat{\rho}^T\right)\left(\hat{\sigma}^2\right)^T, \quad \hat{\rho}^T = \hat{\alpha}^T + \hat{\beta}^T. \tag{15}$$

Francq *et al.* [4] prove that these estimates are consistent and asymptotically normal provided that the fourth moment of $\mathcal{D}$ and the unconditional kurtosis are finite. In the case of GJR, the formulas change accordingly.

Sometimes, a researcher is willing to apply the method of maximum likelihood in place of normal QML. This is important when the objective is to construct value-at-risk predictions (Mittnik and Paolella [22], Bekaert *et al.* [23]). Kuester *et al.* [24], in particular, conclude, as a result of extensive comparison of various VaR prediction methods, that assuming a flexible parametric distribution for standardized returns is one of the most important factors in prediction accuracy. We consider a situation when the actual skewed Student distribution is used and a situation when the (underspecified) Student t distribution is used that ignores the asymmetry of the actual shape. In the latter case, the log-likelihood for one observation equals:

$$\ell_t^{NT}\left(\omega, \alpha, \beta, \eta\right) = \ln c - \frac{1}{2}\ln h_t - \frac{\eta + 1}{2}\ln\left(1 + \frac{r_t^2}{(\eta - 2)\,h_t}\right) \tag{16}$$

with:

$$c = \frac{\Gamma\left(\frac{\eta+1}{2}\right)}{\Gamma\left(\frac{\eta}{2}\right)\sqrt{\pi\left(\eta - 2\right)}}. \tag{17}$$

In the former case, it is:

$$\ell_t^{NT}(\omega,\alpha,\beta,\eta,\lambda) = \ln(bc) - \frac{1}{2}\ln h_t - \frac{\eta+1}{2}\ln\left(1 + \frac{1}{\eta-2}\left(\frac{a + bh_t^{-1/2}r_t}{1 + \mathrm{sgn}\left(h_t^{-1/2}r_t + a/b\right)\lambda}\right)^2\right), \quad (18)$$

where $2 < \eta < \infty$ and $-1 < \lambda < 1$. The constants a and b are given by:

$$a = 4\lambda c\frac{\eta-2}{\eta-1}, \quad b^2 = 1 + 3\lambda^2 - a^2, \quad (19)$$

and constant c is the same as in the case of the symmetric Student distribution.

2.4. Criteria

We first of all track percentage biases in parameter estimates:

$$B_\theta = \frac{\hat{\theta}-\theta}{\theta}, \quad (20)$$

where θ is one of the following parameters: the intercept parameter ω, the feedback parameter β, the news impact parameter α, the persistence parameter ρ, the unconditional variance σ^2 and, if the volatility equation contains a leverage effect, also the leverage parameter γ.

Second, we look at biases of two VaR predictors. The VaR predictions are functions of the estimates of the parameters listed above, but the evidence of them is meant to give an idea of how biases of parameter estimates translate into those of practically useful measures.[11]

In particular, we consider the one period in advance 5%-level VaR prediction conditional on $\mathcal{F}_t$, information available at time t,

$$\mathrm{VaR}_{t+1}^{5\%} \equiv \inf\left\{x \in \mathbb{R} \;:\; P\left(r_{t+1} \le x \,|\, \mathcal{F}_t\right) \ge 5\%\right\} = Q_{5\%}\sqrt{h_{t+1}}, \quad (21)$$

where $Q_{5\%}$ is a 5%-quantile of the distribution of standardized returns, which are typically measured in percentages. We look at the difference between its prediction and true value given $r_t^2 = h_t = 1$:

$$B_{+1}^{5\%} = \widehat{\mathrm{VaR}}_{t+1}^{5\%} - \mathrm{VaR}_{t+1}^{5\%} = \left.Q_{5\%}\left(\sqrt{\hat{h}_{t+1}} - \sqrt{h_{t+1}}\right)\right|_{r_t^2=h_t=1}. \quad (22)$$

The one-step volatility prediction is $h_{t+1}\big|_{r_t^2=h_t=1} = \omega + \rho$ (in case of GJR, parameter ρ is an average persistence over negative and positive past return outcomes), so the bias in the VaR prediction equals:

$$B_{+1}^{5\%} = Q_{5\%}\left(\sqrt{\hat{\omega}+\hat{\rho}} - 1\right). \quad (23)$$

When QML is used for estimation, $Q_{5\%} = Z_{5\%} \approx -1.645$ is a 5%-quantile of the standard normal distribution, and the prediction is 'naive' because one uses a quantile of the normal distribution instead of that of the unknown true one. Note however, that the prediction is proportional to that corresponding

---

[11] The predictions of value-at-risk and predictions of volatility itself are functions of the same parameter combinations of possibly different shapes and, hence, lead to qualitatively similar results. Therefore, we entirely focus on the former.

to the true distribution and so is its bias. When a researcher uses ML based on a non-normal distribution, the quantile $Q_{5\%}$ is a quantile of the distribution of ε_t used in ML estimation.

Another predictor we consider is the long-run per unit of time VaR predictor:

$$B_{+\infty}^{5\%} = Z_{5\%} \left(\hat{\sigma} - 1 \right), \tag{24}$$

where $Z_{5\%} \approx -1.645$ is a 5%-quantile of the standard normal distribution irrespective of the true return distribution. This value is multiplied by 100% to reflect the fact that returns are measured in percentages. See Lemma 3.1 in Francq *et al.* [4], who prove that the long-run, or infinite horizon, VaR prediction is well defined and is represented by the minuend.

When r_t^2 and/or h_t are different from unity, some parameters carry more weight relative to others. For example, when current volatility is higher than the unconditional variance and the current return is smaller than its standard deviation, the loading on β will be higher than that on α; when current volatility is higher than the unconditional variance and the current return is larger than its standard deviation, the loading on the suitable linear combination of α and β will overweight the loading on ω, the unity. Note also that long-run predictions $B_{+\infty}^{5\%}$ explicitly contain (an estimate of) the unconditional variance, while short-run predictions $B_{+1}^{5\%}$ do not.

3. Results

We report distributions of biases instead of means, variances and/or mean squared errors because of the possible non-existence of the second (and even first) moments for these distributions. Appendix B 'Selected Bias Distributions' shows noteworthy graphs representative of the results; an online Appendix at is.gd/vartarget contains a more complete collection of graphs (see the table with the description of experiments in its preamble), as well as tables with their numerical representation. Of primary interest is a median bias as the most appealing characteristic of the bias of estimated objects. See Andrews [25] who made forceful arguments about the practical appeal of the notion of median-unbiasedness, especially "when the parameter space is bounded or when the distributions of estimators are skewed and/or kurtotic," the circumstances we face here.

