TI 2017-111/III
Tinbergen Institute Discussion Paper

Finite Sample Optimality of Score-Driven Volatility Models

Francisco (F.) Blasques[1]
André (A.) Lucas[1]
Andries van Vlodrop[1]

1: VU Amsterdam; Tinbergen Institute, The Netherlands

Finite Sample Optimality of Score-Driven Volatility Models[☆]

Francisco Blasques[a], André Lucas[a], Andries van Vlodrop[a]

[a] *Vrije Universiteit Amsterdam and Tinbergen Institute*

Abstract

We study optimality properties in finite samples for time-varying volatility models driven by the score of the predictive likelihood function. Available optimality results for this class of models suffer from two drawbacks. First, they are only asymptotically valid when evaluated at the pseudo-true parameter. Second, they only provide an optimality result 'on average' and do not provide conditions under which such optimality prevails. We show in a finite sample setting that score-driven volatility models have optimality properties when they matter most. Score-driven models perform best when the data is fat-tailed and robustness is important. Moreover, they perform better when filtered volatilities differ most across alternative models, such as in periods of financial distress. These results are confirmed by an empirical application based on U.S. stock returns.

Keywords: volatility models, score-driven dynamics, finite samples, Kullback-Leibler divergence, optimality.

1. Introduction

The class of score-driven models introduced in Creal et al. (2011, 2013) and Harvey (2013) has gained considerable popularity in the recent statistical literature. Score-driven models are typically appreciated for their robustness properties since the models flexibly adapt themselves to the distribution of the innovations. Despite being relatively new, a wide range of applications of score-driven models is already available in the literature; see e.g. Harvey and Luati (2014) for location and scale for fat-tailed data; Creal et al. (2014) for mixed measurement dynamic factor models; Opschoor et al. (2017) for a multivariate dynamic covariance matrix; Blasques et al. (2016) and Catania and Billé (2017) for applications to spatial models; and Janus et al. (2014), Lucas et al. (2014, 2017), Salvatierra and Patton (2015), and Oh and Patton (2017) for recent applications to score-driven copula models.

[☆] Van Vlodrop and Lucas thank the Dutch National Science Foundation (NWO, grant VICI453-09-005) for financial support. Corresponding author: Andries van Vlodrop, SBE/FIN Vrije Universiteit, De Boelelaan 1105, 1081HV Amsterdam. Phone +31 20 598 3739.

Email addresses: `f.blasques@vu.nl` (Francisco Blasques), `a.lucas@vu.nl` (André Lucas), `a.c.van.vlodrop@vu.nl` (Andries van Vlodrop)

Blasques, Koopman, and Lucas (2015) – from now on referred to as BKL2015 – provide one of the theoretical motivations for score-driven models. They highlight the unique properties of score-driven models by showing that these models are optimal in an information theoretic sense. Their results, however, are only valid *asymptotically* when the model is evaluated at the limiting pseudo-true parameter. Furthermore, their supporting Monte Carlo findings only illustrate that score-driven models are optimal *on average*. Therefore, little is known about the finite sample optimality properties of score-driven models.

In this paper, we try to fill this gap by answering two main questions. First, we study how score-driven models perform when analyzed at a (finite-sample) parameter estimate rather than at the pseudo-true parameter. Second, we investigate whether the better performance of a score-driven volatility model depends on the properties of the data and the level of the time-varying parameter itself: *does the model perform better during either periods of high (or low) volatility, and in the presence or absence of outliers and fat tails?* As our benchmark we use the well-known Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model of Engle (1982) and Bollerslev (1986), both with normally and Student's t distributed error terms. We find that the score-driven volatility model not only outperforms the GARCH models asymptotically, but also in finite samples when evaluated at estimated parameter values, despite the considerable variation in these estimates. We also find that the score-driven model performs better precisely when it matters most, i.e., when volatility is high, data are fat-tailed, and robustness is most important. Finally, we find empirically that score-driven volatility models perform best in periods of financial distress. This finding is particularly useful in practice since it indicates that the optimality results apply when we need them most.

Section 2 provides a further motivation and establishes the main concepts. Section 3 studies the finite sample properties of score-driven volatility models evaluated at estimated parameter values. Section 4 characterizes the volatility scenarios for which score-driven models perform best. Section 5 applies the results to U.S. stock returns. Section 6 concludes.

2. Score-driven volatility models and optimality

Let $y_1, y_2, \ldots, y_T$ be a sample of financial returns with true time-varying conditional density $p_t^0(y_t \mid Y_{t-1})$, where $Y_{t-1} = \{y_1, \ldots, y_{t-1}\}$. While the true unknown conditional density p_t^0 may be potentially complex and nonparametric, we attempt to describe the distribution of the data using a parametric model that generates a much smaller and simpler class of conditional densities. In particular, we consider observation-driven models with a time-varying volatility component,

$$y_t = \sigma_t \, \varepsilon_t, \qquad \sigma_{t+1}^2 = h\big(\sigma_t^2, y_t; \boldsymbol{\theta}\big), \tag{1}$$

where $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ is a sequence of independent and identically distributed innovations with zero mean and unit variance, $h(\cdot)$ is some measurable function, and $\boldsymbol{\theta}$ is a vector of parameters. The function h in (1) depends on one lag of σ_t^2 and y_t only, but can easily be generalized to include more lags. The key feature of (1) is that σ_t^2 depends on past values of y_t only, which facilitates estimation of $\boldsymbol{\theta}$ by maximum likelihood (ML) using a standard prediction error decomposition. Let $\hat{\boldsymbol{\theta}}_T$ denote the ML estimator (MLE) based on a sample of size T.

The class of models in (1) includes many well-known models such as the ARCH model of Engle (1982), the GARCH model of Bollerslev (1986), the EGARCH model of Nelson (1991), the TGARCH model of Zakoian (1994), the QGARCH model of Sentana (1995), and many more. Each of these models implies a parametric model density $p(y_t \mid \sigma_t^2; \hat{\boldsymbol{\theta}}_T) = p(y_t \mid Y_{t-1}; \hat{\boldsymbol{\theta}}_T)$ that aims to approximate the true unknown conditional density $p_t^0(y_t \mid Y_{t-1})$. The typical concrete objective is to minimize a discrepancy $D(\,\cdot\,,\,\cdot\,)$ between the two distributions,

$$D\Big(\ p_t^0(\,\cdot\,\mid Y_{t-1})\ ,\ p(\,\cdot\,\mid Y_{t-1}; \hat{\boldsymbol{\theta}}_T)\ \Big). \tag{2}$$

Next, different models with different specifications for h in (1) can be compared in terms of their ability to reduce the discrepancy in (2). BKL2015 showed that if one uses the familiar Kullback-Leibler (KL) divergence as the discrepancy measure in (2), then we can identify a class of observation-driven models that possesses unique fundamental approximation properties that distinguish them from other models. This class of models is called score-driven models.

Score-driven models were introduced by Creal et al. (2011, 2013) and Harvey (2013).[1] They define the conditional density of y_t given Y_{t-1} implicitly through a time-varying parameter f_t that is updated using the derivative of the time t log-likelihood function, also known as *the score*:

$$p(y_t \mid Y_{t-1}; \boldsymbol{\theta}) = p(y_t \mid f_t; \boldsymbol{\theta})\ , \quad f_{t+1} = \omega + \alpha s_t + \beta f_t\ , \tag{3}$$
$$s_t := S_t \cdot \frac{\partial \log p(y_t \mid f_t; \boldsymbol{\theta})}{\partial f},$$

where S_t is a scaling matrix that is known given the data up to time $t-1$. The model can be generalized further by including more lags of f_t and/or s_t, or by scaling the score by an estimate of its local curvature via the matrix S_t. A typical concrete example of this model that we use throughout this paper is based on a setting with $f_t = \sigma_t^2$ and assuming a Student's t distribution with $\tau > 2$ degrees of freedom and unit variance for ε_t in (1). Scaling the score by a factor proportional to the time t inverse conditional Fisher information matrix to account for the curvature, we then obtain a non-linear recurrence $h(\cdot)$ in (1) of the form

$$f_{t+1} = \omega + \alpha\,(w_t\, y_t^2 - f_t) + \beta\, f_t, \qquad w_t = \frac{(1+\tau^{-1})}{1 - 2\tau^{-1} + \tau^{-1}\, y_t^2/f_t}, \tag{4}$$

[1] See www.gasmodel.com for a compendium of papers using score-driven dynamics.

where $\boldsymbol{\theta} = (\omega, \alpha, \beta, \tau)$; see Creal et al. (2013) for further details. To ensure positivity of σ_t^2 at all times, we assume $\sigma_1^2 > 0$, $\omega > 0$, and $\beta > \alpha > 0$. If ε_t is normally distributed, $\tau^{-1} = 0$ and $w_t = 1$, such that (4) reduces to the standard GARCH(1,1) recursion $f_{t+1} = \omega + \alpha\,(y_t^2 - f_t) + \beta\, f_t = \omega + \alpha\, y_t^2 + (\beta - \alpha)\, f_t$. If $\tau^{-1} > 0$, however, ε_t is modeled as a fat-tailed Student's t distribution and the recursion in (4) is fundamentally different from the GARCH(1,1) model with Student's t distributed innovations. In particular, outlying observations y_t^2/f_t are downweighted via w_t, thus lending the score-driven model a robustness feature. Intuitively, the score-driven dynamics account for the fact that ε_t is fat-tailed. A large value of y_t thus need not automatically be attributed to an increase in volatility, but can also be due to the fat-tailed nature of the innovation distribution. This results in a mitigated impact of such observations on the volatility dynamics in the score-driven approach.

