AISM (2007) 59: 325–348
DOI 10.1007/s10463-006-0052-x

Ramsés H. Mena · Stephen G. Walker

On the stationary version of the generalized hyperbolic ARCH model

Received: 24 February 2005 / Revised: 13 September 2005 / Published online: 11 May 2006

Abstract This paper finds conditions under which the generalized hyperbolic ARCH-type model is strictly stationary. Properties of the model are investigated and in particular an estimation procedure is proposed. The resulting stationary model provides with a robust non-Gaussian ARCH-type alternative.

Keywords ARCH · EM algorithm · Generalized hyperbolic distributions · Stationary processes

1 Introduction

Since the classical *autoregressive conditional heteroskedastic* (ARCH) model was presented by Engle (1982), many other ARCH type models have been introduced to fit certain features for the distribution of equity returns. In particular, features such as thick tails, volatility clustering, leptokurtosis and skewness have been considered. For reviews see Bollerslev et al. (1992) and Shephard (1996).

An ARCH(p) model able to include all these features was presented by Barndorff-Nielsen (1997). He considered the following *observation-driven state space* model:

$$X_{t+p} \mid Y_{t+p} \sim \mathrm{N}\left(\mu + \beta\, Y_{t+p}, Y_{t+p}\right), \quad \text{for} \quad t = 1, 2, \ldots \text{(observation density)}$$
$$Y_{t+p} \mid X^{(t,\, p-1)} \sim \mathrm{GIG}\left(1/2, r(X^{(t,\, p-1)}; \theta), \alpha^2\right) \qquad \text{(state density)},$$

R.H. Mena (✉)
Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas (IIMAS), Universidad Nacional Autonoma de México (UNAM). A.P. 20-726. 01000 Mexico, D.F. México
E-mail: ramses@sigma.iimas.unam.mx

S.G. Walker
Institute of Mathematics, Statistics and Actuarial Sciences. University of Kent, Canterbury, Kent, CT2 7NZ, UK

for $p = 1, 2, \ldots, \mu$, $\alpha > 0, 0 \leq |\beta| < \alpha$ and $X^{(t,\,p-1)} = \{X_t, X_{t+1}, \ldots, X_{t+p-1}\}$. Here GIG denotes a *generalized inverse Gaussian* distribution and $r(\cdot\,; \theta)$ refers to a function specifying the observation-driven part of the model, which in turn can depend on certain parameters, generically denoted by θ. We refer to Cox (1981) for more on classification of state space models. Barndorff-Nielsen (1997) suggested the following specification for such observation-driven component

$$r(X^{(t,i)}; \theta) = \left(\varepsilon + \sum_{j=0}^{i} \rho_j X_{t+j}^2 \right)^k,$$

where $\varepsilon, \rho_j, k > 0$, $j = 0, 1, \ldots$. Within the econometric literature the (observable) variables $\{X_t\}$ are typically regarded as the log-equity returns whereas the (not directly observable) variables $\{Y_t\}$ as the underlying squared volatility. Other specifications of the function $r(\cdot\,; \theta)$ with econometric arguments are studied in Andersson (2001) and Jensen and Lunde (2001). However, none of these ARCH type models lead always to strict stationarity. The aim of this paper is to establish conditions under which the Barndorff-Nielsen model is strictly stationary. This is a well motivated issue for ARCH, GARCH type models; see for example Nelson (1990) who finds conditions under which the GARCH(1, 1) is strictly stationary. We achieve our conditions by appropriate specification of the function $r(X^{(t,i)}; \theta)$. In order to do this we use the approach introduced by Pitt et al. (2002).

In Sect. 2 we give a brief introduction to generalized hyperbolic distributions including the general form of all their non-central moments. Section 3 reviews the construction of Pitt et al. (2002), to be used in the construction of the stationary ARCH-type models with GH marginal distributions. Sections 4 and 5 are devoted to the construction of GH-ARCH(1) and GH-ARCH(p) respectively. A correlation property, available from the underlying stationarity, is also presented. Some issues concerning the estimation of stationary GH-ARCH models are discussed in Sect. 6. For the sake of exposition proofs are deferred to the Appendix.

2 Preliminaries

Before presenting our results, we give a brief review of the Generalized hyperbolic and GIG distributions. Generalized hyperbolic distributions were first introduced by Barndorff-Nielsen (1977) in order to model grain size distributions of wind blown sands. Nowadays, GH distributions play an important role in financial applications; see, for instance, Prause (1999), Raible (2000), Barndorff-Nielsen and Shephard (2001) and Mencía and Sentana (2004).

Definition 1 *The random variable X has a* generalized hyperbolic distribution (GH) if its density is defined as

$$\begin{aligned} \mathrm{GH}(x; \lambda, \alpha, \beta, \delta, \mu) = a(\lambda, \alpha, \beta, \delta) \left\{\delta^2 + (x-\mu)^2\right\}^{(\lambda-1/2)/2} \\ \times K_{\lambda-1/2}\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right) \exp\{\beta(x-\mu)\} \end{aligned}$$

with

$$a(\lambda, \alpha, \beta, \delta) = \frac{\left(\alpha^2 - \beta^2\right)^{\lambda/2}}{\sqrt{2\pi}\alpha^{\lambda-\frac{1}{2}}\delta^{\lambda} K_{\lambda}\left(\delta\sqrt{\alpha^2 - \beta^2}\right)},$$

where $x \in \mathbb{R}$ and K_{ν} is the modified Bessel function of the third kind with index ν, see Abramowitz and Stegun (1992). The above density depends on five parameters: $\alpha > 0$ as a shape parameter, β with $0 \leq |\beta| < \alpha$ determines the skewness, $\mu \in \mathbb{R}$ the location, $\delta > 0$ as a scaling factor and the parameter $\lambda \in \mathbb{R}$ which characterizes certain sub-classes and is related to the amount of mass in the tails. Here and henceforth, we use the notation $\mathrm{D}(x; \theta)$ to denote the density function corresponding to the random variable $X \sim \mathrm{D}(\theta)$.

GH distributions can arise via the following mean-variance mixture

$$\mathrm{GH}(x; \lambda, \alpha, \beta, \delta, \mu) = \int_0^{\infty} \mathrm{N}(x; \mu + \beta y, y)\, \mathrm{GIG}(y; \lambda, \delta^2, \alpha^2 - \beta^2)\, \mathrm{d}y,$$

where the mixing distribution is *generalized inverse Gaussian* (GIG), that is, a distribution with density given by

$$\mathrm{GIG}\,(x; \lambda, \delta, \gamma) = \frac{\left(\frac{\gamma}{\delta}\right)^{\lambda/2}}{2K_{\lambda}(\sqrt{\delta\gamma})}\, x^{\lambda-1} \exp\left\{-\frac{1}{2}\left(\delta x^{-1} + \gamma x\right)\right\}, \qquad x > 0, \tag{1}$$

where $\lambda \in \mathbb{R}$, $(\delta, \gamma) \in \Theta_{\lambda}$ and

$$\Theta_{\lambda} = \begin{cases} \delta \geq 0, \gamma > 0, & \text{if } \lambda > 0, \\ \delta > 0, \gamma > 0, & \text{if } \lambda = 0, \\ \delta > 0, \gamma \geq 0, & \text{if } \lambda < 0. \end{cases}$$

The cases $\delta = 0$ and $\gamma = 0$ are interpreted as the limiting cases. Using the asymptotic expansion 1 in Appendix F we can see that if $\lambda > 0$ and $\delta \downarrow 0$, then density (1) reduces to

$$\mathrm{GIG}\,(x; \lambda, 0, \gamma) = \frac{(\gamma/2)^{\lambda}}{\Gamma(\lambda)}\, x^{\lambda-1}\, \mathrm{e}^{-(\gamma/2x)} = \mathrm{Ga}\left(x; \lambda, \frac{\gamma}{2}\right),$$

characterizing, as a special case, the gamma distribution. Analogously, when $\lambda < 0$, $\gamma \downarrow 0$ and using the asymptotic expansion 2 in Appendix F we get

$$\mathrm{GIG}\,(x; \lambda, \delta, 0) = \frac{(2/\delta)^{\lambda}}{\Gamma(-\lambda)}\, x^{\lambda-1}\, \mathrm{e}^{-(\delta/2x)} = \mathrm{Iga}\left(x; -\lambda, \frac{\delta}{2}\right),$$

characterizing the inverse gamma distribution. With this, the Student's t distribution can be seen as a particular case of the GH distribution by noticing the following

$$\mathrm{St}(x; \mu, \beta^2, \nu) = \int_0^{\infty} N(x; \mu, y^{-1})\, \mathrm{Ga}\left(y; \nu/2, \nu\beta^2/2\right)\, \mathrm{d}y, \tag{2}$$

where $N(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2, and $\mathrm{St}(\mu, \sigma^2, \nu)$ denotes a non-central Student's t distribution with location parameter μ, dispersion σ and ν degrees of freedom. Many other cases are encompassed by GIG distributions and therefore by GH distributions. We refer to Jørgensen (1982) and Eberlein and Hammerstein (2004) for further results on GIG and GH distributions.

Proposition 1 (Moments) *If $X \sim GH(\lambda, \alpha, \beta, \delta, \mu)$ then the n-th non-central moments, $M_{\mathrm{GH}}^{(n)}$, are characterized through the following recursion formula*

$$M_{\mathrm{GH}}^{(n)} = \sum_{i=0}^{n} \binom{n}{i} \mu^{n-i} M_{\mathrm{GIG}}(i, \beta, \omega, \eta), \qquad n = 1, 2, \ldots$$

where $M_{\mathrm{GIG}}(0, \beta, \omega, \eta) = 1$ and

$$M_{\mathrm{GIG}}(n, \beta, \omega, \eta) = \begin{cases} \sum_{i=1}^{r} \frac{(2r-1)!\beta^{2i-1}}{(2i-1)!(r-i)!2^{r-i}} \frac{K_{\lambda+r+i-1}(\omega)\eta^{r+i-1}}{K_\lambda(\omega)}, & n = 2r-1 \\ \sum_{i=0}^{r} \frac{(2r)!\beta^{2i}}{(2i)!(r-i)!2^{r-i}} \frac{K_{\lambda+r+i}(\omega)\eta^{i+r}}{K_\lambda(\omega)}, & n = 2r \end{cases}$$

with $\omega = \delta\sqrt{\alpha^2 - \beta^2}$ and $\eta = \delta/\sqrt{\alpha^2 - \beta^2}$.

