

NOTICE: this is the author's version of a work that was accepted for publication in Computational Statistics and Data Analysis. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Computational Statistics and Data Analysis, vol. 56, issue 11 (November 2012), pp. 3035–3054. DOI information: [10.1016/j.csda.2011.10.021](https://doi.org/10.1016/j.csda.2011.10.021)

BAYESIAN ESTIMATION OF GENERALIZED HYPERBOLIC SKEWED STUDENT GARCH MODELS

PHILIPPE J. DESCHAMPS

Université de Fribourg

October 2011

ABSTRACT. Efficient posterior simulators for two GARCH models with generalized hyperbolic disturbances are presented. The first model, GHt-GARCH, is a threshold GARCH with a skewed and heavy-tailed error distribution; in this model, the latent variables that account for skewness and heavy tails are identically and independently distributed. The second model, ODLV-GARCH, is formulated in terms of observation-driven latent variables; it automatically incorporates a risk premium effect. Both models nest the ordinary threshold t-GARCH as a limiting case. The GHt-GARCH and ODLV-GARCH models are compared with each other and with the threshold t-GARCH using five publicly available asset return data sets, by means of Bayes factors, information criteria, and classical forecast evaluation tools. The GHt-GARCH and ODLV-GARCH models both strongly dominate the threshold t-GARCH, and the Bayes factors generally favor GHt-GARCH over ODLV-GARCH. A Markov switching extension of GHt-GARCH is also presented. This extension is found to be an empirical improvement over the single-regime model for one of the five data sets.

SÉMINAIRE D'ÉCONOMÉTRIE, BOULEVARD DE PÉROLLES 90, CH-1700 FRIBOURG, SWITZERLAND. TELEPHONE: +41-26-300-8252. TELEFAX: +41-26-300-9781.

E-mail address: philippe.deschamps@unifr.ch

Key words and phrases. Autoregressive conditional heteroskedasticity; Markov chain Monte Carlo; bridge sampling; heavy-tailed skewed distributions; generalized hyperbolic distribution; generalized inverse Gaussian distribution.

Forthcoming in Computational Statistics and Data Analysis. doi:10.1016/j.csda.2011.10.021

1. INTRODUCTION

The conditional distributions of asset returns are well-known to be leptokurtic. They are also known to exhibit the “leverage effect”: a negative past innovation on asset returns tends to increase the current volatility. The leptokurticity justifies the introduction of heavy-tailed (e.g. Student or GED) disturbances in the GARCH class of models, and the “leverage effect” has motivated consideration of asymmetric extensions of the basic GARCH variance function introduced by Bollerslev (1986). An example of such an extension is proposed by Glosten *et al.* (1993). An excellent survey of ARCH and GARCH models can be found in Bollerslev *et al.* (1994).

There can be no certainty that the GARCH variance function will capture all the asymmetry present in an asset return distribution, even when this function incorporates a leverage effect. Presumably for this reason, several GARCH models with skewed error distributions can be found in recent literature. Examples are Mittnik and Paoletta (2000), Giot and Laurent (2003), Bauwens and Laurent (2005), Aas and Haff (2006) and Dark (2010). Aas and Haff (2006) provide additional references. However, none of these contributions propose a Bayesian treatment of the estimation problem: the models are estimated by maximum likelihood or quasi-maximum likelihood. This absence of Bayesian treatments is unfortunate, since the Bayesian paradigm offers a natural way of taking both parameter uncertainty and model uncertainty into account. Geweke and Amisano (2010) show that this is important in the context of forecast evaluation; it is therefore also likely to have an impact in the context of value at risk or expected shortfall estimation.

On the other hand, the Bayesian Markov chain Monte Carlo (MCMC) estimation of t-GARCH models with symmetric errors, but possibly asymmetric variance functions is now well-established; an efficient method, based on the previous contribution of Nakatsuma (2000), is fully described in Ardia (2008, chap. 5). This method relies on the fact that a Student-t variate can be represented as a Normal variate with stochastic variance; see, e.g., Geweke (1993). It is indeed this fact which allows an efficient Bayesian posterior simulator to be designed, using the technique of data augmentation.

As shown by Barndorff-Nielsen (1977), a Normal distribution can also be extended by taking *both* moments of the Normal to be functions of an inverted

Gamma variate. The resulting distribution is leptokurtic and skewed. It is known as the Generalized Hyperbolic (GH) distribution, and has been extensively discussed by Prause (1999); see also Paoletta (2007, chap. 9). However, empirical applications of the GH distribution have been few, perhaps due to the fact that its parameters can be difficult to identify in general.

Recently, however, Aas and Haff (2006) investigated a special case of the GH distribution that may considerably alleviate the identification problem mentioned above. It has the advantage of nesting the ordinary Student-t as a limiting case, and can therefore be called the Generalized Hyperbolic Skewed Student-t (GHSST). Aas and Haff (2006) show that the GHSST can exhibit unequal thickness in both tails, contrary to other skewed extensions of the Student-t, and argue that this offers more flexibility.

In this paper, we will propose an efficient MCMC posterior simulator, based on data augmentation, that can be used with models having a GHSST error distribution. It will be applied to two GARCH formulations. The first one (called GHt-GARCH for short) is a threshold version of the GARCH model in Aas and Haff (2006). In this first model, the inverted Gamma latent variables are identically distributed. In the second model, by contrast, these latent variables can be interpreted as stochastic volatilities, since their conditional distributions depend on past observations. The second model can therefore be called an “observation-driven stochastic volatility model” in the sense of Barndorff-Nielsen (1997), and does not appear to have been estimated before by any method. In order to avoid possible confusion between this model and the state-space volatility models that have been proposed in the literature, we will refer to this second model by the acronym ODLV-GARCH, where ODLV stands for “observation-driven latent variables”.

Both the ODLV-GARCH and GHt-GARCH models nest the ordinary t-GARCH as a limiting case. So, they belong to a different class than the state-space formulation used by Nakajima and Omori (2010), who used the GHSST distribution in conjunction with an evolution equation implying Lognormal volatilities.

The t-GARCH, ODLV-GARCH, and GHt-GARCH models will be compared using asset return data, by means of Bayes factors, Bayes information criteria, and classical forecast evaluation tools.

An outline of the paper follows. In Section 2, we state the GHT-GARCH model. Section 3 discusses the ODLV-GARCH model and the differences between this model and GHT-GARCH. Section 4 describes the posterior simulator. Section 5 presents empirical results based on five publicly available asset return data sets. Section 6 discusses a Markov switching extension, and Section 7 concludes. In Appendix A, we discuss at length an efficient algorithm for drawing the latent variables used in data augmentation.

2. THE GHT-GARCH MODEL

This section will present an AR(1)-GARCH model with an asymmetric variance function and the skewed heavy-tailed error distribution discussed in Aas and Haff (2006). We model the log-return y_t of an asset at time t as:

$$y_t = \phi_1 + \phi_2 y_{t-1} + u_t \quad \text{for } t = 1, \dots, T, \quad (1)$$

where $u_t = \sigma_t \eta_t$ and η_t follows a GHSST distribution with zero expectation and unit variance. The variance equation is an asymmetric GARCH model of the type proposed by Glosten *et al.* (1993):

$$\sigma_t^2 = \alpha_0^* + [\alpha_1^* \mathcal{I}(u_{t-1} \geq 0) + \alpha_2^* \mathcal{I}(u_{t-1} < 0)] u_{t-1}^2 + \beta^* \sigma_{t-1}^2, \quad (2)$$

where \mathcal{I} denotes an indicator function. For simplicity, we take y_0 and y_{-1} as fixed, and let $u_0 = y_0 - \phi_1 - \phi_2 y_{-1}$, $\sigma_0^2 = y_0^2$.

The density of u_t has a complicated analytical form which involves a Bessel function of u_t ; see Aas and Haff (2006, Eq. 8). Evaluating the Bessel function is very time-consuming. Fortunately, u_t can be shown to have the following mixture representation:

$$u_t = \sigma_t \left(\beta \left[Z_t - \frac{\delta^2}{\nu - 2} \right] + \sqrt{Z_t} \epsilon_t \right), \quad (3)$$

where ϵ_t is a standard Normal random variable, and where Z_t is independent of ϵ_t and has the following inverted Gamma density:

$$f(Z_t) = \frac{(\frac{\delta^2}{2})^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} Z_t^{-\nu/2-1} \exp \left[-\frac{\delta^2}{2Z_t} \right], \quad (4)$$

with:

$$\delta^2 = \frac{(\nu - 2)(\nu - 4)}{4\beta^2} \left(-1 + \sqrt{1 + \frac{8\beta^2}{\nu - 4}} \right). \quad (5)$$

For a given σ_t , the marginal density of u_t in (3) thus depends on two free parameters, ν and β . Its first two moments exist if $\nu > 4$. When $\beta \rightarrow 0$, the limit of δ^2 is easily seen to be $(\nu-2)$, and the GHSST distribution converges to the central Student-t with ν degrees of freedom and unit variance. In the GHSST, ν can therefore be interpreted as a kurtosis parameter and β as a skewness parameter. Figure 1 plots various shapes of this density; it is shown to be left-skewed when $\beta < 0$ and right-skewed when $\beta > 0$.

3. THE ODLV-GARCH MODEL

In this version, the conditional density of Z_t will depend on the conditional variance of y_t . Equation (3) above is modified as:

$$u_t = \beta \left[Z_t - \frac{\delta_t^2}{\nu - 2} \right] + \sqrt{Z_t} \epsilon_t, \quad \text{with:} \quad (6)$$

$$f_t(Z_t) = \frac{(\frac{\delta_t^2}{2})^{\frac{\nu}{2}}}{\Gamma(\frac{\nu}{2})} Z_t^{-\nu/2-1} \exp \left[-\frac{\delta_t^2}{2Z_t} \right], \quad (7)$$

and

$$\delta_t^2 = \frac{(\nu - 2)(\nu - 4)}{4\beta^2} \left(-1 + \sqrt{1 + \frac{8\beta^2\sigma_t^2}{\nu - 4}} \right). \quad (8)$$

The new error term in (6) also has an expectation of zero and a variance equal to σ_t^2 . Its density is given by:

$$f_t(u_t) = \frac{2^{\frac{1-\nu}{2}} \delta_t^\nu |\beta|^{\frac{\nu+1}{2}}}{\Gamma(\frac{\nu}{2}) \sqrt{\pi}} K_{\frac{\nu+1}{2}} \left(|\beta| \sqrt{\delta_t^2 + x_t^2(u_t)} \right) \exp [\beta x_t(u_t)] \left(\sqrt{\delta_t^2 + x_t^2(u_t)} \right)^{-\frac{\nu+1}{2}}, \quad (9)$$

where $K_\lambda(\cdot)$ is the modified Bessel function of the third kind with index λ , and:

$$x_t(u_t) = u_t + \frac{\beta \delta_t^2}{\nu - 2}.$$

In the ODLV-GARCH model, the Z_t can be interpreted as stochastic volatilities, and $\beta[Z_t - \delta_t^2/(\nu - 2)]$ in (6) as a risk premium effect that affects the mode

(but not the mean) of the distribution of u_t . A similar model, where the distribution of Z_t depends on past observations, was proposed by Barndorff-Nielsen (1997). Since $\delta_t^2/(\nu - 2) = E_{t-1}(Z_t)$, $Z_t - \delta_t^2/(\nu - 2)$ can be interpreted as the unanticipated component of volatility. It was argued by French *et al.* (1987) that this component is usually negatively correlated with asset returns. One would therefore expect the coefficient β to be negative, so that u_t has a positive mode (see Figure 1).

It is straightforward (but tedious) to obtain the coefficients of conditional skewness and kurtosis, given respectively by $m_{3t} = \sigma_t^{-3} E_{t-1}(u_t^3)$ and $m_{4t} = \sigma_t^{-4} E_{t-1}(u_t^4)$, as functions of β , ν , and σ_t , using equations (6) and (7) and the well-known expressions for the moments of ϵ_t and Z_t ; the resulting expressions are quite complicated but can be obtained with standard analytical software such as *Mathematica*. The third moment of u_t exists if $\nu > 6$, and the fourth moment exists if $\nu > 8$. Figure 2 plots m_{3t} and m_{4t} as functions of σ_t , for $\nu = 10$ degrees of freedom and $\beta = -0.5$. As will be seen in Section 5, these values are typical of some asset return series. Figure 2 shows that skewness and kurtosis increase with volatility in the ODLV-GARCH model. In the GHt-GARCH model, by contrast, $E_{t-1}(u_t^3)$ and $E_{t-1}(u_t^4)$ are constants, so that $|m_{3t}|$ and m_{4t} are monotonically decreasing functions of σ_t . So, the two models have very different properties.