Let $\boldsymbol{\theta}_0^*$ denote the MLE's pseudo-true parameter, i.e., the probability limit of the MLE for $\boldsymbol{\theta}$ in (1) under the true (unknown) distribution $p_t^0(y_t \mid Y_{t-1})$. Then BKL2015 derive an optimality property for score-driven models like that in equation (4). They show that under appropiate regularity conditions an observation-driven time-varying parameter model improves the local KL divergence from time t to $t+1$ if and only if the volatility filter behaves like a score-driven filter, i.e.,

$$D\Big(p_t^0(\,\cdot \mid Y_{t-1})\ ,\ p(\,\cdot \mid f_{t+1}; \boldsymbol{\theta}_0^*)\Big) < D\Big(p_t^0(\,\cdot \mid Y_{t-1})\ ,\ p(\,\cdot \mid f_t; \boldsymbol{\theta}_0^*)\Big), \tag{5}$$

if and only if f_t is updated in the direction suggested by the score s_t as in (3), where $D(\cdot,\cdot)$ denotes the local KL divergence.

The theoretical optimality results in BKL2015 are supported by a Monte Carlo study for the score-driven volatility model in (4). These results show that the score-driven Student's t based volatility model outperforms the popular GARCH and t-GARCH specifications not only in a local sense, but even in terms of global expected KL divergence. This suggests that a score-driven approach can lead to a better description of the conditional data density, globally, and over time.

The results reported in BKL2015, however, are subject to two important limitations. First, the model comparisons are performed by evaluating each parametric model at its (asymptotically) best possible parameter, namely, the *pseudo-true* parameter $\boldsymbol{\theta}_0^*$. There is no reason to assume that the results continue to hold if the models are evaluated at the estimated parameter $\hat{\boldsymbol{\theta}}_T$ in a finite sample. To illustrate the potential extent of this problem, consider Figure 1. The figure shows the results of an extensive simulation experiment where we generated $T = 1,000$ data points from a stochastic volatility data generating process (DGP),

$$y_t = \sigma_t\, \varepsilon_t = \sqrt{f_t}\, \varepsilon_t, \qquad \log f_{t+1} = a + b \log f_t + \upsilon_t, \tag{6}$$

where $\varepsilon_t \sim t(\tau)$, i.e., ε_t follows a Student's t distribution with unit variance and τ degrees of freedom, and $\upsilon_t \sim \mathrm{N}(0, \sigma_\upsilon^2)$ is serially independent and independent from ε_t. For this stochastic

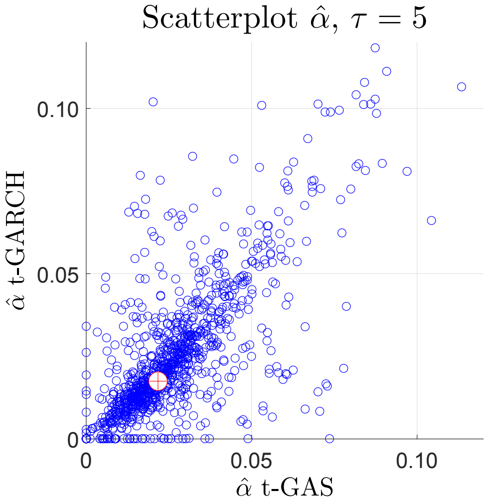

Figure 1: Estimated α parameters of the Student's t score-driven volatility model (t-GAS) against the standard t-GARCH model. Pseudo-true values are plotted as a circle with plus.

volatility DGP we took the same parametrization as in BKL2015, $(a, b, \sigma_\upsilon) = (0.00, 0.98, 0.065)$, and considered $\tau \in \{3, 4, 5, 7, 9\}$. For each generated series, we estimated $\boldsymbol{\theta} = (\omega, \alpha, \beta, \tau)$ by the MLE $\hat{\boldsymbol{\theta}}_T$ for both the score-driven model (t-GAS) and the t-GARCH model. We repeated the process $N = 1,000$ times for each DGP.

Figure 1 presents the results for $\hat{\alpha}_T$ for both models for the case $\tau = 5$.[2] The figure also plots the pseudo-true values α_0^* for both models as the circle with a plus around the point $(0.020, 0.015)$. We clearly see that there is substantial variability in the MLE point estimates obtained from a t-GARCH and a t-GAS model. In particular, the point estimates in finite samples may differ substantially from the asymptotic pseudo-true value as used in BKL2015. Sometimes the finite sample estimate of the t-GARCH model is closer to its pseudo-true value, while at other times the t-GAS estimate is closer. It is therefore difficult to say from the figure whether the asymptotic optimality results carry over to the finite sample setting, either on average or conditionally. In Section 3 we therefore demonstrate that the score-driven model still outperforms its competitors in a KL divergence sense in finite samples.

A second drawback of the simulation results in BKL2015 is that they only show that the score-driven model outperforms its competitors *on average*. It is not clear whether the outperformance is due to the less interesting low-volatility periods, or to the more interesting high-volatility periods. Particularly in periods of distress / high-volatility, we would appreciate a model that fits the true DGP better. The top panel in Figure 2 shows that we can obtain more articulate results on

[2] Results for the other parameters ($\hat{\omega}_T$, $\hat{\beta}_T$, $\hat{\tau}_T$) and for other degrees of freedom (τ) for the DGP reveal a similar message and can be found in Figure 10 in the Appendix.

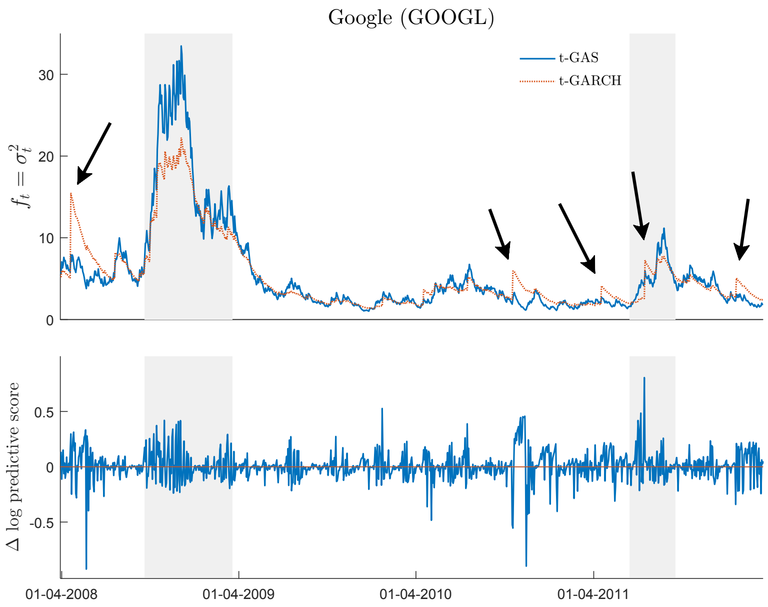

Figure 2: The top panel shows the estimated conditional volatility paths for the t-GAS and t-GARCH models. The arrows show the time points at which outliers impact the t-GARCH paths. The shaded regions show the periods of volatility clustering. The bottom panel shows the difference in in-sample log predictive score for each observation.

when the score-driven volatility model and the GARCH alternatives differ. The figure plots the filtered volatility paths for Google stock returns for the t-GARCH and t-GAS models. The shaded areas highlight the periods of high volatility. The bottom panel shows the difference in in-sample log predictive score between the two models. Positive values indicate a preference for the score-driven t-GAS model. It is clear that the filtered volatilities differ mostly during high-volatility periods. Also, the difference in log predictive scores are on average more positive during highly volatile periods. Both features suggest that the score-driven approach may be most valuable during periods when it matters most. The figure also highlights a number of outliers by means of arrows. These observations appear to distort the filtered volatilities of the t-GARCH model much more than their t-GAS counterparts, again highlighting that the asymptotic optimality properties of the score-driven approach may extend to finite samples, and may be sharpened to include statements conditional on the volatility level. We investigate this in detail in Section 4.

