Semiparametric Score Driven Volatility Models*

Francisco Blasques^(a), Jiangyu Ji^(a), André Lucas^(a)

^(a) VU University Amsterdam and Tinbergen Institute

Abstract

A new semiparametric observation-driven volatility model is proposed. In contrast to the standard semiparametric generalized autoregressive conditional heteroskedasticity (GARCH) model, the form of the error density has a direct influence on both the semiparametric likelihood and the volatility dynamics. The estimator is shown to consistently estimate the conditional pseudo true parameters of the model. Simulation-based evidence and an empirical application to stock return data confirm that the new statistical model realizes substantial improvements compared to GARCH type models and quasi-maximum likelihood estimation if errors are fat-tailed and possibly skewed.

Keywords: time-varying volatility, generalized autoregressive score model, observation driven models, kernel density estimation.

1 Introduction

Models for time varying volatility are very popular in statistics and economics as parsimonious descriptions of a range of empirical stylized facts. Starting with the seminal (G)ARCH papers by Engle (1982) and Bollerslev (1986), many variations and extensions of the GARCH class of models have been proposed in the literature, such as the GARCH-M model of Engle et al. (1987), the EGARCH model of Nelson (1991), the GJR-GARCH model of Glosten et al. (1993), and the Threshold GARCH (TGARCH) model by Zakoian (1994). For a recent textbook treatment of GARCH models, see for example Francq and Zakoian (2010). Interestingly, Quasi

*Corresponding author: Francisco Blasques, VU University Amsterdam, FEWEB/FIN, de Boelelaan 1105, 1081 HV Amsterdam, Netherlands, phone: +31 205 985 621, email: f.blasques@vu.nl.

Maximum Likelihood Estimation (QMLE) of GARCH models based on the (possibly incorrect) assumption of conditional normality for the error terms still yields $\sqrt{T}$ -consistent estimators in many cases, with T denoting the sample size; see Weiss (1986), Lee and Hansen (1994), and Lumsdaine (1996). This is important, as the conditional normality assumption is often rejected by empirical data. The statistical efficiency loss, however, can be considerable. Therefore, many empirical studies assume non-Gaussian conditional distributions, such as for example a (skewed) Student's t distribution, see Baillie and Bollerslev (1989) and Bauwens and Laurent (2005).

Semiparametric GARCH models have been proposed to avoid the potential efficiency loss and bias problems caused by assuming an incorrect parametric family of conditional distributions. Engle and Gonzalez-Rivera (1991) show that more efficient parameter estimates can be obtained by estimating the error density non-parametrically, though some efficiency loss remains. Drost and Klaassen (1997) and Sun and Stengos (2006) develop kernel-based estimators and establish the semiparametric efficiency bounds for parameter estimation. They characterize the conditions under which one can adaptively estimate a subset of the model's parameters, i.e., one can achieve the same asymptotic efficiency as if the true error density were known.

Our main contribution is to provide a new semiparametric model for time-varying volatility in which the form of the error distribution is directly linked to the volatility dynamics using the generalized autoregressive score (GAS) dynamics of Creal, Koopman, and Lucas (2011, 2013) and Harvey (2013). By doing so, we extend the class of GAS models to the semiparametric context. So far, all GAS applications have been fully parametric. We find that if the true data generating process is conditionally fat-tailed and skewed, the semiparametric model regains a substantial part of the efficiency loss due to the use of an incorrectly assumed Gaussian or Student's t distribution.

In a fully parametric context, the GAS framework of Creal et al. (2013) has been applied successfully to a range of empirical applications, including dynamic location models, dynamic correlation and copula and dynamic count data models; see for example Creal et al. (2011), Oh and Patton (2013), Harvey and Luati (2014), Lucas et al. (2014), Andres (2014), and Creal et al. (2014). Parameter estimation

for GAS models is straightforward, since the likelihood function can be specified in closed form. The asymptotic properties of GAS processes and of the maximum likelihood estimator for GAS models were studied in Blasques et al. (2014a,b). Our current model estimates the conditional density of the innovations non-parametrically, and uses the estimate to construct both the semiparametric maximum likelihood estimator, as well as the dynamics of the time varying volatility using the score of the density estimate. The information theoretic optimality properties of the score of the conditional density as a driver for the time varying parameters were shown in Blasques et al. (2015).

Parameter estimation for our model is straightforward and carried out in two steps. First, we estimate the parameters by assuming a normal or Student's t conditional quasi density for the error terms. The assumed quasi density also influences the volatility dynamics through the assumed GAS dynamics. From this we obtain the standardized residuals and the kernel density estimate of the error density. Second, the kernel density estimate is used to re-estimate the model's static parameters, resulting in the Semiparametric Maximum Likelihood Estimator. Again, the nonparametric estimate of the density provides new dynamics for the dynamic volatility parameter. This process can be iterated further until the semiparametric likelihood estimate does not increase further. We prove the theoretical properties of this approach. It is useful at this stage to note that the semiparametric nature of our model relates to the nonparametric estimation of the conditional error density as in Drost and Klaassen (1997) and Sun and Stengos (2006), and not to the nonparametric estimation of the volatility dynamics as in Pagan and Schwert (1990), Linton and Mammen (2005) and Yang (2006). The latter would be interesting in its own right, but is not pursued here.

The remainder of this paper is organized as follows. Section 2 introduces the model. Section 3 lays down the basic theoretical conditions for maximum likelihood estimation of this model. Section 4 provides Monte Carlo evidence of the new model's performance. Section 5 presents an empirical application to daily euro-dollar exchange rate data. Section 6 concludes.

2 The semiparametric volatility model

Consider the GARCH(1,1) model

$$y_t = \mu + \xi_t = \mu + \tilde{f}_t^{1/2} \varepsilon_t, \quad \varepsilon_t \sim q(\varepsilon_t), \quad (1)$$

$$\tilde{f}_{t+1} = \tilde{\omega} \cdot (1 - \tilde{\alpha} - \tilde{\beta}) + \tilde{\alpha} \xi_t^2 + \tilde{\beta} \tilde{f}_t, \quad (2)$$

with $y_t \in \mathbb{R}$, $t = 1, \dots, T$, a univariate time series with conditional mean $\mu \in \mathbb{R}$, $q(\cdot)$ the density of the standardized error term ε_t , $\tilde{f}_t \in \mathbb{R}^+$ the conditional variance, and $\tilde{\omega}, \tilde{\alpha}, \tilde{\beta} > 0$ static parameters satisfying $\tilde{\alpha} + \tilde{\beta} < 1$. The standardized density $q(\cdot)$ does not depend on $\tilde{f}_t$. Note that we can easily replace the conditional mean μ by a non-constant conditional mean μ_t involving exogenous regressors or autoregressive moving average components. Moreover, we can include lags of ξ_t and $\tilde{f}_t$ in (2).

If the distribution $q(\cdot)$ of ε_t is unknown, one can still proceed by assuming the normal distribution as a quasi density and obtain consistent estimates, see Francq and Zakoïan (2010). As mentioned in the introduction, however, the efficiency loss of this approach can be substantial. If $q(\cdot)$ itself needs to be estimated (nonparametrically), model (1)–(2) is called a semiparametric GARCH(1,1) model. Estimation is usually performed in two steps. First, one estimates the static parameters $(\mu, \tilde{\omega}, \tilde{\alpha}, \tilde{\beta})$ using a Quasi Maximum Likelihood estimator (QMLE) based on the standard normal distribution. Using the standardized residuals $\hat{\tilde{f}}_t^{-1/2} \cdot (y_t - \hat{\mu})$ and standard nonparametric density estimation techniques, one then obtains an estimate $\hat{q}(\cdot)$ of the standardized error density. The estimated density is used to estimate the parameters $(\mu, \tilde{\omega}, \tilde{\alpha}, \tilde{\beta})$ in a second step by semiparametric ML, assuming the estimated density $\hat{q}(\cdot)$ is the true density. Engle and Gonzalez-Rivera (1991) use the discrete maximum penalized likelihood estimator (DMPLE) of Tapia and Thompson (1978) for density estimation, followed by a BHHH algorithm to maximize the semiparametric likelihood. Drost and Klaassen (1997) and Sun and Stengos (2006) use standard kernel density methods to estimate the error density and develop a Newton-Raphson algorithm to maximize the semiparametric likelihood. Their estimator achieves the same asymptotic efficiency for $(\tilde{\alpha}, \tilde{\beta})$ as the maximum likelihood estimator (MLE).