Proof See Appendix A. □

By applying Proposition 1 we can compute some useful moments, for example, if $X \sim \mathrm{GH}(x; \lambda, \alpha, \beta, \delta, \mu)$ then

$$\mathbb{E}[X] = \mu + \beta\,\eta\, \frac{K_{\lambda+1}(\omega)}{K_\lambda(\omega)}, \tag{3}$$

$$\mathbb{E}[X^2] = \mu^2 + \frac{K_{\lambda+1}(\omega)}{K_\lambda(\omega)}\,\eta(2\mu\beta + 1) + \eta^2\beta^2\, \frac{K_{\lambda+2}(\omega)}{K_\lambda(\omega)}, \tag{4}$$

$$\mathbb{E}[X^3] = \mu^3 + \frac{K_{\lambda+1}(\omega)\,\eta}{K_\lambda(\omega)}\left(3\mu^2\beta + 3\mu\right) + \frac{K_{\lambda+2}(\omega)\,\eta^2}{K_\lambda(\omega)}\left(3\mu\beta^2 + 3\beta\right) + \frac{\beta^3\eta^3 K_{\lambda+3}(\omega)}{K_\lambda(\omega)}, \tag{5}$$

$$\mathbb{E}[X^4] = \mu^4 + \frac{K_{\lambda+1}(\omega)\,\eta}{K_\lambda(\omega)}\left(4\mu^3\beta + 6\mu^2\right) + \frac{\eta^2 K_{\lambda+2}(\omega)}{K_\lambda(\omega)}\left(6\mu^2\beta^2 + 12\mu\beta + 3\right), + \frac{K_{\lambda+3}(\omega)\,\eta^3}{K_\lambda(\omega)}\left(4\mu\beta^3 + 6\beta^2\right) + \frac{\beta^4\eta^4 K_{\lambda+4}(\omega)}{K_\lambda(\omega)}. \tag{6}$$

3 Construction of first order stationary models via latent variables

The first accounts of the construction of stationary time series models, with non-Gaussian marginal distributions, were given by Lawrance and Lewis (1977) and Jacobs and Lewis (1977). There, moving average models and autoregressive-moving average models with exponential marginal distributions, were presented. The general exposition of these models resulted in an exponential autoregressive-moving average (EARMA(p, q)) model presented in Lawrance and Lewis (1980). Other models with marginal distributions belonging to other families are found in Gaver and Lewis (1980) and Lawrance (1982) for mixed exponential distributions and for gamma distributions; McKenzie (1986, 1988) for negative binomial and Poisson distributions; Joe (1996) for infinitely divisible convolution-closed distributions; and Jørgensen and Song (1998) for convolution-closed exponential dispersion distributions. Some other accounts are found in Lawrance (1991) and Jørgensen and Song (1998).

In Pitt et al. (2002) a novel approach to constructing stationary AR(1) type models with given marginal distributions was presented. In general, their approach goes as follows: Assume that we are interested on a stationary AR(1) model $\{X_t\}_{t=1}^{\infty}$ with marginal distribution given by Q_X. Furthermore, let us introduce a latent variable Y through the conditional distribution $F_{Y|X}$. With the knowledge of Q_X and $F_{Y|X}$, an application of Bayes theorem leads to a form for $F_{X|Y}$. The method of Pitt et al. (2002) consists of introducing a latent process $\{Y_t\}_{t=1}^{\infty}$ via the updating mechanism

$$\{Y_{t+1} \mid X_t = x\} \sim F_{Y|X}(\cdot \mid x), \qquad \{X_{t+1} \mid Y_{t+1} = y\} \sim F_{X|Y}(\cdot \mid y),$$

for $t = 1, 2, \ldots$ The constructed processes $\{X_t\}_{t=1}^{\infty}$ and $\{Y_t\}_{t=1}^{\infty}$ inherit the properties of a Markov chain generated through a Gibbs sampler algorithm, with the characteristic difference that the chains generated by the above mechanism are always in stationarity, provided $X_t \sim Q_X$, for all $t = 1, 2, \ldots$. In particular, both processes $\{X_t\}_{t=1}^{\infty}$ and $\{Y_t\}_{t=1}^{\infty}$ are reversible and satisfy the *independence structure*

$$\mathrm{P}\left(Y_{t+1} \mid \mathcal{Y}^{(t)}, \mathcal{X}^{(t)}\right) = \mathrm{P}(Y_{t+1} \mid \sigma(X_t)), \tag{7}$$

$$\mathrm{P}\left(X_{t+1}, \mid \mathcal{Y}^{(t+1)}, \mathcal{X}^{(t)}\right) = \mathrm{P}(X_{t+1}, \mid \sigma(Y_{t+1})), \tag{8}$$

where $\mathcal{X}^{(t)} := \sigma(X_1, X_2, \ldots, X_t)$ and $\mathcal{Y}^{(t)} := \sigma(Y_1, Y_2, \ldots, Y_t)$. This independence structure can be seen as a direct consequence of the *interleaving property* underlying to such chains; see Liu et al. (1994). When regarded as model assumptions, equalities (7) and (8) are known as the state distribution and the observation distribution of an observation-driven model. Assuming the existence of densities q_X, $f_{Y|X}$ and $f_{X|Y}$, a well-defined Markovian (one-step ahead) transition density is given as

$$p(x_t, x_{t+1}) = \int f_{X|Y}(x_t \mid y)\, f_{Y|X}(y \mid x_{t+1})\, \lambda(\mathrm{d}y), \tag{9}$$

where $\lambda(\cdot)$ is a σ-finite measure on the space of latent variables Y. It is straightforward to check that this transition density defines a stationary and reversible process, for any choice of Q_X. See, for instance, Robert and Casella (2002). Notice that different choices of $F_{Y|X}$ lead to different models with the same marginal distribution. In order to narrow this choice Pitt et al. (2002) took the approach of imposing the following linearity constraint

$$\mathbb{E}[X_{t+1} \mid X_t = x] = (1 - \rho)\,\mu + \rho\, x,$$

where $|\rho| < 1$ and $\mu = \mathbb{E}[X]$. Notice that, a choice of $F_{Y|X}$ can be found by assuming forms for $F_{X|Y}$ and Q_Y, that is

$$f_{Y|X}(y \mid x) \propto f_{X|Y}(x \mid y)\, q_Y(y).$$

However, this will depend on whether the target stationary distribution satisfies

$$q_X(x) = \int f_{X|Y}(x \mid y)\, q_Y(y) \lambda(\mathrm{d}y).$$

For some situations, where conjugacy of $F_{X|Y}$ with Q_Y is available, the above integration is simplified, making possible the constructions of stationary models with marginal distributions being Q_X. Such is the case of the ARCH(1) type stationary model with Student-t innovations considered in Pitt and Walker (2005), where they made use of the mixture representation (2). More precisely, let us assume that we want to construct a stationary model with marginal density $q_X(x) = \mathrm{St}(x; 0, \beta^2, \nu)$. Furthermore, assume that $f_{X|Y}(x \mid y) = N(x; 0, y^{-1})$ and $q_Y(y) = \mathrm{Ga}\left(y; \nu/2, \nu\beta^2/2\right)$. This is up to say, due to the underlying conjugacy, that

$$f_{Y|X}(y \mid x) = \mathrm{Ga}\left(y; \frac{(\nu+1)}{2}, \frac{(x^2+\nu\beta^2)}{2}\right).$$

With these components it is possible to construct a Markovian model $\{X_t\}_{t=1}^{\infty}$ with transition density

$$\begin{aligned} p(x_t, x_{t+1}) &= \int_0^{\infty} \mathrm{N}(x_{t+1}; 0, y^{-1})\, \mathrm{Ga}\left(y; \frac{(\nu+1)}{2}, \frac{(x_t^2+\nu\beta^2)}{2}\right) \mathrm{d}y \\ &= \mathrm{St}\left(x_{t+1}; 0, \frac{{x_t}^2+\nu\beta^2}{\nu+1}, \nu+1\right). \end{aligned}$$

This model can be rewritten as

$$X_{t+1} = \sqrt{\frac{X_t^2+\nu\beta^2}{\nu+1}}\, \varepsilon_{t+1}, \qquad \varepsilon_{t+1} \sim \mathrm{St}(0, 1, \nu+1), \tag{10}$$

which can be recognized as the stationary version of the t-ARCH(1) of Bollerslev (1987).

4 GH-ARCH(1) stationary model

Following Pitt and Walker (2005), we will consider

$$F_{X|Y}(\cdot \mid y) = \mathrm{N}(\mu + \beta y, y) \qquad \text{and} \qquad Q_Y(\cdot) = \mathrm{GIG}(\lambda, \delta^2, \alpha^2 - \beta^2), \tag{11}$$

where $\mu, \beta, \lambda \in \mathbb{R}$ and $0 \le |\beta| < \alpha$. An application of Bayes theorem leads to

$$f_{Y|X}(y \mid x) = \mathrm{GIG}\left(y; \lambda - \frac{1}{2}, (x-\mu)^2 + \delta^2, \alpha^2\right).$$

With the given conditional probabilities, the one-step transition probability is given through (9) by

$$p(x_t, x_{t+1}) = \mathrm{GH}\left(x_{t+1}; \lambda - \frac{1}{2}, \sqrt{\beta^2 + \alpha^2}, \beta, \sqrt{(x_t - \mu)^2 + \delta^2}, \mu\right). \qquad (12)$$

Notice that, for this transition density, we have

$$q_X(x_{t+1}) = \int_{\mathbb{R}} p(x_t, x_{t+1}) q_X(x_t) \lambda(\mathrm{d}x_t), \qquad (13)$$

where $q_X(x) = \mathrm{GH}(x; \lambda, \alpha, \beta, \delta, \mu)$.

Definition 2 *A strictly stationary generalized hyperbolic ARCH-type model, hereafter referred as the stationary GH-ARCH (1) model, is a Markov process $\{X_t\}_{t=1}^{\infty}$ with transition distribution given by (12) and marginal distribution given by $GH(\lambda, \alpha, \beta, \delta, \mu)$ for all $t = 1, \ldots, \infty$.*

This model belongs to the class of ARCH-type models proposed by Barndorff-Nielsen (1997) when $r(x_t; \theta) = (\delta^2 + (x_t - \mu)^2)^{1/2}$ and $\lambda^* = \lambda - 1/2$, where λ^* corresponds to the Barndorff-Nielsen (1997) specification. Property (13), a feature of this construction, clearly leads to a strictly stationary model. Some useful quantities are at order.