3. Optimality in finite samples

To investigate whether the asymptotic optimality properties of the score-driven approach characterized in BKL2015 carry over to finite samples, we conduct an extensive Monte Carlo study. In this study, we compare the standard GARCH model of Engle (1982) and Bollerslev (1986), the

Table 1: Models used in the Monte Carlo comparison

Model	Observation Eq	Innovation Density	Transition Eq
GARCH	$y_t = \sqrt{\tilde{f}_t}\epsilon_t$	$\epsilon_t \sim N(0,1)$	$\tilde{f}_{t+1} = \omega(1-\beta) + \alpha(y_t^2 - \tilde{f}_t) + \beta\tilde{f}_t$
t-GARCH	$y_t = \sqrt{\tilde{f}_t}\epsilon_t$	$\epsilon_t \sim t(\tilde{\tau})$	$\tilde{f}_{t+1} = \omega(1-\beta) + \alpha(y_t^2 - \tilde{f}_t) + \beta\tilde{f}_t$
t-GAS	$y_t = \sqrt{\tilde{f}_t}\epsilon_t$	$\epsilon_t \sim t(\tilde{\tau})$	$\tilde{f}_{t+1} = \omega(1-\beta) + \alpha s_t + \beta\tilde{f}_t$
log-GAS	$y_t = \exp(\tilde{f}_t/2)\epsilon_t$	$\epsilon_t \sim t(\tilde{\tau})$	$\tilde{f}_{t+1} = \omega(1-\beta) + \alpha s_t + \beta\tilde{f}_t$

fat-tailed t-GARCH model of Bollerslev (1987), the t-GAS model of Equation (4) as introduced by Creal et al. (2011, 2013), and the log-GAS model introduced in Harvey (2013). The latter uses an updating equation for log volatility $f_t = \log(\sigma_t^2)$. The different models and their specifications are summarized in Table 1. For notational purposes, we indicate the (model) filtered volatility by $\tilde{f}_t = \tilde{f}_t(\boldsymbol{\theta})$ and distinguish it from the true DGP volatility f_t. We also use this notational convention for the degrees of freedom parameter, which is $\tilde{\tau}$ for the model, and τ for the DGP. The Monte Carlo set-up is as described for Figure 1 in Section 2, with a stochastic volatility DGP. None of the statistical models thus coincides with the DGP.

Following BKL2015, we are foremost interested in the performance of the different models in terms of their KL divergence. We evaluate the KL divergence between the true conditional density $p_t^0(\cdot|Y_{t-1})$ and the model-implied conditional density $p(y_t|f_t, \hat{\boldsymbol{\theta}}_T)$, where the latter is evaluated at the point estimate $\hat{\boldsymbol{\theta}}_T$,

$$D\Big(p_t^0(\,\cdot \mid Y_{t-1})\ ,\ p(\,\cdot \mid f_t; \hat{\boldsymbol{\theta}})\Big) = \int \Big(\log p_t^0(y \mid Y_{t-1}) - \log p(y \mid f_t; \hat{\boldsymbol{\theta}}_T)\Big)\ \ p_t^0(y_t \mid Y_{t-1})\ \ \mathrm{d}y, \quad (7)$$

where both expectations are taken with respect to the true conditional density p_t^0. Given the stationarity and ergodicity of the DGP, we can approximate the KL divergence using a sample average over a very long path of $H = 500,000$ observations,

$$\frac{1}{H}\sum_{t=1}^{H} \log p_t^0(y_t \mid Y_{t-1}) - \frac{1}{H}\sum_{t=1}^{H} \log p(y_t \mid f_t; \hat{\boldsymbol{\theta}}_T), \quad (8)$$

where the sample (with H observations) used to evaluate the KL divergence is different from the sample (with $T = 1,000$ observations) used to compute the MLE $\hat{\boldsymbol{\theta}}_T$. We repeat this calculation for each of the $N = 1,000$ simulated series of length $T = 1,000$ with corresponding point estimates $\hat{\boldsymbol{\theta}}_T$. As the average involving $\log p_t^0(y_t \mid Y_{t-1})$ in (8) does not depend on the estimated parameter values, we only need to compute it once. We do so using a particle filter based on the DGP specified for the simulation.

Besides the KL divergence we also consider the pathwise average root mean squared error

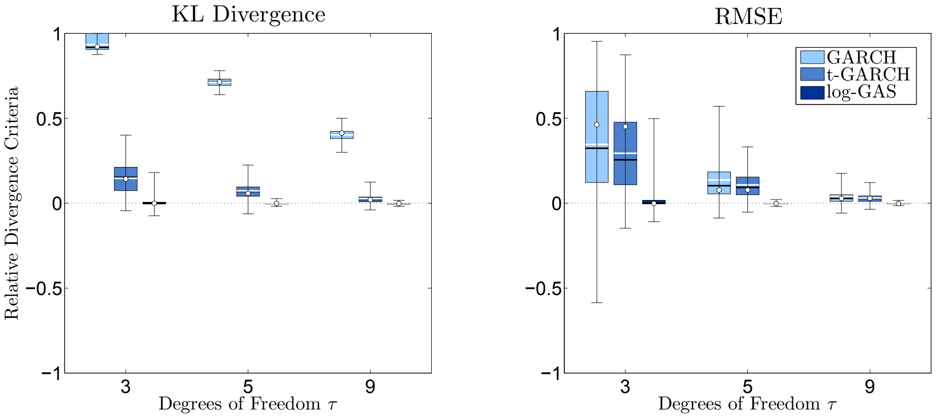

Figure 3: Divergence criteria of the GARCH, t-GARCH, and log-GAS models (from left to right) relative to the t-GAS model. The left panel displays the relative Kullback-Leibler (KL) divergences and the right panel displays the RMSE for the DGP with degrees of freedom parameter $\tau \in \{3, 5, 9\}$. The box plots show the interquartile range (in the box), the median (white bar), the mean (black bar) and the middle 95% (from end to end). The divergence criteria for the pseudo-true parameter values are shown in white circles. The sample consists of $N = 1,000$ Monte Carlo simulations for each $\tau \in \{3, 5, 9\}$. Positive values indicate the t-GAS model performs better.

(RMSE). We compute the RMSE based on the filtered volatility paths as

$$\widehat{\text{RMSE}}(f, \tilde{f}) = \sqrt{\frac{1}{H}\sum_{t=1}^{H}\Big(f_t - \tilde{f}_t(\hat{\boldsymbol{\theta}}_T)\Big)^2}, \tag{9}$$

where f_t denotes the true volatility parameter from the DGP, and $\tilde{f}_t(\hat{\boldsymbol{\theta}}_T)$ denotes the filtered volatility evaluated at the estimated parameter value $\hat{\boldsymbol{\theta}}_T$ for a specific model from Table 1. The RMSE thus measures the model discrepancy in terms of the volatility parameter only, whereas the KL divergence measures the discrepancy in terms of all distributional properties.

The results for this Monte Carlo study are depicted in Figure 3. The left-hand panel reports the distribution of the average KL divergence and RMSE of the GARCH, t-GARCH and log-GAS models relative to the t-GAS model. The relative KL divergence is approximately zero if the t-GAS model and its competitor provide a similar description of the DGP. It is positive if the t-GAS model provides a better description, and negative if the t-GAS model performs worse than its competitor. The same holds for the RMSE (right-hand panel). The maximum relative KL divergence is one, in which case the t-GAS model performs infinitely better. The results are presented as box-and-whisker plots based on the $N = 1,000$ simulated time series. There are three boxes for each of the degrees of freedom considered, namely for $\tau = 3, 5, 9$.