Our new semiparametric GAS(1,1) model is given by

$$y_t = \mu + \xi_t = \mu + \exp(f_t/2)\varepsilon_t, \quad \varepsilon_t \sim q(\varepsilon_t), \quad (1')$$

$$f_{t+1} = \omega \cdot (1 - \beta) + \alpha s_t + \beta f_t, \quad (2')$$

$$s_t = \frac{\partial \ln p(y_t|f_t)}{\partial f_t} = -\frac{1}{2} (e^{-f_t/2} \xi_t \cdot \nabla q(e^{-f_t/2} \xi_t) - 1), \quad (3')$$

where $p(y_t|f_t)$ is the conditional observation density given the value of the filtered time-varying parameter f_t , when it is initialized at a fixed point $f_1 = \bar{f}_1 \in \mathbb{R}$, and $\omega, \alpha \in \mathbb{R}$ and $|\beta| < 1$ are static parameters. Furthermore, $\nabla q(\varepsilon)$ denotes the derivative $\nabla q(\varepsilon) = \partial \ln q(\varepsilon) / \partial \varepsilon$. Further lags of s_t and f_t as well as terms capturing asymmetric leverage type reactions can easily be included in (2').

Equations (2) and (2') have two main differences. First, equation (2') uses the score s_t rather than the squared lagged residuals ξ_t^2 as in equation (2). Second, equation (2') parameterizes the model in terms of the log-variance f_t rather than the variance $\tilde{f}_t$, as in equation (2). Arguments for the use of the score are discussed in Creal et al. (2013), Harvey (2013), and Blasques et al. (2015). For example, if ε_t is right skewed, an observation in the right tail of the distribution of ξ_t should be less informative about an increase in volatility than an observation in the left tail. Therefore, in such cases we expect an asymmetric reaction of volatility to values of ξ_t . Such asymmetries are directly embedded into (2') through the use of the score (3'), whereas no such effects are present in the standard GARCH equation (2). A similar argument holds if ε_t is fat-tailed.

The use of the parameterization of the log-variance f_t resembles the EGARCH approach of Nelson (1991). One of the advantages of modeling log volatilities is that we need not impose parameter restrictions on ω , α , or β to ensure a positive variance for each time t . In addition, the conditional second moment of the score is given by

$$\mathbb{E}_{t-1} [s_t^2] = \frac{1}{4} \mathbb{E}_{t-1} [\varepsilon_t^2 \cdot (\nabla q(\varepsilon_t))^2 - 1], \quad (4)$$

which is independent of the time varying parameter f_t . As a result, we need not worry about an appropriate scaling the score function, as discussed extensively in Creal et al. (2013). See also the arguments in Harvey (2013).

Parameter estimation for GAS models using Maximum Likelihood (ML) estimation is straightforward given the models' observation driven nature. In particular,

we obtain parameter estimates by maximizing the scaled log likelihood function

$$L_T(\boldsymbol{\theta}, q) = \frac{1}{T} \sum_{t=1}^T -\frac{1}{2} f_t(\boldsymbol{\theta}) + \log q(e^{-f_t(\boldsymbol{\theta})/2}(y_t - \mu)), \quad (5)$$

where $\boldsymbol{\theta} := (\mu, \omega, \alpha, \beta)$ and $f_t(\boldsymbol{\theta})$ is defined recursively through (2') for any given $\boldsymbol{\theta} \in \Theta$. If the error density is assumed to be of a certain parametric form such as the standard normal or (skewed) Student's t distribution, we obtain a parametric model as in Creal et al. (2011) or Lucas et al. (2014). To avoid the introduction of an incorrect parametric specification of $q(\cdot)$, we estimate $q(\cdot)$ nonparametrically using a kernel density estimator as in Drost and Klaassen (1997) and Sun and Stengos (2006). To do so, we first estimate $\boldsymbol{\theta}$ by adopting a specific parametric family for $q(\cdot)$ as a quasi density.

It is important to note here that the quasi density $q(\cdot)$ not only affects the quasi likelihood (5), but also the transition dynamics of f_t through equations (2') and (3'). We use the estimates obtained with this Quasi Maximum Likelihood Estimator (QMLE) to compute standardized residuals $\hat{\varepsilon}_t = \exp(-f_t(\hat{\boldsymbol{\theta}})/2)(y_t - \hat{\mu})$. The latter are the inputs for the kernel density estimator of $q(\cdot)$,

$$\hat{q}_T(x) = \frac{1}{T \cdot b_T} \sum_{i=1}^T K\left(\frac{\hat{\varepsilon}_t - x}{b_T}\right), \quad (6)$$

where b_T is a bandwidth parameter satisfying standard regularity conditions, and $K(\cdot)$ is a kernel. For example, one can use a kernel based on the standard normal distribution, $K(v) = \exp(v^2/2)/\sqrt{2\pi}$. We rescale and relocate $\hat{q}_T(\cdot)$ to have mean zero and unit variance and use it in (5) to estimate the parameters $\boldsymbol{\theta}$. We label the resulting estimated model SP-GAS(1). In particular, SP-GAS(1) is the model obtained after estimating the parameter vector $\boldsymbol{\theta}$ for the first time using semiparametric maximum likelihood estimation. Once the SP-GAS(1) is obtained, we can recalculate the standardized residuals and obtain a new kernel density estimate of $q(\cdot)$ to be used in (5). Maximizing this new loglikelihood function with respect to $\boldsymbol{\theta}$, we get new parameter estimates and denote the new estimated model as SP-GAS(2). In general, SP-GAS(i) denotes the model obtained after the i -th estimation of $\boldsymbol{\theta}$.

Note that, conditional on the estimated density $\hat{q}_T(\cdot)$ and the parameter estimates $\hat{\boldsymbol{\theta}}_T$, the conditional volatility f_t can be forecast in the usual way by iterating

the volatility filter forward. In particular, since the conditional mean of the score is zero, the n -step-ahead volatility forecast is given by

$$f_{T+n} = \hat{\omega}_T \cdot (1 - \hat{\beta}_T) \sum_{i=0}^{n-2} \hat{\beta}_T^i + \hat{\beta}_T^{n-1} f_{T+1}.$$

As in any observation-driven model, f_{T+1} is the last filtered value of volatility produced by the observed sample $y_1, \dots, y_T$. The forecast f_{T+n} is thus written as a function of f_{T+1} . As expected, if the filtered volatility is stable, i.e. $|\hat{\beta}_T| < 1$, then the volatility forecast converges to the unconditional mean over the forecast horizon,

$$\lim_{n \rightarrow \infty} f_{T+n} = \lim_{n \rightarrow \infty} \hat{\omega}_T \cdot (1 - \hat{\beta}_T) \sum_{i=0}^{\infty} \hat{\beta}_T^i + o_p(1) = \hat{\omega}_T.$$

3 Model Estimation

In what follows we establish the properties of the SP-GAS(i) estimator for any order i . In particular, we show that the SP-GAS(1) possesses important local optimality properties under weak regularity conditions. We also lay down a likelihood-based strategy for selecting the order i for the SP-GAS procedure. The method is based on the fact that SP-GAS(i) for $i > 1$ can improve the likelihood compared to the SP-GAS(1).