Proposition 2 *The one-lag autocorrelation for the stationary GH-ARCH* (1) *model is given by*

$$\mathrm{Corr}(X_{t+1}, X_t) = \frac{\mathrm{Cov}(X_{t+1}, X_t) K_\lambda(\omega)}{\eta K_{\lambda+1}(\omega) + \mathrm{Cov}(X_{t+1}, X_t) K_\lambda(\omega)},$$

where

$$\mathrm{Cov}(X_{t+1}, X_t) = \beta^2 \eta^2 \left\{ \frac{K_{\lambda+2}(\omega)}{K_\lambda(\omega)} - \left(\frac{K_{\lambda+1}(\omega)}{K_\lambda(\omega)} \right)^2 \right\}$$

and $\omega = \delta\sqrt{\alpha^2 - \beta^2}$ and $\eta = \delta/\sqrt{\alpha^2 - \beta^2}$.

Proof See Appendix B. □

For the symmetric case, that is when $\beta = 0$, the correlation is clearly 0. This agrees with the empirical observation that financial log-returns are zero correlated. An important property to underline here is that the general form of our GH-ARCH(1) model allows for asymmetric marginal distributions, a property that is not common among many of the reversible models available in the literature. When we allow for asymmetry the correlation structure gets more complicated and even the non-squared process shows correlation.

Another quantity of interest is the autocorrelation in the squares, $\mathrm{Corr}(X_{t+1}^2, X_t^2)$. In order to compute this quantity we can proceed as in the proof for Proposition 2. That is, we can compute

$$\mathrm{Corr}(X_{t+1}^2, X_t^2) = \frac{\mathbb{E}[X_{t+1}^2 X_t^2] - \mathbb{E}[X_t^2]^2}{\mathbb{E}[X_t^4] - \mathbb{E}[X_t^2]^2}, \qquad (14)$$

where the second and fourth moments are given by equations (4) and (6) respectively. The underlying cross moment can be found as follows

$$\begin{aligned}\mathbb{E}[X_{t+1}^2 X_t^2] &= \mathbb{E}\left\{X_t^2 \mathbb{E}[X_{t+1}^2 \mid X_t]\right\}\\ &= \mathbb{E}\left\{X_t^2\left(\mu^2 + \frac{K_{\lambda+(1/2)}(\omega_{X_t})}{K_{\lambda-(1/2)}(\omega_{X_t})}\eta_{X_t}(2\mu\beta+1) + \frac{K_{\lambda+(3/2)}(\omega_{X_t})}{K_{\lambda-(1/2)}(\omega_{X_t})}\right.\right.\\ &\qquad \left.\left.\times\ \eta_{X_t}^2\beta^2\right)\right\}\\ &= \mu^2\,\mathbb{E}_\lambda[X_t^2] + (2\mu\beta+1)\eta\,\frac{K_{\lambda+1}(\omega)}{K_\lambda(\omega)}\,\mathbb{E}_{\lambda+1}[X_t^2]\\ &\quad + \beta^2\eta^2\,\frac{K_{\lambda+2}(\omega)}{K_\lambda(\omega)}\,\mathbb{E}_{\lambda+2}[X_t^2],\end{aligned} \tag{15}$$

where $\mathbb{E}_{\lambda+i}$ is taken with respect to $\mathrm{GH}(\lambda+i,\alpha,\beta,\delta,\mu)$.

The general expression for the one-lag autocorrelation (14) does not have a short form, therefore we have not displayed it here. However, we turn to the simpler, though no less important, symmetric case, for which the correlation is given by

$$\begin{aligned}&\mathrm{Corr}_{GH(\lambda,\alpha,0,\delta,\mu)}\left(X_{t+1}^2, X_t^2\right)\\ &\quad= \frac{\delta\left\{K_{\lambda+2}(\alpha\delta)\,K_\lambda(\alpha\delta) - (K_{\lambda+1}(\alpha\delta))^2\right\}}{4\mu^2\alpha\,K_{\lambda+1}(\alpha\delta)\,K_\lambda(\alpha\delta) + 3\delta\,K_{\lambda+2}(\alpha\delta)\,K_\lambda(\alpha\delta) - (K_{\lambda+1}(\alpha\delta))^2}.\end{aligned}$$

In particular, for $\lambda = -1/2$

$$\mathrm{Corr}_{\mathrm{NIG}(\alpha,0,\delta,\mu)}\left(X_{t+1}^2, X_t^2\right) = \frac{1}{4\mu^2\alpha^2 + 2\delta\alpha + 3}.$$

Figure 1 gives an example of how the existence of a skewness parameter (β) may allow for larger autocorrelations. The graph is presented for different values of λ.

5 GH-ARCH(p) stationary model

So far, we have only constructed time series models with one lag-dependence. Now we turn to the higher lag-dependence case. Let us consider the $(p+1)$-dimensional distributions given by

$$p(X^{(t,p)}) = q(X_t)\prod_{i=1}^{p} p(X_{t+i} \mid X^{(t,i-1)}), \tag{16}$$

where $X^{(t,i)} := (X_t, \ldots, X_{t+i})$ for any $t \in \mathbb{N}$. As before, in order to keep the strict stationarity of the sequence $\{X_t\}_{t=1}^{\infty}$ with joint distributions given by Eq. (16), we need to impose further conditions on the updating mechanism $p(X_{t+i} \mid X^{(t,i-1)})$. That is, we need to ensure the symmetry of the p-dimensional distributions given

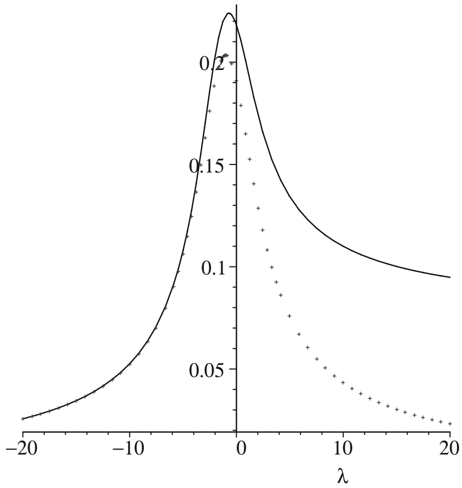

Fig. 1 One-lag correlation for the squares in a stationary GH-ARCH(1) model. The *solid line* corresponds to a model with parameters $\alpha = 1$, $\beta = -0.2$, $\delta = 1$, $\mu = 0$ and the *other line* to a symmetric model with parameters $\alpha = 1$, $\beta = 0$, $\delta = 1$, $\mu = 0$

the marginal distribution Q_X. This can be equally done via the following Gibbs sampler type updating mechanism

(i) $\{Y_{t+i} \mid X^{(t,i-1)} = x^{(t,i-1)}, Y^{(t,i-1)} = y^{(t,i-1)}\} \sim F_{Y|X^i}(\cdot \mid x^{(t,i-1)})$

(ii) $\{X_{t+i} \mid Y_{t+i} = y, X^{(t,i-1)} = x^{(t,i-1)}, Y^{(t,i-1)} = y^{(t,i-1)}\}$
$\sim F_{X|Y}(\cdot \mid y_{t+i})$,

where $i = 1, \ldots, p$, X^i denotes an i-dimensional random vector and $x^{(t,i-1)}$ an i-dimensional vector denoting the time-space values corresponding to $X^{(t,i-1)}$. Due to the underlying conditional independence and under the knowledge of Q_Y, the density corresponding to $F_{Y|X^i}$ can be found as follows

$$f_{Y|X^i}\left(y_{t+i} \mid x^{(t,i-1)}\right) \propto q_Y(y_{t+i}) \prod_{j=1}^{i} f_{X|Y}(x_{t+j-1} \mid y_{t+j}). \tag{17}$$

Therefore, assuming that marginally $X_t \sim Q_X$ and using the conditional independence structure assumed in (i) and (ii), we only need to specify the forms of $F_{Y|X^i}$ for $i = 1, \ldots, p$ in order to have a stationary process with marginal distribution Q_X. As before, the specification of the functional forms for $F_{Y|X^i}$ is quite open. The associated one-step transition density for this model is given by

$$p(x_{t+i} \mid x^{(t,i-1)}) = \int_E f_{X|Y}(x_{t+i} \mid y)\, f_{Y|X^i}(y \mid x^{(t,i-1)})\, \lambda(\mathrm{d}y). \tag{18}$$

Following Pitt and Walker (2005), a similar form for the conditional distributions can be introduced. The objective here is to find a function $r(\cdot,\cdot)$ that leads to a stationary GH-ARCH(p) model. In order to construct the conditional distribution (17) we assume that (11) holds, as we did for the GH-ARCH(1). With this specification and using construction (17) we see that

$$f_{X^p|Y}(x^p \mid y) = \mathrm{N}_p\left(x^p; \mathrm{M} + y\,\mathrm{B},\, y\,\mathrm{I}\right), \tag{19}$$

where M, B $\in \mathbb{R}^p$ denote the corresponding mean and skewness parameter vectors and I denotes the identity matrix.

In fact, from the outset, a more general model can be considered

$$f_{X^p|Y}(x^p \mid y) = \mathrm{N}_p\left(x^p; \mathrm{M} + y\Delta\,\mathrm{B},\, y\Delta\right), \tag{20}$$

where $\Delta \in \mathbb{R}^{p\times p}$ is a positive definite matrix with $|\Delta| = 1$. The specification of Q_Y is given by

$$Q_Y(\cdot) = \mathrm{GIG}(\lambda, \delta^2, \alpha^2 - \mathrm{B}^{\mathrm{T}}\Delta\,\mathrm{B}), \tag{21}$$

where $\lambda \in \mathbb{R}$, $\delta > 0$, $\Delta \in \mathbb{R}^{p\times p}$ and $\alpha^2 > B^{\mathrm{T}}\Delta B$. With the conditional (20) and Eq. (21) we can compute the conditional density (17), getting the following:

$$f_{Y|X^p}(y \mid x^p) = \mathrm{GIG}\left(y; \lambda - \frac{p}{2}, {r_p}^2, \alpha^2\right),$$

where $r_p = \sqrt{(x^p - \mathrm{M})^T\Delta^{-1}(x^p - \mathrm{M}) + \delta^2}$.

Using the updating mechanism (i), (ii), we can construct an i-order Markov transition probability as follows

$$\begin{aligned} p(x_{t+i} \mid x^{(t,i-1)}) &= \int_{\mathbb{R}_+} \mathrm{N}\left(x_{t+i}; \mu_{t+i} + y\beta_{t+i}, y\right)\, \mathrm{GIG}\left(y; \lambda - \frac{i}{2}, r^2_{(t,i-1)}, \alpha^2\right) \mathrm{d}y \\ &= \mathrm{GH}\left(x_{t+i}; \lambda - \frac{i}{2}, \sqrt{\alpha^2 + \beta^2_{t+i}}, \beta_{t+i}, r_{(t,i-1)}, \mu_{t+i}\right), \end{aligned} \tag{22}$$

where $r_{(t,i-1)} = \sqrt{(x^{(t,i-1)} - \mathrm{M})^{\mathrm{T}}\Delta^{-1}(x^{(t,i-1)} - \mathrm{M}) + \delta^2}$. Here, we have treated the case where $|\Delta| = 1$, however the assumption that

$$F_{X|Y}(\cdot, y) = \mathrm{N}\left(\mu_{t+i} + y\beta_{t+i}, y\right)$$

in transition probability (22) imposes that the diagonal elements of Δ must be equal to one, that is the identity matrix, as stated in Eq. (19).