We see that the median relative performance in finite samples coincides with the asymptotic performance (at the pseudo-true value) studied in BKL2015. There is, however, considerable spread around this median (or asymptotic) relative performance. Still, we see that the t-GAS performs considerably better in terms of KL divergence than the GARCH model, particularly if the DGP is

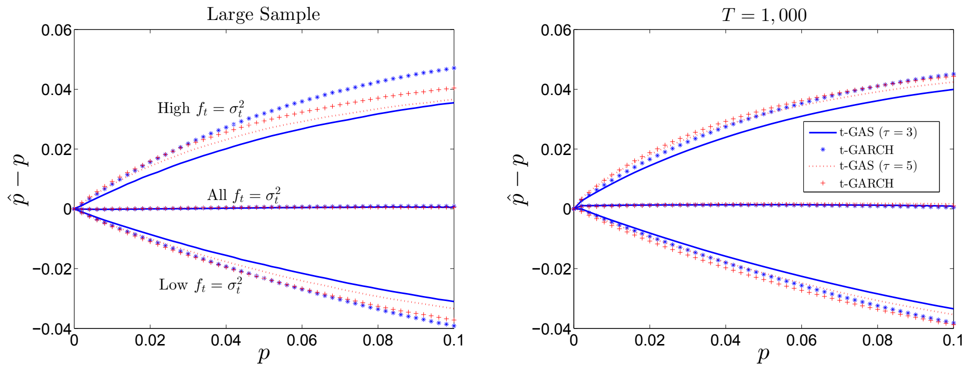

Figure 4: Unconditional deviation of $\mathrm{VaR}_t(p)$-break probabilities $\hat{p} - p$ for the t-GAS and t-GARCH model for nominal VaR levels $p \in (0,\ 0.1]$. The DGP has $\tau = 3$ or $\tau = 5$ degrees of freedom. The VaR is computed under the model density $p(y_t \mid f_t; \boldsymbol{\theta})$, and the VaR-break probability $\hat{p}$ is computed under the true DGP as $\hat{p} = H^{-1} \sum_{t=1}^{H} \mathbf{1}\{y_t < -\mathrm{VaR}_t(p)\}$ for $H = 500,000$. The left panel displays the results for the models evaluated at the pseudo-true parameters ($\boldsymbol{\theta} = \boldsymbol{\theta}_0^*$). The right panel displays the averages over the $N = 1,000$ replications of the deviations evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_T$. Results are provided for all f_t and for high and low f_t, respectively.

fat-tailed ($\tau = 3$). The median relative KL divergence is far above zero, and even accounting for the spread across $N = 1,000$ different samples, we still see that the entire box (inclusive of whiskers) is far above the horizontal axis. If the DGP has lighter tails, the t-GAS model continues to perform better, but the margin of outperformance decreases somewhat. This decrease is to be expected, as for large τ also $\hat{\tau}_T$ is large, and the t-GAS and GARCH specifications coincide more and more. The t-GAS model also has better KL divergence properties than its t-GARCH counterpart. The outperformance is more modest, but still clearly visible. Though not uniform for every sample as in the case of t-GAS versus GARCH, the KL divergence is better for t-GAS versus t-GARCH for the vast majority of samples considered, at least for $\tau = 3$ and $\tau = 5$. Finally, the performance of the t-GAS and log-GAS models are highly similar, such that we can conclude that the precise form of parameterizing the time-varying parameter matters less in this case. All findings are corroborated if we consider the relative RMSEs rather than the relative KL divergences.[3]

Volatility models like that in (1) are typically used to estimate economic or financial risk. Therefore, as a final exercise we also compare the different models in direct economic terms. A well-known risk measure is Value-at-Risk, or VaR, which is defined as a quantile of the conditional distribution $p(y_t \mid f_t; \hat{\boldsymbol{\theta}}_T)$ for profits or returns y_t, such that larger losses than $\mathrm{VaR}_t(p)$ only occur

[3] We note that if one estimates the t-GARCH model parameters through (Gaussian) QML as is typically done in the literature, one obtains the same estimates of ω, α, β as for the GARCH model. As a result, the RMSE in the right panel of Figure 3 for GARCH coincides with the RMSE of the t-GARCH model estimated using Gaussian QML. We see that ML estimation of the t-GARCH model performs slightly better than Gaussian QML estimation in terms of RMSE performance.

with probability (at most) p, i.e.,

$$\mathbb{P}\left[y_t \ < \ -\mathrm{VaR}_t(p) \mid Y_{t-1}\right] = p.$$

We compute the $\mathrm{VaR}_t(p)$ based on the model distribution $p(y_t \mid f_t; \hat{\boldsymbol{\theta}}_T)$ and then evaluate its accuracy under the true DGP. Note that all models considered are mis-specified for the DGP. We approximate the true probability of a loss larger than $\mathrm{VaR}_t(p)$ by the time series average of indicator functions over a long sample H, i.e., $\hat{p} = H^{-1}\sum_{t=1}^{H} \mathbf{1}\{y_t < -\mathrm{VaR}_t(p)\}$. We do this for different nominal probability levels $p \in (0,\ 0.1]$ and two different degrees of freedom for the DGP, $\tau = 3, 5$. Figure 4 presents the results.

The unconditional (All $f_t = \sigma_t^2$) VaR-break probabilities $\hat{p}$ roughly coincide with the nominal levels p for both degrees of fat-tailedness ($\tau = 3, 5$) considered, both for the t-GAS and t-GARCH model, and both at the pseudo-true parameter (large sample) and at the finite sample estimates ($T = 1,000$). This is already interesting: despite the fact that both the score-driven GAS model and the GARCH model are mis-specified for the stochastic volatility DGP, they are able to reliably estimate the tail shape of the distribution by the approximating statistical model. The conditional results for high and low (true) volatility levels, however, reveal a clear difference between the performance of the t-GAS and t-GARCH model.[4] Both for high and low volatility levels the difference $|\hat{p}-p|$ is closer to zero for the t-GAS model, indicating that for extreme volatilty outcomes the score-driven t-GAS specification captures the tails of the distribution more accurately. This result holds at the pseudo-true value, but carries over to the finite sample setting as well. More specifically, the t-GARCH has too high a $\hat{p}$ compared to p vis-à-vis the t-GAS model. This means that the t-GARCH quantile is not far enough out into the tails, implying that the t-GARCH underestimates volatility in case volatility and thus risk is high. Similarly, for low volatility levels the t-GARCH quantile is too far out into the tail and overestimates volatility and thus risk. In both cases, the t-GAS model positions the appropriate risk quantile more accurately. We investigate the origin of these differences in more detail in the next section.

4. Performance over different volatility states

Both Figures 2 and 4 are suggestive that the t-GAS model outperforms the t-GARCH model particularly in the tails of the volatility distribution. This includes the distressed periods during which outperformance matters most in economic terms. In this section we study this phenomenon in more detail to uncover its origins.

[4] High volatility levels refer to values of f_t above the 90^{th} percentile of the unconditional distribution of f_t in (6). Similarly low volatility levels refer to values below the 10^{th} percentile.

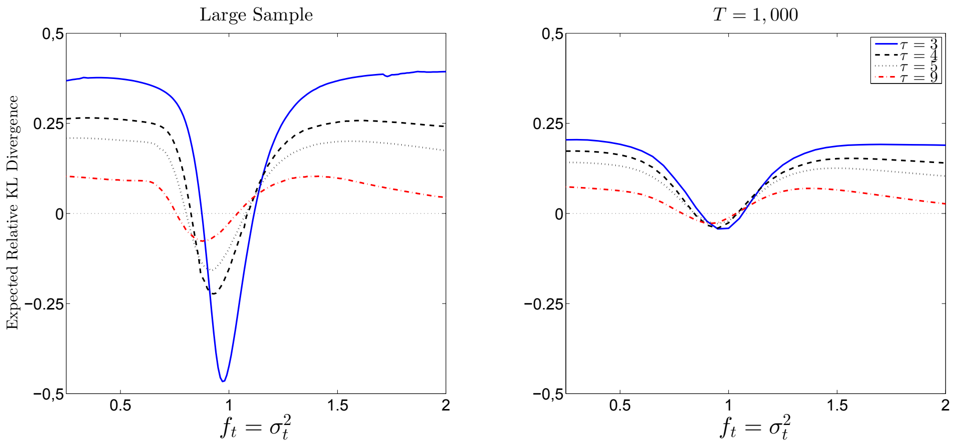

Figure 5: Relative KL divergences between the t-GAS and t-GARCH models conditional on a true variance f_T at time T for the stochastic volatility model DGP. Positive values indicate the t-GAS performs better. The left panel shows the results when evaluated at the pseudo-true parameter values, while the right panel shows the results for the average over the finite sample estimates $\hat{\boldsymbol{\theta}}_T$. Number of replications is $N = 1,000$ for each of the parameter draws $\hat{\boldsymbol{\theta}}_T$.