Let P_0 denote the true probability measure of $\{y_t\}_{t \in \mathbb{Z}}$ and let $P(\boldsymbol{\theta}, q)$ denote the model-implied measure for $\{y_t\}_{t \in \mathbb{Z}}$ under the parametric update (2') with parameter $\boldsymbol{\theta}$ and probability density function (pdf) q . The model implied measure for $\{y_t(\bar{f}_1)\}_{t \in \mathbb{N}}$ depends also on the initialization point $\bar{f}_1$ for the time-varying parameter $\{f_t\}$. Under appropriate regularity conditions, the fading memory properties of y_t and f_t ensure that $\{y_t(\bar{f}_1)\}_{t \in \mathbb{N}}$ converges to a unique limit process $\{y_t\}_{t \in \mathbb{Z}}$ for any fixed initial point $\bar{f}_1$; see Blasques et al. (2014a). In other words, the sequence $\{y_t(\bar{f}_1)\}_{t \in \mathbb{N}}$ satisfies $|y_t(\bar{f}_1) - y_t| \xrightarrow{a.s.} 0$ as $t \rightarrow \infty \quad \forall \bar{f}_1 \in \mathbb{R}$. Furthermore, let $\mathbb{E}_0$ denote an expectation under the true measure P_0 , let $L_T(\boldsymbol{\theta}, q) = T^{-1} \sum_{t=1}^T \ell_t(\boldsymbol{\theta}, q)$ denote the average log-likelihood function evaluated at $(\boldsymbol{\theta}, q)$, and $\ell_\infty(\boldsymbol{\theta}, q) := \mathbb{E}_0 \ell_t(\boldsymbol{\theta}, q)$ denote the corresponding limit likelihood. Using the terminology of Gouriéroux et al. (1983), the following definition distinguishes between several types of pseudo-true parameters. In particular, it defines ‘*conditional*’ pseudo-true parameters and the notion of ‘*dominant*’ pseudo-true parameters.

DEFINITION 1. A parameter θ^* is said to be a conditional pseudo-true parameter for given q if it satisfies $\ell_\infty(\theta^*, q) \geq \ell_\infty(\theta, q)$ for all θ . A conditional pseudo-true parameter θ^* given q is said to dominate another conditional pseudo-true parameter θ^{**} given q' , if $\ell_\infty(\theta^*, q) \geq \ell_\infty(\theta^{**}, q')$.

Under a correct specification axiom one assumes that $\exists(\theta_0, q_0)$ such that $P_0 = P(\theta_0, q_0)$. A Gibbs' inequality then implies immediately that, conditional on q_0 , the pseudo-true parameter θ_0^* corresponds to the true parameter, i.e. $\theta_0^* = \theta_0$. The same Gibbs inequality can be used to show that θ_0 dominates all other pseudo-true parameters. However, in general, if $q \neq q_0$, then θ_0 is no longer the conditional pseudo-true parameter.

As explained in Section 2, estimation in the SP GAS framework proceeds in at least two steps. First, an estimate of θ is obtained conditional on an initial density q_1 . To fix notation, we denote this conditional estimator as $\hat{\theta}_T(q_1)$. Second, an estimate of q is obtained conditional on the realized value of $\hat{\theta}_T(q_1)$. We denote this second estimator as $\hat{q}_T(\hat{\theta}_T(q_1))$. Under well known classical conditions, each of these conditional estimators behaves in a way that ensures convergence to a conditional pseudo-true parameter; see for example White (1994).

Let $\theta^*(q_1)$ denote the pseudo-true parameter of θ conditional on q_1 ; i.e. let $\theta^*(q_1) := \arg \min_{\theta} \ell_\infty(\theta, q_1)$. The following result follows directly from Theorem 2 in Blasques et al. (2014a). Below, we let $\mathbb{C}^4(\mathbb{R})$ denote the space of four-times continuously differentiable real-valued functions defined on $\mathbb{R}$.

PROPOSITION 1. Let $\{y_t\}_{t \in \mathbb{Z}}$ be a strictly stationary and ergodic sequence satisfying $\mathbb{E}|y_t|^2 < \infty$, let $q_1 \in \mathbb{C}^4(\mathbb{R})$ and Θ be compact subset of $\mathbb{R}^4$ such that $\mathbb{E} \sup_{\mu \in \Lambda} |s(f, y_t, \mu)|^2 < \infty$ and $\sup_{(f^*, y, \theta) \in \mathcal{F} \times \mathcal{Y} \times \Theta} |\beta + \alpha \partial s(f^*, y; \mu) / \partial f| < 1$. Then, $\hat{\theta}_T(q_1) \xrightarrow{a.s.} \theta^*(q_1)$ as $T \rightarrow \infty$.

Proposition 1 requires (i) observed data to be stationary and ergodic (SE) with the first two moments bounded; (ii) the score function to have two bounded moments and satisfy a contraction condition that is common for nonlinear autoregressive processes; see e.g. Bougerol (1993) and Pötscher and Prucha (1997).

Similarly, let $q^*(\theta^*(q_1))$ denote the density of ε_t conditional on $\theta^*(q_1)$, then the next proposition shows that $\hat{q}_T(\theta^*(q_1)) \xrightarrow{a.s.} q^*(\theta^*(q_1))$ and $\hat{q}_T(\hat{\theta}_T(q_1)) \xrightarrow{a.s.} q^*(\theta^*(q_1))$. The first claim follows immediately from the fact that the conditions of Theorem

2.13 in Pagan and Ullah (1999) are satisfied, while the second claim follows from Theorem 2.14 in Pagan and Ullah (1999).

PROPOSITION 2. *Let $\{y_t\}_{t \in \mathbb{Z}}$ be an SE sequence and let the kernel $\mathcal{K}$ be the class of all Borel measurable bounded real-valued functions k satisfying $\int K(x)dx = 1$, $\int |K(x)|dx < \infty$, $|x|K(x) \rightarrow 0$ as $|x| \rightarrow \infty$, $\sup_x |K(x)| < \infty$ and $\int K(x)^2 dx < \infty$. Furthermore, let*

$$|2T^{-2} \sum_{t=1}^{T-1} (T-t) \text{Cov}(w_1, w_{t+1})| \leq \frac{2\lambda_T}{T} \left(\int |K(x)|dx \right)^2,$$

where $w_t = K((\hat{\varepsilon}_t - x)/b_T)$, and let the bandwidth $b_T \rightarrow 0$ and $Tb_T \rightarrow \infty$ as $T \rightarrow \infty$. Then $\hat{q}_T(q^*(\theta^*(q_1))) \xrightarrow{a.s.} q^*(\theta^*(q_1))$ as $T \rightarrow \infty$.

If furthermore, $\lambda_T = O(1)$, then $q^*(\theta)$ is continuous in θ in a neighborhood of $\theta^*(q_1)$ and $\hat{q}_T(\hat{\theta}_T(q_1)) \xrightarrow{a.s.} q^*(\theta^*(q_1))$ as $T \rightarrow \infty$.

Note that $\hat{q}_T$ is, in general, not an ML estimator and hence does not necessarily satisfy

$$\hat{q}_T(q^*(\theta^*(q_1))) := \arg \max_q \ell_T(\hat{\theta}_T(q_1), q).$$

Furthermore, note that since the density estimate $\hat{q}_T$ determines the GAS parameter update equation through the score in equation (2'), the likelihood of the model can actually decrease when taking an SP-GAS iteration from SP-GAS(i) to SP-GAS($i+1$). In particular, it is possible that the pseudo-true parameter $\theta_0^*(q_i)$ associated with the i th iteration does not dominate the pseudo-true parameter $\theta_0^*(q_{i+1})$ such that

$$\ell_\infty(\theta_0^*(q_i), q_i) \geq \ell_\infty(\theta_0^*(q_{i+1}), q_{i+1}).$$

For this reason, the number of iterations should be chosen so as to ensure that we obtain a dominating pseudo-true parameter asymptotically. This naturally leads to an iterative procedure that increases the number of iterations if and only if the likelihood of the model increases from one iteration to the next. The following iterative procedure is based on Akaike's information criterion (AIC) and ensures that we select a dominating pseudo-true parameter θ_0^* across the first k of iterations.