Proposition 3 *Assume that $\Delta = I$, $B = (\beta, \beta, \ldots, \beta)^{\mathrm{T}}$, $\beta_{t+i} = \beta$, $M = (\mu, \mu, \ldots, \mu)^{\mathrm{T}}$ and $\mu_{t+i} = \mu$, for $i = 1, \ldots, p$. If $X_t \sim GH(\lambda, \alpha, \beta, \delta, \mu)$ and $X_{t+i} \mid X^{(t,i-1)}$ follows distribution (22) for $i = 1, \ldots, p$, then marginally $X_{t+p} \sim GH(\lambda, \alpha, \beta, \delta, \mu)$. The resulting model will be termed the stationary generalized hyperbolic ARCH(p) model, stationary GH-ARCH(p).*

Proof The proof follows by construction of the updating mechanism (i) and (ii). □

When conditions in Proposition 3 are satisfied the resulting GH-ARCH(p) model can be seen as a stationary version of the model proposed by Barndorff-Nielsen (1997) when

$$r(X^{(t,i)}; \delta, \mu) = \left(\delta^2 + \sum_{j=0}^{i}(X_{t+j} - \mu)^2\right)^{1/2}. \tag{23}$$

However, in Barndorff-Nielsen (1997) the existence of the marginal (invariant) density leading to a stationary model was not considered. In other words, the parameter restrictions stated in Proposition 3 can be thought as sufficient conditions for stationarity.

In this case the stationary GH-ARCH(p) does not have an easy stochastic representation as in (10) since it considers the behavior in the initial block of p-lagged values, $x_1, \ldots, x_p$. However, after given such values, the model can be written in the following form

$$X_{t+p} = \mu + \sqrt{\delta^2 + \sum_{j=1}^{p}(X_{t+p-j} - \mu)^2}\ \varepsilon_{t+p},$$
$$\varepsilon_{t+p} \sim \mathrm{GH}\left(\lambda - \frac{p}{2}, \sqrt{\alpha^2 + \beta^2}, \beta, 1, 0\right). \tag{24}$$

One aspect of interest, for the sake of comparison with other ARCH models, is that the model (24) can be expressed in terms of a variable ε with zero mean and unit variance. However, in the general case, for any choice of λ, this is neither analytically nor numerically simple. See Andersson (2001) and Jensen and Lunde (2001) for applications of non-stationary models when the mentioned standardization is simpler ($\lambda = -1/2$). In Andersson (2001), the case when $\beta = 0$ and $\lambda = -1/2$ (NIG case) is studied and compared with some other ARCH models. In Jensen and Lunde (2001) a more complete comparison study is done, although the assumption of $\lambda = -1/2$ was still made.

The correlation structure for the stationary GH-ARCH(p) model gets more complicated than the one-lag case ($p = 1$) already presented. However, in Fig. 2 we have showed, based on simulated data via representation (i) and (ii), how the autocorrelation of $\{X_t\}$ behaves as the order changes. The higher the lag-dependence is the higher the correlation observed. Here, we can also observe that $\beta \neq 0$ allows for high (not-squared) autocorrelation.

6 Estimation of the GH-ARCH(p) stationary model

Estimation of the parameters in the stationary GH-ARCH(p) model can be done via numeric maximum likelihood estimation (MLE). The corresponding likelihood in terms of a p-lagged model and a given sample $\mathbf{x} = (x_1, x_2, \ldots, x_T)$ is given by

$$L_{\mathbf{x}}(\theta) = q_X^{\theta}(x_1)\left\{\prod_{i=1}^{p-1} p(x_{i+1} \mid x^{(1,i-1)})\right\}\left\{\prod_{i=1}^{T-p} p(x_{i+p} \mid x^{(i,p-1)})\right\}, \tag{25}$$

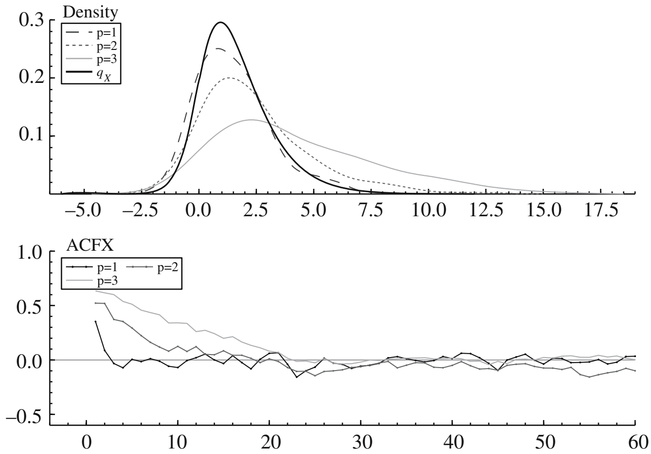

Fig. 2 Densities and ACF's for 500 simulated data from stationary GH-ARCH(p) models. The parameters are $\lambda = 2, \alpha = 2, \beta = 1, \delta = 1$ and $\mu = 0$. The density in bold represents the marginal distribution. For the simulated data, a burn in of 2,000 simulations was applied

where $\theta = (\lambda, \alpha, \beta, \delta, \mu)$. In the case of a stationary GH-ARCH(p) model, the transition probabilities in (25) are given by (22). For this numeric procedure the gradient ∇l_θ, where $l_\theta = \log L_{\mathbf{x}}(\theta)$, is typically required. We have provided this gradient in Appendix C.

Analytical maximization for the complete set of parameters $\theta = (\alpha, \beta, \delta, \mu, \lambda)$ is not possible. However, numerical maximization is possible with procedures such as the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm; See Press et al. (1992). On the other hand, using decomposition (17), the estimation procedure can be done via the *expectation maximization* (EM) algorithm. The EM algorithm consists of iteratively computing the expectation and maximization steps given by

$$Q(\theta \mid \theta_{(j)}, \mathbf{x}) = \mathbb{E}_{\theta_{(j)}}\left[\log L^{\text{aug}}_{\mathbf{x},\mathbf{y}}(\theta)\right] \qquad \text{and} \qquad \theta_{(j+1)} = \arg\max_{\theta} Q(\theta \mid \theta_{(j)}, \mathbf{x}),$$

respectively. Here $L^{\text{aug}}_{\mathbf{x},\mathbf{y}}(\theta)$ denotes the augmented likelihood and $\theta_{(j)}, j = 1, 2, \ldots$ stands for the current parameter value (or initial value if $j = 0$). The expectation $\mathbb{E}_{\theta_{(j)}}[\cdot]$ is taken with respect to $\mathrm{F}^{\theta_{(j)}}_{\mathbf{Y}|\mathbf{X}}$, that is, the distribution of the latent information given the observations and the current parameter values.

The augmented likelihood is given by

$$L^{\text{aug}}_{\mathbf{x},\mathbf{y}}(\theta) = q^\theta_X(x_1) \prod_{i=1}^{p-1} f^\theta_{X|Y}(x_{i+1} \mid y_{i+1})\; f^\theta_{Y|X}\left(y_{i+1} \mid x^{(1,i-1)}\right)$$
$$\times \prod_{i=1}^{T-p} f^\theta_{X|Y}\left(x_{i+p} \mid y_{i+p}\right) f^\theta_{Y|X}(y_{i+p} \mid x^{(i,p-1)}\right).$$

Therefore, for the GH-ARCH(p) case, we have

$$l^{\text{aug}}_{\mathbf{x},\mathbf{y}}(\theta) = \log\left(\text{GH}(x_1, \lambda, \alpha, \beta, \delta, \mu)\right) + \sum_{i=1}^{T-1} \log\left(\text{N}\left(x_{i+1}; \mu + \beta y_{i+1}, y_{i+1}\right)\right)$$
$$+ \sum_{i=1}^{p-1} \log\left(\text{GIG}\left(y_{i+1}; \lambda - \frac{i}{2}, r^2_{(1,i-1)}, \alpha^2\right)\right)$$
$$+ \sum_{i=1}^{T-p} \log\left(\text{GIG}\left(y_{i+p}; \lambda - \frac{p}{2}, r^2_{(i,p-1)}, \alpha^2\right)\right),$$

where $l^{\text{aug}}_{\mathbf{x},\mathbf{y}}(\theta) = \log L^{\text{aug}}_{\mathbf{x},\mathbf{y}}(\theta)$. The gradient for the augmented likelihood is given in Appendix D.

The main difficulty with the EM algorithm is to obtain the expectation step, since the distribution $\text{F}^{\theta_{(j)}}_{\mathbf{Y}|\mathbf{X}}$ might not have a simple form. However, in this case, we can build the distribution of $\text{F}^{\theta}_{\mathbf{Y}|\mathbf{X}}$ by considering component-wise the conditional independent variables with density function

$$f(y_{t+i} \mid x_{t+i}, x^{(t,i-1)}) \propto f^{\theta}_{X|Y}(x_{t+i} \mid y_{t+i})\, f^{\theta}_{Y|X}(y_{t+i} \mid x^{(t,i-1)}),$$

where $x^{(t,i-1)} = (x_t, x_{t+i,\dots,x_{t+i-1}})$. In this case, we can see that

$$f(y_{t+i} \mid x_{t+i}, x^{(t,i-1)}) = \text{GIG}\left(y_{t+i}; \lambda - \frac{i+1}{2}, r^2_{(t,i)}, \alpha^2 + \beta^2\right),$$

for $i = 1, \dots, T-1$. Therefore, for a stationary GH-ARCH(p) model, we can depict the random vector $\{\mathbf{Y} \mid \mathbf{X}\}$ by considering the following conditionally independent random variables

$$\begin{aligned}
Y_1 &\sim \text{GIG}(\lambda, \delta^2, \alpha^2 - \beta^2),\\
Y_{i+1} \mid x^{(1,i)} &\sim \text{GIG}\left(\lambda - \frac{i+1}{2}, r^2_{(1,i)}, \alpha^2 + \beta^2\right), \quad \text{for } i = 1, \dots, p-1\\
Y_{i+p} \mid x^{(i,p)} &\sim \text{GIG}\left(\lambda - \frac{p+1}{2}, r^2_{(i,p)}, \alpha^2 + \beta^2\right), \quad \text{for } i = 1, \dots, T-p.
\end{aligned} \tag{26}$$

Using this decomposition, it is possible to compute the density $f^{\theta_{(j)}}_{\mathbf{Y}|\mathbf{X}}$ for the required distribution by simply multiplying the densities corresponding to the above random variables. This decomposition is also useful to simulate from $\{\mathbf{Y} \mid \mathbf{X}\}$, which can be used to implement a Monte Carlo EM scheme.