3.1. Parameter Biases

In this subsection, we present the results on parameter biases B_θ given by (20) and expressed in percentages, from QML estimation in the skewed Student GARCH model in the following order: B_ω, B_α, B_β, B_ρ and B_{σ^2}. Their distributions are partly shown in Figure 3; more graphs are available in the online Appendix as Figures A.1–A.5. The axes are expressed in percentages to true parameter values. In Figure 3, the two columns correspond to two selected values of the degrees of freedom parameter η representing almost mesokurtic ($\eta = 30$) and highly leptokurtic ($\eta = 4$) standardized return distributions, while the five rows correspond to five parameters; the value of the feedback parameter β is fixed at 0.8. In Figures A.1–A.5, the two columns correspond to two values of the feedback parameter β, while the five rows correspond to five values of the degrees of freedom parameter η. The blue lines correspond to variance targeting, the green lines to "regular" QML estimation. The vertical lines of corresponding color are placed at the medians of distributions; in addition, gray vertical lines are placed at zero, the ideal value

of a bias. Correspondingly, Tables A.1–A.5 contain 5%, 25%, 50%, 75% and 95% percentiles, expressed in percentages to the true parameter value, of bias distributions. Recall that throughout, $\sigma^2 = 1$ and $\rho = 0.99$ (implying $\omega = 0.01$).

The estimates of the intercept ω are not severely median biased: the maximal reported median bias corresponds to an about 13% deviation from the true value. The median bias is positive, smaller for the smaller value of β and always bigger when variance targeting is used, although by a narrow margin. The shape of the bias distribution is similar in the cases of variance targeting and of no variance targeting and so is its dispersion. The median bias stays stable (or goes up very slowly) when the return distribution becomes heavier tailed, but sharply drops down when the tails become so heavy that the higher moments of the standardized error distribution fail to exist. The dispersion goes up significantly with heavy tailedness for both methods: the $[5\%, 95\%]$ inter-percentile range approximately doubles when the return distribution turns from mesokurtic to strongly leptokurtic. The maximal reported 95% percentile corresponds to an estimate that is higher than triple the true value.

The median bias for the news impact parameter α is negative and is again larger in absolute value when variance targeting is used, up to twice as large, although the discrepancy in relative terms falls with the degree of heavy tailedness. In contrast to median biases for ω, it is larger for the smaller value of β (and hence, the bigger value of α). The median bias, importantly, increases (in absolute value) with the degree of return leptokurtosity pretty quickly. However, even for highly tailed return distributions, the median bias is relatively moderate (maximum 20%) if contrasted to the bias dispersion, which is big in both cases: the 95% percentile can exceed 120%. It tends to be bigger when no variance targeting is used, though not significantly, and bigger for the smaller value of α.

The median bias of estimates of the feedback parameter β is practically non-existent (which is natural given its high identifiability), whether the return distribution is thin or thick tailed and whether variance targeting is used or not; it amounts to less than half a percent in absolute value, even when the distribution is very leptokurtic. However, the tails of the distributions and their dispersion steadily rise with the degree of return leptokurtosity. The left tail is about twice as long as the right tail.

The relative bias of estimates of the persistence parameter ρ is of a similar magnitude as that for β, the median bias being of order of a few percent, despite the fact that estimation of the other ingredient of the sum, α, may be subject to substantial biases. This happens because in absolute terms, α is (as typically) much smaller than β, whose estimation is subject to a very small median bias. The median bias is always negative, larger (in absolute value) for the smaller value of β and moderately increasing in the degree of return leptokurtosity. It is larger when variance targeting is used than when it is not, though not appreciably. As far as the dispersion is concerned, it is inherited from those in the estimation of both β and α in such a way that most probability weight is put on the left tail, which is quite long and heavy. The right tail of bias distributions are bounded because of an implicitly-embedded condition on ρ not to exceed unity.

Last, but not least, and in fact most importantly, one can see that estimates of the unconditional variance σ^2 exhibit severe biases[12] compared to those of the other parameters. Even the median bias reaches and well exceeds (in absolute value) 50% in some cases. It steadily goes up with the degree of return leptokurtosity and is noticeably larger for the smaller value of β. The bias distribution quickly shifts leftward as the degree of return leptokurtosity increases. When variance targeting is used, the median bias is much larger, up to twice as much, although the discrepancy in relative terms falls with the degree of heavy tailedness, and when the tails are very thick, the difference is not so impressive. The larger median bias under variance targeting may sound surprising, as the estimate in this case is just that of the method-of-moments, which is mean unbiased (but highly skewed and, thus, severely median biased) and does not use the GARCH model at all. Exploitation of the model imposing the connection between the model parameters in the likelihood function significantly reduces the median bias, though making the bias distribution much more dispersed, primarily in the right tail. The latter fact is explained by a high frequency of estimates of ρ that get close to unity and make the estimate of σ^2 unstable; this phenomenon, naturally, becomes weaker when the persistence is not so strong.

To conclude, parameter estimate biases are strongly and positively related to the heaviness of the tails of the return distribution. Among the model parameters, it is the unconditional variance that experiences the severity of estimation bias most, with the intercept coming next. However, that severity is qualitatively different when variance targeting is used or not.

3.2. VaR Prediction Biases

Now, we analyze the bias of one-period and long-run value-at-risk predictions $B_{+1}^{5\%}$ and $B_{+\infty}^{5\%}$ defined in Equations (23) and (24), respectively. The biases of VaR predictions are inherited from the biases in the intercept ω and the persistence parameter ρ in the case of short-run prediction and from the biases in the unconditional variance σ^2 in the case of long-run prediction. From the previous analysis, it is clear that the behavior of biases is expected to be qualitatively different in the two cases. Their distributions are shown in Figure 4 and more fully in Figures A.6–A.7, with numerical presentation given in Tables A.6–A.7. The format of the graphs and tables is the same as described before. In Figure 4, the left column contains distributions for $B_{+1}^{5\%}$ and the right column for $B_{+\infty}^{5\%}$.

One-period VaR predictions $B_{+1}^{5\%}$ are the most frequently overpredicted, although the bias is quite small, except when the standardized return distribution is highly leptokurtic. This bias is driven mostly by the bias in the persistence parameter ρ, hence it is larger for the smaller value of β. The median bias rises with return leptokurtosity and becomes perceptible only for very heavy-tailed distributions of standardized returns. It is systematically larger when variance targeting is used, but by small amounts; the discrepancy almost vanishes when the standardized error distribution gets very thick tailed. The inter-percentile ranges are also quite similar.

However, the typical figures of the maximum mere a few percent, characterizing that the median biases for short-run VaR are incomparable to those for the long-run VaR, which can reach 50% and higher. This bias is driven entirely by the bias in the unconditional variance σ^2, hence such a result.

---

[12] The example in the Introduction corresponds to one of these cases.

The median bias is appreciably higher for the smaller value of feedback β. Under variance targeting, for mesokurtic standardized returns, the median bias is about twice as much as that without variance targeting. The median bias is increasing with the degree of thick tailedness and so do the inter-percentile ranges. An interesting phenomenon occurs when the degrees of freedom parameter η changes its value from four (implying the existence of conditional skewness, but the non-existence of conditional kurtosis) to three (implying the non-existence of either). When this happens, the distributions of VaR predictions (in addition to spreading out) shift rightward, leading to significant increases in either median bias. Overall, the bias is more dispersed when no variance targeting is used.