4. POSTERIOR SIMULATION

We will first discuss the GHt-GARCH model; the slight modifications needed in the context of the ODLV-GARCH model will be given at the end of this section. The posterior simulator will be a Metropolis-Hastings (MH) algorithm involving the three blocks $\theta_1 = (\nu, \beta, \alpha_0^*, \alpha_1^*, \alpha_2^*, \beta^*)$, $\theta_2 = (\phi_1, \phi_2)$, and $\theta_3 = (Z_1, \dots, Z_T)$. Let $\theta = (\theta_1, \theta_2, \theta_3)$, $y = (y_1, \dots, y_T)$, and let θ_{-i} contain all the parameters in θ except θ_i . We now detail the three steps involved in a single sweep of the MH algorithm.

Step 1. Drawing θ_1 conditional on (θ_{-1}, y) .

This step will use the tailored MH proposal, which was introduced by Chib and Greenberg (1994, 1995). We assume that the prior support of θ_1 is:

$$\{(\nu, \beta, \alpha_0^*, \alpha_1^*, \alpha_2^*, \beta^*) \mid \nu > 4, \alpha_i^* > 0 \text{ for all } i \text{ and } 0 < \beta^* < 1\}.$$

The candidate density will have a location vector obtained by constrained maximization of the posterior log-kernel, and a scale matrix given by a positive

definite approximation to minus the inverse Hessian. It is important that the tails of the candidate density be heavier than those of the target density; this will be the case if the candidate density is Student with a low degrees of freedom parameter. Using (3) and (4), the conditional posterior log-kernel can be written as:

$$k_1(\theta_1) = \ln p_1(\theta_1) + \sum_{t=1}^T k_{1t}(\theta_1), \quad \text{with:}$$

$$k_{1t}(\theta_1) = - \frac{\left(\frac{u_t}{\sigma_t} - \beta \left[Z_t - \frac{\delta^2}{\nu-2} \right] \right)^2}{2Z_t} + \frac{\nu}{2} \ln \left(\frac{\delta^2}{2} \right) - \ln \Gamma \left(\frac{\nu}{2} \right) - \left(\frac{\nu}{2} + 1 \right) \ln Z_t - \frac{\delta^2}{2Z_t} - \ln \sigma_t, \quad (10)$$

where $p_1(\theta_1)$ is the prior kernel, δ^2 is evaluated using (5), and σ_t^2 is evaluated using (2) with $u_t = y_t - \phi_1 - \phi_2 y_{t-1}$. Define:

$$\tilde{\theta}_1 = h(\theta_1) = \left[\ln(\nu - 4), \beta, \ln \alpha_0^*, \ln \alpha_1^*, \ln \alpha_2^*, \ln \left(\frac{\beta^*}{1 - \beta^*} \right) \right], \quad (11)$$

$$\tilde{k}_{1t}(\tilde{\theta}_1) = k_{1t}[h^{-1}(\tilde{\theta}_1)], \quad (12)$$

and let $\tilde{\theta}_1^*$ be an approximate maximizer of $\tilde{k}_1(\tilde{\theta}_1) = k_1[h^{-1}(\tilde{\theta}_1)]$, with $h^{-1}(\tilde{\theta}_1)$ restricted to a compact subset of the prior support. A vector $\tilde{\theta}_1$ is drawn from a multivariate Student distribution with 3 degrees of freedom, expectation $\tilde{\theta}_1^*$, and scale matrix Σ_1 , where:

$$\Sigma_1^{-1} = \sum_{t=1}^T \left(\frac{\partial \tilde{k}_{1t}(\tilde{\theta}_1)}{\partial \tilde{\theta}_1} \frac{\partial \tilde{k}_{1t}(\tilde{\theta}_1)}{\partial \tilde{\theta}_1'} \right)_{\tilde{\theta}_1^*}. \quad (13)$$

The candidate $\theta_1 = h^{-1}(\tilde{\theta}_1)$ is accepted with probability:

$$\min \left[1, \exp \{ k_1(\theta_1) - k_1(\theta_1^{\text{old}}) + f_1[h(\theta_1^{\text{old}})] + J(\theta_1^{\text{old}}) - f_1[h(\theta_1)] - J(\theta_1) \} \right], \quad (14)$$

where θ_1^{old} is the previously drawn vector, $f_1(\cdot)$ is the log-kernel of the multivariate Student density described above, and $J(\theta_1)$ is the log-Jacobian:

$$J(\theta_1) = -\ln(\nu - 4) - \sum_{i=0}^2 \ln \alpha_i^* - \ln[\beta^*(1 - \beta^*)]. \quad (15)$$

If θ_1 is rejected, θ_1^{old} is retained.

The derivatives in (13) can be computed analytically.

Step 2. Drawing θ_2 conditional on (θ_{-2}, y) .

The prior on θ_2 will be the multivariate Normal density:

$$p_2(\theta_2) = f_N(\theta_2; \mu_2, \Omega_2).$$

The MH candidate density will exploit the fact that the observation equation (1) combined with (3) is a linear Gaussian regression model conditional on θ_{-2} and σ_t^2 . We define:

$$\tilde{y}_t(\theta_2) = y_t - \sigma_t(\theta_2)\beta \left[Z_t - \frac{\delta^2}{\nu - 2} \right], \quad (16)$$

where δ^2 is evaluated using (5) and $\sigma_t^2(\theta_2)$ is evaluated using (2) with $u_t = u_t(\theta_2) = y_t - \phi_1 - \phi_2 y_{t-1}$. The candidate density will be multivariate Student with 3 degrees of freedom and scale matrix $\Sigma_2(\theta_2^{\text{old}})$, where:

$$\Sigma_2^{-1}(\theta_2^{\text{old}}) = \Omega_2^{-1} + \sum_{t=1}^T \frac{1}{\sigma_t^2(\theta_2^{\text{old}})Z_t} \begin{pmatrix} 1 & y_{t-1} \\ y_{t-1} & y_{t-1}^2 \end{pmatrix}, \quad (17)$$

and with expectation vector:

$$\theta_2^*(\theta_2^{\text{old}}) = \Sigma_2(\theta_2^{\text{old}}) \left[\Omega_2^{-1} \mu_2 + \sum_{t=1}^T \frac{1}{\sigma_t^2(\theta_2^{\text{old}})Z_t} \begin{pmatrix} \tilde{y}_t(\theta_2^{\text{old}}) \\ y_{t-1} \tilde{y}_t(\theta_2^{\text{old}}) \end{pmatrix} \right], \quad (18)$$

where θ_2^{old} is the previous draw. The candidate θ_2 is accepted with probability:

$$\min \left[1, \exp \{ k_2(\theta_2) - k_2(\theta_2^{\text{old}}) + f_2(\theta_2^{\text{old}}, \theta_2) - f_2(\theta_2, \theta_2^{\text{old}}) \} \right], \quad (19)$$

where $k_2(\theta_2)$ is the conditional posterior log-kernel:

$$k_2(\theta_2) = \ln p_2(\theta_2) + \sum_{t=1}^T k_{2t}(\theta_2), \quad \text{with:}$$

$$k_{2t}(\theta_2) = - \frac{\left(\frac{u_t(\theta_2)}{\sigma_t(\theta_2)} - \beta \left[Z_t - \frac{\delta^2}{\nu - 2} \right] \right)^2}{2Z_t} - \ln \sigma_t(\theta_2), \quad (20)$$

and where $f_2(\theta_2, \theta_2^{\text{old}})$ is the log-kernel of the candidate density with expectation $\theta_2^*(\theta_2^{\text{old}})$ and scale matrix $\Sigma_2(\theta_2^{\text{old}})$. If θ_2 is rejected, θ_2^{old} is retained.

Step 3. Drawing θ_3 conditional on (θ_{-3}, y) .

An important property of Equation (3), mentioned by Barndorff-Nielsen (1997), is that the conditional distribution of Z_t given u_t is a Generalized Inverse Gaussian (GIG) distribution. The GIG distribution is extensively discussed in Jørgensen (1982). Using Bayes' theorem, it is easily seen that the full conditional posterior of Z_t is a GIG with parameters:

$$\lambda = \frac{\nu + 1}{2}, \quad \chi_t = \delta^2 + \left(\frac{u_t}{\sigma_t} + \frac{\beta\delta^2}{\nu - 2} \right)^2 \quad \text{and} \quad \psi = \beta^2. \quad (21)$$

In Appendix A, we will state the explicit form of this distribution, present an efficient algorithm for simulating Z_t , and compare this algorithm with the one proposed by Dagpunar (2007).

In the ODLV-GARCH model, the above equations (10), (16), (17), (18), (20), and (21) are modified as follows:

$$k_{1t}(\theta_1) = -\frac{\left(u_t - \beta \left[Z_t - \frac{\delta_t^2}{\nu - 2} \right] \right)^2}{2Z_t} + \frac{\nu}{2} \ln \left(\frac{\delta_t^2}{2} \right) - \ln \Gamma \left(\frac{\nu}{2} \right) - \left(\frac{\nu}{2} + 1 \right) \ln Z_t - \frac{\delta_t^2}{2Z_t}, \quad (10')$$

$$\tilde{y}_t(\theta_2) = y_t - \beta \left[Z_t - \frac{\delta_t^2}{\nu - 2} \right], \quad (16')$$

$$\Sigma_2^{-1} = \Omega_2^{-1} + \sum_{t=1}^T \frac{1}{Z_t} \begin{pmatrix} 1 & y_{t-1} \\ y_{t-1} & y_{t-1}^2 \end{pmatrix}, \quad (17')$$

$$\theta_2^*(\theta_2^{\text{old}}) = \Sigma_2 \left[\Omega_2^{-1} \mu_2 + \sum_{t=1}^T \frac{1}{Z_t} \begin{pmatrix} \tilde{y}_t(\theta_2^{\text{old}}) \\ y_{t-1} \tilde{y}_t(\theta_2^{\text{old}}) \end{pmatrix} \right], \quad (18')$$

$$k_{2t}(\theta_2) = -\frac{\left(u_t - \beta \left[Z_t - \frac{\delta_t^2}{\nu - 2} \right] \right)^2}{2Z_t} + \frac{\nu}{2} \ln \left(\frac{\delta_t^2}{2} \right) - \frac{\delta_t^2}{2Z_t}, \quad (20')$$

$$\lambda = \frac{\nu + 1}{2}, \quad \chi_t = \delta_t^2 + \left(u_t + \frac{\beta\delta_t^2}{\nu - 2} \right)^2 \quad \text{and} \quad \psi = \beta^2. \quad (21')$$

Note that in (17'), Σ_2 no longer depends on θ_2 , and that δ_t^2 in (10'), (16'), (20'), and (21') is evaluated using (8) with $\sigma_t^2 = \sigma_t^2(\theta_2)$.

5. EMPIRICAL RESULTS

5.1 Data and models.

This section will present empirical results for the models described in Sections 2 and 3, using five samples of daily asset log-returns. In each case, the dependent variable is defined as $y_t = 100 \ln(P_t/P_{t-1})$, where P_t is the closing price index. The data are the S&P500 returns from January 6, 1970 to May 5, 2010 (10180 observations); the FTSE100 returns from April 5, 1984 to May 5, 2010 (6588 observations); the CAC40 returns from March 6, 1990 to May 5, 2010 (5094 observations); the NIKKEI225 returns from January 9, 1984 to April 30, 2010 (6471 observations); and the Swiss Market Index (SMI) returns from November 14, 1990 to May 5, 2010 (4904 observations). All five data sets are freely available on finance.yahoo.com.

In each case, four models will be investigated:

- (1) An AR(1) GHt-GARCH model;
- (2) An AR(0) GHt-GARCH model, where $\phi_1 = \phi_2 = 0$ is imposed in Equation (1);
- (3) An ODLV-GARCH model, where the conditional mean is either AR(0) or AR(1), depending on the relative values of information criteria obtained for the first and second models above;
- (4) The usual threshold t-GARCH model ($\beta = 0$), where the conditional mean is the same as in the ODLV-GARCH model above.