In Walker (1996), it was shown that the M-step involved in the EM algorithm can be simplified as follows; if we let $\theta^{-i}_{(j)} = (\theta^1_{(j)}, \dots, \theta^{i-1}_{(j)}, \theta^{i+1}_{(j)}, \dots, \theta^d_{(j)})$, hence the M-step can be simplified by component-wise solving

$$\frac{\partial Q(\theta^i \mid \theta^{-i}_{(j)}, \mathbf{x})}{\partial \theta^i} = \mathbb{E}_{\theta_{(j)}}\left[\frac{\partial l^{\text{aug}}_{\mathbf{x},\mathbf{y}}(\theta^i)}{\partial \theta^i}\right]\Big|_{\theta^i = \theta^i_{(j+1)}} = 0 \tag{27}$$

for $i = 1, \dots, d$ and where the expectation $\mathbb{E}_{\theta_{(j)}}[\cdot]$ is taken with respect to $F^{\theta_{(j)}}_{\mathbf{Y}|\mathbf{X}}$. See also Louis (1982).

In Walker (1996), the expectation in (27) was computed with Monte Carlo simulations from $\{\mathbf{Y} \mid \mathbf{X}\}$. In this case such expectation can be taken analytically by using decomposition (26). The following result will allow us to estimate the parameters of the stationary GH-ARCH(p) via a sequential MLE.

Proposition 4 *Let* $\mathbf{x} = (x_1, x_2, \dots, x_T)$ *be a sample from a stationary GH-ARCH*(p) *model, then*

$$\mathbb{E}_{\theta_{(j)}}\left[\nabla l^{\text{aug}}_{\mathbf{x},\mathbf{y}}(\theta)\right] = \nabla l_{\mathbf{x}}(\theta). \tag{28}$$

Proof See Appendix E. □

Proposition 4 allows us to estimate the parameters of the stationary GH-ARCH(p) model as follows: Given a set of initial values $\theta_0 = (\lambda_0, \alpha_0, \beta_0, \delta_0, \mu_0)$, the first update (first iteration of the EM algorithm) is given by individually solving

$$\left.\frac{\partial l_{\mathbf{x}}(\lambda, \alpha_0, \beta_0, \delta_0, \mu_0)}{\partial \lambda}\right|_{\lambda=\lambda_1} = 0, \quad \left.\frac{\partial l_{\mathbf{x}}(\lambda_1, \alpha, \beta_0, \delta_0, \mu_0)}{\partial \alpha}\right|_{\alpha=\alpha_1} = 0$$

$$\left.\frac{\partial l_{\mathbf{x}}(\lambda_1, \alpha_1, \beta, \delta_0, \mu_0)}{\partial \beta}\right|_{\beta=\beta_1} = 0, \quad \left.\frac{\partial l_{\mathbf{x}}(\lambda_1, \alpha_1, \beta_1, \delta, \mu_0)}{\partial \delta}\right|_{\delta=\delta_1} = 0$$

$$\left.\frac{\partial l_{\mathbf{x}}(\lambda_1, \alpha_1, \beta_1, \delta_1, \mu)}{\partial \mu}\right|_{\mu=\mu_1} = 0. \tag{29}$$

Hence, we use θ_1 to get θ_2 and so on, until a required convergence criterion is satisfied. Ideally we would like Eq. (29) to have an analytical solution. For the general parameter domain of stationary GH-ARCH(p) this is not possible. However, numerical solutions for the above equations are much cheaper, in terms of time and efficiency, than the joint maximization procedure for the complete likelihood.

Notice that in general the above sequential MLE approach does not guaranty the convergence to the optimum value. However, in this case, Proposition 4 uses the EM algorithm to ensure such convergence.

Example 1 In order to illustrate the estimation mechanism described above, we have simulated a sample path with 2000 values of a stationary GH-ARCH(2) model with parameters $\lambda = 2$, $\alpha = 2$, $\beta = 1$, $\delta = 1$ and $\mu = 0$. Therefore, applying the sequential MLE algorithm described in (29), we obtain the results displayed in Table 1. The non-linear equations involved in (29) require the computation of modified Bessel functions, in this case we have used the routines described in Press et al. (1992) to compute them. The time required to get an overall relative error less or equal than 10^{-6} was 38.45 sec. Figure 3 shows the smoothed densities corresponding to the parameter estimates of 500 sample paths. Notice how the modes concentrate around the true values. ○

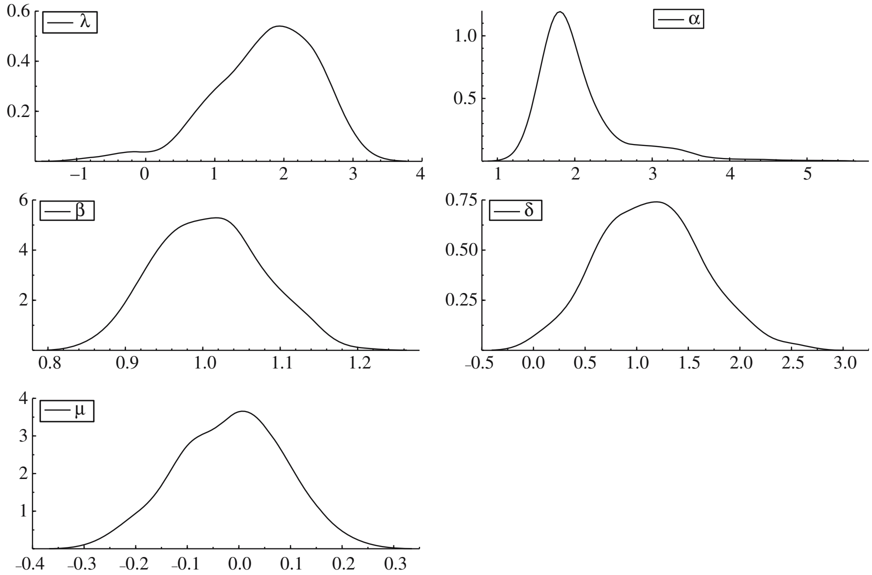

Fig. 3 Smoothed densities for the parameter estimates corresponding to 500 sample paths of size 2000 each of a GH-ARCH(2) model with parameters $\lambda = 2, \alpha = 2, \beta = 1, \delta = 1$ and $\mu = 0$

Table 1 Iteration results of the EM algorithm corresponding to Example 1

Iterations	λ	α	β	δ	μ	l_θ
Model	2.00000	2.00000	1.00000	1.00000	0.00000	−3740.490000
Initial	3.00000	1.50000	1.30000	3.00000	1.00000	−5660.629660
50	−0.2627527	1.517131	1.007450	2.257465	0.000649	−3717.160942
100	0.7007581	1.627385	0.990147	1.828023	0.032054	−3716.186550
300	1.587935	1.734899	0.978064	1.378675	0.052922	−3715.821308
500	1.622195	1.739547	0.977901	1.360448	0.053250	−3715.820679
800	1.623464	1.739720	0.977894	1.359774	0.053263	−3715.820678
1,000	1.623473	1.739723	0.977895	1.359771	0.053262	−3715.820678
1,280	1.623497	1.739730	0.977897	1.359765	0.053260	−3715.820678
1,300	1.623519	1.739737	0.977901	1.359760	0.053255	−3715.820677

Appendix A

Proof of Proposition 1

First, let us notice that if $\mathcal{L}(\theta) = \mathbb{E}[e^{\theta X}]$ then $\mathrm{M}_{GH}^{(n)} = \mathcal{L}_{GH}^{(n)}(0)$, where $\mathcal{L}_{GH}(\theta) = \mathrm{e}^{\mu\theta}\mathcal{L}_{\mathrm{GIG}}(g(\theta))$ and $g(\theta) = \theta^2/2 + \beta\theta$. Notice that $g'(\theta) = \theta + \beta$, $g'{'}(\theta) = 1$ and $g^{(n)}(\theta) = 0$ for $n \geq 3$. Using a simple binomial induction, we get

$$\mathcal{L}_{GH}^{(1)}(\theta) = \mathrm{e}^{\mu\theta}\left[\mu\mathcal{L}_{\mathrm{GIG}}(g(\theta)) + \left(\frac{\partial}{\partial\theta}\right)\mathcal{L}_{\mathrm{GIG}}(g(\theta))\right],$$

$$\mathcal{L}^{(2)}_{GH}(\theta) = e^{\mu\theta}\left[\mu^2 \mathcal{L}_{\text{GIG}}(g(\theta)) + 2\mu\left(\frac{\partial}{\partial\theta}\right)\mathcal{L}_{\text{GIG}}(g(\theta)) + \left(\frac{\partial}{\partial\theta}\right)^2 \mathcal{L}_{\text{GIG}}(g(\theta))\right], \tag{30}$$

$$\vdots$$

$$\mathcal{L}^{n}_{GH}(\theta) = e^{\mu\theta}\sum_{i=0}^{n}\binom{n}{i}\mu^{n-i}\left(\frac{\partial}{\partial\theta}\right)^i \mathcal{L}_{\text{GIG}}(g(\theta)).$$

In order to compute the differential operator in Eq. (31) notice that the moments corresponding to a normal distribution $\text{N}(\beta y, y)$ are given by

$$\mu'_n(y) = \begin{cases} \sum_{i=1}^{r} \frac{(2r-1)!\beta^{2i-1} y^{i+r-1}}{(2i-1)(r-i)2^{r-i}}, & n = 2r-1, \\ \sum_{i=0}^{r} \frac{(2r)\beta^{2i} y^{i+r}}{(2i)(r-i)2^{r-i}}, & n = 2r. \end{cases} \tag{31}$$

See Bain (1969). Therefore, we can define the operator $\text{D}_n y$ by replacing in (31) all the powers of y with the corresponding differentiation order, that is we replace y^i with $y^{(i)}$, and obtain

$$\left(\frac{\partial}{\partial\theta}\right)^n \mathcal{L}_{\text{GIG}}(g(\theta)) = \text{D}_n \mathcal{L}_{\text{GIG}}(\theta).$$

Hence, by noticing that

$$\mathcal{L}^{(n)}_{\text{GIG}}(0) = \frac{K_{\lambda+n}(\omega)}{K_\lambda(\omega)}\eta^n, \tag{32}$$

and evaluating $\theta = 0$ in (31) the stated result follows. See Jørgensen (1982) for the Laplace transform of GIG distributions.