Next we analyze the impact of skewness of the return distribution on the bias. The evidence for the VaR criteria is shown in Figure 5, while Figures and Tables B.1–B.7 give a more complete picture, including that for parameter estimates. The five rows now correspond to five values of the asymmetry parameter λ; the degrees of freedom parameter is fixed at $\eta = 4$, implying a highly leptokurtic distribution for standardized returns. As the degree of distributional asymmetry varies from zero up to extreme values, the distribution of the bias does change, but not very significantly. Biases in some parameters double and so do biases in short-run VaR predictions as a result. The unconditional variance σ^2 and, as a result, the long-run VaR predictions are subject to higher estimation biases when the return distribution becomes asymmetric, but this change is not too dramatic: the median bias goes up by less than 50% of its value.

Recall that all previous evidence is given when the sample size T equals 2000. Now, we analyze the impact of a sample size on the bias. The evidence for the VaR criteria is shown in Figure 6, while Figures and Tables C.1–C.7 give more complete evidence, including that for parameter estimates. The five rows now correspond to five values of T; the shape parameters are fixed at $\lambda = 0$, $\eta = 4$ implying a highly leptokurtic symmetric distribution for standardized returns. Note that for a very small (for typical financial series, about two years of daily data) sample size of 500, our previous conclusions are exacerbated. As T increases, all biases naturally become substantially smaller, both in median and inter-percentile terms. However, this happens at a different rate depending on whether short- or long-run VaR predictions are considered and whether variance targeting is used or not. Overall, biases of short-run predictions tend to shrink faster than long-run counterparts and faster when no variance targeting is used than when it is used. Indeed, as T rises by 20 times (note that $\sqrt{20} \approx 4.5$), while the median bias drops by 6–8 times or by 2–3 times, depending on the value of the feedback parameter under QML, it drops only by 3–5 times or 1.5–2 times, respectively, under variance targeting. However, QML predictions of long-run VaR predictions bear a significant risk of underprediction, even when sample sizes are quite large.

3.3. GJR in Place of GARCH

Next, we verify whether and how the presence of leverage effects in the volatility equation affects bias. The distributions of biases for selected parameters and VaR predictions are shown in Figure 7. Figures and Tables D.1–D.8 contain more complete information; one can compare these to Figures and Tables A.1–A.8 and Figure 7 to Figures 3 and 4.

The bias properties of estimates of the news impact parameter α when leverage is present naturally deteriorate, as serial correlation features in squared returns (together with return signs) now need to identify two parameters α and γ at the same time; effectively, only (approximately) half of the sample is used for either. As a result, the (negative) median bias increases and so does bias variability. This effect is especially pronounced for the larger value of β and, hence, smaller value of α. However, the bias properties for the persistence parameter ρ do not significantly change, even though it contains both α and γ; evidently, persistence is identifiable as adequately with leverage effects as without them. The discrepancy in bias properties for α and γ between variance targeting and no variance targeting are similar, though noticeably magnified. The intercept parameter ω and feedback parameter β experience bias properties practically indistinguishable from those when there is no leverage effect. The invariance of properties of estimates of ω and ρ to the presence of leverage in the volatility equation makes the bias properties of one-period VaR predictions also invariant to it.

Interestingly, the bias properties of estimates of the unconditional variance σ^2 do not change much either as a result of adding the leverage effect. The median bias tends to shrink a little when no variance targeting is employed and tends to expand a little when variance targeting is employed; in addition, in the former case, the upper tail becomes even more extreme. These tendencies pass over to the bias of long-run VaR predictions (with the change that an upper tail becomes a lower tail).

3.4. ML in Place of QML

If a researcher uses a non-normal distribution to construct estimates and, more importantly, VaR predictions, the bias properties of the latter are expected to significantly improve; see Mittnik and Paolella [22] and Kuester *et al.* [24] for the evidence in favor. The evidence when the (underspecified) Student t distribution is used is shown in Figure 8, while Figures and Tables E.6–E.7 give a more complete picture. Likewise, when the (correct) skewed Student distribution is used, the results are contained in Figure 9 and Figures and Tables F.6–F.7.

Comparison of these to Figure 4 and Figures and Tables A.6–A.7 reveals that for one-period VaR predictions, the expectations are met in median terms, except, possibly, for the case when the tails of the standardized return distribution are so thick that the conditional kurtosis fails to exist. However, perhaps surprisingly, in inter-percentile terms, the 'naive' one-period VaR predictions are less biased when one uses predictions based on an erroneous normal distribution than when one uses ML based on correct (nearly or entirely) correct distributional shapes. This can be explained by additional noise arising from a higher number of parameters that enter into the construction of quantiles under ML. The prediction median bias under variance targeting is still larger, often by a factor of two and more, than when no variance targeting is employed.

As far as the long-run VaR predictions are concerned, these exploit the same quantiles of the normal distribution irrespective of the shape of the return distribution. Naturally, then, the bias properties are not expected to vary much with which estimation method is used. Indeed, this is the case when variance targeting is employed. However, interestingly, when no variance targeting is used, the median bias systematically declines as the tails of the conditional return distribution get thicker, and so, the discrepancy in median bias across the techniques sharply goes up when ML is used (recall that it goes

down and seems to vanish when QML is used). Such a property, of course, is inherited from a similar behavior of the median bias of estimates of the unconditional variance σ^2, which, in turn, is inherited from that for the persistence parameter ρ; this can be verified from Tables and Figures A.4–A.5, E.4–E.5 and F.4–F.5. This intriguing phenomenon, evidently, is explained by a much better identifiability of persistence when the shape of the error distribution, especially leptokurtosity, is taken into account during estimation.

4. Conclusions

Variance targeting is a computationally-attractive detour in the estimation of GARCH models. However, our simulations indicate that in practice, variance targeting may lead to bigger biases in parameter estimates and associated prediction measures than when no targeting is used in the QML procedure. Under variance targeting, the bias properties deteriorate with return heavy tailedness faster, and the bias is a bit more sensitive to return skewness and sample sizes than in the 'regular' QML. This is especially true when a statistic of interest is built on the estimate of unconditional variance, which under variance targeting is a robust measure. Some of these effects further intensify if one uses ML based on a leptokurtic distribution in place of normal QML. Thus, variance targeting should probably be avoided in those cases when computational burden is not prohibitive, which certainly includes the one-dimensional case.

Acknowledgments

Author Contributions

The authors contributed equally to this work.