We first investigate the KL divergence behavior of the different models conditional on the true volatility state. For this, we set up a new set of simulations. As we want to condition on a particular volatility level, we want to simulate data in such a way that the final volatility level is high (or low), as desired. Therefore, for a grid of *final* volatility levels f_T, we simulate backwards a range of volatility paths $f_T, f_{T-1}, \ldots, f_1$ that could have let up to the current volatility level. Given the Gaussian autoregressive structure of the stochastic volatility DGP in (6), the log-volatility process is time reversible and we can simulate past volatility scenarios by the simple time-reversed recursion

$$\log f_{t-1} = a + b \log f_t + \upsilon_t, \qquad \upsilon_t \sim \mathrm{N}(0, \sigma_\upsilon^2),$$

for $(a, b) = (0.00, 0.98)$; see Ōsawa (1988). We use 1,000 (backward) simulated volatility paths $f_T, \ldots, f_1$ to draw paths of y_t as in (6). This allows us to construct estimates of the expected KL divergences at time T conditional on f_T for both the t-GAS and t-GARCH model. For each model and each parameter draw $\hat{\boldsymbol{\theta}}_T$ obtained earlier, the conditional (on f_T) expected relative KL divergence is estimated by the average over the KL divergences at time T for the 1,000 backward simulated volatility paths.

Figure 5 presents the average relative KL divergences of the t-GAS versus the t-GARCH model conditional on the value of f_T. Positive values indicate that the t-GAS model does better. Both panels in Figure 5 illustrate that the t-GAS indeed performs better in the tails of the volatility distribution (high and low values of f_T), both at the pseudo-true parameter (left panel) and at the finite sample estimates (right panel). The fatter tailed the DGP, the more pronounced is the

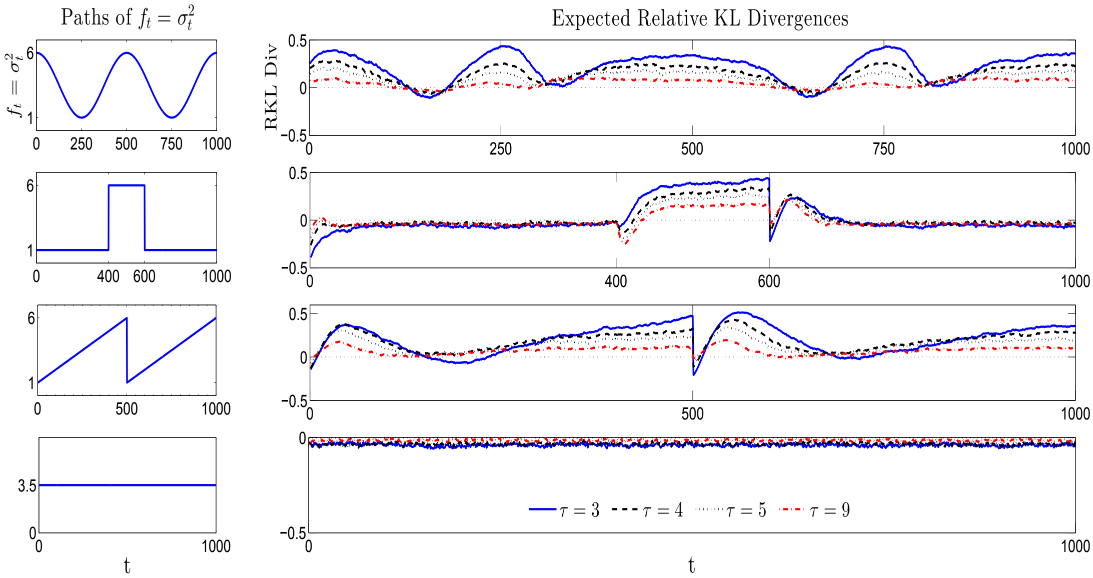

Figure 6: Relative KL divergences between the t-GAS and t-GARCH models for four pre-specified paths of $f_t = \sigma_t^2$. Positive values indicate the t-GAS performs better. The left panels visualize the volatility paths. The right panels show the time series of expected relative KL divergences. For each path we take $\tau \in \{3, 4, 5, 9\}$. The expectations are obtained using $N_f = 2,500$ Monte Carlo simulations.

difference. The pattern is stronger at the pseudo-true parameter, because the finite sample results are averaged across a wide range of finite sample estimates, see Figure 1. To sum up, Figure 5 clearly indicates the better performance of the t-GAS model in periods of high (and low) volatility. This holds both in terms of accuracy of the fit in the tails of f_t (Figure 4) and in terms of relative KL divergence (Figure 5).

The results are further corroborated in Figure 6. Here we plot the relative KL divergence over time for a number of deterministic volatility paths. Similar paths were used in Engle (2002) to show how well a model keeps track of the true time-varying parameters. The figure shows that the t-GAS outperforms the t-GARCH model in periods of high and low volatility. For example, for the sine wave pattern, we see peaks in relative KL divergence around $t = 0, 250, 500, 750, 1000$, which are precisely the times at which volatility is at a peak or a trough. Similar patterns emerge for the pulse and sawtooth case. We see a short underperformance of the t-GAS precisely at the time of the volatility break in those cases, but already quickly after the t-GAS picks up again and outperforms the t-GARCH in terms of relative KL divergence. The results are consistent across different degrees of fat-tailedness (τ) of the DGP.

The better performance of the t-GAS model should be closely related to the t-GAS model stepping more often in the correct direction than the t-GARCH model, or, at least, stepping less aggressively in the wrong direction. Figure 7 visualizes that this is indeed the case, particularly

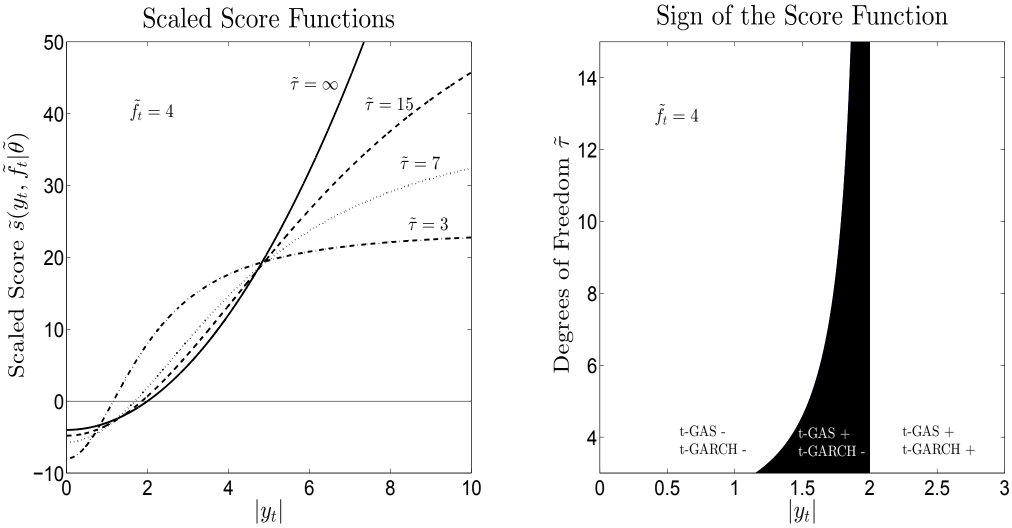

Figure 7: Illustration of the differences in the scaled score functions of the t-GAS and t-GARCH models. The left panel displays the shape of the scaled score functions for different values of $\tilde{\tau}$, a fixed $\tilde{f}_t = \tilde{\sigma}_t^2 = 4$, and an incoming $|y_t|$. The right panel shows the sign regions of these scaled score functions.

for more fat-tailed data, i.e., for lower values of τ. The left panel plots the scaled score steps of the t-GAS model for a range of estimated degrees of freedom $\tilde{\tau}$. For $\tilde{\tau} \to \infty$, the t-GAS and the t-GARCH model coincide. The left panel shows that the t-GAS model is increasingly robust if the innovations in the model are fatter tailed (small $\tilde{\tau}$). This robustness property is what drives the better performance under fat-tailedness: outliers have a reduced impact, which causes the t-GAS model to step less quickly in the wrong direction.

The right panel in Figure 7 illustrates when the t-GAS and t-GARCH models step in different directions. The white regions indicate where the t-GAS and the t-GARCH models step in the same direction, either by decreasing volatility (left-hand white region) or by increasing it (right-hand white region). The black region in the middle indicates where the two models step in different directions. If $\tilde{\tau}$ is large (vertical axis), the region of $|y_t|$ where the two models take a differently signed step is small. If $\tilde{\tau}$ is small, i.e., if we account for a high degree of fat-tailedness, the region is substantial and explains part of the difference in performance between the two models, as the models step into different directions more often.

