Univariate Generalized Autoregressive Score Volatility Models

Drew Creal[a],* Siem Jan Koopman[b,d], André Lucas[c,d]

(a) *Booth School of Business, University of Chicago*

(b) *Department of Econometrics, VU University Amsterdam*

(c) *Department of Finance, VU University Amsterdam & Duisenberg School of Finance*

(d) *Tinbergen Institute, Amsterdam*

September 23, 2012

PRELIMINARY DRAFT: COMMENTS WELCOME

Abstract

This short note develops the generalized autoregressive score volatility models for different parametric distributions including the normal, Laplace, Student's t, and several special cases of the generalized hyperbolic distribution. This note is written to accompany publicly available Matlab code that implements the models.

Keywords: generalized autoregressive score model, GARCH model, volatility.

JEL classification codes: C32.

---

*Lucas thanks the Dutch National Science Foundation (NWO) for financial support.

Corresponding author: Drew Creal, University of Chicago, Booth School of Business, 5807 S. Woodlawn Ave., Chicago, IL 60637; Tel 773.834.5249, Email: dcreal@chicagobooth.edu.

1 Generalized autoregressive score models

The generalized autoregressive score (GAS) model is specified as

$$y_t \sim p(y_t|f_t;\theta), \tag{1}$$

$$f_{t+1} = \omega + \sum_{i=1}^{p} A_i s_{t-i+1} + \sum_{j=1}^{q} B_j f_{t-j+1}, \tag{2}$$

where θ is a vector of unknown parameters and f_t represents a time-varying parameter that characterizes some aspect of the conditional distribution $p(y_t|f_t;\theta)$. The conditional distribution may depend on additional observable variables (covariates) but we omit this from the notation here for simplicity. The factors are driven through time by the scaled score function

$$s_t = S_t \nabla_t, \tag{3}$$

$$\nabla_t = \frac{\partial \log p(y_t|f_t;\theta)}{\partial f_t} \tag{4}$$

where S_t is a user-defined scaling matrix such as the Fisher information matrix

$$S_t = -\mathrm{E}_{t-1}\left[\nabla_t \nabla_t'\right]^{-1}. \tag{5}$$

Further details on the properties of GAS models can be found in Creal, Koopman, and Lucas (2012), Creal, Koopman, and Lucas (2011), and Blasques, Koopman, and Lucas (2012).

In this document, we focus on univariate volatility models where the time-varying factor f_t determines the variance (or log-variance) of the conditional distribution $p(y_t|f_t;\theta)$. The document is intended as an extended "read me" file to accompany Matlab code that implements the models. Throughout this document, we assume that the conditional mean of the data has been appropriately modeled. For simplicity, we assume that y_t has mean zero with variance σ_t^2. We will consider the parameterizations $f_t = \sigma_t^2$ and $f_t = \log(\sigma_t^2)$.

When deriving GAS models it is easiest to take derivatives w.r.t. σ_t^2 and correct for any changes in the parameterization in a second step. The score density with respect to f_t is then

$$\nabla_t = \frac{\partial \log p(y_t|f_t;\theta)}{\partial \sigma_t^2} \frac{\partial \sigma_t^2}{\partial f_t}$$

In the following, we derive the score function for the first term. In cases where the factor is parameterized as $f_t = \log(\sigma_t^2)$, the second term is $\frac{\partial \sigma_t^2}{\partial f_t} = \exp(f_t)$ and when $f_t = \sigma_t^2$ then this second term is of course just equal to one.

For those interested only in the Matlab code, please skip to Section 4. This section includes some details on how the models are implemented.

2 Basic GAS volatility models

In this section, we cover the basic volatility models including the symmetric Gaussian, Student's t, and Laplace distributions.