5.2 Model comparison.

The first issue to be considered is model choice. In the Bayesian paradigm, the usual model selection criterion is the marginal likelihood:

$$p(y | M) = \int_{\Theta} f(y | \theta, M) p(\theta | M) d\theta, \quad (22)$$

where $y = (y_1, \dots, y_T)$ is the data vector, θ is the vector of unknown parameters in Model M , $f(y | \theta, M)$ is the likelihood, and $p(\theta | M)$ is the prior. The likelihood of the GHt-GARCH model can be obtained from the prediction error decomposition, upon noting that the marginal density of u_t in (3) can be written as:

$$g_t(u_t) = \frac{2^{\frac{1-\nu}{2}} \delta^\nu |\beta|^{\frac{\nu+1}{2}}}{\sigma_t \Gamma(\frac{\nu}{2}) \sqrt{\pi}} K_{\frac{\nu+1}{2}} \left(|\beta| \sqrt{\delta^2 + z_t^2(u_t)} \right) \exp[\beta z_t(u_t)] \left(\sqrt{\delta^2 + z_t^2(u_t)} \right)^{-\frac{\nu+1}{2}}, \quad (23)$$

where δ^2 is given by (5), and:

$$z_t(u_t) = \frac{u_t}{\sigma_t} + \frac{\beta \delta^2}{\nu - 2}. \quad (24)$$

The marginal likelihood can be used to compute the Bayes factor for Model M_i against Model M_j :

$$BF_{ij} = \frac{p(y | M_i)}{p(y | M_j)}. \quad (25)$$

Jeffreys (1961, Appendix B) proposes a rule of thumb where the evidence against M_j is treated as strong if $\log_{10}(BF_{ij}) < -1$, and decisive if $\log_{10}(BF_{ij}) < -2$.

Another widely used benchmark for model comparison is the Schwarz information criterion, defined here as:

$$BIC(\bar{\theta}) = -2 \ln f(y | \bar{\theta}) + q \ln T, \quad (26)$$

where $\bar{\theta}$ is the average of the MCMC replications of θ and where q is the number of parameters in θ . The lowest value of (26) corresponds to the best model. The magnitude of the difference between Schwarz information criteria is difficult to interpret, but this criterion has the advantage of being insensitive to the prior in large samples; we will therefore report its value along with Bayes factors.

The marginal likelihood in (22) was estimated by the bridge sampling method of Meng and Wong (1996), as implemented in Deschamps (2008). This implementation uses posterior replications as well as an importance sampling density $q(\theta)$, chosen here as a multivariate Normal on:

$$\left(\ln(\nu - 4), \beta, \ln \alpha_0^*, \ln \alpha_1^*, \ln \alpha_2^*, \ln \left(\frac{\beta^*}{1 - \beta^*} \right), \phi_1, \phi_2 \right), \quad (27)$$

with parameters chosen to match the empirical first- and second-order posterior moments obtained by MCMC. The priors were standard Normal for β , ϕ_1 , and ϕ_2 , Gamma(2, 16) for the ARCH parameters α_0^* , α_1^* , and α_2^* , Beta(8, 2) for the

GARCH parameter β^* , and Gamma(10, 1) on $\nu - 4$, the latter implying a prior expectation of 14 for ν . Prior independence was assumed for all parameters. This prior implies a mass of approximately 70% on the parameter subset where the GHt-GARCH variance process is covariance-stationary, with a modal value of $\beta^* + (\alpha_1^* + \alpha_2^*)/2 - 1$ close to zero.

Table 1 reports the logarithmic Bayes factors and Schwarz information criteria. In all cases but one (the S&P500 data), the AR(0) model is clearly preferred, and the Bayes factor evidence in favor of a skewed error distribution ranges from strong (for the S&P500 data) to decisive (for all the other samples). The GHt-GARCH model is preferred to the ODLV-GARCH in all cases, the evidence being very weak for the S&P500 data (with posterior odds less than three to one in favor of GHt-GARCH), strong for the NIKKEI225 and FTSE100 data, and decisive for the SMI and CAC40 data. The fact that the Schwarz criterion yields identical model rankings gives credence to a claim that the rankings obtained by Bayes factors do not depend on our choice of prior. The numerical standard errors (NSE) in Table 1 were computed by the method described in Frühwirth-Schnatter (2004); repeated estimations of the marginal likelihood yielded variability consistent with the reported NSE values.

For comparison purposes, marginal likelihood estimates were also obtained by the method of Chib and Jeliazkov (2001). This method is easiest to implement in the AR(0) case. Let $q(\theta_1 | Z, y)$ be the MH proposal density described in Step 1 of Section 4, let $f^*(Z | y, \theta_1)$ be the conditional posterior of $Z = \theta_3$ in Step 3, and let $\alpha(\theta_1^{\text{old}}, \theta_1 | Z, y)$ be the acceptance probability (14). From Chib (1995) and Chib and Jeliazkov (2001), the marginal likelihood can be estimated by:

$$\ln \hat{p}(y) = \ln p(\bar{\theta}_1) + \ln f(y | \bar{\theta}_1) - \ln \hat{p}(\bar{\theta}_1 | y),$$

where $p(\bar{\theta}_1)$ and $f(y | \bar{\theta}_1)$ are the normalized prior and likelihood evaluated at the posterior mean $\bar{\theta}_1$, and where:

$$\hat{p}(\bar{\theta}_1 | y) = \frac{\frac{1}{m} \sum_{i=1}^m \alpha(\theta_1^i, \bar{\theta}_1 | Z^i, y) q(\bar{\theta}_1 | Z^i, y)}{\frac{1}{m} \sum_{j=1}^m \alpha(\bar{\theta}_1, \theta_1^j | Z^j, y)},$$

in the last expression, the θ_1^i and Z^i are draws from the posterior, and the θ_1^j and Z^j are draws from the density:

$$q^*(\theta_1^j, Z^j | y, \bar{\theta}_1) = q(\theta_1^j | Z^j, y) f^*(Z^j | y, \bar{\theta}_1).$$

TABLE 1. *Bayes factors and information criteria*

Data	Model	BIC	$\log_{10}(BF)$	NSE
S&P500	AR(1), GHt-GARCH	26155.18	0.00	
	AR(0), GHt-GARCH	26195.71	-8.74	0.003
	AR(1), ODLV-GARCH	26157.70	-0.43	0.004
	AR(1), t-GARCH	26159.82	-1.68	0.004
FTSE100	AR(1), GHt-GARCH	17734.46	-1.66	0.002
	AR(0), GHt-GARCH	17727.00	0.00	
	AR(0), ODLV-GARCH	17732.98	-1.31	0.004
	AR(0), t-GARCH	17756.58	-7.24	0.005
NIKKEI225	AR(1), GHt-GARCH	20766.39	-2.41	0.002
	AR(0), GHt-GARCH	20755.16	0.00	
	AR(0), ODLV-GARCH	20761.61	-1.61	0.004
	AR(0), t-GARCH	20767.47	-3.29	0.002
CAC40	AR(1), GHt-GARCH	16412.31	-3.41	0.004
	AR(0), GHt-GARCH	16396.45	0.00	
	AR(0), ODLV-GARCH	16405.93	-2.18	0.005
	AR(0), t-GARCH	16412.66	-4.29	0.004
SMI	AR(1), GHt-GARCH	13732.47	-1.81	0.003
	AR(0), GHt-GARCH	13724.17	0.00	
	AR(0), ODLV-GARCH	13742.44	-3.91	0.004
	AR(0), t-GARCH	13766.99	-10.08	0.004

BIC : Schwarz information criterion.

$\log_{10}(BF)$: Decimal logarithm of Bayes factor against preferred model.

NSE: Numerical standard error of $\log_{10}(BF)$.

Table 2 compares the results obtained by the two methods for the AR(0) GHt-GARCH model. The results of the two methods are almost identical; however, the method of Chib and Jeliazkov (2001), which does not require the evaluation

TABLE 2. *Comparison of bridge sampling and Chib-Jeliazkov marginal likelihood estimates*

Data	Method	$\ln \hat{p}(y)$	NSE	Time
SP500	Chib-Jeliazkov	-13091.8	0.046	831
	Bridge sampling	-13091.6	0.004	1235
FTSE100	Chib-Jeliazkov	-8852.5	0.068	557
	Bridge sampling	-8852.5	0.004	850
NIKKEI225	Chib-Jeliazkov	-10369.6	0.052	528
	Bridge sampling	-10369.8	0.003	756
CAC40	Chib-Jeliazkov	-8187.9	0.073	420
	Bridge sampling	-8187.9	0.007	637
SMI	Chib-Jeliazkov	-6851.8	0.060	427
	Bridge sampling	-6851.8	0.005	633

$\hat{p}(y)$: marginal likelihood estimate for AR(0) GHt-GARCH.

NSE : numerical standard error. Time : execution time (seconds).

of Bessel functions, is approximately 50% faster in this case.

It is instructive to plot the evolution of logarithmic Bayes factors over time when the models are estimated from successive expanding windows. Such plots can serve to identify those observations that contribute the most to the evidence for GHt-GARCH against ODLV-GARCH. The slope of the curve at a particular point gives essentially the same information as the predictive Bayes factor defined by Geweke (2005, p. 66). Figure 3 plots the logarithmic (base 10) Bayes factors of the ODLV-GARCH and GHt-GARCH models against t-GARCH, estimated from such expanding windows; the first window consisted of the first 2000 observations and each succeeding window was obtained by adding 50 observations to the previous one. The last $T - 2000$ sample observations are reported along with the Bayes factor curves. The stepwise evolution of the curves reflects the fact that the posteriors are not updated continuously, but rather by increments of 50 days. The vertical distance between the curves reflects the evidence in favor of ODLV-GARCH against GHt-GARCH (or conversely) accumulated at a

particular point in time, and an ordinate below zero signifies that the cumulative evidence favors t-GARCH.

It is seen in Figure 3 that for the S&P500 and FTSE100 data, the evidence slightly favors the ODLV-GARCH model for most of the sample period; it is only at the very end (corresponding to the onset of the recent financial crisis) that the evidence begins to favor GHt-GARCH. Also, for the S&P500 data, there is weak evidence in favor of the t-GARCH model before mid-2007. For the NIKKEI225 data, the cumulative evidence uniformly favors GHt-GARCH over ODLV-GARCH, and GHt-GARCH is preferred to t-GARCH over the entire sample. For the SMI and CAC40 data, it appears that the dominance of GHt-GARCH over ODLV-GARCH is due to a few observations occurring in 2002 and 2008. Both models are uniformly preferred to t-GARCH, the dominance being very strong for the SMI data.

5.3 Estimation results.

Table 3 reports estimated posterior quantiles and other statistics for the five data sets described in Section 5.1, using the models chosen according to the evidence in Table 1 with the prior described in Section 5.2. The posterior samples were obtained by combining two independent chains, each obtained by running the algorithm of Section 4 for 35000 passes of which the first 10000 were discarded. The final sample was obtained by selecting every fifth replication of the combined chains, yielding a posterior sample of size $(2 \times 25000)/5 = 10000$. The rejection rates ranged from 0.42 to 0.52 in Step 1, and the rate in Step 2 was equal to 0.38 for the S&P500 data. These rates can be reduced by choosing larger values for the candidate degrees of freedom parameter. Convergence was tested by comparing the two chains, using the method of Gelman and Rubin (1992). The validity of the code was tested by the joint distribution method of Geweke (2004). In each case, the credible set for the asymmetry parameter β only covers negative values, implying left-hand skewness of the error distribution. All the estimated variance equations exhibit strong asymmetry ($\alpha_1^* < \alpha_2^*$). They also suggest high persistence, but covariance stationarity: the posterior mass on $\beta^* + (\alpha_1^* + \alpha_2^*)/2 > 1$ equals zero for the S&P500, CAC40, and FTSE100 data, 1.98% for the NIKKEI225 data, and 0.03% for the SMI data. This compares with a prior mass of 30%, as seen in Section 5.2. It is noteworthy that the estimated marginal posterior densities are all far from Normality, except in the case of ϕ_1 and ϕ_2 .