For notational convenience, we will often denote the SP-GAS model with parameter estimate $\hat{\theta}_T(\hat{q}_i)$ obtained in the i th iteration as SP-GAS(i). In this way, SP-GAS(1) denotes the model estimated using only 1 iteration, while SP-GAS(3)

Figure 1: Iterative estimation algorithm

-
1. Given $\hat{q}_i$, obtain the ML estimate $\hat{\theta}_T(\hat{q}_i)$.
 2. Given $\hat{\theta}_T(\hat{q}_i)$, obtain the kernel estimate $\hat{q}_{i+1}$.
 3. Given $\hat{q}_{i+1}$, obtain the ML estimate $\hat{\theta}_T(\hat{q}_{i+1})$.
 4. If $\ell_T(\hat{\theta}_T(\hat{q}_{i+1}), \hat{q}_{i+1}) > \ell_T(\hat{\theta}_T(\hat{q}_i), \hat{q}_i)$: set $i = i + 1$ and proceed to step 1. Otherwise, keep $(\hat{\theta}_T(\hat{q}_i), \hat{q}_i)$ as the final estimate and stop iterating.
-

denotes the model obtained after 3 iterations. To see why this iterative procedure ensures the selection of the estimate with the lowest AIC, note that the likelihood of the first k iterations must satisfy by construction

$$\ell_T(\hat{\theta}_T(q_1), q_1) < \ell_T(\hat{\theta}_T(\hat{q}_2), \hat{q}_2) < \dots < \ell_T(\hat{\theta}_T(\hat{q}_k), \hat{q}_k). \quad (7)$$

Conditional on $\hat{q}_1, \dots, \hat{q}_k$, comparing the k likelihoods in (7) corresponds to comparing the likelihoods of k different non-nested models (with different distributions for the innovations and different updating equations). Since these models have the same number of parameters, it follows immediately that

$$\text{AIC}(\hat{\theta}_T(q_1), q_1) > \text{AIC}(\hat{\theta}_T(\hat{q}_2), \hat{q}_2) > \dots > \text{AIC}(\hat{\theta}_T(\hat{q}_k), \hat{q}_k),$$

so that we are sure to select the model with lowest AIC, and hence, asymptotically, the dominating pseudo-true parameter; see Akaike (1973, 1974, 1981).

Finally, we note that this iterative procedure may, in general, lead to a conditional pseudo-true parameter that depends on the choice of q_1 . In this sense, q_1 should be seen as an intrinsic part of the model. Indeed, just as the pseudo-true parameter depends on the choice of observation equation and updating equation, so may the choice of q_1 determine the best possible approximation to the true unknown conditional density of the data in Kullback-Leibler divergence. Luckily, there are a number of alternative conditions that ensure the convergence of the iterative procedure to a unique limit that does not depend on q_1 . When these conditions are satisfied, the model proposed here has a more definite semiparametric nature,

as the model specification becomes ‘independent’ of q_1 . As a concrete example, Proposition 3 formulates a set of such sufficient conditions.

PROPOSITION 3. *Let the conditions of Propositions 1 and 2 hold and assume that*

$$\sup_{(\boldsymbol{\theta}, \boldsymbol{\theta}'): \|\boldsymbol{\theta} - \boldsymbol{\theta}'\| > 0} \frac{\|\boldsymbol{\theta}^*(q^*(\boldsymbol{\theta})) - \boldsymbol{\theta}^*(q^*(\boldsymbol{\theta}'))\|}{\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|} < 1.$$

Then $\hat{\boldsymbol{\theta}}_T(\hat{q}_k) \xrightarrow{a.s.} \boldsymbol{\theta}_0^$ as $T \rightarrow \infty$ and $k \rightarrow \infty$ for every q_1 , where $\boldsymbol{\theta}_0^*$ does not depend on q_1 . If moreover*

$$\sup_{(q, q'): \|q - q'\| > 0} \frac{\|q^*(\boldsymbol{\theta}^*(q)) - q^*(\boldsymbol{\theta}^*(q'))\|}{\|q - q'\|} < 1,$$

then

$$(\hat{\boldsymbol{\theta}}_T(\hat{q}_k), \hat{q}_k) \xrightarrow{a.s.} (\boldsymbol{\theta}_0^*, q_0^*) \quad \text{as } T \rightarrow \infty \text{ and } k \rightarrow \infty,$$

for every q_1 , where $(\boldsymbol{\theta}_0^, q_0^*)$ does not depend on q_1 .*

Both conditions used in Proposition 3 are ‘contraction conditions’ that ensure the existence of a unique global attractor for a dynamic system defined by a difference equation. Note that, for any given model, these conditions can in principle be verified numerically. As such, the researcher can establish if indeed the estimator converges asymptotically to a unique point regardless of q_1 .

4 Monte Carlo evidence

To study the model’s behavior in finite samples, we carry out two sets of simulation experiments. First, we consider a setting where the GAS model (1’)–(3’) is the actual data generating process (DGP). In this setting we focus on how well the semiparametric model performs in terms of estimation efficiency and volatility forecast accuracy compared to the MLE for the correctly specified model. Next, we consider a setting where the model is mis-specified, such that we can investigate the robustness of the model and estimator.

4.1 Results under correct specification

We use the DGP (1’)–(2’) for four different choices of $q(\cdot)$, namely the standard normal, the Student’s t distribution with $\nu = 3$ and $\nu = 5$ degrees of freedom, and

a mixture of normals. The mixture of normals is chosen such that it is close (in a Kullback-Leibler sense) to a chi-squared distribution with 6 degrees of freedom, thus exhibiting substantial skewness. Each of the above densities is rescaled and relocated to have zero mean and unit variance. For each DGP, we simulate $B = 1000$ return series of length $T = 2000$. The first 1000 observations are labeled the in-sample observations and are used to estimate the parameters. The last 1000 observations are called out-of-sample and are used to assess the model's forecast adequacy. We set $\mu = 0$, $\omega = 2$, $\alpha = 0.3$, and $\beta = 0.9$. Reasonable changes of the parameter values do not alter the conclusions below. Note that β takes the role of $\tilde{\alpha} + \tilde{\beta}$ from the familiar GARCH model in equation (2). Stationarity and ergodicity of the GAS process are thus ensured as long as $\beta < 1$, even in cases where $\alpha + \beta = 0.3 + 0.9 > 1$; see also Blasques et al. (2014a,b).

For each simulated sample, we compute the parameter estimates using five different statistical models: MLE, QML-GARCH, $t(\nu)$ -GAS, SP-GAS(1), and SP-GAS(AIC). The SP-GAS(AIC) uses the algorithm explained in Figure 1 to determine the number of iterations for the kernel density estimator; see also Section 3. QML-GARCH uses the normal distribution as a quasi-likelihood for computing the objective function, as well as for computing the transition dynamics of the variance in equations (2') and (3'). Similarly, the $t(\nu)$ -GAS models use the Student's t density with ν degrees of freedom both for the (quasi)-likelihood and for the transition dynamics. We estimate the degrees of freedom parameter ν jointly with the other static parameters.

The semiparametric estimators require the choice of the bandwidth parameter b_T . In our current setting, we use $b_T = 0.5$ as in Drost and Klaassen (1997). Other reasonable choices $0.3 \leq b_T \leq 0.8$ again do not alter our conclusions. Similar as for the Student's t distribution, the choice of the semiparametric (quasi)-density matters for both the likelihood (5) and the volatility dynamics in (2').

We evaluate the small-sample properties of the estimators for the parameters $(\mu, \omega, \alpha, \beta)$ in terms of root mean squared error (RMSE). We also evaluate the forecast accuracy of the different models by comparing RMSEs of the fitted values $\exp(\hat{f}_t/2)$, both in-sample and out-of-sample. In the current simulation setting, we can compute the RMSE as the data generating process is known and the true $\exp(f_t/2)$ s are observed.

For the semiparametric approaches SP-GAS(1) and SP-GAS(AIC), we use the $t(\nu)$ -GAS estimate as our starting value. We do not use the MLE estimate, as ML estimation is not feasible in practice. The computations are performed using a unit variance Student's $t(5)$ kernel in equation (6). This further stabilizes the simulations, particularly for fat-tailed DGPs.

Table 1 reports the results. All RMSEs are normalized with respect to those for QML-GARCH, which acts as our benchmark. We concentrate our discussion on $\tilde{\alpha}$ and $\tilde{\beta}$, as $\tilde{\mu}$ and $\tilde{\omega}$ cannot be estimated adaptively; see Drost and Klaassen (1997). For reasons of comparability, we also normalize the out-of-sample RMSE with respect to the in-sample RMSE of the QML-GARCH (RMSE in).