2.1 Gaussian GAS models

Suppose the observations y_t have a conditional normal distribution with log-density

$$\log p(y_t|f_t;\theta) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log\sigma_t^2 - \frac{1}{2}\frac{y_t^2}{\sigma_t^2}$$

where σ_t^2 is the conditional variance of y_t. When $f_t = \sigma_t^2$, the score, inverse of the Fisher information matrix, and GAS(1,1) factor recursion are

$$\begin{aligned}\nabla_t &= \frac{y_t^2}{2f_t^2} - \frac{1}{2f_t}\\ -\mathrm{E}_{t-1}\left[\nabla_t\nabla_t'\right]^{-1} &= 2f_t^2\\ f_{t+1} &= \omega + A\left(y_t^2 - f_t\right) + Bf_t\end{aligned}$$

When the observation density has a normal distribution and $f_t = \sigma_t^2$, the GAS(1,1) model reduces to the GARCH(1,1) model of Engle (1982) and Bollerslev (1986). The factor recursion is however parameterized in a different way. Typically, the GARCH(1,1) model is parameterized as

$$f_{t+1} = \omega + \alpha y_t^2 + \beta f_t$$

such that $\alpha = A$ and $\beta = B - A$.

When $f_t = \log(\sigma_t^2)$, the score, inverse of the Fisher information matrix, and GAS(1,1) factor recursion are

$$\begin{aligned}
\nabla_t &= \frac{y_t^2}{2\exp(f_t)} - \frac{1}{2} \\
-\mathrm{E}_{t-1}\left[\nabla_t \nabla_t'\right]^{-1} &= 2 \\
f_{t+1} &= \omega + A\left(\frac{y_t^2}{\exp(f_t)} - 1\right) + Bf_t
\end{aligned}$$

When an alternative parameterization is desired, the factor recursion in the GAS model naturally adapts.

2.2 Student's t GAS models

The log-density of the Student's t distribution is given by

$$\begin{aligned}
\log p(y_t|f_t;\theta) &= \log\Gamma\left(\frac{\nu+1}{2}\right) - \log\Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2}\log(\pi) - \frac{1}{2}\log(\nu-2) \\
&\quad -\frac{1}{2}\log(\sigma_t^2) - \frac{(\nu+1)}{2}\log\left(1+\frac{y_t^2}{(\nu-2)\sigma_t^2}\right),
\end{aligned}$$

where σ_t^2 is the variance (and not the scale parameter). The score w.r.t. σ_t^2 is

$$\frac{\partial \log p(y_t|f_t;\theta)}{\partial \sigma_t^2} = \frac{(\nu+1)}{2}\left(1+\frac{y_t^2}{(\nu-2)\sigma_t^2}\right)^{-1}\frac{y_t^2}{(\nu-2)\sigma_t^4} - \frac{1}{2\sigma_t^2}$$

The negative of the inverse of the information matrix is

$$-\mathrm{E}_{t-1}\left[\frac{\partial^2 \ln p(y_t|f_t;\theta)}{\partial\sigma_t^2\partial\sigma_t^2}\right]^{-1} = \frac{2\sigma_t^4(\nu+3)}{\nu}.$$

When $f_t = \sigma_t^2$, the score, inverse of the information matrix, and factor recursion are

$$\begin{aligned}
\nabla_t &= \frac{(\nu+1)}{2}\left(1+\frac{y_t^2}{(\nu-2)f_t}\right)^{-1}\frac{y_t^2}{(\nu-2)f_t^2}-\frac{1}{2f_t} \\
-\mathrm{E}_{t-1}\left[\nabla_t\nabla_t'\right]^{-1} &= \frac{2f_t^2\,(\nu+3)}{\nu} \\
f_{t+1} &= \omega + A\left[\frac{(\nu+3)}{\nu}\left(\left(1+\frac{y_t^2}{(\nu-2)}\right)^{-1}\frac{(\nu+1)\,y_t^2}{(\nu-2)}-f_t\right)\right]+Bf_t
\end{aligned}$$

A key feature of the Student's t GAS(1,1) model that differentiates it from the Gaussian model is the weighting term

$$\begin{aligned}
w_t &= \left(1+\frac{y_t^2}{(\nu-2)}\right)^{-1}\frac{(\nu+1)}{(\nu-2)} \\
f_{t+1} &= \omega + A\frac{(\nu+3)}{\nu}\left(w_t y_t^2 - f_t\right)+Bf_t
\end{aligned}$$

The weight function w_t lessens the impact of occasional extreme observations y_t. The intuition is simple. As the data are heavy-tailed, the model expects there to be rare large values of y_t that are not necessarily caused by a large variance. We note that as $\nu \to \infty$, the weight converges to one and the Student's t GAS(1,1) reduces to the Gaussian model above.