Conflicts of Interest

The authors declare no conflict of interest.
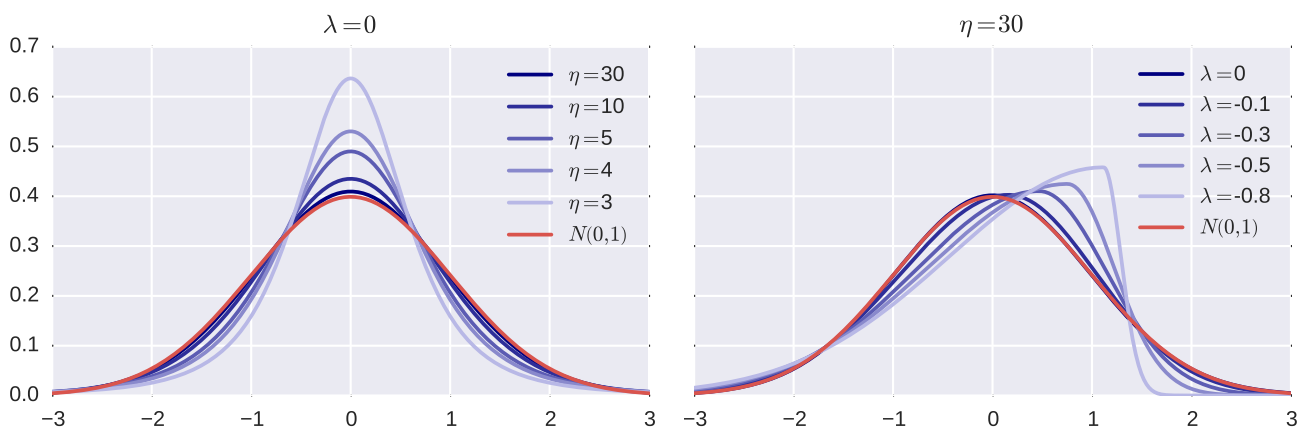
Appendix

A. Illustrations of DGP

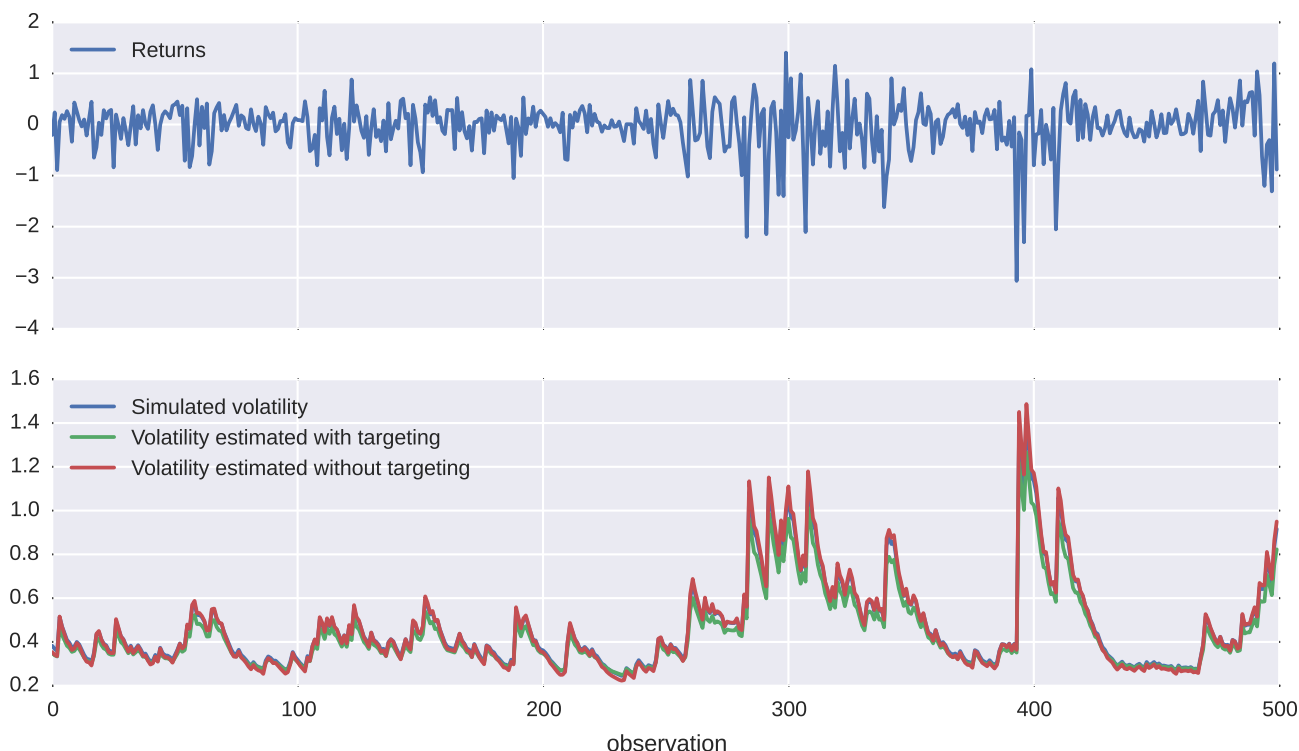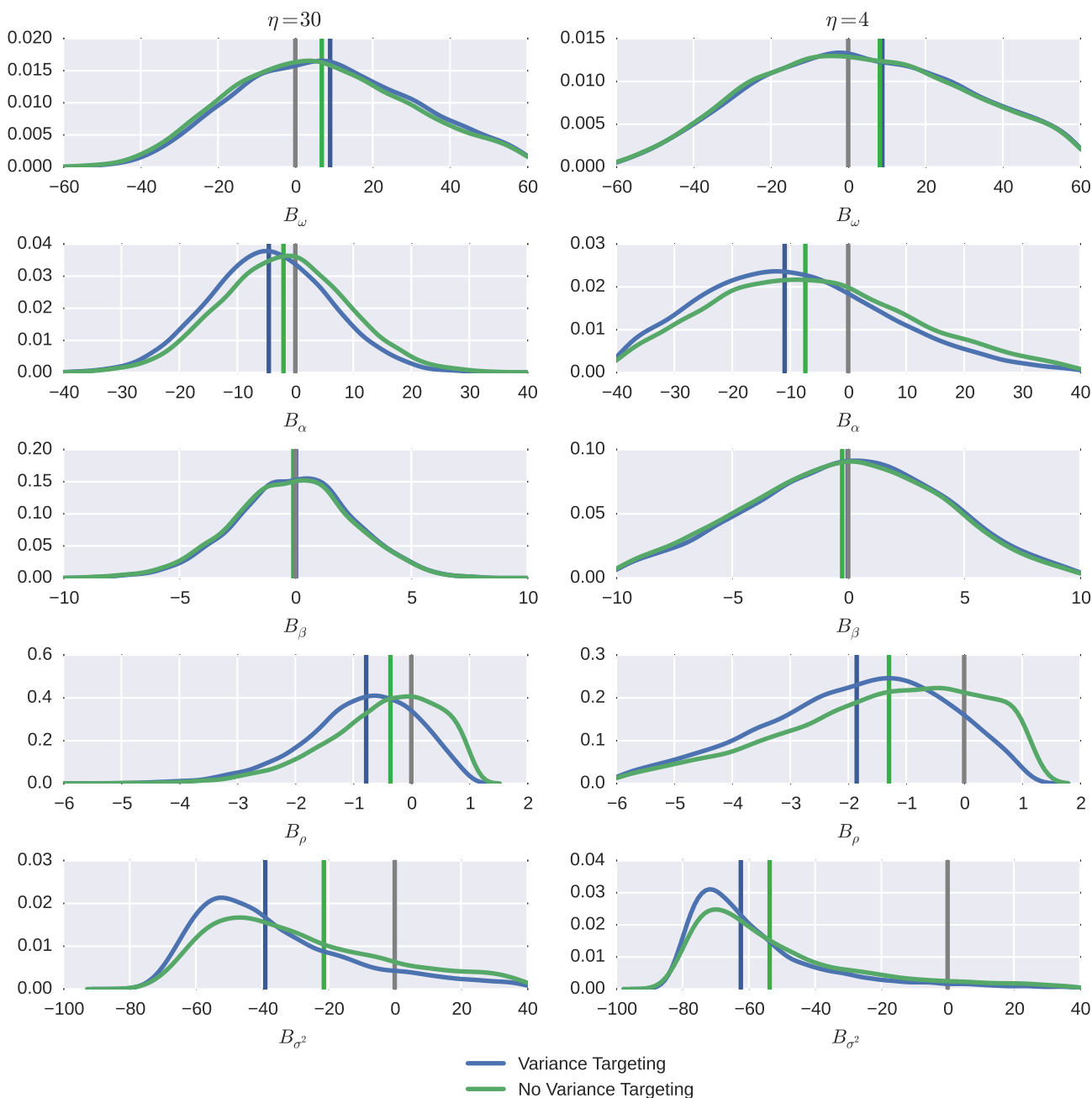**Figure 1.** Skewed Student distribution of Hansen [12].