The different performance of the two models becomes even clearer if we introduce the concept of local improvements (LI) and the corresponding conditional probability of a local improvement (CPLI).

Definition 1 (LI)**.** *An update from $\tilde{f}_t$ to $\tilde{f}_{t+1}$ is called a Local Improvement (LI) if*

$$D\Big(p_t^0(y_t \mid Y_{t-1})\ ,\ p(y_t \mid \tilde{f}_{t+1}; \boldsymbol{\theta}_0^*)\ ,\ \boldsymbol{Y}\Big) < D\Big(p_t^0(y_t \mid Y_{t-1})\ ,\ p(y_t \mid \tilde{f}_t; \boldsymbol{\theta}_0^*)\ ,\ \boldsymbol{Y}\Big),$$

where

$$D\Big(p_t^0(\,\cdot\mid Y_{t-1})\;,\;p(\,\cdot\mid \tilde{f}_t;\boldsymbol{\theta}_0^*)\;,\;\boldsymbol{Y}\Big)=\int_{\boldsymbol{Y}}\;\Big(\log p_t^0(y\mid Y_{t-1})-\log p(y\mid \tilde{f}_t;\boldsymbol{\theta}_0^*)\Big)\;\;p_t^0(y\mid Y_{t-1})\;\;dy,$$

is the local KL divergence over the set $\boldsymbol{Y}$*, and where* $\boldsymbol{Y}=\{y:|y-y_t|<\delta\}$*,* $\delta\to 0$*.*

Definition 2 (CPLI)**.** *The Conditional Probability of a Local Improvement (CPLI) of a parameter update at time t given* $\tilde{f}_t$ *and* y_t *is defined as*

$$\mathit{CPLI}(\tilde{f}_t,y_t)=\mathbb{P}\left[D\Big(p_t^0(y_t\mid Y_{t-1})\;,\;p(y_t\mid \tilde{f}_{t+1};\hat{\boldsymbol{\theta}}_T)\;,\;\boldsymbol{Y}\Big)<D\Big(p_t^0(y_t\mid Y_{t-1})\;,\;p(y_t\mid \tilde{f}_t;\hat{\boldsymbol{\theta}}_T)\;,\;\boldsymbol{Y}\Big)\;\middle|\;y_t,\tilde{f}_t\right],$$

where $\boldsymbol{Y}=\{y:|y-y_t|<\delta\}$*,* $\delta\to 0$*, and where the probability* $\mathbb{P}[\,\cdot\mid y_t,\tilde{f}_t]$ *is taken with respect to the randomness in* $\hat{\boldsymbol{\theta}}_T$*.*

The concept of LI is intuitive. At time t, conditional on the filtered $\tilde{f}_t$, the incoming y_t contains new information about the true unknown conditional density. Therefore, in order to track this true unknown density by the statistical model, the best we can do is to aim for an increase in the model density at y_t using the update step for the time-varying parameter from $\tilde{f}_t$ to $\tilde{f}_{t+1}$. The LI is evaluated at the pseudo-true parameter $\boldsymbol{\theta}_0^*$. The CPLI can be used to assess the improvement behavior at the finite sample estimate $\hat{\boldsymbol{\theta}}_T$. As $\hat{\boldsymbol{\theta}}_T$ is random, the binary check in Definition 1 is replaced by the probability statement in Definition 2, where the randomness in the estimate $\hat{\boldsymbol{\theta}}_T$ is integrated out.

The top panels of Figure 8 show the LI results for the t-GARCH (left panel) and t-GAS (right panel) models evaluated at their pseudo-true parameters under the stochastic volatility DGP with $\tau=3$. For each pair $(\tilde{f}_t,y_t)$, white indicates that the model has an LI, whereas black indicates that the model does not have an LI. It is clear that the black region is much larger for the t-GARCH than for the t-GAS model. Particularly for very high volatility levels, the region where the t-GARCH model does not attain an LI becomes very wide. The black region for the t-GAS model at high volatility levels, however, remains small. This supports our earlier findings that the t-GAS performs very well in such (economically relevant) cases.

To test whether the results in the top panels carry over to the finite sample setting, the lower two panels in Figure 8 show the results for the CPLIs. The CPLI indicates how often an update delivers an LI in finite samples across many different possible outcomes for $\hat{\boldsymbol{\theta}}_T$. Of course, the concept of a CPLI is complicated by the fact that the distribution of $\hat{\boldsymbol{\theta}}_{ML}$ is not independent of $\tilde{f}_t$. In the Appendix we give a detailed description of our weighting procedure to obtain the CPLI values, accounting for the dependence between $\hat{\boldsymbol{\theta}}_T$ and $\tilde{f}_t$. The bottom panels of Figure 8 present the CPLI results for samples of size $T=1,000$ and a fat-tailed ($\tau=3$) DGP using $N=10,000$ estimates for the t-GARCH and t-GAS parameters $\hat{\boldsymbol{\theta}}_T$. The pattern for the LI is supported by

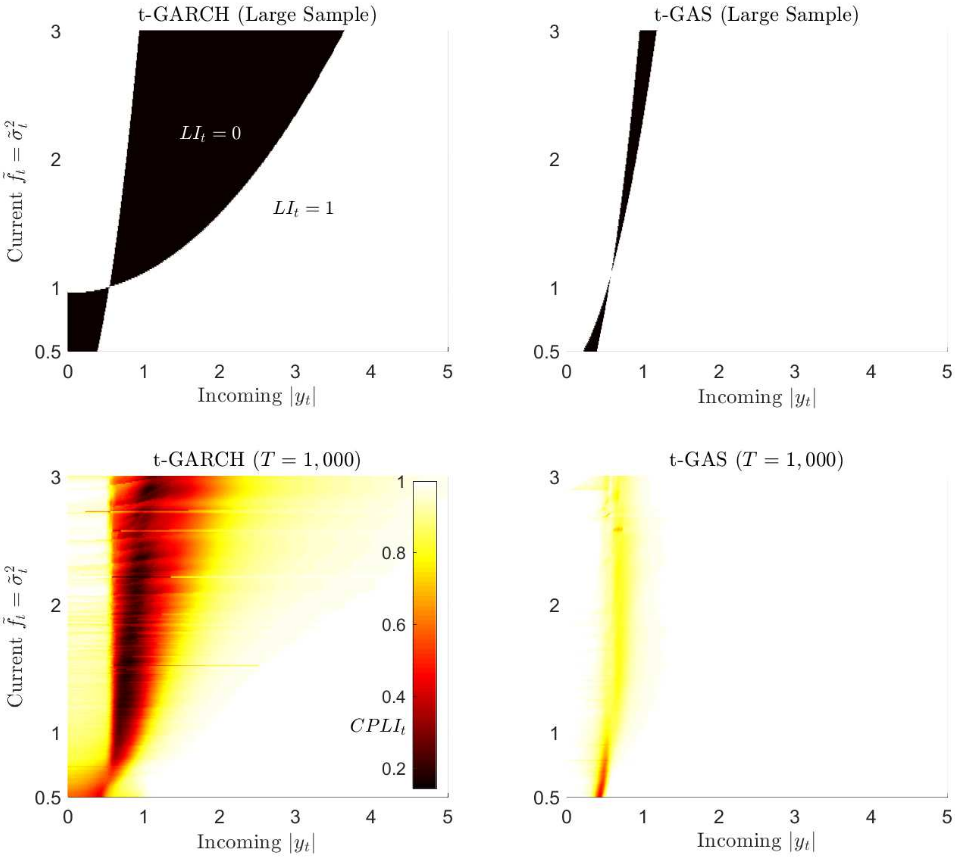

Figure 8: Local Improvement (LI) and Conditional Probability of a Local Improvement (CPLI) results. The top-left panel provides the $(|y_t|, \tilde{f}_t)$ region for which the t-GARCH model attains an LI (white area) or not (black area) when evaluated at the pseudo-true parameter under the stochastic volatility DGP from Section 2 (for $\tau = 3$). The top-right panel provides these regions for the t-GAS model. The bottom panels display the CPLIs for both models as a heat map, lighter colors indicating a higher probability of an LI. The CPLIs are calculated based on a weighted average of the LIs for $N = 10,000$ estimated parameters under the stochastic volatility DGP (for $\tau = 3$).

the CPLI results. The CPLIs of the t-GAS model are considerably higher than those of the t-GARCH model over the entire $(\tilde{f}_t, |y_t|)$ plane, resulting in a much smaller region for the t-GAS where no improvements are obtained. It is also clear that particularly for high volatility levels $\tilde{f}_t$ the difference between both models is most pronounced. This is relevant in practice since it is precisely during periods of financial distress that different filters are more likely to give different estimates of the conditional volatility and thus of risk.