The expected best performing method in all cases is the MLE, which provides our efficiency bound for each parameter. For the normal distribution (N), the MLE corresponds to the QML-GARCH model, and the results are identical. The RMSE results obtained for the $t(\nu)$ -GAS model are very similar to those of the MLE and QML-GARCH under the normal distribution. The $t(\nu)$ -GAS model contains the QML-GARCH model as a special case if we set $\nu = \infty$. The estimation of ν results in a small efficiency loss. The largest (relative) efficiency loss appears for α and equals about 6% relative to the QML-GARCH benchmark. The relative RMSEs for SP-GAS(1) for the parameters μ , ω , α , and β are only slightly larger than that of $t(\nu)$ -GAS. The efficiency loss of the parameters with respect to the QML-GARCH benchmark thus appears limited. We note that the relative increase in RMSE for α for SP-GAS(1) is relative to a low absolute level of the RMSE for the QML-GARCH in this case. The RMSEs of the QML-GARCH for the non-Gaussian distributions that are discussed next are considerably higher and show an advantage of the SP-GAS framework compared to the QML-GARCH.

For the fat-tailed $t(3)$ and $t(5)$ distributions, the QML-GARCH performs poorly in terms of RMSE compared to the MLE. For β , the RMSE of the QML-GARCH is 3 ($t(5)$) or even 6 ($t(3)$) times as high as that of the MLE or $t(\nu)$ -GAS. For α , these factors are 2.5 and 3, respectively. The $t(\nu)$ -GAS has a similar performance as the MLE for both α and β . The estimation of ν apparently does not cause a substantial efficiency loss. Interestingly, also the SP-GAS(1) does remarkably well in terms of RMSE for α and β . It is also interesting to see that the biggest improvement in RMSE for the SP-GAS method is already made in the first step, as the SP-GAS(1)

Table 1: Simulation results for generalized autoregressive score processes

This table presents the relative Root Mean Square Error (RMSE) of parameter estimates with respect to QML-GARCH and the relative median Root Mean Square Error of in-sample volatility estimates (RMSE in) and out-of-sample volatility forecasts (RMSE out) with respect to the in-sample RMSE of volatility estimates for the QML-GARCH. Results are based on 1,000 Monte Carlo replications of time series of 2,000 observations (1,000 in-sample and 1,000 out-of-sample). The underlying data generating process is given by equations (1')–(3'). We consider four distributions: the standard normal (N), Student's $t(3)$, $t(5)$, and a mixture of normals resembling a $\chi^2(6)$ distribution (MN($\chi^2(6)$)). The statistical models are the based on the true model (MLE), the normal distribution with volatility dynamics specified in equation (2') (QML-GARCH), the Student's t distribution with estimated degrees of freedom ($t(\nu)$ -GAS), and on the semiparametric approach with one iteration (SP-GAS(1)) or with the number of iterations determined by the AIC (SP-GAS(AIC)). The SP-GAS(1) approach is initialized using the $t(\nu)$ -GAS estimates.

DGP	Model	μ	ω	α	β	RMSE	
						in	out
N	MLE	1.000	1.000	1.000	1.000	1.000	1.010
	QML-GARCH	1.000	1.000	1.000	1.000	1.000	1.010
	$t(\nu)$ -GAS	1.004	1.000	1.058	1.001	1.043	1.026
	SP-GAS(1)	1.025	1.020	1.072	1.039	1.333	1.437
	SP-GAS(AIC)	1.025	1.018	1.078	1.046	1.336	1.441
$t(3)$	MLE	0.656	0.523	0.372	0.178	0.261	0.259
	QML-GARCH	1.000	1.000	1.000	1.000	1.000	1.143
	$t(\nu)$ -GAS	0.657	0.957	0.374	0.209	0.393	0.400
	SP-GAS(1)	0.867	1.025	0.587	0.373	0.799	0.768
	SP-GAS(AIC)	0.869	1.001	0.599	0.389	0.796	0.772
$t(5)$	MLE	0.858	0.849	0.426	0.352	0.321	0.324
	QML-GARCH	1.000	1.000	1.000	1.000	1.000	1.128
	$t(\nu)$ -GAS	0.860	0.933	0.434	0.352	0.362	0.365
	SP-GAS(1)	0.986	1.197	0.580	0.485	0.739	0.716
	SP-GAS(AIC)	0.993	1.294	0.588	0.509	0.719	0.741
MN	MLE	0.534	0.939	0.642	0.635	0.254	0.259
	QML-GARCH	1.000	1.000	1.000	1.000	1.000	1.011
	$t(\nu)$ -GAS	1.798	1.072	1.411	0.967	1.002	1.019
	SP-GAS(1)	0.671	1.092	0.958	0.718	0.572	0.578
	SP-GAS(AIC)	0.644	1.032	0.860	0.699	0.543	0.545

and SP-GAS(AIC) behave similarly.

The in-sample and out-of-sample forecast accuracies of the $t(\nu)$ -GAS model are slightly inferior to those of the MLE, followed by those of the SP-GAS methods. Both the SP-GAS and $t(\nu)$ -GAS model, however, have substantially lower in-sample and out-of-sample RMSEs than the QML-GARCH. The inferior performance of the QML-GARCH should not come as a surprise: the QML-GARCH model is doubly mis-specified in this case. Not only does it incorrectly impose the conditional Gaussian distribution on ε_t , it also assumes the incorrect specification for the volatility dynamics in (2'). In particular, large values of $\xi_t^2 = (y_t - \mu)^2$ receive a higher impact in the QML-GARCH model than warranted under the true data generating process.

For the skewed mixture of normals (MN), the estimators for α and β for the $t(\nu)$ -GAS model do not improve upon the QML-GARCH model. In fact, for α the performance is even worse. Interestingly, the performance of the SP-GAS(1) and SP-GAS(AIC) is substantially better than that of the $t(\nu)$ -GAS model. This is caused by the fact that the Student's t distribution cannot account for skewness, whereas the nonparametric kernel density estimate in the SP-GAS model can. For β , the RMSE of SP-GAS(1) even comes close to that of the MLE. Interestingly, for the skewed distribution it appears useful to iterate more when estimating the density function q nonparameterically. This is seen by the overall improvements in terms of RMSE when moving from SP-GAS(1) to SNP-GAS(AIC). The improvements seem to apply for all parameters. We also note that the SP-GAS(AIC) model has the best in-sample and out-of-sample RMSE performance after the (infeasible) MLE. The RMSE of both the QML-GARCH and $t(\nu)$ -GAS are twice the size of that of SP-GAS.

4.2 Results under mis-specification

We now turn from the correctly specified setting of Section 4.1 to a setting where all models are mis-specified. Our first experiment is similar to that of Engle (2002). We choose a number of deterministic volatility patterns and estimate each of the statistical models discussed earlier. As our performance metric, we only choose the in-sample and out-of-sample RMSEs of the volatility. The RMSEs of the parameters are not useful here, as the statistical model never embeds the true data generating process. The specific deterministic volatility patterns are presented in Figure 2.

Figure 2: Deterministic volatility patterns

The specifications are: Sine ($\sigma_t = 2.5 + \cos(2\pi t/200)$), Long regimes (of 500 observations with $\sigma_t = 1.5, 3.5$), Short regimes (of 200 observations with $\sigma_t = 1.5, 3.5$), Burst ($\sigma_t = 1.5$ for 150 days, growing to $\sigma_t = 3.5$ over 5 days, and then linearly decreasing to $\sigma_t = 1.5$ over the next 95 days).

They include slowly moving patterns of volatility (sine), long-lived and short-lived structural changes in volatility, and regular bursts of volatility that gradually taper off.

The simulation results are presented in Table 2. We normalize each RMSE with respect to the in-sample RMSE of QML-GARCH. Note that even the MLE model is mis-specified here: though it uses the correct distributional assumptions, the volatility dynamics in equation (2') are incorrect under the volatility patterns plotted in Figure 2.