TABLE 3. *Posterior replication summaries (Ght-GARCH)*

Data	θ	$\theta_{0.025}$	$\theta_{0.5}$	$\theta_{0.975}$	RNE	s_θ	m_3	m_4
S&P500	ϕ_1	0.008	0.023	0.039	0.742	0.008	0.043	3.067
	ϕ_2	0.049	0.069	0.089	0.936	0.010	0.011	3.012
	α_0^*	0.006	0.009	0.012	0.761	0.001	0.349	3.214
	α_1^*	0.013	0.021	0.029	0.890	0.004	0.199	3.071
	α_2^*	0.084	0.099	0.115	0.755	0.008	0.199	3.009
	β^*	0.920	0.931	0.941	0.771	0.005	-0.206	3.018
	β	-0.240	-0.149	-0.069	0.251	0.043	-0.198	3.087
	ν	8.130	9.415	11.128	0.026	0.766	0.383	3.117
FTSE100	α_0^*	0.017	0.023	0.030	0.735	0.003	0.261	3.156
	α_1^*	0.027	0.041	0.057	0.871	0.008	0.254	3.116
	α_2^*	0.110	0.131	0.154	0.806	0.011	0.220	3.125
	β^*	0.876	0.894	0.910	0.762	0.009	-0.211	3.085
	β	-0.708	-0.483	-0.309	0.079	0.102	-0.438	3.463
	ν	11.481	14.373	18.696	0.021	1.859	0.685	4.025
NIKKEI225	α_0^*	0.018	0.025	0.032	0.813	0.004	0.289	3.125
	α_1^*	0.025	0.037	0.052	0.880	0.007	0.228	3.041
	α_2^*	0.138	0.163	0.191	0.791	0.014	0.209	3.166
	β^*	0.876	0.892	0.907	0.798	0.008	-0.187	3.122
	β	-0.302	-0.200	-0.112	0.274	0.048	-0.236	3.265
	ν	7.123	8.383	9.919	0.036	0.704	0.320	3.238
CAC40	α_0^*	0.019	0.027	0.038	0.750	0.005	0.346	3.267
	α_1^*	0.006	0.017	0.031	0.785	0.006	0.374	3.253
	α_2^*	0.102	0.123	0.149	0.734	0.012	0.260	3.181
	β^*	0.898	0.915	0.931	0.727	0.008	-0.219	3.240
	β	-0.691	-0.426	-0.232	0.071	0.117	-0.583	3.994
	ν	11.112	14.290	19.016	0.020	1.982	0.614	3.605
SMI	α_0^*	0.020	0.028	0.037	0.778	0.004	0.358	3.182
	α_1^*	0.012	0.027	0.045	0.823	0.009	0.333	3.158
	α_2^*	0.139	0.168	0.201	0.757	0.016	0.280	3.192
	β^*	0.862	0.884	0.904	0.751	0.011	-0.249	3.129
	β	-0.748	-0.516	-0.349	0.082	0.101	-0.541	3.653
	ν	9.333	11.449	14.684	0.026	1.388	0.794	4.574

θ_α : posterior quantile at probability α ; RNE: relative numerical efficiency;
 s_θ : posterior standard deviation; m_3 : posterior skewness; m_4 : posterior kurtosis.

TABLE 4. *Correlation matrix of posterior replications (Ght-GARCH model, S&P500 data)*

	ϕ_1	ϕ_2	α_0^*	α_1^*	α_2^*	β^*	β	ν
ϕ_1	1.00	-0.06	-0.14	0.04	-0.14	0.02	0.20	0.03
ϕ_2		1.00	0.01	-0.03	0.07	-0.04	0.03	0.07
α_0^*			1.00	0.06	0.49	-0.71	-0.02	-0.03
α_1^*				1.00	0.17	-0.50	-0.03	-0.04
α_2^*					1.00	-0.83	-0.02	-0.09
β^*						1.00	0.02	0.01
β							1.00	-0.36
ν								1.00

The posteriors of α_0^* , α_1^* , α_2^* and ν exhibit significant right-hand skewness, and those of β and β^* exhibit significant left-hand skewness. So, Normal asymptotic approximations would not appear to be reliable in spite of the very large sample sizes.

Table 4 reports the posterior correlations between parameters for the S&P500 data; these correlations range from low to moderate, suggesting that the model is not over-parameterized. Figure 4 shows the correlograms, which indicate that selecting every fifth replication eliminates most of the autocorrelation for the AR coefficients ϕ_1 and ϕ_2 , for the ARCH coefficients α_0^* , α_1^* , α_2^* , and for the GARCH coefficient β^* . The autocorrelation of the degrees of freedom parameter ν remains strong, however, and that for the asymmetry parameter β remains moderate.

5.4 Forecast evaluation.

5.4.1 Probability integral transforms. Following Rosenblatt (1952), Kim *et al.* (1998) proposed basing forecast evaluation in GARCH and stochastic volatility models on the one-step ahead predictive distribution function. Normal deviates can be obtained from out of sample predictive probability integral transforms, as:

$$v_t = \Phi^{-1} (P[Y_t \leq y_t | y_{-1}, \dots, y_{t-1}]), \quad (28)$$

where $\Phi(\cdot)$ is the standard Normal integral. An indication that the v_t in (28) are

not independent standard Normal can be interpreted as evidence that the model is misspecified; see, e.g., Diebold *et al.* (1998) and Berkowitz (2001).

Using the mixture representation in (3), the predictive probability in (28) can be estimated from a posterior sample:

$$\left(\nu^{(i)}, \beta^{(i)}, \alpha_0^{*(i)}, \alpha_1^{*(i)}, \alpha_2^{*(i)}, \beta^{*(i)}, \phi_1^{(i)}, \phi_2^{(i)} \right) \quad \text{for } i = 1, \dots, m, \quad (29)$$

by the following average:

$$\frac{1}{m} \sum_{i=1}^m \frac{1}{\ell} \sum_{j=1}^{\ell} \Phi \left(\frac{y_t - \sigma_t^{(i)} \beta^{(i)} \left[Z_t^{(j)} - \delta^{2(i)} / (\nu^{(i)} - 2) \right] - \phi_1^{(i)} - \phi_2^{(i)} y_{t-1}}{\sigma_t^{(i)} \sqrt{Z_t^{(j)}}} \right), \quad (30)$$

where $\sigma_t^{(i)}$ and $\delta^{2(i)}$ are computed using (2) and (5) evaluated at a posterior replication, and where $Z_t^{(j)}$ is drawn from the inverted Gamma distribution in (4) with $\nu = \nu^{(i)}$, $\delta^2 = \delta^{2(i)}$. With $m = 10000$ posterior replications, a number $\ell = 100$ of draws $Z_t^{(j)}$ was found to be adequate.

For each data set, posterior samples based on M rolling estimation windows were obtained. Each estimation window consisted of 2000 observations, and was obtained by advancing the start and end of the previous window by 50 days. Each posterior sample was then used for estimating 50 one-day ahead probability integral transforms, using (30). The $N = 50M$ estimates of the v_t in (28) were then combined and assessed using the following diagnostics:

- (1) An F -statistic, denoted by AR, for testing the nullity of the autoregression coefficients in an AR(6) model of the v_t .
- (2) An F -statistic, denoted by ARCH, for testing the nullity of the autoregression coefficients in an AR(6) model of the v_t^2 .
- (3) The Bera-Jarque statistic, denoted by BJ, for testing the normality of the v_t .
- (4) A likelihood ratio statistic, denoted by LR, for testing the $N(0, 1)$ null against an unrestricted Normal alternative.

Table 5 lists the p-values for the specifications yielding the highest marginal likelihoods and lowest Schwarz criteria in Table 1; for comparison, the p-values obtained with the ODLV-GARCH and t-GARCH formulations are also given. With several thousand observations, significance at the 0.01 level can be viewed as evidence of misspecification. The ODLV-GARCH and GHt-GARCH models

TABLE 5. *Forecast evaluation based on probability integral transforms (p-values)*

Model	Data	N	AR	ARCH	BJ	LR
GHt-GARCH	S&P500	8150	0.02	0.44	0.00	0.58
	FTSE100	4550	0.06	0.22	0.89	0.35
	NIKKEI225	4450	0.93	0.04	0.80	0.38
	CAC40	3050	0.12	0.02	0.52	0.98
	SMI	2900	0.17	0.01	0.59	0.79
ODLV-GARCH	S&P500	8150	0.02	0.27	0.00	0.34
	FTSE100	4550	0.05	0.21	0.55	0.35
	NIKKEI225	4450	0.93	0.03	0.35	0.46
	CAC40	3050	0.12	0.02	0.62	0.97
	SMI	2900	0.18	0.03	0.84	0.81
t-GARCH	S&P500	8150	0.02	0.36	0.00	0.40
	FTSE100	4550	0.08	0.20	0.00	0.26
	NIKKEI225	4450	0.91	0.03	0.03	0.54
	CAC40	3050	0.17	0.03	0.00	0.90
	SMI	2900	0.20	0.03	0.00	0.88

N : sample size. AR: autocorrelation. ARCH: autocorrelation of squares. BJ: Bera-Jarque. LR: LR test for $N(0, 1)$. Based on the v_t in (28). Conditional mean is AR(1) for S&P500, zero for all the other samples.

pass all diagnostic tests, with the exception of the BJ test for the S&P500 data. The significance in this case, however, is entirely due to two outliers occurring on October 19, 1987 and February 27, 2007. When these two observations (in a total of $N = 8150$) are removed, the p-value of the BJ statistic rises to 0.02 for GHt-GARCH and to 0.22 for ODLV-GARCH. In contrast, the BJ statistic in the t-GARCH model is strongly significant in four out of five cases (and remains significant when the two S&P500 outliers are removed), confirming that the t-GARCH model does not fully account for the skewness in the data in spite of the

TABLE 6. *Forecast evaluation based on coverage statistics (p-values)*

Model	Data	N	α							
			0.01 UC	0.025 UC	0.05 UC	0.05 IND	0.05 CC	0.10 UC	0.10 IND	0.10 CC
GHt-GARCH	S&P500	8150	0.87	0.98	0.82	0.35	0.63	0.56	0.34	0.54
	FTSE100	4550	0.50	0.40	0.66	0.69	0.84	0.30	0.35	0.37
	NIKKEI225	4450	0.82	0.23	0.04	0.05	0.02	0.03	0.06	0.01
	CAC40	3050	0.40	0.66	0.27	0.71	0.50	0.37	0.48	0.52
	SMI	2900	0.45	0.16	0.40	0.63	0.62	0.81	0.64	0.87
ODLV-GARCH	S&P500	8150	0.46	0.58	0.94	0.28	0.56	0.46	0.42	0.55
	FTSE100	4550	0.94	0.22	0.66	0.69	0.84	0.23	0.48	0.38
	NIKKEI225	4450	0.49	0.58	0.06	0.03	0.02	0.03	0.13	0.03
	CAC40	3050	0.93	0.93	0.48	0.58	0.67	0.63	0.67	0.81
	SMI	2900	0.34	0.20	0.73	0.28	0.53	0.86	0.61	0.87
t-GARCH	S&P500	8150	0.35	0.32	0.63	0.55	0.75	0.77	0.41	0.68
	FTSE100	4550	0.07	0.25	0.27	0.79	0.52	0.70	0.50	0.74
	NIKKEI225	4450	0.71	0.12	0.03	0.02	0.01	0.02	0.03	0.01
	CAC40	3050	0.25	0.10	0.03	0.59	0.07	0.28	0.41	0.40
	SMI	2900	0.36	0.27	0.01	0.32	0.02	0.62	0.74	0.84

N : sample size. UC: unconditional coverage. IND: independence.

CC: conditional coverage. α : VaR probability.

asymmetric formulation of the variance equation.

5.4.2 Coverage. Forecast evaluation can also be based on the methodology of Christoffersen (1998). The previously described M posterior samples can be used for estimating $N = 50M$ values at risk by the empirical quantiles of the simulated one-day ahead predictives at probability α , denoted by q_t^α . Upon defining:

$$\begin{aligned} \mathcal{I}_t^\alpha &= 1 && \text{if } y_t < q_t^\alpha, \\ &= 0 && \text{otherwise,} \end{aligned}$$

the following tests are proposed:

- (1) UC : unconditional coverage : test of $p = \alpha$, where p is the probability that \mathcal{I}_t^α equals one, assuming that the \mathcal{I}_t^α have independent Bernoulli distributions.
- (2) IND : test of independence versus the alternative that the \mathcal{I}_t^α follow a Markov chain.
- (3) CC : conditional coverage: joint test of unconditional coverage and independence.

Table 6 lists the p-values of coverage and independence statistics. For $\alpha = 0.01$ and $\alpha = 0.025$, the IND and CC statistics cannot always be computed due to insufficient numbers of observations; for these values of α , we therefore only report the UC statistic. All three models appear to be acceptable from a coverage standpoint. There are, however, substantial differences between the 5% value at risk estimates from symmetric and asymmetric versions of threshold t-GARCH. As expected, the t-GARCH estimates are generally more conservative (i.e. less negative).