5. Empirical application: Clusters of volatility in stock returns

In this section, we show that the results obtained thus far also apply to empirical data. We consider two large company stocks, namely Goldmann Sachs (GS) and Google (GOOGL). For both stocks we use the total return index, which we obtain from Datastream. We take daily data for the period from 01-04-2008 until 19-03-2012, which provides us two samples consisting of $T = 1,000$ observations each. Figure 9 shows the log-returns for both series. Two periods of high variability for both stocks are evident, namely the financial crisis (2008-2009) and part of European sovereign debt crisis (2011-2012). Volatility appears low in the rest of the sample.

The bottom rows in each of the two panels of Figure 9 show that the differences in filtered volatility paths are most noticeable during periods of high volatility and outliers. This is in line with the results from Section 4. During the financial crisis we see that the filtered path of the t-GAS model indicates higher levels of the conditional variance compared to the t-GARCH model. We also see that the filtered t-GARCH conditional variance can strongly increase with the occurance of an outlier. For instance, for Google there is a peak with exponential decay in the volatility pattern at the beginning of the sample. Similar peaks with subsequent decay patterns are also seen at other times for both stocks for the t-GARCH model. The filtered conditional variances for the t-GAS model on the other hand are much more robust. This robustness property of the t-GAS is in line with the good performance of the model in the simulation setting, particularly for fat-tailed data; see Figures 5–7.

To make a statement about the relative performance of the t-GAS and t-GARCH models for empirical data, we consider the local improvements. The second and third row in each panel of Figure 9 show the indicator time series for LI and the time series of the CPLI. As the kernel weights used to compute the CPLIs in Figure 8 are not available for a single realized empirical time series, we use the asymptotic approximation to the distribution of $\hat{\boldsymbol{\theta}}_T$ and equal weights to compute the CPLI in Figure 9. The diamonds on the horizontal axis indicate the times when there was *no* local improvement (LI) of the model. Overall for the whole sample period, it is clear that the t-GAS model performs better than its t-GARCH counterpart. The majority of the update steps for the t-GAS model yield a local improvement. The CPLIs are correspondingly high, such that this result is not sensitive to the uncertainty in the maximum likelihood estimator $\hat{\boldsymbol{\theta}}_T$. The opposite holds for the t-GARCH model. There are many periods in which the model does not yield a local improvement, and the CPLI values are correspondingly low.

When looking into the properties of the periods during which the t-GAS model does better than the t-GARCH, we confirm the findings of our simulation study in Sections 3 and 4. During periods of high volatility we see that the t-GAS model continues to make local improvements, while the t-GARCH performs poorly. There are also some periods during which the t-GAS model does not

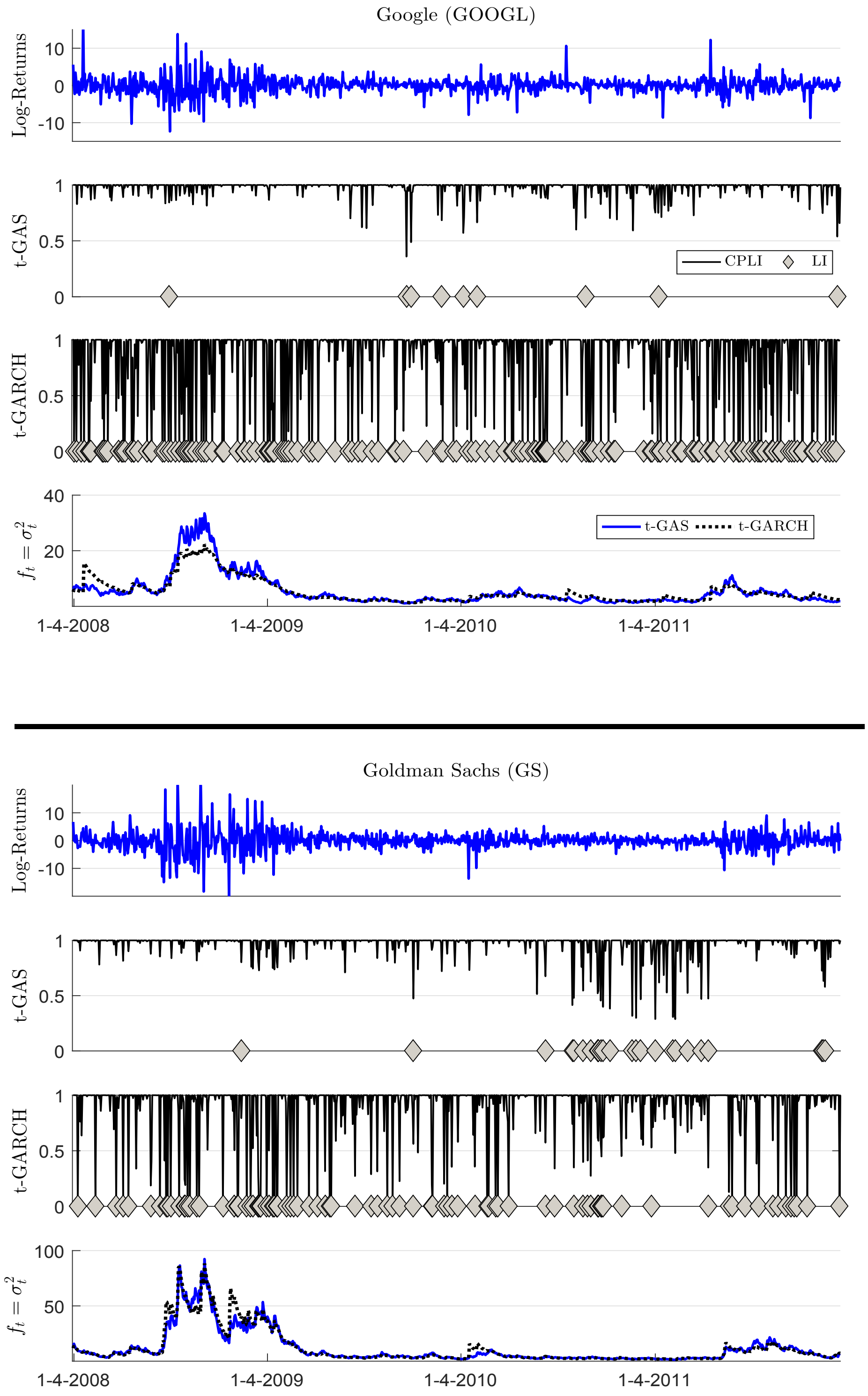

Figure 9: Local performance for the t-GAS and t-GARCH models for log-returns for Google (top part) and Goldman Sachs (bottom part). The sample period is 01-04-2008 until 19-03-2012. The top row in each panel shows the series of log-returns. The second and third row display the time series for the CPLI and the times when the model did *not* have an LI (◇ on the x-axis) for the t-GAS and t-GARCH model, respectively. The fourth row displays the filtered variances $\tilde{f}_t = \tilde{\sigma}_t^2$ for the t-GAS and t-GARCH models.

make local improvements. These periods, however, are particularly during episodes of moderate to low volatility. Such periods are economically less interesting than the periods of financial distress. So in line with the results for simulated data, the empirical results support the conclusion that the t-GAS model outperforms the t-GARCH model when it matters most.

6. Conclusion

The results in this paper further characterize the optimality properties of score-driven models and considerably extend the findings of Blasques et al. (2015). In particular, we have shown that the optimality properties of score-driven volatility models carry over from the asymptotic setting to finite samples. This is remarkable given that finite samples show considerable dispersion of the maximum likelihood estimates due to the strong non-linearity of score-driven models.

Moreover, we found that score-driven models are optimal when it matters most: (a) when the data is fat-tailed and robustness is important, and (b) in periods of financial distress when true volatilities are high and filtered volatilities differ mostly across models. It is precisely then that differences are also large in terms of economic risk measures such as distribution quantiles, or Value-at-Risk. Using new simulations, we were able to pinpoint the origin of the differences. Score-driven models step in the correct direction when updating the time-varying parameter more often than traditional competing models, such as the GARCH model. Defining the concept of the conditional probability of a local improvement (CPLI), we were able to show that the CPLI is substantially larger for the score-driven model compared to the GARCH model. This holds particularly when volatility levels are high.