Appendix B

Proof of Proposition 2

The stationarity of the process $\{X_t\}_{t=1}^{\infty}$ implies

$$\text{Cov}(X_{t+1}, X_t) = \mathbb{E}\{X_t \mathbb{E}[X_{t+1} \mid X_t]\} - \mathbb{E}[X_t]^2. \tag{33}$$

For notation simplicity we let $r_x = \sqrt{(x-\mu)^2 + \delta^2}$, $\omega_x = \alpha r_x$ and $\eta_x = r_x/\alpha$ be a re-parametrization corresponding to the transition probability (12).

First let us work out the cross moment in expression (33)

$$\mathbb{E}\{X_t \mathbb{E}[X_{t+1} \mid X_t]\} = \mu\,\mathbb{E}[X_t] + \beta\,\mathbb{E}\left\{X_t \eta_{X_t}\frac{K_{\lambda+(1/2)}(\omega_{X_t})}{K_{\lambda-(1/2)}(\omega_{X_t})}\right\},$$

where,

$$\mathbb{E}\left[X_t \eta_{X_t} \frac{K_{\lambda+(1/2)}(\omega_{X_t})}{K_{\lambda-(1/2)}(\omega_{X_t})}\right]$$
$$= \frac{a(\lambda, \alpha, \beta, \delta)}{\alpha a(\lambda+1, \alpha, \beta, \delta)} \int_{\mathbb{R}} x_t \underbrace{a(\lambda+1, \alpha, \beta, \delta)\, r_{x_t}^{\lambda+(1/2)} K_{\lambda+(1/2)}(\omega_{x_t})\, \mathrm{e}^{\beta(x_t-\mu)}}_{\mathrm{GH}(\lambda+1,\alpha,\beta,\delta,\mu)} \tag{34}$$
$$= \eta \frac{K_{\lambda+1}(\omega)}{K_\lambda(\omega)} \mathbb{E}_{\lambda+1}[X_t]$$
$$= \mu\eta \frac{K_{\lambda+1}(\omega)}{K_\lambda(\omega)} + \beta\eta^2 \frac{K_{\lambda+2}(\omega)}{K_\lambda(\omega)}. \tag{35}$$

In Eq. (34) the expectation $\mathbb{E}_{\lambda+1}[\cdot]$ is taken with respect to $\mathrm{GH}(\lambda+1, \alpha, \beta, \delta, \mu)$. Here is worth noticing that in general for $n = 0, 1, 2, \ldots$ and any integrable function f we have

$$\mathbb{E}_\lambda \left[f(X_t)\eta_{X_t}^{n+1} \frac{K_{\lambda+n+(1/2)}(\omega_{X_t})}{K_{\lambda-(1/2)}(\omega_{X_t})} \right] = \eta^{n+1} \frac{K_{\lambda+n+1}(\omega)}{K_\lambda(\omega)} \mathbb{E}_{\lambda+n+1}[f(X_t)].$$

Therefore, we can write

$$\mathbb{E}[X_{t+1}X_t] = \mu\, \mathbb{E}[X_t] + \beta\eta \frac{K_{\lambda+1}(\omega)}{K_\lambda(\omega)} \mathbb{E}_{\lambda+1}[X_t]. \tag{36}$$

Substituting the marginal expectation $\mathbb{E}[X_t]$ (Eq. 3) into Eq. (33) we get the stated result for the covariance $\mathrm{Cov}(X_{t+1}, X_t)$. Finally, using the second marginal moment (4) we get

$$\begin{aligned} \mathrm{Var}(X_t) &= \mathbb{E}[X_t^2] - \mathbb{E}[X_t]^2 \\ &= \frac{K_{\lambda+1}(\omega)}{K_\lambda(\omega)}\eta + \beta^2\eta^2 \left\{ \frac{K_{\lambda+2}(\omega)}{K_\lambda(\omega)} - \left(\frac{K_{\lambda+1}(\omega)}{K_\lambda(\omega)} \right)^2 \right\} \\ &= \frac{K_{\lambda+1}(\omega)}{K_\lambda(\omega)}\eta + \mathrm{Cov}(X_{t+1}, X_t). \end{aligned} \tag{37}$$

The underlying stationarity implies

$$\mathrm{Corr}(X_{t+1}, X_t) = \frac{\mathrm{Cov}(X_{t+1}, X_t)}{\mathrm{Var}(X_t)}. \tag{38}$$

A direct substitution of Eq. (37) into (38) leads to the result.

Appendix C

Gradient for MLE

The first step towards the maximum likelihood estimation is given by obtaining the gradient ∇l_θ, where $l_\theta = \log L_{\mathbf{x}}(\theta)$. To simplify notation let us assume that

$$\zeta = \sqrt{\alpha^2+\beta^2}, \qquad \omega = \delta\sqrt{\alpha^2-\beta^2}, \qquad \eta = \delta/\sqrt{\alpha^2-\beta^2},$$

and $r_{(t,i)}$ is given as in Eq. (23). With this notation and the differentiation rules given in Appendix F, the following expressions for the components of the gradient ∇l_θ are available

$$\begin{aligned}\frac{\partial l_\theta}{\partial \alpha} &= \left\{R_\lambda(\omega)\alpha\eta - R_{\lambda-(1/2)}(\alpha r_{(1,0)})r_{(1,0)}\right\}\\ &\quad+\sum_{i=1}^{p-1}\left\{R_{\lambda-(i/2)}(\alpha r_{(1,i-1)})r_{(1,i-1)} - \frac{\alpha}{\zeta}R_{\lambda-((i+1)/2)}(\zeta r_{(1,i)})r_{(1,i)}\right\}\\ &\quad+\sum_{i=1}^{T-p}\left\{R_{\lambda-(p/2)}(\alpha r_{(i,p-1)})r_{(i,p-1)} - \frac{\alpha}{\zeta}R_{\lambda-((p+1)/2)}(\zeta r_{(i,p)})r_{(i,p)}\right\},\end{aligned} \tag{39}$$

$$\begin{aligned}\frac{\partial l_\theta}{\partial \beta} &= \{(x_1-\mu) - \beta\eta R_\lambda(\omega)\}\\ &\quad+\sum_{i=1}^{p-1}\left\{(x_{i+1}-\mu) - \frac{\beta}{\zeta}R_{\lambda-((i+1)/2)}(\zeta r_{(1,i)})r_{(1,i)}\right\}\\ &\quad+\sum_{i=1}^{T-p}\left\{(x_{i+p}-\mu) - \frac{\beta}{\zeta}R_{\lambda-((p+1)/2)}(\zeta r_{(i,p)})r_{(i,p)}\right\},\end{aligned} \tag{40}$$

$$\begin{aligned}\frac{\partial l_\theta}{\partial \delta} &= \left\{\frac{2\delta\,(\lambda-(1/2))}{r_{(1,0)}^2} - \frac{2\lambda}{\delta} + \frac{\omega}{\delta}R_\lambda(\omega) - \frac{\delta\alpha}{r_{(1,0)}}R_{\lambda-(1/2)}(\alpha r_{(1,0)})\right\}\\ &\quad+\sum_{i=1}^{p-1}\left\{\frac{2\delta\,(\lambda-((i+1)/2))}{r_{(1,i)}^2} - \frac{2\delta\,(\lambda-(i/2))}{r_{(1,i-1)}^2} + \frac{\alpha\delta R_{\lambda-(i/2)}(\alpha r_{(1,i-1)})}{r_{(1,i-1)}}\right.\\ &\qquad\left.-\frac{\zeta\delta R_{\lambda-((i+1)/2)}(\zeta r_{(1,i)})}{r_{(1,i)}}\right\}\\ &\quad+\sum_{i=1}^{T-p}\left\{\frac{2\delta\,(\lambda-((p+1)/2))}{r_{(i,p)}^2} - \frac{2\delta\,(\lambda-(p/2))}{r_{(i,p-1)}^2} + \frac{\alpha\delta R_{\lambda-(p/2)}(\alpha r_{(i,p-1)})}{r_{(i,p-1)}}\right.\\ &\qquad\left.-\frac{\zeta\delta R_{\lambda-((p+1)/2)}(\zeta r_{(i,p)})}{r_{(i,p)}}\right\},\end{aligned} \tag{41}$$

$$
\begin{aligned}
\frac{\partial l_\theta}{\partial \mu} &= \left\{ \frac{\alpha(x_1-\mu)}{r_{(1,0)}} R_{\lambda-(1/2)}(\alpha r_{(1,0)}) - \frac{2\,(\lambda-(1/2))\,(x_1-\mu)}{r^2_{(1,0)}} - \beta T \right\} \\
&\quad + \left\{ \sum_{i=1}^{p-1} \frac{\mathbf{1}(x^{(1,i)}-\mu)}{r_{(1,i)}} \left(R_{\lambda-((i+1)/2)}(\zeta r_{(1,i)})\zeta - \frac{2\,(\lambda-((i+1)/2))}{r_{(1,i)}} \right) \right. \\
&\quad \left. - \frac{\mathbf{1}(x^{(1,i-1)}-\mu)}{r_{(1,i-1)}} \left(R_{\lambda-(i/2)}(\alpha r_{(1,i-1)})\alpha - \frac{2\,(\lambda-(i/2))}{r_{(1,i-1)}} \right) \right\} \\
&\quad + \left\{ \sum_{i=1}^{T-p} \frac{\mathbf{1}(x^{(i,p)}-\mu)}{r_{(i,p)}} \left(R_{\lambda-((p+1)/2)}(\zeta r_{(i,p)})\zeta - \frac{2\,(\lambda-((p+1)/2))}{r_{(i,p)}} \right) \right. \\
&\quad \left. - \frac{\mathbf{1}(x^{(i,p-1)}-\mu)}{r_{(i,p-1)}} \left(R_{\lambda-(p/2)}(\alpha r_{(i,p-1)})\alpha - \frac{2\,(\lambda-(p/2))}{r_{(i,p-1)}} \right) \right\}, \qquad (42)
\end{aligned}
$$

where $\mathbf{1}(x^{(i,t)}-\mu) = \sum_{j=i}^{t+i}(x_j-\mu)$. Finally,

$$
\begin{aligned}
\frac{\partial l_\theta}{\partial \lambda} &= \left\{ \log\left(\frac{r_{(1,0)}}{\alpha\eta}\right) - \frac{\dot{K}_{\lambda-(1/2)}(\alpha r_{(1,0)})}{K_{\lambda-(1/2)}(\alpha r_{(1,0)})} - \frac{\dot{K}_\lambda(\omega)}{K_\lambda(\omega)} \right\} \\
&\quad + \sum_{i=1}^{p-1} \left\{ \log\left(\frac{\alpha r_{(1,i)}}{\zeta r_{(1,i-1)}}\right) + \frac{\dot{K}_{\lambda-((i+1)/2)}(\zeta r_{(1,i)})}{K_{\lambda-((i+1)/2)}(\zeta r_{(1,i)})} - \frac{\dot{K}_{\lambda-(i/2)}(\alpha r_{(1,i-1)})}{K_{\lambda-(i/2)}(\alpha r_{(1,i-1)})} \right\} \\
&\quad + \sum_{i=1}^{T-p} \left\{ \log\left(\frac{\alpha r_{(i,p)}}{\zeta r_{(i,p-1)}}\right) + \frac{\dot{K}_{\lambda-((p+1)/2)}(\zeta r_{(i,p)})}{K_{\lambda-\frac{p+1}{2}}(\zeta r_{(i,p)})} - \frac{\dot{K}_{\lambda-(p/2)}(\alpha r_{(i,p-1)})}{K_{\lambda-(p/2)}(\alpha r_{(i,p-1)})} \right\},
\end{aligned}
\qquad (43)
$$

where $\dot{K}_\nu(x) = \partial K_\nu/\partial \nu$. For the differentials of Bessel functions with respect to their order we refer to Abramowitz and Stegun (1992) (expressions 9.6.43 and 9.6.45).