Figure 2. An example of simulated returns and conditional variance estimates from GARCH(1,1). Data (500 observations) is simulated from GARCH(1,1) model with persistence $\alpha + \beta = 0.99$, feedback $\beta = 0.8$, degrees of freedom $\eta = 5$ and asymmetry $\lambda = -0.3$.
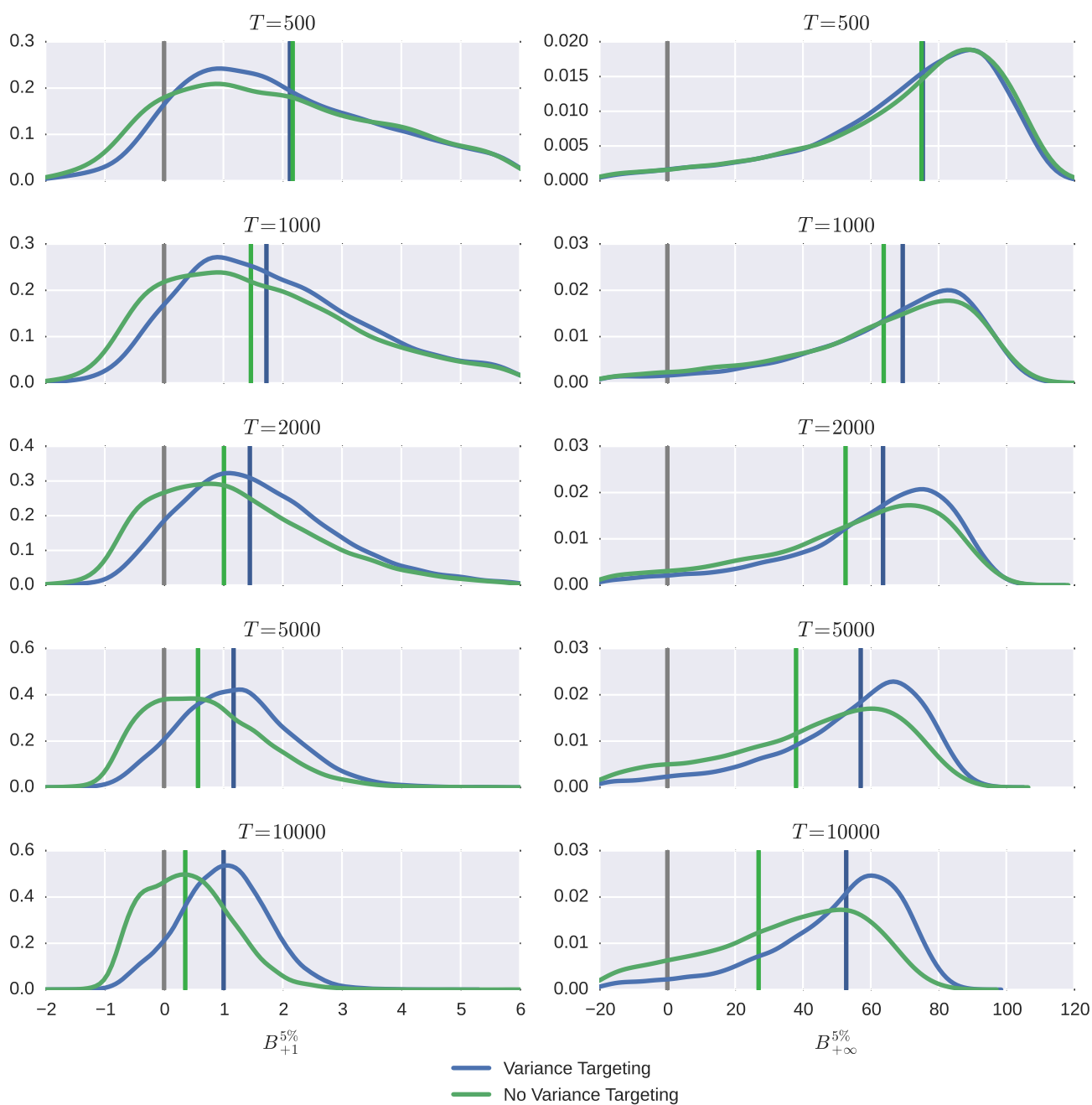
B. Selected bias distributions