The results were corroborated for empirical data on U.S. stock returns. Also there, the empirical CPLIs were much higher for the score-driven model than for its GARCH counterpart. Particularly during highly volatile periods such as the 2008 financial crisis, the score-driven model stepped more often into the right direction. Combined, the results point to a further underpinning of the use of score-driven models also in an empirical, finite sample setting.

References

Blasques, F., S. J. Koopman, and A. Lucas (2015). Information-theoretic optimality of observation-driven time series models for continuous responses. *Biometrika 102*(2), 325–343.

Blasques, F., S. J. Koopman, A. Lucas, and J. Schaumburg (2016). Spillover dynamics for systemic risk measurement using spatial financial time series models. *Journal of Econometrics 195*(2), 211–223.

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

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The review of economics and statistics*, 542–547.

Catania, L. and A. G. Billé (2017). Dynamic spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Applied Econometrics*.

Creal, D., S. J. Koopman, and A. Lucas (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business & Economic Statistics 29*(4), 552–563.

Creal, D., S. J. Koopman, and A. Lucas (2013). Generalized autoregressive score models with applications. *Journal of Applied Econometrics 28*(5), 777–795.

Creal, D., B. Schwaab, S. J. Koopman, and A. Lucas (2014). Observation-driven mixed-measurement dynamic factor models with an application to credit risk. *Review of Economics and Statistics 96*(5), 898–915.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics 20*(3), 339–350.

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

Harvey, A. and A. Luati (2014). Filtering with heavy tails. *Journal of the American Statistical Association 109*(507), 1112–1122.

Harvey, A. C. (2013). *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*, Volume 52. Cambridge University Press.

Janus, P., S. J. Koopman, and A. Lucas (2014). Long memory dynamics for multivariate dependence under heavy tails. *Journal of Empirical Finance 29*, 187–206.

Lucas, A., B. Schwaab, and X. Zhang (2014). Conditional euro area sovereign default risk. *Journal of Business & Economic Statistics 32*(2), 271–284.

Lucas, A., B. Schwaab, and X. Zhang (2017). Modeling financial sector joint tail risk in the euro area. *Journal of Applied Econometrics 32*(1), 171–191.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347–370.

Oh, D. H. and A. J. Patton (2017). Modeling dependence in high dimensions with factor copulas. *Journal of Business & Economic Statistics 35*(1), 139–154.

Opschoor, A., P. Janus, A. Lucas, and D. Van Dijk (2017). New heavy models for fat-tailed realized covariances and returns. *Journal of Business & Economic Statistics*, 1–15.

Ōsawa, H. (1988). Reversibility of first-order autoregressive processes. *Stochastic processes and their applications 28*(1), 61–69.

Salvatierra, I. D. L. and A. J. Patton (2015). Dynamic copula models and high frequency data. *Journal of Empirical Finance 30*, 120–135.

Sentana, E. (1995). Quadratic arch models. *The Review of Economic Studies 62*(4), 639–661.

Zakoian, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and control 18*(5), 931–955.

Appendix

Computing the Conditional Probability of a Local Improvement (CPLI)

We express the CPLI as the probability of obtaining a Local Improvement (LI) conditional on y_t and $\tilde{f}_t$, i.e., the probability that $LI = LI(y_t, \tilde{f}_t, \hat{\boldsymbol{\theta}}_T) = 1$. We have

$$\begin{aligned} CPLI_t = \mathbb{P}[LI_t \mid y_t, \tilde{f}_t] &= \int LI(y_t, \tilde{f}_t, \hat{\boldsymbol{\theta}}_T)\ p(\hat{\boldsymbol{\theta}}_T \mid y_t, \tilde{f}_t) \mathrm{d}\hat{\boldsymbol{\theta}}_T \\ &= \int LI(y_t, \tilde{f}_t, \hat{\boldsymbol{\theta}}_T)\ \frac{p(\tilde{f}_t, y_t \mid \hat{\boldsymbol{\theta}}_T)\ p(\hat{\boldsymbol{\theta}}_T)}{\int p(\tilde{f}_t, y_t \mid \hat{\boldsymbol{\theta}}_T)\ p(\hat{\boldsymbol{\theta}}_T) \mathrm{d}\hat{\boldsymbol{\theta}}_T}\ \mathrm{d}\hat{\boldsymbol{\theta}}_T \\ &\approx \sum_{i=1}^{M} w_t^{(i)}\ LI(y_t, \tilde{f}_t, \hat{\boldsymbol{\theta}}_T^{(i)}), \\ w_t^{(i)} &= \frac{p(\tilde{f}_t, y_t \mid \hat{\boldsymbol{\theta}}_T^{(i)})}{\sum_{j=1}^{M} p(\tilde{f}_t, y_t \mid \hat{\boldsymbol{\theta}}_T^{(j)})}, \end{aligned} \tag{10}$$

where $\hat{\boldsymbol{\theta}}_T^{(i)}$, $i = 1, ..., M$, denotes a set of draws from the density $p(\hat{\boldsymbol{\theta}}_T)$. In our computations we use $M = 10,000$. To estimate the density $p(\tilde{f}_t, y_t \mid \hat{\boldsymbol{\theta}}_T^{(i)})$ for each $\hat{\boldsymbol{\theta}}_T^{(i)}$, we simulate a long series $\{y_t\}$ from the DGP and compute the corresponding series $\{\tilde{f}_t(\hat{\boldsymbol{\theta}}_T^{(i)})\}$. We use these two series as inputs for a standard bivariate kernel density estimate, which is plugged back into (10). The kernel density estimates are (nearly) singular for low values of $\hat{\alpha}$, i.e., in cases where the filtered variance $\tilde{f}_t$ is estimated to be (almost) constant. These cases cause numerical instabilities and are, moreover, not interesting for the concept of the CPLI. We therefore exclude cases with $\hat{\alpha} < 0.0001$ when computing the lower panel in Figure 8.

Parameter estimates t-GAS vs t-GARCH

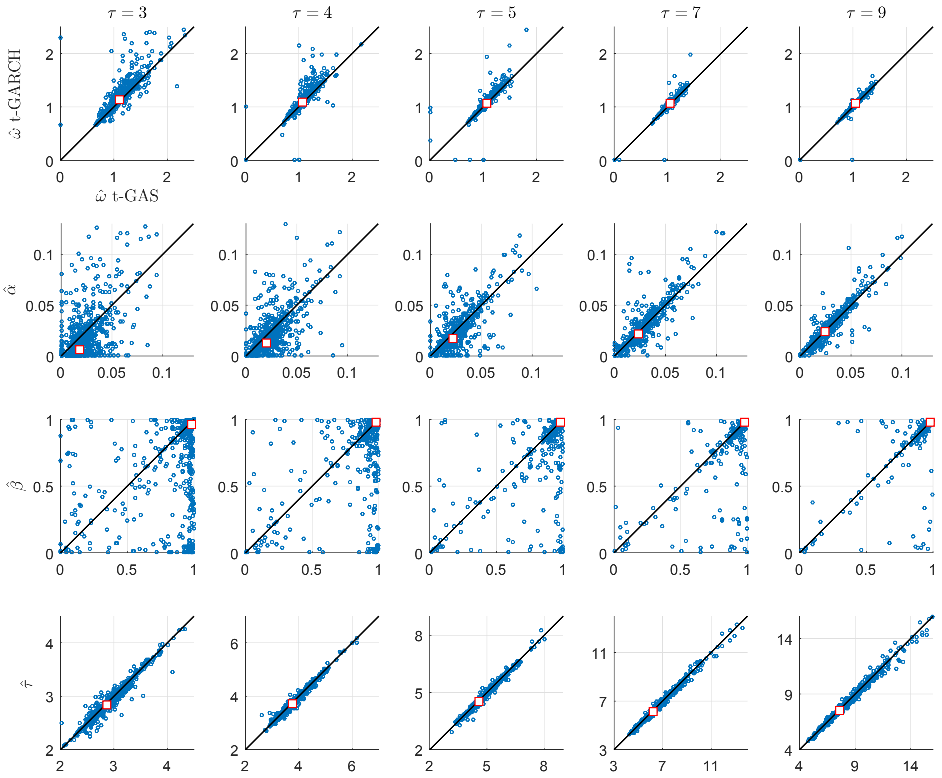

Figure 10: The figure shows the estimated parameters of the t-GAS against the t-GARCH model under the SV model in Equation (6). The rows show the different parameters, while the columns show the various degrees of freedom τ in the DGP. The pseudo-true parameters are shown as red squares.