Table 2 clearly illustrates the excellent performance of the SP-GAS approach. Particularly for the skewed mixture of normals, the SP-GAS results are close to those of the MLE. Also for the fat-tailed DGPs, the SP-GAS typically does much better than the QML-GARCH, while the differences between SP-GAS(1) and SP-GAS(AIC) are small. We also see that for volatility processes that are gradually and continuously changing over time, such as the sine and long regime pattern, SP-GAS performs well. This is interesting, as such data generating processes are close in spirit to a typical stochastic volatility process; see Shephard (2005) for specifications of these models. To conclude our simulations, we therefore perform

Table 2: Simulation results: mis-specified volatility dynamics

This table presents the relative median of Root Mean Square Error of in-sample volatility estimates (RMSE in) and out-of-sample volatility forecasts (RMSE out) with respect to in-sample RMSE of the QML-GARCH estimates over 1,000 Monte Carlo replications and 2,000 time series observations (1,000 in-sample and 1,000 out-of-sample). There are four distributions: normal, Student's $t(3)$, $t(5)$, and a mixture of normals (MN, matched to resemble a $\chi^2(6)$ distribution). The SP-GAS(1) approach is initialized using the $t(\nu)$ -GAS estimates.

DGP	Model	sine		short regime		long regime		burst	
		in	out	in	out	in	out	in	out
N	MLE	1.000	0.987	1.000	1.098	1.000	1.023	1.000	0.998
	QML-GARCH	1.000	0.987	1.000	1.098	1.000	1.023	1.000	0.998
	$t(\nu)$ -GAS	0.986	0.974	0.987	1.104	0.992	1.006	1.014	1.010
	SP-GAS(1)	1.003	0.991	1.001	1.122	1.007	1.027	1.020	1.034
	SP-GAS(AIC)	1.004	0.992	1.003	1.122	1.009	1.028	1.020	1.035
$t(3)$	MLE	0.700	0.679	0.633	0.754	0.646	0.639	0.770	0.767
	QML-GARCH	1.000	1.133	1.000	1.278	1.000	1.129	1.000	1.057
	$t(\nu)$ -GAS	0.751	0.736	0.713	0.814	0.717	0.726	0.813	0.808
	SP-GAS(1)	0.850	0.847	0.743	0.912	0.770	0.736	0.881	0.892
	SP-GAS(AIC)	0.847	0.847	0.744	0.901	0.762	0.744	0.877	0.894
$t(5)$	MLE	0.811	0.799	0.787	0.926	0.791	0.783	0.879	0.877
	QML-GARCH	1.000	1.014	1.000	1.115	1.000	1.024	1.000	1.011
	$t(\nu)$ -GAS	0.833	0.826	0.835	0.961	0.830	0.834	0.898	0.891
	SP-GAS(1)	0.901	0.886	0.853	0.987	0.874	0.867	0.932	0.941
	SP-GAS(AIC)	0.899	0.901	0.851	0.989	0.866	0.889	0.929	0.946
MN	MLE	0.886	0.874	0.909	0.989	0.904	0.923	0.897	0.896
	QML-GARCH	1.000	0.988	1.000	1.088	1.000	1.008	1.000	0.994
	$t(\nu)$ -GAS	0.997	0.994	1.023	1.134	1.012	1.018	1.051	1.043
	SP-GAS(1)	0.890	0.874	0.871	0.994	0.890	0.898	0.932	0.938
	SP-GAS(AIC)	0.893	0.878	0.873	0.994	0.890	0.900	0.924	0.938

a second experiment to investigate the performance of the SP-GAS model if the data are generated by a stochastic volatility process. Such an experiment may be closer to typical empirical data generating processes than the deterministic volatility patterns investigated so far.

Our stochastic volatility (SV) DGP is specified as $y_t \sim p(0, \sigma_t^2)$ with $\sigma_t^2 = \exp(\alpha_t)$ and $\alpha_t = 0.01 + 0.98\alpha_{t-1} + \eta_t$, where $\eta_t \sim N(0, 0.1^2)$, for $t = 1, \dots, n$. The distribution p can be any of the four distributions used in Section 4.1. We

also performed simulations with a less persistent SV process. The results of that experiment were similar in terms of the relative RMSEs of the different models, and are hence not reported here.

To differentiate the current simulation experiment from the previous one, we also compare the value-added of the transition dynamics of volatility. We consider GARCH dynamics as given in equation (2) versus GAS dynamics as specified in equation (2').

For each DGP and model, we perform a test to see whether median over the Monte Carlo simulations of the RMSEs of the GAS specification are significantly lower than those of the GARCH specification. Note that even under the normal distribution the GAS and GARCH specifications do not coincide, as the former parameterizes the log volatility, whereas the latter parameterizes the variance itself.

Table 3 presents the results. The table shows that for the normal distribution (N) the performance of all models is similar, both in terms of in-sample and out-of-sample RMSE. In-sample relative differences are not larger than 0.7%, and most of the time smaller than 0.5%. The QML favors the GARCH type parameterization rather than the GAS alternative. Given the use of the normal distribution for the QML, the differences between GARCH and GAS boil down to modeling the variance or the log-variance, respectively. Also looking at the results for the other distributions (t3, t5, MN), modeling variances directly seems the preferred approach if one adopts the Gaussian QML approach for estimation.

The picture changes substantially if we consider the fat-tailed distributions $t(3)$ and $t(5)$. The $t(\nu)$ -GAS, SP-GAS(1), and SP-GAS(AIC) approach all improve upon QML-GAS. The relative differences in the GARCH column are less than those in the GAS columns. This finding can be explained by looking at the SV specification: under the SV dynamics, volatility evolves slowly over time as a Gaussian process with a high persistence parameter of 0.98. The measurement equation, however, is conditionally fat-tailed, such that tail-observations occur frequently. Tail observations have a high impact on the GARCH dynamics in equation (2) through the use of squared residuals ξ_t^2 . The impact of such tail-observations using the score of the Student's t density ($t(\nu)$ -GAS) or its nonparametrically estimated counterpart (SP-GAS) is much less, resulting in a better estimate of the underlying true volatilities. The improvement in RMSE of GAS versus GARCH dynamics is significant for both

Table 3: Simulation results: stochastic volatility

This table presents the relative median of the Root Mean Square Error (RMSE) of in-sample volatility estimates (RMSE in) and out-of-sample volatility forecasts (RMSE out) with respect to the GAS model's in-sample RMSE of QML. For each density we use 1,000 Monte Carlo replications and 2,000 time series observations (1,000 in-sample and 1,000 out-of-sample). The stochastic volatility (SV) DGP is specified as $y_t \sim p(0, \sigma_t^2)$ with $\sigma_t^2 = \exp(\alpha_t)$ and $\alpha_t = 0.01 + 0.98\alpha_{t-1} + \eta_t$, where $\eta_t \sim N(0, 0.1^2)$, for $t = 1, \dots, T$. Four distributions are used to generate y_t : normal, Student's $t(3)$, $t(5)$, and a mixture of normals MN (resembling a $\chi^2(6)$). We report the p -value of a one-sided sign test for the average RMSE for the GAS dynamics being significantly smaller than that of the GARCH specification. A p -value near zero indicates that GAS dynamics are significantly better than GARCH dynamics.

DGP	Model	GAS		GARCH		sign test (GAS<GARCH)	
		in	out	in	out	in	out
N	QML-GARCH	1.000	1.040	1.002	1.037	0.134	0.996
	$t(\nu)$ -GAS	0.993	1.033	1.001	1.036	0.000	0.000
	SP-GAS(1)	1.006	1.046	1.005	1.042	0.999	1.000
	SP-GAS(AIC)	1.006	1.046	1.004	1.042	0.000	0.000
$t(3)$	QML-GARCH	1.000	1.218	1.002	1.080	1.000	1.000
	$t(\nu)$ -GAS	0.830	0.872	0.951	1.057	0.000	0.000
	SP-GAS(1)	0.905	0.947	0.962	1.040	0.000	0.000
	SP-GAS(AIC)	0.896	0.944	0.949	1.038	0.000	0.000
$t(5)$	QML-GARCH	1.000	1.100	0.985	1.039	1.000	1.000
	$t(\nu)$ -GAS	0.912	0.940	0.973	1.037	0.000	0.000
	SP-GAS(1)	0.947	0.997	0.975	1.029	0.000	0.000
	SP-GAS(AIC)	0.950	1.002	0.975	1.031	0.000	0.000
MN	QML-GARCH	1.000	1.050	0.995	1.037	1.000	1.000
	$t(\nu)$ -GAS	1.026	1.071	1.042	1.080	0.000	0.000
	SP-GAS(1)	0.933	0.971	0.982	1.034	0.000	0.000
	SP-GAS(AIC)	0.938	0.971	0.982	1.032	0.000	0.000

the $t(\nu)$ -GAS and SP-GAS models, as we can see from the reported p -values.