Figure 3. Parameter biases for varying degrees of freedom η. Horizontal axis is in percentages of true parameter value. Vertical solid lines are medians and zero (gray). Number of observations is 2000. True value of persistence ρ is 0.99, feedback is 0.8, asymmetry λ is 0, and $\omega = 1 - \rho$.
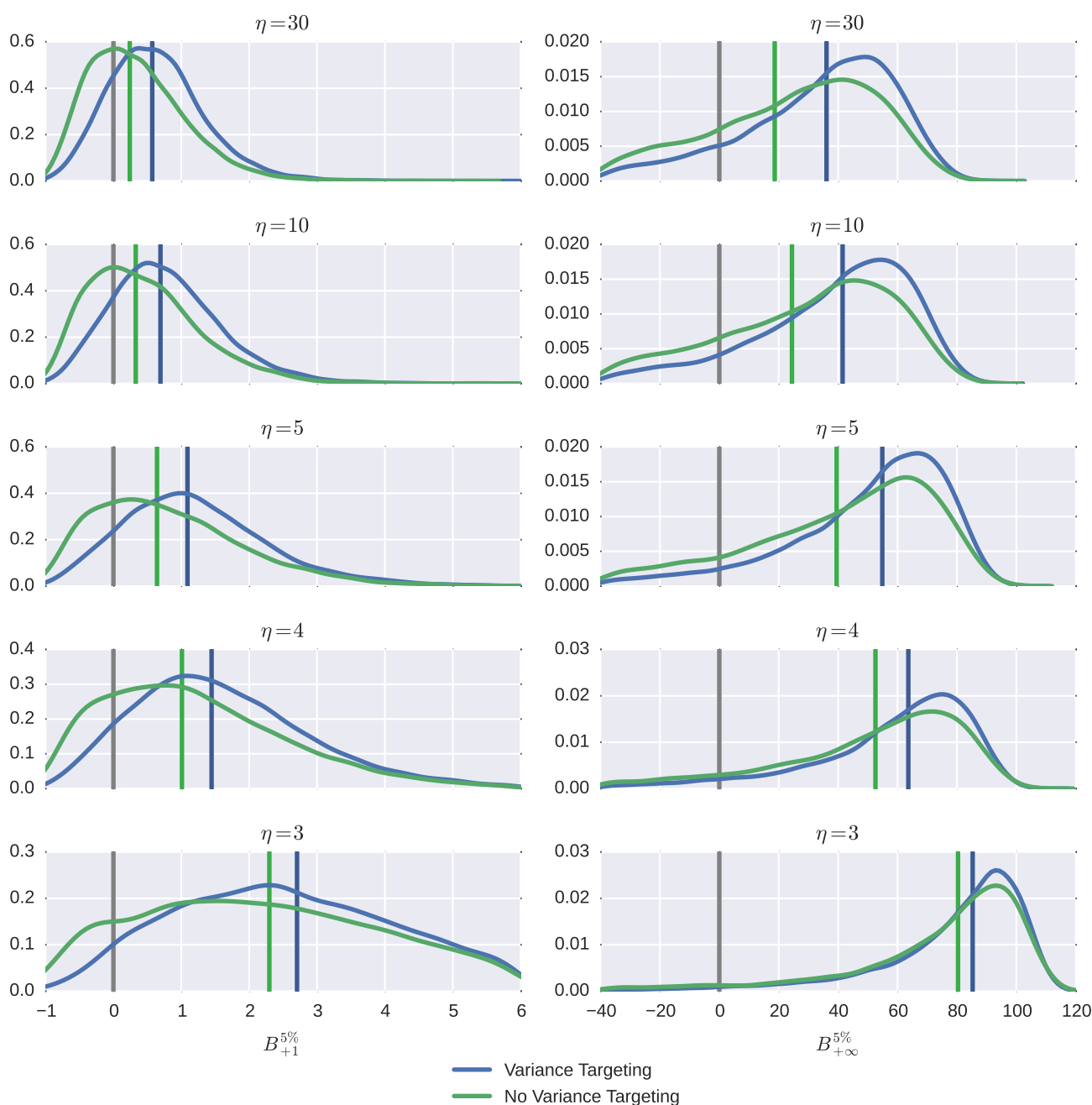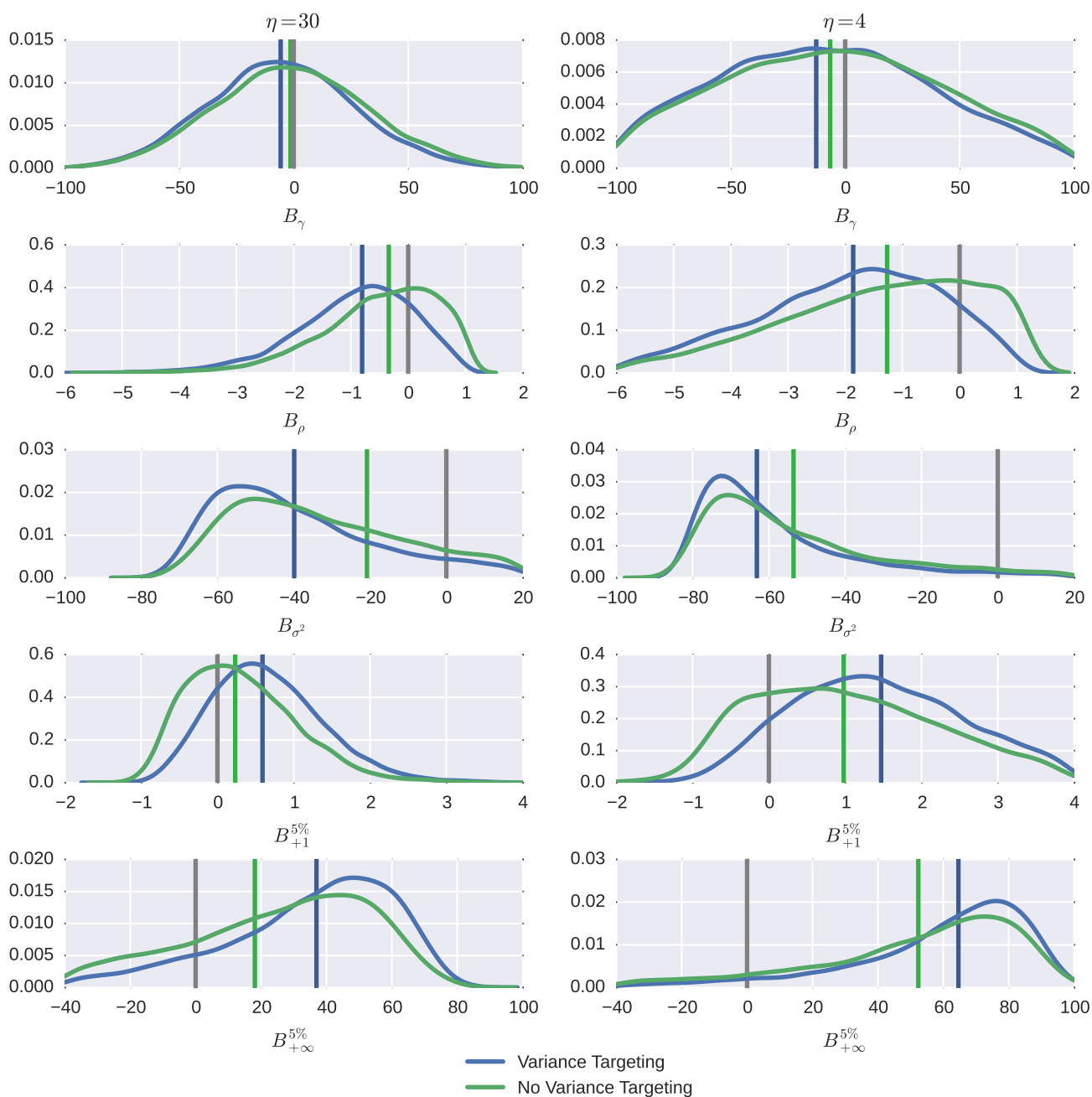
Figure 4. VaR prediction biases for varying degrees of freedom η. Horizontal axis is difference between predicted and true VaRs which correspond to distribution of returns measured in percentages. Vertical solid lines are medians and zero (gray). Number of observations is 2000. True value of persistence ρ is 0.99, feedback is 0.8, asymmetry λ is 0, and $\omega = 1 - \rho$.