6. A MARKOV SWITCHING EXTENSION

6.1 The model.

In this section, we will investigate the empirical performance of a Markov switching extension of the GHt-GARCH model. In order to avoid the well-known problem of path dependence, the methodology proposed by Haas *et al.* (2004) can be adopted: a separate GARCH process is associated with each latent regime. For $t = 1, \dots, T$, we let $s_t \in \{1, \dots, K\}$ follow a first-order Markov chain, with:

$$P(s_{t+1} = j \mid s_t = i) = p_{ij} \quad \text{and} \quad (31)$$

$$P(s_1 = j) = \frac{1}{K}. \quad (32)$$

The MS-GHt-GARCH model reads as:

$$y_t = \phi_1 + \phi_2 y_{t-1} + u_t \quad \text{for } t = 1, \dots, T, \quad (33)$$

$$u_t = \sigma_t(s_t) \left[\beta(s_t) \left(Z_t - \frac{\delta^2(s_t)}{\nu(s_t) - 2} \right) + \sqrt{Z_t} \epsilon_t \right], \quad (34)$$

$$\delta^2(s_t) = \frac{(\nu(s_t) - 2)(\nu(s_t) - 4)}{4\beta^2(s_t)} \left[-1 + \sqrt{1 + \frac{8\beta^2(s_t)}{\nu(s_t) - 4}} \right] \quad \text{and} \quad (35)$$

$$\sigma_t^2(j) = \alpha_0^*(j) + [\alpha_1^*(j)\mathcal{I}(u_{t-1} \geq 0) + \alpha_2^*(j)\mathcal{I}(u_{t-1} < 0)] u_{t-1}^2 + \beta^*(j)\sigma_{t-1}^2(j) \quad \text{for } j = 1, \dots, K, \quad (36)$$

where, in (34), ϵ_t is a standard Normal random variable and Z_t has an inverted Gamma distribution with parameters $\nu(s_t)/2$ and $\delta^2(s_t)/2$. Equation (36) is initialized by letting $u_0 = y_0 - \phi_1 - \phi_2 y_{-1}$, and $\sigma_0^2(j) = y_0^2$ for all j . Note that Equation (36) defines K separate GARCH processes; the value j of s_t determines the particular variance process in effect at time t .

6.2 Posterior simulation.

It will be helpful to denote by Θ_1 the $K \times 6$ matrix with row j given by:

$$\theta_1(j) = (\nu(j) \quad \beta(j) \quad \alpha_0^*(j) \quad \alpha_1^*(j) \quad \alpha_2^*(j) \quad \beta^*(j)),$$

and by θ_2 the vector (ϕ_1, ϕ_2) .

A single sweep of the posterior simulator involves the $K + 4$ blocks $S = (s_1, \dots, s_T)$, $P = [p_{ij}]$, $Z = (Z_1, \dots, Z_T)$, $\theta_1(j)$ for $j = 1, \dots, K$, and θ_2 . The idea of involving the random states S in the simulation was introduced by Albert and Chib (1993). Following Frühwirth-Schnatter (2001), each sweep of the simulator is followed by a permutation of the regime definitions: $\theta_1(j)$ is replaced by $\theta_1(\pi(j))$, s_t by $\pi(s_t)$, and p_{ij} by $p_{\pi(i), \pi(j)}$, where $\pi = (\pi(1), \dots, \pi(K))$ is a permutation of $(1, \dots, K)$. The vector π can either be selected at random, or determined by an appropriate identification constraint; see Frühwirth-Schnatter (2001, 2006) for details.

6.2.1. Drawing S and P from their full conditional posterior distributions is straightforward; the methods are fully described in Chib (1996) and Frühwirth-Schnatter (2006, chapter 11). Note that this step does not depend on Z : drawing the latent states s_t uses a forward-filtering backward-sampling algorithm, with filtering based on the predictive density (23) rather than on the mixture representation (3)–(4).

6.2.2. Drawing Z conditional on S , Θ_1 , and θ_2 can be done by the algorithm of Appendix A, upon noting that Z_t has a GIG distribution with parameters:

$$\lambda_t = \frac{\nu(s_t) + 1}{2}, \quad \chi_t = \delta^2(s_t) + \left(\frac{u_t}{\sigma_t(s_t)} + \frac{\beta(s_t)\delta^2(s_t)}{\nu(s_t) - 2} \right)^2 \quad \text{and} \quad \psi_t = \beta^2(s_t),$$

where $u_t = y_t - \phi_1 - \phi_2 y_{t-1}$, and $\delta^2(s_t)$ is obtained by (35). Note that computing the $\sigma_t^2(s_t)$ involves generating the entire path $\{\sigma_t^2(j)\}_{t=1}^T$ for each value of j between 1 and K .

6.2.3. Drawing $\theta_1(j)$ conditional on θ_2 , S , Z , and $\theta_1(k)$ for all $k \neq j$ can be done along the lines of Step 1 in Section 4, upon noting that the conditional posterior log-kernel is given by:

$$\ln p_1(\theta_1(j)) + \sum_{t \in T_j} k_{1t}(\theta_1(j)),$$

where $p_1(\cdot)$ is a prior kernel, $T_j = \{t \mid s_t = j\}$, and $k_{1t}(\theta_1(j))$ is obtained by replacing, in Equation (10), ν by $\nu(j)$, β by $\beta(j)$, δ^2 by $\delta^2(j)$, and σ_t by $\sigma_t(j)$. Indeed, $\sigma_t^2(s_t)$ depends only on $\theta_1(j)$ when $t \in T_j$, and on $\theta_1(k)$ for $k \neq j$ when $t \notin T_j$. So, prior independence also implies the posterior conditional independence of $\theta_1(j)$ and $\theta_1(k)$ for $j \neq k$. The sum in (13) is now over $t \in T_j$, and an identity matrix is added to (13) to guard against the possibility that Σ_1^{-1} is not positive definite.

6.2.4. Finally, θ_2 is drawn conditionally on Θ_1 , Z , and S by a straightforward modification of Step 2 in Section 4. Equations (16), (17), and (18) respectively become:

$$\tilde{y}_t(\theta_2, s_t) = y_t - \sigma_t(\theta_2, s_t)\beta(s_t) \left[Z_t - \frac{\delta^2(s_t)}{\nu(s_t) - 2} \right],$$

$$\Sigma_2^{-1}(\theta_2^{\text{old}}) = \Omega_2^{-1} + \sum_{t=1}^T \frac{1}{\sigma_t^2(\theta_2^{\text{old}}, s_t) Z_t} \begin{pmatrix} 1 & y_{t-1} \\ y_{t-1} & y_{t-1}^2 \end{pmatrix},$$

and

$$\theta_2^*(\theta_2^{\text{old}}) = \Sigma_2(\theta_2^{\text{old}}) \left[\Omega_2^{-1} \mu_2 + \sum_{t=1}^T \frac{1}{\sigma_t^2(\theta_2^{\text{old}}, s_t) Z_t} \begin{pmatrix} \tilde{y}_t(\theta_2^{\text{old}}, s_t) \\ y_{t-1} \tilde{y}_t(\theta_2^{\text{old}}, s_t) \end{pmatrix} \right],$$

where $\sigma_t^2(\theta_2, s_t)$ is evaluated using (36) with $u_t = u_t(\theta_2) = y_t - \phi_1 - \phi_2 y_{t-1}$. In the conditional posterior log-kernel (20), $k_{2t}(\theta_2)$ becomes:

$$-\frac{\left(\frac{u_t(\theta_2)}{\sigma_t(\theta_2, s_t)} - \beta(s_t) \left[Z_t - \frac{\delta^2(s_t)}{\nu(s_t) - 2} \right] \right)^2}{2Z_t} - \ln \sigma_t(\theta_2, s_t).$$

6.3 Empirical results.

Estimating the Markov switching model of this section implies choosing the number K of regimes. This number could be chosen by maximizing the marginal likelihood. In this context, Frühwirth-Schnatter (2004) recommends using bridge sampling in conjunction with the random (i.e. unconstrained) permutation sampler, using an importance sampling density constructed from the MCMC transition kernel. This method appears to be effective in the Gibbsian case; see Frühwirth-Schnatter (2004) and Deschamps (2008). However, its performance with the tailored Metropolis-Hastings proposal densities used in this paper appears to be uncharted territory. It can also be mentioned that the marginal likelihood is not useful for selecting an identification constraint, the unconstrained marginal likelihood being always $K!$ times the constrained one.

When the only objective is the choice of K , a heuristic, but much simpler procedure can be used. The random permutation sampler is run for several values of K , and the largest value of K yielding noticeable replication clustering is chosen. The identification constraint suggested by this clustering is then imposed in a second application of MCMC, and the estimated Markov switching model is compared with its single-regime counterpart using the Schwarz information criterion and misspecification diagnostics. This will be the approach used in the present paper.

For illustrative purposes, we report in Figure 5 some scatter diagrams obtained by running the random permutation sampler on the NIKKEI225 data, for $K = 2$ and $K = 3$. The diagrams for $K = 2$ clearly suggest the identification constraint $\beta(1) < \beta(2)$, and those for $K = 3$ clearly suggest that the latter model is over-parameterized (the results for $K = 4$ were similar).

Table 7 summarizes the Schwarz criteria obtained from the best single-regime models and those obtained from the corresponding two-regime MS-GHt-GARCH, when the identification constraints suggested by the random permutation sampler are imposed. The priors on the persistence probabilities p_{11} and p_{22} were independent Beta(18,2), and those for the other parameters were the same as in Section 5. The information criteria in Table 7 favor the single-regime model in all cases except for the S&P500 and NIKKEI225 data.

In Table 8, in-sample misspecification diagnostics are reported; they were computed using estimates of the v_t in (28) based on posterior replications conditional

TABLE 7. *Schwarz information criteria (single and two-regime models)*

Data	GHt-GARCH		MS-GHt-GARCH ($K = 2$)	
	BIC		Id. var.	BIC
S&P500	26155.18		α_0^*	26143.84
FTSE100	17727.00		α_0^*	17746.06
NIKKEI225	20755.16		β	20734.05
CAC40	16396.45		β^*	16448.22
SMI	13724.17		ν	13749.34

BIC: Schwarz criterion; Id. var: identifying variable.

TABLE 8. *In-sample diagnostics (single and two-regime models, p-values)*

Data	GHt-GARCH				MS-GHt-GARCH ($K = 2$)			
	AR	ARCH	BJ	LR	AR	ARCH	BJ	LR
S&P500	0.15	0.05	0.53	0.99	0.14	0.00	0.65	0.98
FTSE100	0.04	0.23	0.85	0.04	0.05	0.25	0.49	0.05
NIKKEI225	0.59	0.37	0.85	0.83	0.50	0.15	0.25	0.75
CAC40	0.01	0.07	0.88	0.78	0.01	0.04	0.97	0.75
SMI	0.01	0.02	0.66	0.06	0.01	0.00	0.87	0.04

AR: autocorrelation. ARCH: autocorrelation of squares.

BJ: Bera-Jarque. LR: LR test for $N(0, 1)$. Based on the v_t in (28).

Conditional mean is AR(1) for S&P500, zero for all the other samples.

on the entire sample (using rolling windows would have been impractical in the MS-GHt-GARCH model). For comparison, the corresponding single-regime diagnostics are also reported. The ARCH statistics for the MS-GHt-GARCH model estimated on the S&P500 and SMI data are significant, providing some evidence against this model for these data sets.

TABLE 9. *Posterior replication summaries*
(MS-GHt-GARCH, NIKKEI225 data)

	θ	$\theta_{0.025}$	$\theta_{0.5}$	$\theta_{0.975}$	RNE	s_θ	m_3	m_4
Regime 1	α_0^*	0.043	0.068	0.107	0.292	0.017	0.685	3.814
	α_1^*	0.003	0.019	0.058	0.479	0.015	1.247	5.236
	α_2^*	0.152	0.230	0.345	0.343	0.048	0.602	3.598
	β^*	0.680	0.779	0.847	0.246	0.043	-0.589	3.606
	β	-0.895	-0.566	-0.346	0.060	0.139	-0.656	3.968
	ν	6.259	8.053	10.965	0.032	1.210	0.709	3.710
Regime 2	α_0^*	0.046	0.067	0.097	0.596	0.013	0.632	3.764
	α_1^*	0.007	0.018	0.034	0.802	0.007	0.393	3.100
	α_2^*	0.117	0.144	0.175	0.799	0.015	0.377	4.701
	β^*	0.869	0.892	0.911	0.828	0.011	-2.246	44.639
	β	-0.234	-0.083	0.068	0.254	0.075	0.016	3.356
	ν	9.097	11.269	15.102	0.024	1.568	0.835	4.125
Transition	p_{11}	0.990	0.995	0.998	0.927	0.002	-0.718	3.727
Probs.	p_{22}	0.996	0.998	0.999	0.831	0.001	-0.971	4.687

θ_α : posterior quantile at probability α ; RNE: relative numerical efficiency;
 s_θ : posterior standard deviation; m_3 : posterior skewness; m_4 : posterior kurtosis.