Appendix D

Gradient for the augmented likelihood

For the augmented likelihood, the gradient simplifies slightly with respect to gradient of the non-augmented likelihood. Using the same notation as in (39–43) we get the following expressions

$$
\begin{aligned}
\frac{\partial l^{\text{aug}}_{\mathbf{x},\mathbf{y}}(\theta)}{\partial \alpha} &= \left\{ R_\lambda(\omega)\alpha\eta - R_{\lambda-(1/2)}(\alpha r_{(1,0)}) r_{(1,0)} \right\} \\
&\quad + \sum_{i=1}^{p-1} \left\{ R_{\lambda-(i/2)}(\alpha r_{(1,i-1)}) r_{(1,i-1)} \right\} \\
&\quad + \sum_{i=1}^{T-p} \left\{ R_{\lambda-(p/2)}(\alpha r_{(i,p-1)}) r_{(i,p-1)} \right\} - \alpha \sum_{i=1}^{T-1} y_{i+1}, \qquad (44)
\end{aligned}
$$

$$\frac{\partial l^{\text{aug}}_{\mathbf{x},\mathbf{y}}(\theta)}{\partial \beta} = \sum_{i=1}^{T} x_i - \beta \sum_{i=1}^{T-1} y_{i+1} - T\mu - \beta\eta R_\lambda(\omega), \tag{45}$$

$$\begin{aligned}\frac{\partial l^{\text{aug}}_{\mathbf{x},\mathbf{y}}(\theta)}{\partial \delta} &= \left\{ \frac{2\delta\,(\lambda-(1/2))}{r^2_{(1,0)}} - \frac{2\lambda}{\delta} + \frac{\omega}{\delta} R_\lambda(\omega) - \frac{\delta\alpha}{r_{(1,0)}} R_{\lambda-(1/2)}(\alpha r_{(1,0)}) \right\} \\ &+ \sum_{i=1}^{p-1} \left\{ R_{\lambda-(i/2)}(\alpha r_{(1,i-1)}) \frac{\delta\alpha}{r_{(1,i-1)}} - \frac{2\delta\,(\lambda-(i/2))}{r^2_{(1,i-1)}} \right\} \\ &+ \sum_{i=1}^{T-p} \left\{ R_{\lambda-(p/2)}(\alpha r_{(i,p-1)}) \frac{\delta\alpha}{r_{(i,p-1)}} - \frac{2\delta\,(\lambda-(p/2))}{r^2_{(i,p-1)}} \right\} - \delta \sum_{i=1}^{T-1} \frac{1}{y_{i+1}},\end{aligned} \tag{46}$$

$$\begin{aligned}\frac{\partial l^{\text{aug}}_{\mathbf{x},\mathbf{y}}(\theta)}{\partial \mu} &= R_{\lambda-(1/2)}(\alpha r_{(1,0)}) \frac{(x_1-\mu)\alpha}{r_{(1,0)}} - \beta T - \frac{2\,(\lambda-(1/2))\,(x_1-\mu)}{r^2_{(1,0)}} \\ &+ \sum_{i=1}^{T-1} \frac{x_{i+1}-\mu}{y_{i+1}} + \sum_{i=1}^{p-1} \mathbf{1}(x^{(1,i-1)}-\mu) \\ &\times \left\{ \frac{2\,(\lambda-(i/2))}{r^2_{(1,i-1)}} + \frac{1}{y_{i+1}} - \frac{R_{\lambda-(i/2)}(\alpha r_{(1,i-1)})\alpha}{r_{(1,i-1)}} \right\} \\ &+ \sum_{i=1}^{T-p} \mathbf{1}(x^{(i,p-1)}-\mu) \left\{ \frac{2\,(\lambda-(p/2))}{r^2_{(i,p-1)}} + \frac{1}{y_{i+p}} - \frac{R_{\lambda-(p/2)}(\alpha r_{(i,p-1)})\alpha}{r_{(i,p-1)}} \right\},\end{aligned} \tag{47}$$

$$\begin{aligned}\frac{\partial l^{\text{aug}}_{\mathbf{x},\mathbf{y}}(\theta)}{\partial \lambda} &= \left\{ \log\left[\frac{r_{(1,0)}}{\alpha\eta}\right] - \frac{\dot{K}_{\lambda-(1/2)}(\alpha r_{(1,0)})}{K_{\lambda-(1/2)}(\alpha r_{(1,0)})} - \frac{\dot{K}_\lambda(\omega)}{K_\lambda(\omega)} \right\} + \sum_{i=1}^{T-1} \log(y_{i+1}) \\ &+ \sum_{i=1}^{p-1} \left\{ \log\left[\frac{\alpha}{r_{(1,i-1)}}\right] - \frac{\dot{K}_{\lambda-(i/2)}(\alpha r_{(1,i-1)})}{K_{\lambda-(i/2)}(\alpha r_{(1,i-1)})} \right\} \\ &+ \sum_{i=1}^{T-p} \left\{ \log\left[\frac{\alpha}{r_{(i,p-1)}}\right] - \frac{\dot{K}_{\lambda-(p/2)}(\alpha r_{(i,p-1)})}{K_{\lambda-(p/2)}(\alpha r_{(i,p-1)})} \right\}.\end{aligned} \tag{48}$$

Appendix E

Proof of Proposition 4

All we need is to consider the first moments for the statistics involved in the augmented gradient $l_{\mathbf{x},\mathbf{y}}^{\text{aug}}$. First notice that if $Y \sim \text{GIG}(\lambda, \delta, \gamma)$ and $Z = 1/Y$ then $Z \sim \text{GIG}(-\lambda, \gamma, \delta)$. Therefore, using (32), we see that

$$\mathbb{E}\,[1/Y] = \frac{R_{-\lambda}(\omega)}{\eta},$$

where $\omega = \sqrt{\delta\gamma}$ and $\eta = \sqrt{\delta/\gamma}$. Now, using property 4 in Appendix F, we see that

$$R_{-\lambda}(\omega) = R_{\lambda}(\omega) - \frac{2\lambda}{\omega}.$$

Hence,

$$\mathbb{E}\left[\frac{1}{Y}\right] = \frac{\{R_{\lambda}(\omega) - 2\lambda/\omega\}}{\eta}.$$

Applying the above result to the corresponding GIG distributions involved in (26) we get

$$\mathbb{E}\left[{Y_{i+1}}^{-1} \mid x^{(1,i)}\right] = \frac{\zeta\, R_{\lambda-((i+1)/2)}(r_{(1,i)}\zeta)}{r_{(1,i)}} - \frac{2\,(\lambda - ((i+1)/2))}{r_{(1,i)}^2}, \tag{49}$$

when $Y_{i+1} \mid x^{(1,i)} \sim \text{GIG}\left(\lambda - ((i+1)/2), r_{(1,i)}^2, \alpha^2 + \beta^2\right)$, and

$$\mathbb{E}\left[{Y_{i+p}}^{(-1)} \mid x^{(i,p)}\right] = \frac{\zeta\, R_{\lambda-((p+1)/2)}(r_{(i,p)}\zeta)}{r_{(i,p)}} - \frac{2\,(\lambda - ((p+1)/2))}{r_{(i,p)}^2}, \tag{50}$$

when $Y_{i+p} \sim \text{GIG}\left(\lambda - (p+1)/2), r_{(i,p)}^2, \alpha^2 + \beta^2\right)$.

On the other hand, if $Z = \log Y$ then $f_Z(\xi) = \text{GIG}(e^{\xi}; \lambda, \delta, \gamma)\, e^{\xi}$, $\xi \in \mathbb{R}$. With this the Laplace transform for Z is given by

$$\mathcal{L}_Z(\kappa) = \frac{K_{\lambda+\kappa}(\omega)}{K_{\lambda}(\omega)}\,\eta^{\kappa},$$

and so

$$\mathbb{E}[\log(Y)] = \mathcal{L}'_Z(0) = \frac{\dot{K}_{\lambda}(\omega)}{K_{\lambda}(\omega)} + \log(\eta),$$

where $\dot{K}_{\nu}(x) = \partial K_{\nu}(x)/\partial\nu$. Therefore, applying the above result as in (49) and (50) we get

$$\mathbb{E}\left[\log(Y_{i+1}) \mid x^{(1,i)}\right] = \frac{\dot{K}_{\lambda-((i+1/2)}(r_{(1,i)}\zeta)}{K_{\lambda-((i+1)/2)}(r_{(1,i)}\zeta)} + \log\left(\frac{r_{(1,i)}}{\zeta}\right) \tag{51}$$

and

$$\mathbb{E}\left[\log(Y_{i+p}) \mid x^{(i,p)}\right] = \frac{\dot{K}_{\lambda-((p+1)/2)}(r_{(i,p)}\zeta)}{K_{\lambda-((p+1)/2)}(r_{(i,p)}\zeta)} + \log\left(\frac{r_{(i,p)}}{\zeta}\right), \tag{52}$$

respectively. In the same way, using (32), we compute

$$\mathbb{E}[Y_{i+1} \mid x^{(1,i)}] = R_{\lambda-((i+1)/2)}(r_{(1,i)}\zeta)\,\frac{r_{(1,i)}}{\zeta}, \qquad i = 1, \ldots, p-1 \tag{53}$$

and

$$\mathbb{E}[Y_{i+p} \mid x^{(i,p)}] = R_{\lambda-((p+1/2)}(r_{(i,p)}\zeta)\,\frac{r_{(i,p)}}{\zeta}, \qquad i = 1, \ldots, T-p. \tag{54}$$

With the above moments, all the quantities required to compute the expectation $\mathbb{E}_{\theta_{(j)}}\left[\nabla l^{\text{aug}}_{\mathbf{x},\mathbf{y}}(\theta)\right]$ are provided.