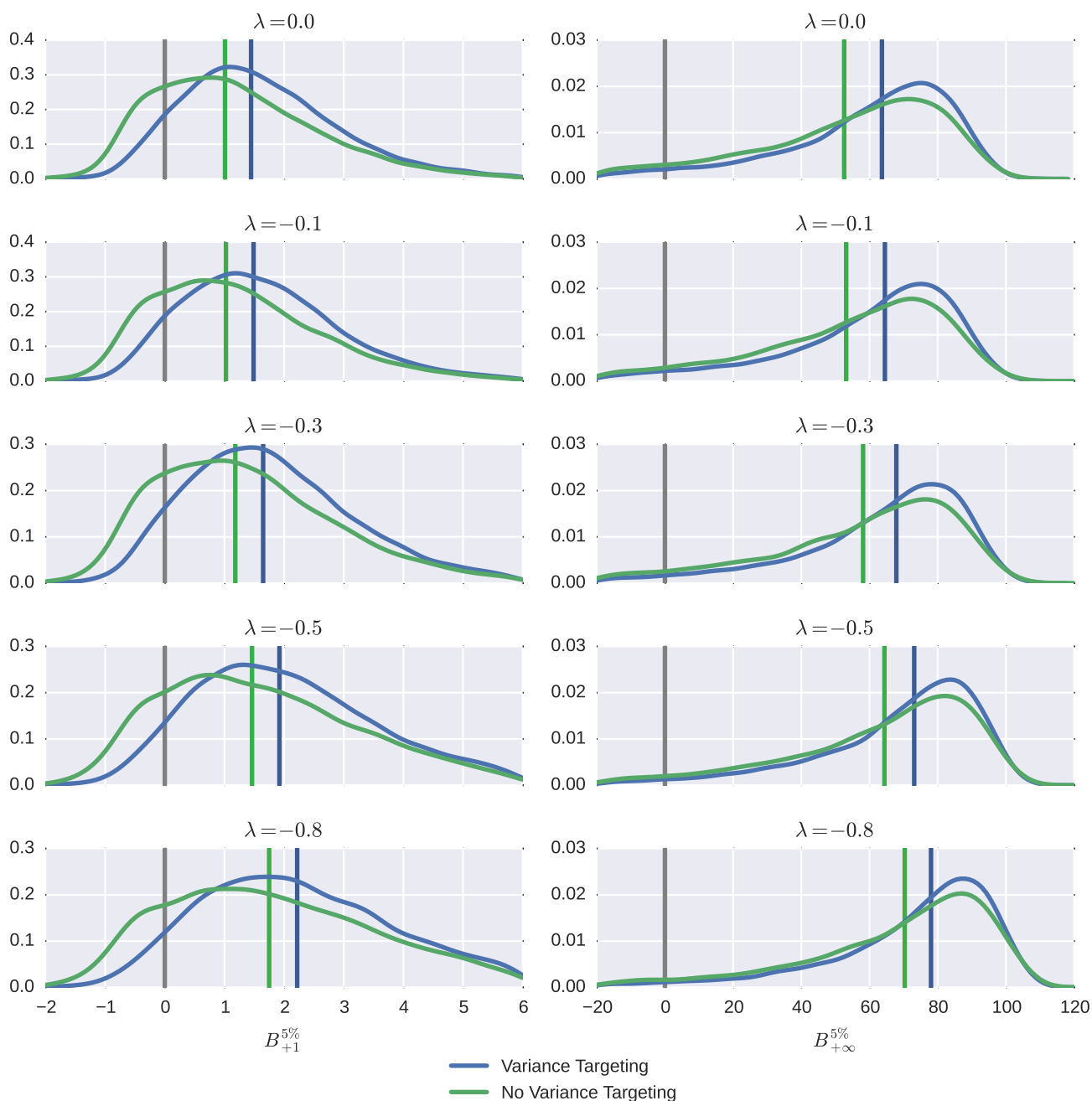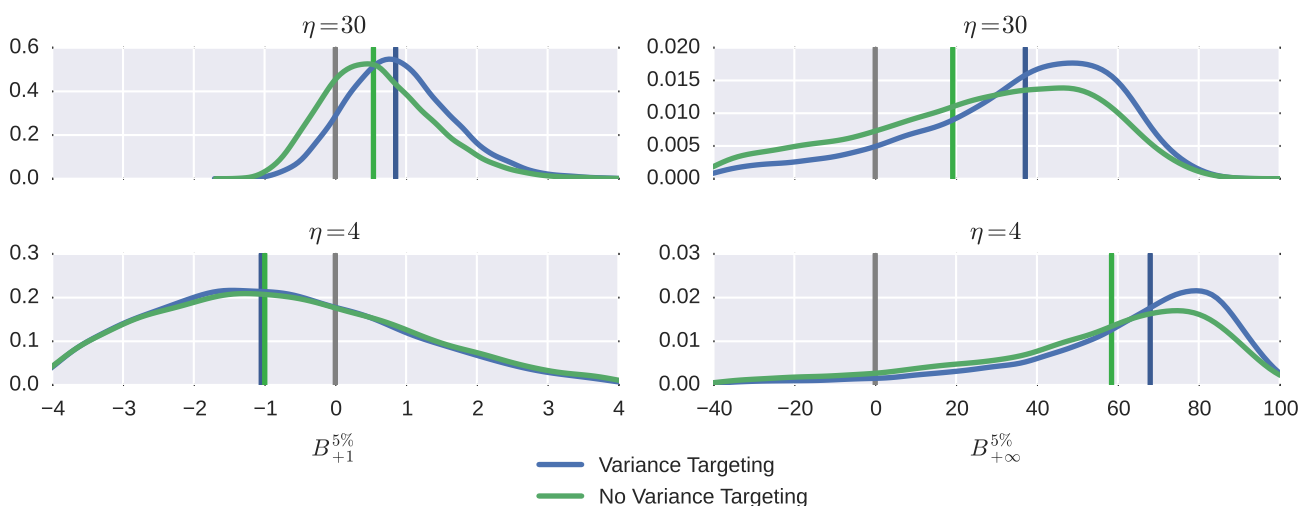
Figure 5. VaR prediction biases for varying skewness λ. Horizontal axis is difference between predicted and true VaRs which correspond to distribution of returns measured in percentages. Vertical solid lines are medians and zero (gray). Number of observations is 2000. True value of persistence ρ is 0.99, degrees of freedom η is 4, and $\omega = 1 - \rho$.
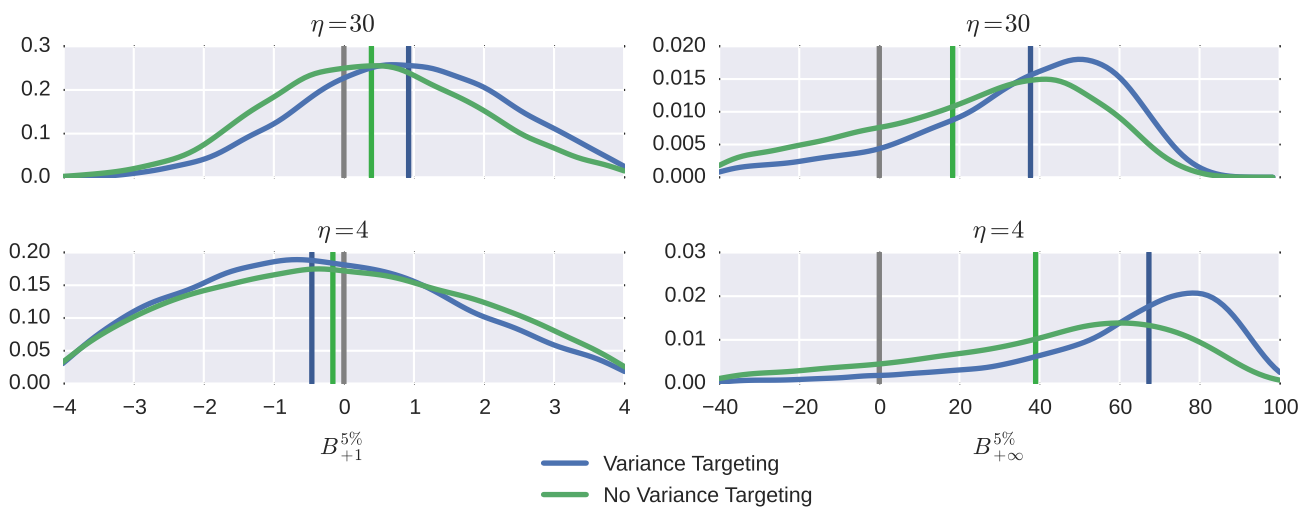
Figure 6. VaR prediction biases for varying sample size T. Horizontal axis is difference between predicted and true VaRs which correspond to distribution of returns measured in percentages. Vertical solid lines are medians and zero (gray). True value of persistence ρ is 0.99, degrees of freedom η is 4, asymmetry λ is 0, and $\omega = 1 - \rho$.