Since the joint evidence provided by Tables 7 and 8 favor the two-regime model only in the NIKKEI225 case, we report in Table 9 posterior replication summaries for this data set only. The first regime is characterized by a larger deviation from Normality (lower values of ν and β), higher ARCH asymmetry (value of α_2^*) and lower GARCH persistence (value of β^*). The smoothed regime probabilities are presented in Figure 6. Figure 6 shows that the first regime is less persistent than the second, since it occurs for five relatively brief spells covering less than 4 years each. The first regime appears to be generally associated with periods of low volatility.

7. DISCUSSION AND CONCLUSIONS

This paper has proposed efficient posterior simulators for two GARCH models having a skewed heavy-tailed error distribution. The first model (GHt-GARCH) is a threshold version of the GARCH model in Aas and Haff (2006), and does not appear to have been estimated before by Bayesian methods. The second (ODLV-GARCH) can be viewed as an observation-driven stochastic volatility model, in the sense of Barndorff-Nielsen (1997); its empirical performance has apparently never been investigated.

Using five publicly available asset return data sets, the new models were found to be significant improvements over classical t-GARCH formulations. In particular, an asymmetric formulation of the conditional variance did not appear sufficient to explain fully the skewness of asset return distributions. The marginal likelihoods for the GHt-GARCH model were found to dominate those of the ODLV-GARCH model; however, for two of the samples, the evidence in favor of GHt-GARCH was not decisive, and only appeared at the very end of the sample (see Figure 3).

A Markov switching extension of the GHt-GARCH model was also presented. Its estimation only required a relatively straightforward modification of the algorithm for the single-regime model. The estimation cost of this MS-GHt-GARCH model is much larger than that of its single-regime counterpart, due to the many evaluations of the Bessel function in (23) that are necessary for drawing the latent variables s_t (see Section 6.2.1). The Markov switching model outperformed the single-regime GHt-GARCH in only one instance, for the NIKKEI225 data.

This paper has not attempted to compare the GHt-GARCH and ODLV-GARCH models with GARCH models based on the other skewed densities that have been proposed in the literature. Apart from the fact that many such densities are known (see Paoletta (2007, Sec. 7.1.2) for a short survey), it is not clear that estimating these models by MCMC would be straightforward. Comparing our models with alternative non-nested formulations would therefore be outside the scope of this paper.

APPENDIX A. EFFICIENT SIMULATION FROM THE
GENERALIZED INVERSE GAUSSIAN DISTRIBUTION

The generalized inverse Gaussian (GIG) density can be written as:

$$f_{\text{GIG}}(x; \lambda, \chi, \psi) = \frac{(\chi/\psi)^{\lambda/2}}{2K_\lambda(\sqrt{\chi\psi})} x^{-(\lambda+1)} \exp\left[-\frac{1}{2}(\chi x^{-1} + \psi x)\right] \quad \text{for } x > 0, \quad (\text{A1})$$

K_λ being the modified Bessel function of the third kind with index λ . The parameters in (A1) are constrained by the inequalities $\chi > 0$, $\psi > 0$, and we will assume that $\lambda > 0$; the case where $\lambda < 0$ can be handled by noting that $x^{-1} \sim \text{GIG}(-\lambda, \psi, \chi)$ whenever $x \sim \text{GIG}(\lambda, \chi, \psi)$.

The kernel of the GIG density can be seen to be bounded above by that of the inverted Gamma (IG) density, since this density can be written as:

$$f_{\text{IG}}(x; \lambda, \chi) = \frac{(\chi/2)^\lambda}{\Gamma(\lambda)} x^{-(\lambda+1)} \exp\left[-\frac{1}{2}(\chi x^{-1})\right] \quad \text{for } x > 0. \quad (\text{A2})$$

A natural rejection method for drawing a GIG variate is therefore based on an IG source density, and can be formulated as:

Algorithm 1: Inverted Gamma rejection sampling (IGRS).

- (1) Draw $x \sim \text{IG}(\lambda, \chi)$.
- (2) Draw $u \sim U(0, 1)$.
- (3) If $u \leq \exp[-(\psi x)/2]$, return x ; otherwise go to Step (1). \square

The unconditional acceptance probability in this method can be seen to be:

$$p(\lambda, \chi, \psi) = \frac{(\chi/2)^\lambda}{\Gamma(\lambda)} \frac{2K_\lambda(\sqrt{\chi\psi})}{(\chi/\psi)^{\lambda/2}} = \frac{2K_\lambda(\omega)(\omega/2)^\lambda}{\Gamma(\lambda)}, \quad (\text{A3})$$

with $\omega = \sqrt{\chi\psi}$.

The probability (A3) is decreasing in ω and increasing in λ . When ω is large relative to λ , it can become unacceptably low, and a general method for drawing from log-concave densities will be more efficient than IGRS.

Even though the GIG density of X is not log-concave when $\lambda > 0$, the density of $Y = \ln X$, which is:

$$f_Y(y) = \frac{1}{2K_\lambda(\omega)} \exp[-\lambda(y - \xi) - \omega \cosh(y - \xi)], \quad (\text{A4})$$

with:

$$\omega = \sqrt{\chi\psi} \quad \text{and} \quad \xi = \frac{1}{2} \ln \left(\frac{\chi}{\psi} \right),$$

is easily seen to be log-concave. When ω is large, (A4) is close to a Normal density with an expectation given by the modal value:

$$y^* = \xi - \ln \left[\frac{\lambda}{\omega} + \sqrt{\frac{\lambda^2}{\omega^2} + 1} \right], \quad (\text{A5})$$

and with a variance smaller than $1/\omega$; see Jørgensen (1982). The approximate Normality of (A4) means that the envelope rejection sampling method of Gilks and Wild (1992) is likely to be effective. In our context, two simplifications of their algorithm can be made: first, it is not necessary to use a squeezing function since the log-kernel in (A4) is easy to compute; second, an adaptive method is not relevant since multiple draws from a single density are not needed. For the sake of completeness, we now describe our implementation of this algorithm. It depends on an integer $k \geq 1$; choosing k is a compromise between the setup cost and the acceptance probability.

Algorithm 2: Envelope rejection sampling (ERS).

(1) Let, for $i = 0, \dots, k$:

$$y_i = y_{\min} + \frac{(y^* - y_{\min})i}{k},$$

and, for $i = k + 1, \dots, 2k$:

$$y_i = y^* + \frac{(y_{\max} - y^*)(i - k)}{k},$$

where y^* is the mode given by (A5) and where $y_{\min} = y^* - 3/\sqrt{\omega}$, $y_{\max} = y^* + 3/\sqrt{\omega}$. The approximate Normality of (A4) ensures that the $2k + 1$ points y_i cover an interval where this density is high.

(2) Compute, for $i = 0, \dots, 2k$, the log-kernels and derivatives:

$$g_i = -\lambda(y_i - \xi) - \omega \cosh(y_i - \xi),$$

and

$$g'_i = -\lambda - \omega \sinh(y_i - \xi).$$

Note that by construction, $g'_k = 0$.

- (3) Compute, for $i = 0, \dots, 2k - 1$:

$$z_i = \frac{g_{i+1} - g_i - y_{i+1}g'_{i+1} + y_i g'_i}{g'_i - g'_{i+1}}.$$

The z_i are the points where the tangents to the log-kernel drawn at consecutive y_i intersect.

- (4) Compute a piecewise linear upper hull of the log-kernel as follows:

$$u(y) = \sum_{i=0}^{2k} (\alpha_i + \beta_i y) \mathcal{I}_{(z_{i-1}, z_i)}(y),$$

where \mathcal{I} denotes an indicator function, and where:

$$\alpha_i = g_i - g'_i y_i, \quad \beta_i = g'_i, \quad z_{-1} = -\infty \quad \text{and} \quad z_{2k} = +\infty.$$

The source density kernel of the ERS method will be $\exp[u(y)]$. Note that $\beta_i > 0$ for $i < k$, $\beta_k = 0$, and $\beta_i < 0$ for $i > k$.

- (5) Compute a discrete probability function (p_0, \dots, p_{2k}) by evaluating:

$$h_i = \alpha_i + \beta_i z_i - \ln |\beta_i| \quad \text{for } i = 0, \dots, 2k - 1 \text{ and } i \neq k,$$

$$h_i^* = \alpha_i + \beta_i z_{i-1} - \ln |\beta_i| \quad \text{for } i = 1, \dots, 2k \text{ and } i \neq k,$$

$$h_k = h_k^* = \alpha_k,$$

$$h_{\max} = \max_{1 \leq i \leq 2k} \max(h_{i-1}, h_i^*),$$

$$I_0 = \exp(h_0 - h_{\max}),$$

$$I_i = |\exp(h_i - h_{\max}) - \exp(h_i^* - h_{\max})| \quad \text{for } i = 1, \dots, 2k - 1 \text{ and } i \neq k,$$

$$I_k = (z_k - z_{k-1}) \exp(\alpha_k - h_{\max}),$$

$$I_{2k} = \exp(h_{2k}^* - h_{\max}),$$

and defining:

$$p_i = \frac{I_i}{\sum_{j=0}^{2k} I_j}.$$

The p_i can be seen to be proportional to the integrals:

$$\int_{z_{i-1}}^{z_i} \exp(\alpha_i + \beta_i y) dy.$$

- (6) Draw an integer ℓ between 0 and $2k$ using the discrete probability function above.
- (7) Draw $u \sim U(0, 1)$. Conditional on the value ℓ drawn in Step (6), let:

$$y = z_\ell + \frac{\ln u}{\beta_\ell} \quad \text{if } \ell = 0,$$

$$y = z_\ell + \frac{\ln[u + (1 - u) \exp[\beta_\ell(z_{\ell-1} - z_\ell)]]}{\beta_\ell} \quad \text{if } 1 \leq \ell \leq k - 1,$$

$$y = z_{k-1} + (z_k - z_{k-1})u \quad \text{if } \ell = k,$$

$$y = z_{\ell-1} + \frac{\ln[u + (1 - u) \exp[\beta_\ell(z_\ell - z_{\ell-1})]]}{\beta_\ell} \quad \text{if } k + 1 \leq \ell \leq 2k - 1,$$

and

$$y = z_{\ell-1} + \frac{\ln u}{\beta_\ell} \quad \text{if } \ell = 2k.$$

- (8) Let ℓ be the value drawn in Step (6) and y be the value drawn in Step (7). Evaluate:

$$g(y) = -\lambda(y - \xi) - \omega \cosh(y - \xi).$$

Draw $u \sim U(0, 1)$. If $u \leq \exp[g(y) - \alpha_\ell - \beta_\ell y]$ return $x = \exp(y)$. Otherwise go to Step (6). \square

We will assume in what follows that $\lambda \geq 2.5$, since this condition is necessary for the existence of the variance of the GHSST distribution in this paper (recall that $\lambda = (\nu + 1)/2$, where ν is the degrees of freedom parameter of this distribution). As mentioned at the beginning of this Appendix, the ERS method will be preferable to IGRS when (A3) yields a low value. Evaluating (A3) is not always practical, since the Bessel function can overflow when λ is large relative to ω . However, for every given value of $\omega = \sqrt{\chi\psi}$, the IGRS acceptance probability attains a minimum over the interval $2.5 \leq \lambda < \infty$ when $\lambda = 2.5$, and it can be checked using standard root-finding algorithms that, in (A3), $p(2.5, \chi, \psi) \geq 0.10$ whenever $\omega \leq 4.95$. This suggests using IGRS whenever $\omega \leq 4.95$ and ERS otherwise. Indeed, the two methods are complementary: as $\omega \rightarrow 0$, (A4) tends to an

improper density and ERS will no longer be efficient, but the IGRS acceptance probability in (A3) tends to one.

The method was tested using 100,000 uniformly distributed random values of λ , χ , and ψ , with $2.5 \leq \lambda \leq 102.5$, $1 \leq \chi \leq 101$, and $0 \leq \psi \leq 100$. Using the low value of $k = 2$, the 100,000 GIG draws required about 600 milliseconds on a 3.2 GHz workstation, using compiled code.