A similar result is seen for the mixture of normals. Though less strong, we find evidence that the RMSE in the semiparametric models improves in a statistically significant way if we move from GARCH dynamics to GAS dynamics. For the asymmetric mixture of normals (MN), this is due to the fact that the score in the GAS specification (2') exploits the feature that left and right tail observations convey dif-

ferent information about changes in volatility if the true distribution is right-skewed. The GARCH dynamics in (2) do not exploit this feature and react symmetrically to observations from the left and right tail, even though the semiparametric GARCH model improves the likelihood estimate using the nonparametrically estimated density. These findings are also in line with the theoretical results in Blasques et al. (2015) that show that the use of the score is optimal in an information theoretic sense.

5 Empirical application

We apply the new semiparametric GAS volatility model to daily euro-dollar exchange rate returns over the period January 4, 1999 to January 19, 2015. We have 4362 observations in total. The data can be freely obtained from ECB Statistical Data Warehouse (<http://sdw.ecb.europa.eu>). For the semiparametric methods, we use the same kernel density estimator as for the simulations in Section 4.

Table 4 presents the estimation results. We report the results for the score driven GAS models as well as for a number of GARCH models. For comparison, we report results in terms of the persistence parameter $\tilde{\beta}$ for the GAS specifications, and $\tilde{\alpha} + \tilde{\beta}$ for the GARCH specifications; see also Creal et al. (2011).

The volatility is highly persistent across all specifications. All values of β and $\tilde{\alpha} + \tilde{\beta}$ are higher than 0.99. The degrees of freedom parameter for the $t(\nu)$ -GAS model is estimated between 9 and 10 for all specifications, which is typical for exchange rate returns and indicates that returns are (conditionally) fat-tailed.

Comparing across models, we see that switching from the normal distribution to the Student's t distribution increases the likelihood by almost 40 points for the GARCH model, and by almost 45 points for the GAS specification. To decompose the effect for the GAS model into its two components, we also estimate the model $t(\nu^*)$ -GAS. This model has the Student's t likelihood, but the volatility dynamics from the QML-GAS model, similar to our simulation results in Table 3. We see that the largest effect (39 points) is due to the change in the conditional density. The remaining improvement of 6 points due to the change in volatility dynamics is smaller, but still non-negligible.

The semiparametric GAS model has similar parameter estimates as the Stu-

Table 4: Empirical Estimation Results

Empirical results for the euro-dollar daily exchange rate returns from January 4, 1999, to January 19, 2015. GAS models are based on (1')–(3') using a normal (QML), $t(\nu)$, or nonparametrically (SP) estimated conditional density. GARCH models use (1)–(2). The GAS $t(\nu^*)$ model uses the $t(\nu)$ density as the likelihood, but uses the score of the normal density in the volatility dynamics in equations (2') and (3'). The starting values for SP-GAS(1) and SP-GARCH are $t(\nu)$ -GAS and $t(\nu)$ -GARCH, respectively.

		log-lik	AIC	$\mu, \tilde{\mu}$	$\omega, \tilde{\omega}$	$\alpha, \tilde{\alpha}$	$\beta, \tilde{\alpha} + \tilde{\beta}$	$\nu, \tilde{\nu}$
GAS	QML	-3719.3	7440.7	0.004 (0.009)	-1.023 (0.184)	0.053 (0.007)	0.996 (0.002)	
GAS	$t(\nu)$	-3674.0	7350.0	0.004 (0.008)	-1.048 (0.190)	0.078 (0.010)	0.996 (0.002)	9.580 (1.283)
GAS	$t(\nu^*)$	-3680.2	7362.4	0.004 (0.008)	-1.040 (0.201)	0.054 (0.006)	0.996 (0.002)	9.763 (1.379)
GAS	SP(1)	-3667.1	7342.2	0.004 (0.008)	-1.023 (0.163)	0.070 (0.009)	0.996 (0.001)	
GAS	SP(2)	-3667.0	7342.1	0.003 (0.008)	-1.031 (0.165)	0.070 (0.009)	0.996 (0.002)	
GAS	SP(3)	-3667.0	7342.0	0.003 (0.008)	-1.031 (0.166)	0.070 (0.009)	0.996 (0.002)	
GARCH	QML	-3718.2	7438.3	0.004 (0.009)	0.423 (0.108)	0.030 (0.004)	0.997 (0.002)	
GARCH	$t(\nu)$	-3679.3	7360.6	0.004 (0.008)	0.418 (0.108)	0.030 (0.004)	0.998 (0.001)	9.959 (1.376)
GARCH	SP(1)	-3672.5	7353.0	0.004 (0.008)	0.421 (0.100)	0.030 (0.004)	0.997 (0.001)	
GARCH	SP(2)	-3672.6	7353.1	0.004 (0.008)	0.421 (0.099)	0.030 (0.004)	0.997 (0.001)	
GARCH	SP(3)	-3672.6	7353.1	0.004 (0.008)	0.421 (0.099)	0.030 (0.004)	0.997 (0.001)	

dent's t based GARCH model. With a likelihood value that is about 7 points higher, the SP-GAS model is preferred in terms of likelihood as well as in terms of AIC for the current dataset. If we iterate the SP-GAS(1) further to SP-GAS(2) or even SP-GAS(3), the changes are very small. Though, strictly speaking, the SP-GAS(3) is the best model in terms of AIC, the main improvement appears to be obtained upon the first iteration. Also note that the SP-GAS specification performs about 11 AIC points better than the equivalent SP-GARCH specification. We conclude that the new SP-GAS model not only performs well in a controlled simulation environment, but also for empirically relevant data sets. For the exchange rate return series considered here, the SP-GAS model performs best in terms of AIC, both compared to GAS specifications with a fixed density, and compared to competing GARCH specifications.

6 Conclusion

We have proposed a new semiparametric observation-driven volatility model based on generalized autoregressive score (GAS) dynamics using a nonparametrically estimated conditional density. The nonparametric density estimate influences both the semiparametric likelihood function and the volatility dynamics and extends the current purely parametric approach to GAS models to the semiparametric domain. We showed that it is straightforward to estimate the model's parameters by standard maximum likelihood methods. The estimators are consistent for the conditional pseudo true parameters of the model. In three simulation experiments we showed that the model has good in-sample and out-of-sample performance and compares favorably to competing models if the data are fat-tailed and/or skewed. Using daily euro-dollar exchange rates, we also showed that the model performs well on empirical data. In particular, the semiparametric GAS model improves upon models driven by GARCH dynamics, as well as on GAS models with a (parametric) normal or symmetric Student's t quasi likelihood.