For example

$$\begin{aligned}
\mathbb{E}_{\theta_{(j)}}\left[\left.\frac{\partial l^{\text{aug}}_{\mathbf{x},\mathbf{y}}(\theta)}{\partial\lambda}\right|\mathbf{x}\right] &= \left\{\log\left(\frac{r_{(1,0)}}{\alpha\eta}\right) - \frac{\dot{K}_{\lambda-(1/2)}(\alpha r_{(1,0)})}{K_{\lambda-(1/2)}(\alpha r_{(1,0)})} - \frac{\dot{K}_{\lambda}(\omega)}{K_{\lambda}(\omega)}\right\} \\
&\quad + \sum_{i=1}^{p-1}\left\{\log\left(\frac{\alpha}{r_{(1,i-1)}}\right) - \frac{\dot{K}_{\lambda-(i/2)}(\alpha r_{(1,i-1)})}{K_{\lambda-(i/2)}(\alpha r_{(1,i-1)})}\right\} \\
&\quad + \sum_{i=1}^{T-p}\left\{\log\left(\frac{\alpha}{r_{(i,p-1)}}\right) - \frac{\dot{K}_{\lambda-(p/2)}(\alpha r_{(i,p-1)})}{K_{\lambda-(p/2)}(\alpha r_{(i,p-1)})}\right\} \\
&\quad + \mathbb{E}_{\theta_{(j)}}\left[\left.\sum_{i=1}^{T-1}\log(Y_{i+1})\right|\mathbf{x}\right],
\end{aligned}$$

where

$$\begin{aligned}
\mathbb{E}_{\theta_{(j)}}\left[\left.\sum_{i=1}^{T-1}\log(Y_{i+1})\right|\mathbf{x}\right] &= \sum_{i=1}^{p-1}\mathbb{E}^{\diamond}_{\theta_{(j)}}\left[\log(Y_{i+1})\right] + \sum_{i=1}^{T-p}\mathbb{E}^{*}_{\theta_{(j)}}\left[\log(Y_{i+p})\right] \\
&= \sum_{i=1}^{p-1}\left\{\frac{\dot{K}_{\lambda-((i+1)/2)}(r_{(1,i)}\zeta)}{K_{\lambda-((i+1)/2)}(r_{(1,i)}\zeta)} + \log\left(\frac{r_{(1,i)}}{\zeta}\right)\right\} \\
&\quad + \sum_{i=1}^{T-p}\left\{\frac{\dot{K}_{\lambda-((p+1)/2)}(r_{(i,p)}\zeta)}{K_{\lambda-((p+1)/2)}(r_{(i,p)}\zeta)} + \log\left(\frac{r_{(i,p)}}{\zeta}\right)\right\}.
\end{aligned}$$

In the first equality the expectations $\mathbb{E}^{\diamond}$ and $\mathbb{E}^{*}$ are given by (51) and (52), respectively. Re-arranging expressions we get Eq. (43). Hence, by simple substituting the required moments while taking the expectations of (44–48), we get all the other equalities in the same way and therefore the stated result (28) follows (Table 2)

Appendix F

Some properties of Modified Bessel functions

Modified bessel functions	
Properties	Asymptotic expansions as $x \downarrow 0$
1. $K_{-\nu}(x) = K_\nu(x)$	1. $K_\nu(x) \sim (1/2)\Gamma(\nu)\,((x/2))^{-\nu}, \quad \nu > 0$
2. $K_{1/2}(x) = \sqrt{(\pi/2x)}\mathrm{e}^{-x}$	2. $K_\nu(x) \sim (1/2)\Gamma(-\nu)\,((x/2))^{\nu}, \quad \nu < 0$
3. $K_{\nu+\varepsilon}(x) > K_\nu(x), \qquad \nu, \varepsilon, x > 0$	3. $K_0(x) \sim -\log(x)$
4. $K_{\nu+1}(x) = (2\nu/x)K_\nu(x) + K_{\nu-1}(x)$	
Integral representation	Asymptotic expansion as $x \uparrow \infty$
$K_\nu(x) = (1/2)\int_0^\infty y^{\nu-1}\exp\left(-(x/2)(y + y^{-1})\right)\mathrm{d}y$	$K_\nu(x) \sim \sqrt{(\pi/2x)}\mathrm{e}^{-x}$
Derivatives $\partial/\partial x$	
1. $K_0'(x) = -K_1(x)$	4. $(\log K_\nu(x))' = \nu/x - R_\nu(x)$
2. $K_\nu'(x) = -(1/2)(K_{\nu+1}(x) + K_{\nu-1}(x))$	
3. $K_\nu'(x) = (\nu/x)K_\nu(x) - K_{\nu+1}(x)$	where $R_\nu(x) := \frac{K_{\nu+1}(x)}{K_\nu(x)}, \quad x > 0$

Reference Abramowitz and Stegun (1992) (Chapter 9) and Eberlein and Hammerstein (2004).

Acknowledgements The research of the first author was partially supported by the National Council for Science and Technology of México (CONACYT) and of the second author by an EPSRC Advanced Research Fellowship. We are thankful to two anonymous referees for their helpful and constructive comments.

References

Abramowitz, M., Stegun, I. A. (Eds.) (1992). *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. New York: Dover Publications Inc. Reprint of the 1972 edition.

Andersson, J. (2001). On the normal inverse Gaussian stochastic volatility model. *Journal of Business Economic Statistics*, *19*(1), 44–54.

Bain, L. J. (1969). Moments of non-central t and non-central F-distribution. *The American Statistician. 23*(4), 33–34.

Barndorff-Nielsen, O. E. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society. London A. 353*, 401–419.

Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics 24*(1), 1–13.

Barndorff-Nielsen, O. E., Shephard, N. (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics (with discussion). *Journal of the Royal Statistical Society. Series B. Statistical Methodology 63*, 167–241.

Bollerslev, T. (1987). A conditionally heteroskedastic time series model for speculative prices and rates of return. *The Review of Economics and Statistics. 69*, 542–47.

Bollerslev, T., Chou, R. C., Kroner, K. F. (1992). ARCH modelling in finace: a review of the theory and empirical evidence. *Journal of Econometrics 52*, 5–56.

Cox, D. R. (1981). Statistical analysis of time series: some recent developments. *Scandinavian Journal of Statistics 8*, 93–115.

Eberlein, E., Hammerstein, E. A. (2004). *Generalized hyperbolic and inverse Gaussian distributions: Limiting cases and approximation of processes.* In: R.C. Dalang, M. Dozzi, F. Russo (Eds.), Seminar on stochastic analysis, random fields and applications IV, progress in probability, vol. 58, Birkhuser, pp. 221–264.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, *50*(4), 987–1008.

Gaver, D. P., Lewis, P. A. W. (1980). First-order autoregressive gamma sequences and point processes. *Advances in Applied Probability 12*, 727–745.

Jacobs, P. A., Lewis, P. A. W. (1977). A mixed autoregressive-moving average exponential sequence and point process (EARMA(1, 1)). *Advances in Applied Probability 9*, 87–104.
Jensen, M. B., Lunde, A. (2001). The NIG-S&ARCH model: a fat tailed, stochastic, and autoregressive conditional heteroskedastic volatility model. *Econometrics Journal 4*, 319–342.
Joe, H. (1996). Time series models with univariate margins in the convolution-closed infinitely divisible class. *Journal of Applied Probability 33*, 664–77.
Jørgensen, B. (1982). *Statistical properties of the generalized inverse Gaussian distribution*. Lecture notes in statistics, vol. 9. Berlin Heidelberg New York: Springer
Jørgensen, B., Song, P. (1998). Stationary time series models with exponentials dispersion model margins. *Journal of Applied Probability 35*, 78–92.
Lawrance, A. J. (1982). The innovation distribution of a gamma distributed autoregressive process. *Scandinavian Journal of Statistics 9*, 234–36.
Lawrance, A. J. (1991). Directionality and reversivility in time series. *International Statistical Review 59*, 67–79.
Lawrance, A. J., Lewis, P. A. W. (1977). An exponential moving-average sequence and point process (EMA1). *Journal of Applied Probability 14*, 98–113.
Lawrance, A. J., Lewis, P. A. W. (1980). The exponential autoregressive-moving average EARMA(p, q) process. *Journal of the Royal Statistical Society. Series B. Statistical Methodology 42*, 150–61.
Liu, J. S., Wong, W. H., Kong, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimation and augmentation schemes. *Biometrika 81*, 27–40.
Louis, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society. Series B. Statistical Methodology 44*, 226–33.
McKenzie, E. (1986). Autoregressive moving-average processes with negative-binomial and geometric marginal distributions. *Advances in Applied Probability 18*, 679–705.
McKenzie, E. (1988). Some ARMA models for dependent sequences of Poisson counts. *Advances in Applied Probability 20*, 822–835.
Mencía, F. J., Sentana, E. (2004). Estimation and testing of dynamic models with generalised hyperbolic innovations. Technical Report 0411, CEMFI.
Nelson, D. B. (1990). Stationarity and persistance in the GARCH(1, 1) model. *Econometric Theory 6*, 318–334.
Pitt, M. K., Walker, S. G. (2005). Constructing stationary time series models using auxiliary variables with applications. *Journal of the American Statistical Association 100*, 554–564.
Pitt, M. K., Chatfield, C., Walker, S. G. (2002). Constructing first order autoregressive models via latent processes. *Scandinavian Journal of Statistics*, *29*, 657–663.
Prause, K. (1999). *The generalized hyperbolic model: estimation, financial derivatives, and risk measures*. Ph.D. Thesis, Universität Freiburg.
Press, W., Teukolsky, S., Vetterling, W., Flannery, B. (1992). *Numerical Recipes. C*. Cambridge University Press.
Raible, S. (2000). *Lévy processes in finance: theory, numerics, and empirical facts*. Ph.D. Thesis, Universität Freiburg.
Robert, C. P., Casella, G. (2002). *Monte Carlo statistical methods*, 3rd edn. Berlin Heidelberg NewYork: Springer.
Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In: D. R. Cox, D. V. Hinkley, O. E. Barndorff-Nielsen, Eds. *Time series models in econometrics finance and other fields*, Chapman and Hall, London, pp. 1–67.
Walker, S. G. (1996). An EM algoritm for nonlinear random effects models. *Biometrics*, **52**, 934–944.