When $\omega > 10^{11}$, (A4) essentially becomes a Dirac mass and the modal value $\exp(y^*)$ can be safely returned.

We will now compare the performance of our algorithm with the one proposed by Dagpunar (2007). Table 10 reports the acceptance probabilities of both methods for combinations of values of λ and ω . The acceptance probability of Dagpunar's method can be computed analytically (Dagpunar, 2007, p. 72). That of the ERS method was estimated by simulation, by drawing 1,000,000 GIG variates and recording, for each draw i , the number N_i of tries (including the successful one). Since $N_i - 1$ follows a geometric distribution, we have $E(N_i) = 1/p$, and the average of the N_i provides a consistent estimate of the inverse of the acceptance probability p .

It is seen in Table 10 that when $\omega \leq 4.95$, Dagpunar's method yields higher acceptance probabilities than the IGRS method described in this paper. However, performing the acceptance test in Dagpunar's method is slightly more elaborate than performing the IGRS counterpart. When $\omega > 4.95$, the ERS probabilities tend to dominate Dagpunar's method, at least for moderate values of λ , and this dominance becomes more obvious as ω increases. It can also be noted that the ERS probabilities could be increased by choosing a value of k greater than two.

In fact, when ω is large, Dagpunar (2007) recommends drawing GIG variates by the ratio of uniforms method in Dagpunar (1989). Implementing this method requires approximating the roots of a polynomial equation by numerical methods, and the acceptance test in Dagpunar (1989, p. 707) is subject to the potential error in this approximation. By contrast, the ERS method is an exact one. For this reason, the author would recommend ERS whenever Dagpunar's (2007) method turns out to be inefficient.

TABLE 10. *Comparison of acceptance probabilities (GIG simulation)*

ω	Method	λ					
		3	4	6	8	10	20
0.01	IGRS	1.000	1.000	1.000	1.000	1.000	1.000
	DAGPUNAR	1.000	1.000	1.000	1.000	1.000	1.000
0.1	IGRS	0.999	0.999	1.000	1.000	1.000	1.000
	DAGPUNAR	1.000	1.000	1.000	1.000	1.000	1.000
0.5	IGRS	0.970	0.980	0.988	0.991	0.993	0.997
	DAGPUNAR	0.990	0.995	0.998	0.999	0.999	1.000
4.95	IGRS	0.134	0.204	0.335	0.439	0.520	0.726
	DAGPUNAR	0.692	0.771	0.865	0.915	0.942	0.985
5	ERS ($k = 2$)	0.885	0.874	0.852	0.824	0.796	0.674
	DAGPUNAR	0.689	0.768	0.864	0.913	0.941	0.984
10	ERS ($k = 2$)	0.898	0.895	0.888	0.878	0.866	0.798
	DAGPUNAR	0.516	0.592	0.703	0.779	0.833	0.943
10^2	ERS ($k = 2$)	0.907	0.906	0.906	0.906	0.906	0.905
	DAGPUNAR	0.168	0.196	0.241	0.279	0.312	0.441
10^3	ERS ($k = 2$)	0.907	0.907	0.907	0.907	0.907	0.907
	DAGPUNAR	0.053	0.062	0.076	0.089	0.099	0.141
10^6	ERS ($k = 2$)	0.907	0.907	0.907	0.907	0.907	0.907
	DAGPUNAR	0.002	0.002	0.002	0.003	0.003	0.004

ACKNOWLEDGMENTS

The author wishes to thank two anonymous referees for their helpful comments on previous versions.

REFERENCES

- Aas, K., Haff, I.H., 2006. The generalized hyperbolic skew Student's t -distribution. *Journal of Financial Econometrics* 4, 275–309.
- Albert, J., Chib, S., 1993. Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business and Economic Statistics* 11, 1–15.
- Ardia, D., 2008. *Financial Risk Management with Bayesian Estimation of GARCH Models: Theory and Applications*. Springer-Verlag, Berlin.
- Barndorff-Nielsen, O.E., 1977. Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London Series A* 353, 401–419.
- Barndorff-Nielsen, O.E., 1997. Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics* 24, 1–13.
- Bauwens, L., Laurent, S., 2005. A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. *Journal of Business and Economic Statistics* 23, 346–354.
- Berkowitz, J., 2001. Testing density forecasts, with applications to risk management. *Journal of Business and Economic Statistics* 19, 465–474.
- Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.
- Bollerslev, T., Engle, R.F., Nelson, D.B., 1994. ARCH models, in: R.F. Engle, McFadden, D.L. (Eds.), *Handbook of Econometrics*, vol. 4. Elsevier, Amsterdam, pp. 2959–3038.
- Chib, S., 1995. Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association* 90, 1313–1321.
- Chib, S., 1996. Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics* 75, 79–97.
- Chib, S., Greenberg, E., 1994. Bayes inference in regression models with ARMA(p, q) errors. *Journal of Econometrics* 64, 183–206.
- Chib, S., Greenberg, E., 1995. Understanding the Metropolis-Hastings algorithm. *The American Statistician* 49, 327–335.
- Chib, S., Jeliazkov, I., 2001. Marginal likelihood from the Metropolis-Hastings output. *Journal of the American Statistical Association* 96, 270–281.
- Christoffersen, P.F., 1998. Evaluating interval forecasts. *International Economic Review* 39, 841–862.

- Dagpunar, J.S., 1989. An easily implemented generalised inverse Gaussian generator. *Communications in Statistics – Simulation and Computation* 18, 703–710.
- Dagpunar, J.S., 2007. *Simulation and Monte Carlo, with Applications in Finance and MCMC*. Wiley, Chichester.
- Dark, J.G., 2010. Estimation of time varying skewness and kurtosis with an application to value at risk. *Studies in Nonlinear Dynamics and Econometrics* 14, issue 2, article 3.
- Deschamps, P.J., 2008. Comparing smooth transition and Markov switching autoregressive models of US unemployment. *Journal of Applied Econometrics* 23, 435–462.
- Diebold, F.X., Gunther, T.A., Tay, A.S., 1998. Evaluating density forecasts with applications to financial risk management. *International Economic Review* 39, 863–883.
- French, K.R., Schwert, G.W., Stambaugh, R.F., 1987. Expected stock returns and volatility. *Journal of Financial Economics* 19, 3–29.
- Frühwirth-Schnatter, S., 2001. Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association* 96, 194–209.
- Frühwirth-Schnatter, S., 2004. Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques. *Econometrics Journal* 7, 143–167.
- Frühwirth-Schnatter, S., 2006. *Finite Mixture and Markov Switching Models*. Springer, New York.
- Gelman, A., Rubin, D., 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7, 457–511.
- Geweke, J., 1993. Bayesian treatment of the independent Student-t linear model. *Journal of Applied Econometrics* 8 (supplement), S19–S40.
- Geweke, J., 2004. Getting it right: Joint distribution tests of posterior simulators. *Journal of the American Statistical Association* 99, 799–804.
- Geweke, J., 2005. *Contemporary Bayesian Statistics and Econometrics*. Wiley, Hoboken.
- Geweke, J., Amisano, G., 2010. Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting* 26, 216–230.
- Gilks, W.R., Wild, P., 1992. Adaptive rejection sampling for Gibbs sampling. *Journal of the Royal Statistical Society Series C (Applied Statistics)* 41, 337–348.
- Giot, P., Laurent, S., 2003. Value-at-risk for long and short trading positions. *Journal of Applied Econometrics* 18, 641–664.

- Glosten, L.R., Jagannathan, R., Runkle, D., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48, 1779–1801.
- Haas, M., Mittnik, S., Paolella, M.S., 2004. A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics* 2, 493–530.
- Jeffreys, H., 1961. *Theory of Probability*, third ed. Oxford University Press, Oxford.
- Jørgensen, B., 1982. *Statistical Properties of the Generalized Inverse Gaussian Distribution*. Lecture Notes in Statistics Vol.9, Springer, Heidelberg.
- Kim, S., Shephard, N., Chib, S., 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65, 361–393.
- Meng, X-L., Wong, W.H., 1996. Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Statistica Sinica* 6, 831–860.
- Mittnik, S., Paolella, M.S., 2000. Conditional density and value-at-risk prediction of Asian currency exchange rates. *Journal of Forecasting* 19, 313–333.
- Nakajima, J., Omori, Y., 2010. Stochastic volatility model with leverage and asymmetrically heavy-tailed error using GH skew Student's t-distribution. *Computational Statistics and Data Analysis*, in press. doi:10.1016/j.csda.2010.07.012.
- Nakatsuma, T., 2000. Bayesian analysis of ARMA-GARCH models: a Markov chain sampling approach. *Journal of Econometrics* 95, 57–69.
- Paolella, M.S., 2007. *Intermediate Probability: A Computational Approach*. Wiley, Chichester.
- Prause, K., 1999. *The Generalized Hyperbolic Model: Estimation, Financial Derivatives, and Risk Measures*. PhD dissertation, University of Freiburg.
- Rosenblatt, M., 1952. Remarks on a multivariate transformation. *Annals of Mathematical Statistics* 23, 470–472.

Figure 1. GHSST densities (unit variance)