A Proof of Proposition 3

The first claim follows from the usual argument for establishing a unique global attractor in a dynamic system described by a difference equation. We proceed first by showing that $\|\boldsymbol{\theta}_{k+1}^*(q_1) - \boldsymbol{\theta}_{k+1}^*(q'_1)\| \rightarrow 0$ as $k \rightarrow \infty$, for any pair (q_1, q'_1) of initial densities, where $\boldsymbol{\theta}_{k+1}^*(q_1)$ and $\boldsymbol{\theta}_{k+1}^*(q'_1)$ denote the $(k+1)$ -iteration estimators under q_1 and q'_1 respectively. This occurs because

$$\begin{aligned} \|\boldsymbol{\theta}_{k+1}^*(q_1) - \boldsymbol{\theta}_{k+1}^*(q'_1)\| &= \frac{\|\boldsymbol{\theta}_{k+1}^*(\boldsymbol{\theta}_k^*(q_1)) - \boldsymbol{\theta}_{k+1}^*(\boldsymbol{\theta}_k^*(q'_1))\|}{\|\boldsymbol{\theta}_k^*(q_1) - \boldsymbol{\theta}_k^*(q'_1)\|} \|\boldsymbol{\theta}_k^*(q_1) - \boldsymbol{\theta}_k^*(q'_1)\| \\ &\leq \sup_{\|\boldsymbol{\theta} - \boldsymbol{\theta}'\| > 0} \left\{ \frac{\|\boldsymbol{\theta}_{k+1}^*(\boldsymbol{\theta}) - \boldsymbol{\theta}_{k+1}^*(\boldsymbol{\theta}')\|}{\|\boldsymbol{\theta} - \boldsymbol{\theta}'\|} \right\} \|\boldsymbol{\theta}_k^*(q_1) - \boldsymbol{\theta}_k^*(q'_1)\| \end{aligned}$$

and hence,

$$\|\boldsymbol{\theta}_{k+1}^*(q_1) - \boldsymbol{\theta}_{k+1}^*(q'_1)\| = c \cdot \|\boldsymbol{\theta}_k^*(q_1) - \boldsymbol{\theta}_k^*(q'_1)\|$$

with $|c| < 1$, which implies naturally that

$$\|\boldsymbol{\theta}_{k+1}^*(q_1) - \boldsymbol{\theta}_{k+1}^*(q'_1)\| \rightarrow 0 \quad \text{as } k \rightarrow \infty \quad \forall (q_1, q'_1).$$

The desired result, is thus obtained by noting that

$$\begin{aligned}\|\hat{\boldsymbol{\theta}}_{k+1}(q_1) - \hat{\boldsymbol{\theta}}_{k+1}(q'_1)\| &\leq \|\hat{\boldsymbol{\theta}}_{k+1}(q_1) - \boldsymbol{\theta}_{k+1}^*(q_1)\| \\ &\quad + \|\hat{\boldsymbol{\theta}}_{k+1}(q'_1) - \boldsymbol{\theta}_{k+1}^*(q'_1)\| + \|\boldsymbol{\theta}_{k+1}^*(q_1) - \boldsymbol{\theta}_{k+1}^*(q'_1)\| \\ &= o_{a.s.}(1) + o_{a.s.}(1) + o(1) \quad \text{as } T \rightarrow \infty \text{ and } k \rightarrow \infty.\end{aligned}$$

The second claim follows by adapting this argument to q .

Acknowledgements

We thank Andrew Harvey, two anonymous referees, and the associate editor for comments on an earlier version of this paper. We thank Bob Brugman for excellent research assistance. All authors thank the Dutch National Science Foundation (NWO under grant VICI453-09-005) for financial support.

References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *International Symposium on Information Theory, 2 nd, Tsahkadsor, Armenian SSR*, pp. 267–281.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19, 716–723.
- Akaike, H. (1981, May). Likelihood of a model and information criteria. *Journal of Econometrics* 16(1), 3–14.
- Andres, P. (2014). Computation of maximum likelihood estimates for score driven models for positive valued observations. *Computational Statistics and Data Analysis*, forthcoming.
- Baillie, R. T. and T. Bollerslev (1989). The message in daily exchange rates: a conditional-variance tale. *Journal of Business and Economic Statistics* 7, 297–305.
- Bauwens, L. and S. Laurent (2005). A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. *Journal of Business and Economic Statistics* 23(3), 346–354.
- Blasques, F., S. J. Koopman, and A. Lucas (2014a). Maximum likelihood estimation for generalized autoregressive score models. *Tinbergen Institute Discussion Paper 14-029/III*.
- Blasques, F., S. J. Koopman, and A. Lucas (2014b). Stationarity and ergodicity of univariate generalized autoregressive score processes. *Electronic Journal of Statistics* 8, 1088–1112.

- Blasques, F., S. J. Koopman, and A. Lucas (2015). Information theoretic optimality of observation driven time series models for continuous responses. *Biometrika*, forthcoming.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 31(3), 307–327.
- Bougerol, P. (1993). *Kalman Filtering with Random Coefficients and Contractions*. Prépublications de l’Institut Elie Cartan. Univ. de Nancy.
- Creal, D., B. Schwaab, S. J. Koopman, and A. Lucas (2014). Observation driven mixed-measurement dynamic factor models. *Review of Economics and Statistics* 96(5), 898–915.
- Creal, D. D., S. J. Koopman, and A. Lucas (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business and Economic Statistics* 29, 552 – 563.
- Creal, D. D., S. J. Koopman, and A. Lucas (2013). Generalized Autoregressive Score models with applications. *Journal of Applied Econometrics* 28(5), 777–795.
- Drost, F. C. and C. A. J. Klaassen (1997). Efficient estimation in semiparametric GARCH models. *Journal of Econometrics* 81(1), 193–221.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica* 50(4), 987–1007.
- Engle, R. F. (2002). Dynamic Conditional Correlation. *Journal of Business and Economic Statistics* 20(3), 339–350.
- Engle, R. F. and G. Gonzalez-Rivera (1991). Semiparametric ARCH models. *Journal of Business and Economic Statistics* 9, 345–359.
- Engle, R. F., D. Lilien, and R. Robins (1987). Estimating time varying risk premia in the term structure: the ARCH-M model. *Econometrica* 55, 391–407.
- Francq, C. and J.-M. Zakoïan (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications*. Wiley.
- Glosten, L., R. Jagannathan, and D. Runkle (1993). On the relation between the expected value and the volatility of the normal excess return on stocks. *The Journal of Finance* 48, 1779–1801.
- Gourieroux, C., A. Monfort, and A. Trognon (1983). Testing nested or non nested hypotheses. *Journal of Econometrics*, 83–115.
- Harvey, A. C. (2013). *Dynamic Models for Volatility and Heavy Tails*. Cambridge University Press.

- Harvey, A. C. and A. Luati (2014). Filtering with heavy tails. *Journal of the American Statistical Association*, forthcoming.
- Lee, S.-W. and B. E. Hansen (1994). Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory* 10(1), 29–52.
- Linton, O. and E. Mammen (2005). Estimating semiparametric GARCH(∞) model by kernel smoothing methods. *Econometrica* 73(3), 771–836.
- Lucas, A., B. Schwaab, and X. Zhang (2014). Conditional euro area sovereign default risk. *Journal of Business and Economic Statistics*, forthcoming.
- Lumsdaine, R. L. (1996). Consistency and asymptotic normality of the quasi-maximum likelihood estimator in igarch (1, 1) and covariance stationary garch (1, 1) models. *Econometrica: Journal of the Econometric Society* 64(3), 575–596.
- Nelson, D. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica* 59, 347–370.
- Oh, D. H. and A. Patton (2013). Time-varying systemic risk: Evidence from a dynamic copula model of cds spreads. *Duke University Discussion Paper*.
- Pagan, A. and A. Ullah (1999). *Nonparametric econometrics*. Cambridge University Press.
- Pagan, A. R. and G. W. Schwert (1990). Alternative models for conditional stock volatility. *Journal of Econometrics* 45, 267–290.
- Pötscher, B. M. and I. R. Prucha (1997). *Dynamic Nonlinear Econometric Models: Asymptotic Theory*. Springer-Verlag.
- Shephard, N. (2005). *Stochastic Volatility: Selected Readings*. Oxford: Oxford University Press.
- Sun, Y. and T. Stengos (2006). Semiparametric efficient adaptive estimation of asymmetric GARCH models. *Journal of Econometrics* 133(1), 373–386.
- Tapia, R. A. and J. R. Thompson (1978). *Nonparametric Probability Density Estimation*. Baltimore: Johns Hopkins University Press.
- Weiss, A. A. (1986). Asymptotic theory for ARCH models: estimation and testing. *Econometric Theory* 2(1), 107–131.
- White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge Books. Cambridge University Press.
- Yang, L. (2006). Semiparametric GARCH model for foreign exchange volatility. *Journal of Econometrics* 130(2), 365–384.

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