Figure 7. Parameter and VaR prediction biases with GJR model. Horizontal axis is in percentages of true parameter value. Vertical solid lines are medians and zero (gray). Number of observations is 2000. True value of persistence ρ is 0.99, feedback is 0.8, leverage γ is 0.1, asymmetry λ is 0, and $\omega = 1 - \rho$.

Figure 8. VaR prediction biases with ML based on symmetric Student distribution. Horizontal axis is difference between predicted and true VaRs which correspond to distribution of returns measured in percentages. Vertical solid lines are medians and zero (gray). Number of observations is 2000. True value of persistence ρ is 0.99, feedback is 0.8, asymmetry λ is -0.3, and $\omega = 1 - \rho$.

Figure 9. VaR prediction biases with ML based on Skewed Student distribution. Horizontal axis is difference between predicted and true VaRs which correspond to distribution of returns measured in percentages. Vertical solid lines are medians and zero (gray). Number of observations is 2000. True value of persistence ρ is 0.99, feedback is 0.8, asymmetry λ is -0.3, and $\omega = 1 - \rho$.

References

1. Engle, R.F.; Mezrich, J. GARCH for Groups. *Risk* **1996**, *9*, 36–40.
2. Pedersen, R.S.; Rahbek, A. Multivariate Variance Targeting in the BEKK-GARCH Model. *Econom. J.* **2014**, *17*, 24–55.
3. Francq, C.; Horvath, L.; Zakoian, J.M. Variance Targeting Estimation of Multivariate GARCH Models. *J. Financ. Econom.* **2014**, doi:10.1093/jjfinec/nbu030.
4. Francq, C.; Horvath, L.; Zakoian, J.M. Merits and Drawbacks of Variance Targeting in GARCH Models. *J. Financ. Econom.* **2011**, *9*, 619–656.
5. Kristensen, D.; Linton, O. Consistent Standard Errors for Target Variance Approach to GARCH Estimation – Solution. *Econom. Theory* **2004**, *20*, 990–993.
6. Vaynman, I.; Beare, B.K. Stable Limit Theory for the Variance Targeting Estimator. In *Essays in Honor of Peter C. B. Phillips (Advances in Econometrics, Volume 33)*; Chang, Y., Fomby, T.B., Park, J.Y., Eds.; Emerald Group Publishing Limited: Bingley, UK, 2014; pp. 639–672.
7. Engle, R.F. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* **1982**, *50*, 987–1007.
8. Bollerslev, T. Generalized Autoregressive Conditional Heteroskedasticity. *J. Econom.* **1986**, *31*, 307–327.
9. Bollerslev, T. Glossary to ARCH (GARCH). In *Volatility and Time Series Econometrics: Essays in Honor of Robert Engle*; Bollerslev, T., Russell, J., Watson, M., Eds.; Oxford University Press: Oxford, UK, 2010; Chapter 8, pp. 137–163.
10. Bauwens, L.; Laurent, S.; Rombouts, J.V.K. Multivariate GARCH models: A survey. *J. Appl. Econom.* **2006**, *21*, 79–109.
11. Glosten, L.R.; Jagannathan, R.; Runkle, D.E. On the relation between the expected value and the volatility of the nominal excess return on stocks. *J. Financ.* **1993**, *48*, 1779–1801.
12. Hansen, B.E. Autoregressive Conditional Density Estimation. *Int. Econ. Rev.* **1994**, *35*, 705–730.
13. Lumsdaine, R.L. Finite-Sample Properties of the Maximum Likelihood Estimator in GARCH(1,1) and IGARCH(1,1). *J. Bus. Econ. Stat.* **1995**, *13*, 1–10.
14. Linton, O. An Asymptotic Expansion in the GARCH(1,1) Model. *Econom. Theory* **1997**, *13*, 558–581.
15. Anatolyev, S.; Gospodinov, N. *Methods for Estimation and Inference in Modern Econometrics*; CRC Press, Taylor & Francis: Boca Raton, FL, USA, 2011; p. 236.
16. Rodríguez, M.J.; Ruiz, E. Revisiting several popular GARCH models with leverage effect: Differences and similarities. *J. Financ. Econom.* **2012**, *10*, 637–668.
17. McAleer, M. Asymmetry and Leverage in Conditional Volatility Models. *Econometrics* **2014**, *2*, 145–150.
18. Bollerslev, T. A Conditionally Hetroskedastic Time Series Model for Speculative Prices and Rates of Return. *Rev. Econ. Stat.* **1987**, *69*, 542–547.
19. Lee, S.W.; Hansen, B.E. Asymptotic Theory for the Garch(1,1) Quasi-Maximum Likelihood Estimator. *Econom. Theory* **1994**, *10*, 29–52.

20. Jondeau, E.; Rockinger, M. Conditional Volatility, Skewness, and Kurtosis: Existence, Persistence, and Comovements. *J. Econom. Dyn. Control* **2003**, *27*, 1699–1737.

21. Hill, J.B.; Renault, E. *Variance Targeting for Heavy Tailed Time Series*; working paper; University of North Carolina at Chapel Hill: Chapel Hill, NC, USA, 2012; pp. 1–42.

22. Mittnik, S.; Paolella, M.S. Conditional density and value-at-risk prediction of Asian currency exchange rates. *J. Forecast.* **2000**, *19*, 313–333.

23. Bekaert, G.; Engstrom, E.; Ermolov, A. Bad Environments, Good Environments: A Non-Gaussian Asymmetric Volatility Model. *J. Econom.* **2015**, *186*, 258–275.

24. Kuester, K.; Mittnik, S.; Paolella, M.S. Value-at-Risk Prediction: A Comparison of Alternative Strategies. *J. Financ. Econom.* **2006**, *4*, 53–89.

25. Andrews, D.W.K. Exactly Median-Unbiased Estimation of First Order Autoregressive/Unit Root Models. *Econometrica* **1993**, *61*, 139–165.