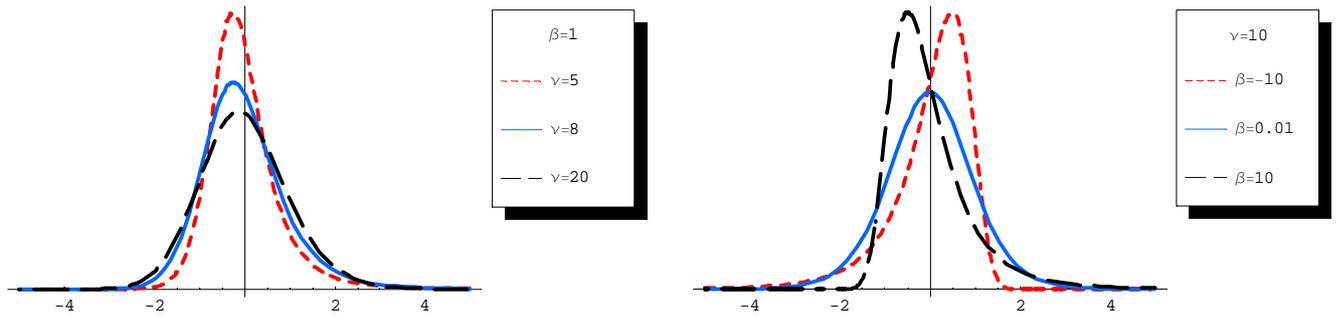


Figure 2. Error skewness and kurtosis as functions of standard deviation

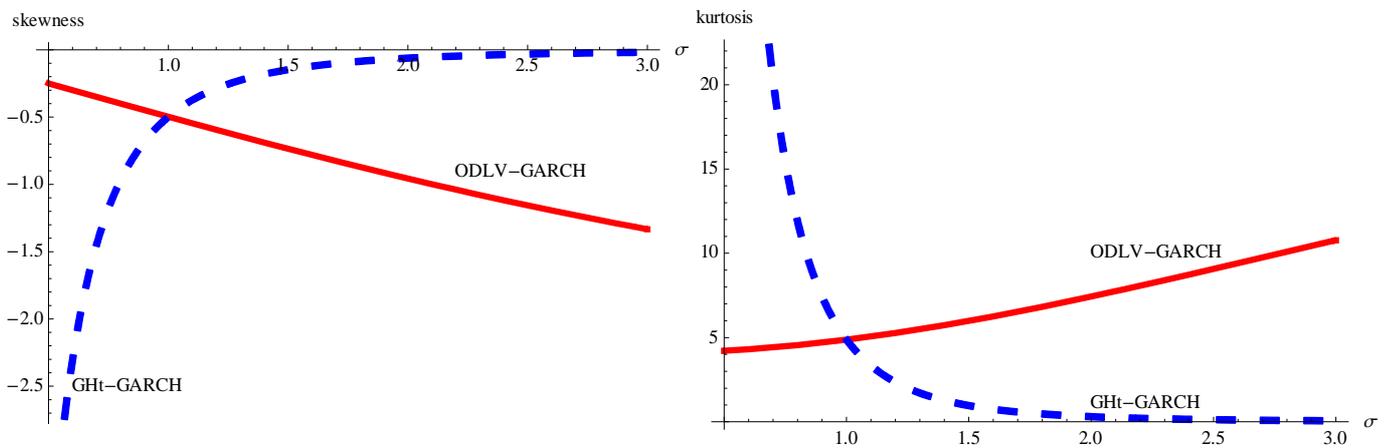


Figure 3. Evolution of logarithmic Bayes factors over time

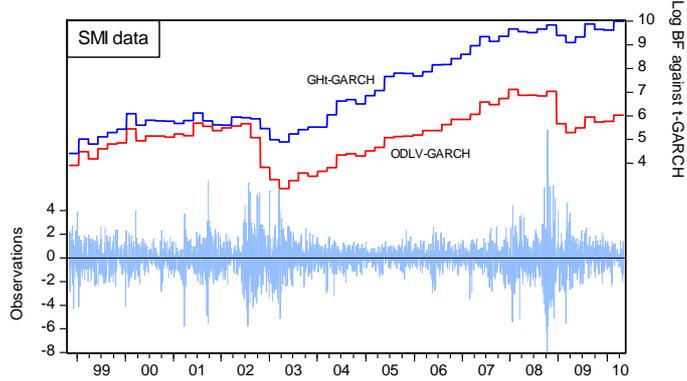
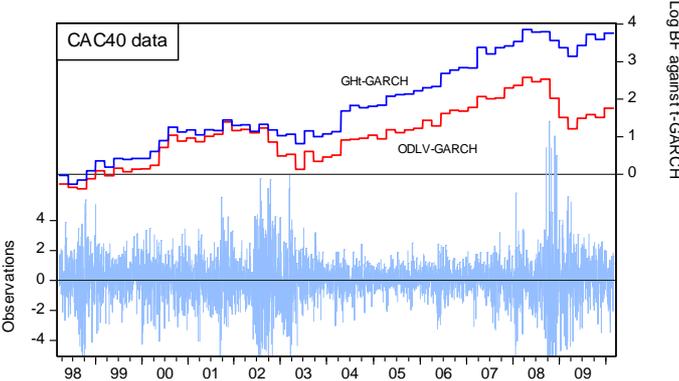
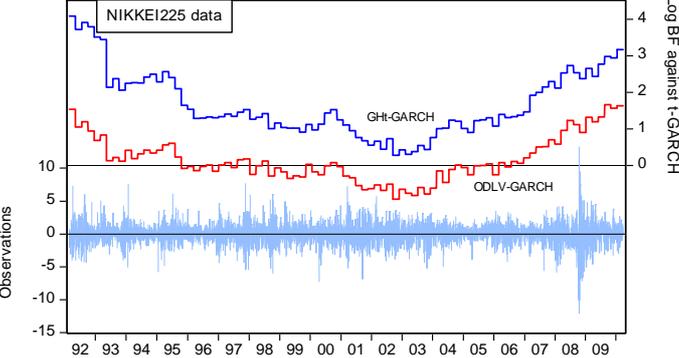
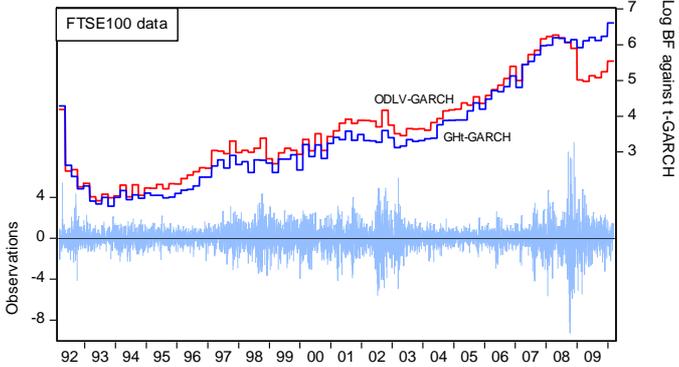
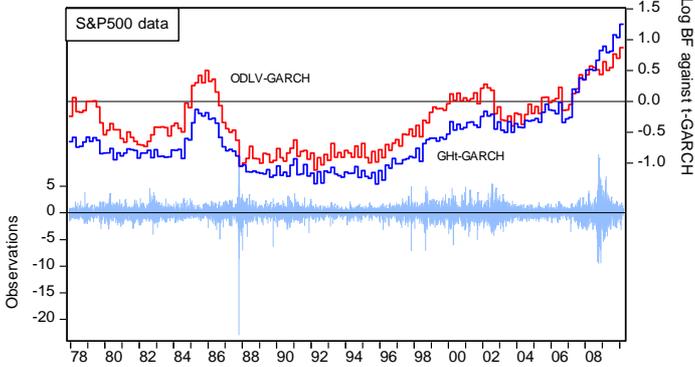


Figure 4. Correlograms of posterior replications (S&P500 data)

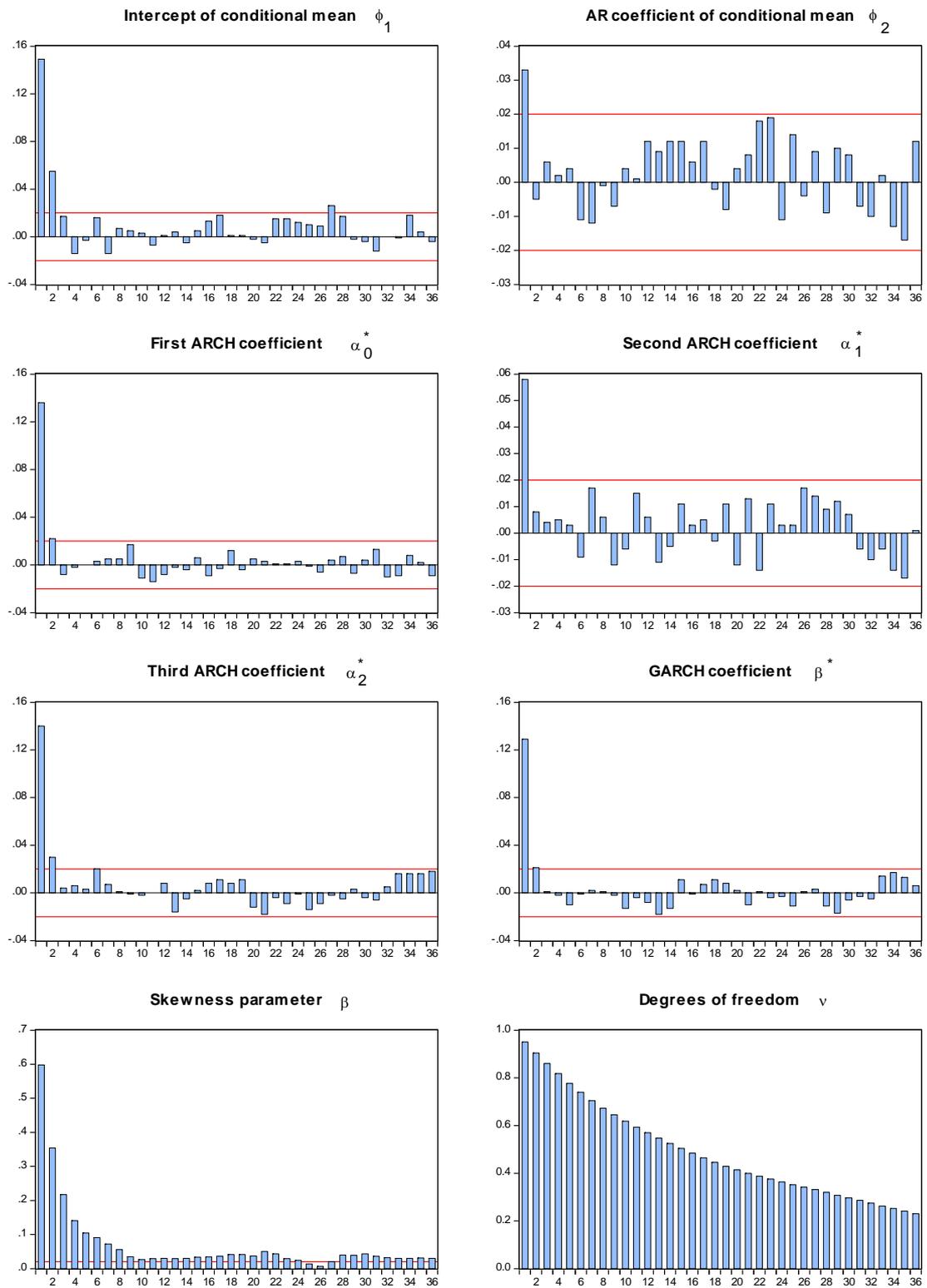


Figure 5. Random permutation sampler replications (NIKKEI225 data)

K=2

K=3

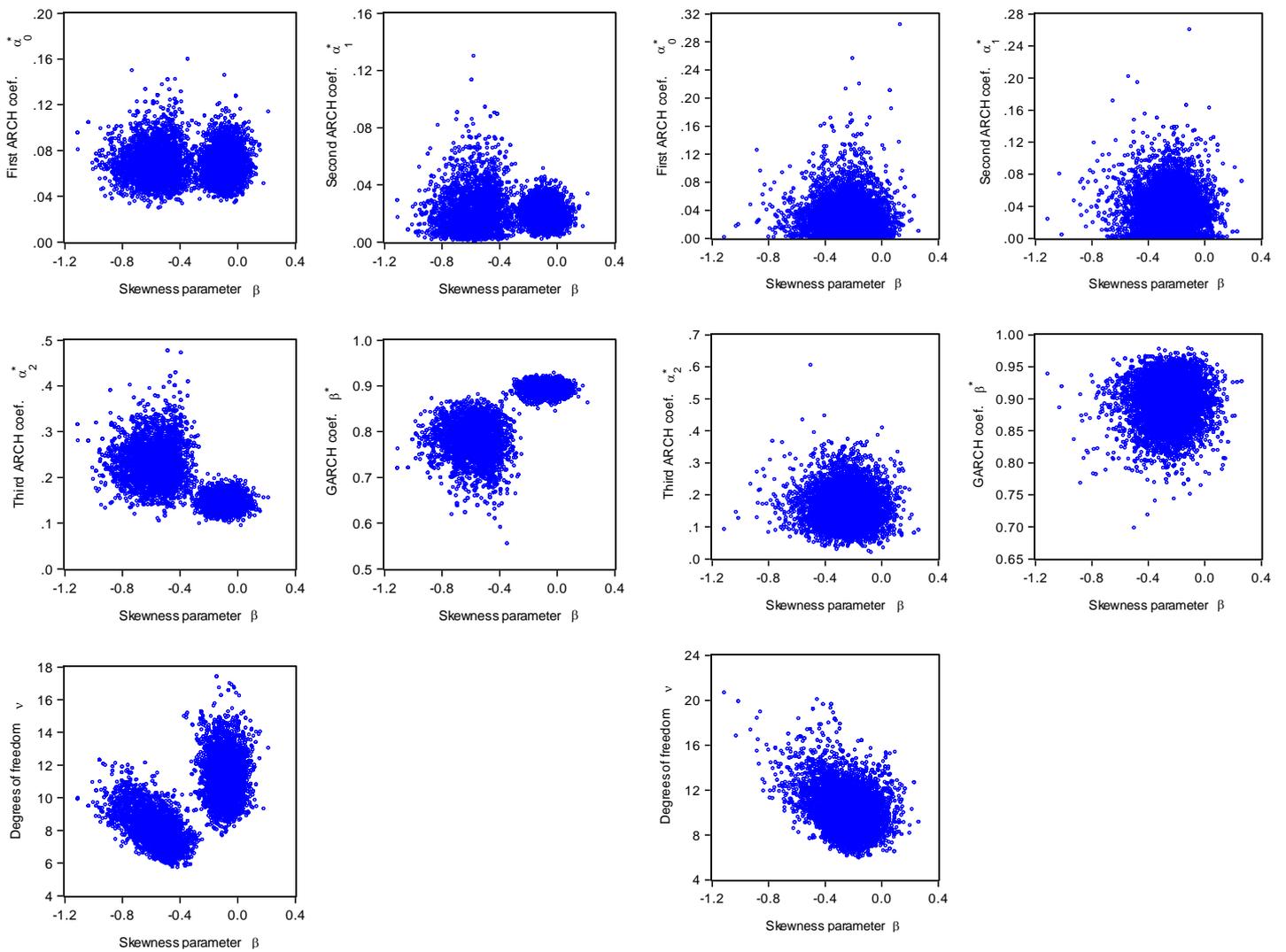


Figure 6. Smoothed regime probabilities (NIKKEI225 data)