When $f_t = \log(\sigma_t^2)$, the score, inverse of the information matrix, and factor recursion are

$$\begin{aligned}
\nabla_t &= \frac{(\nu+1)}{2}\left(1+\frac{y_t^2}{(\nu-2)\exp(f_t)}\right)^{-1}\frac{y_t^2}{(\nu-2)\exp(f_t)}-\frac{1}{2} \\
-\mathrm{E}_{t-1}\left[\nabla_t\nabla_t'\right]^{-1} &= \frac{2\,(\nu+3)}{\nu} \\
f_{t+1} &= \omega + A\left[\frac{(\nu+3)}{\nu}\left(\left(1+\frac{y_t^2}{(\nu-2)}\right)^{-1}\frac{(\nu+1)\,y_t^2}{(\nu-2)\exp(f_t)}-1\right)\right]+Bf_t
\end{aligned}$$

Multivariate versions of the Student's t model have been developed by Creal, Koopman, and Lucas (2011).

2.3 Laplace GAS models

The density of the Laplace distribution (sometimes called the double exponential distribution) is given by

$$p(y_t|f_t;\theta) \quad = \quad \frac{\sqrt{2}}{2\sqrt{\sigma_t^2}} \exp\left(-\frac{\sqrt{2}|y_t|}{\sqrt{\sigma_t^2}}\right)$$

where σ_t^2 is the variance of y_t. The log-density of the Laplace distribution is

$$\ln p(y_t|f_t;\theta) \quad = \quad -\frac{1}{2}\ln(2) - \frac{1}{2}\ln \sigma_t^2 - \frac{\sqrt{2}}{\sqrt{\sigma_t^2}}|y_t|.$$

The score vector with respect to σ_t^2 is

$$\nabla_t \quad = \quad -\frac{1}{2\sigma_t^2} + \frac{\sqrt{2}}{2}\left(\sigma_t^2\right)^{-3/2}|y_t|.$$

The second derivative with respect to σ_t^2 is given by

$$\frac{\partial^2 \ln p(y_t|f_t;\theta)}{\partial\sigma_t^2\partial\sigma_t^2} \quad = \quad \frac{1}{2\sigma_t^4} - \frac{3\sqrt{2}}{4}\left\{\sigma_t^2\right\}^{-5/2}|y_t|$$

and taking expectations we have

$$\mathrm{E}_{t-1}\left[\frac{\partial^2 \ln p(y_t|f_t;\theta)}{\partial\sigma_t^2\partial\sigma_t^2}\right] \quad = \quad \frac{1}{2\sigma_t^4} - \frac{3\sqrt{2}}{4}\left\{\sigma_t^2\right\}^{-5/2}\mathrm{E}_{t-1}\left[|y_t|\right].$$

It can be shown that $\mathrm{E}_{t-1}\left[|y_t|\right] = \sqrt{\frac{\sigma_t^2}{2}}$. Plugging this in we get

$$\mathrm{E}_{t-1}\left[\frac{\partial^2 \ln p(y_t|\mu,\sigma_t^2)}{\partial\sigma_t^2\partial\sigma_t^2}\right] \quad = \quad \frac{1}{2\sigma_t^4} - \frac{3}{4}\left\{\sigma_t^2\right\}^{-5/2}\left\{\sigma_t^2\right\}^{1/2} \quad = \quad -\frac{1}{4\sigma_t^4}.$$

When $f_t = \sigma_t^2$, the score, inverse information matrix, and GAS(1,1) dynamics are given by

$$\begin{aligned}\nabla_t &= \frac{\sqrt{2}}{2} f_t^{-3/2} |y_t| - \frac{1}{2 f_t} \\ -\mathrm{E}_{t-1}\left[\nabla_t \nabla_t'\right]^{-1} &= 4 f_t^2 \\ f_{t+1} &= \omega + A\left(\frac{2\sqrt{2}}{\sqrt{f_t}} |y_t| - 2 f_t\right) + B f_t\end{aligned}$$

When $f_t = \log(\sigma_t^2)$, the score, inverse information matrix, and GAS(1,1) dynamics are given by

$$\begin{aligned}\nabla_t &= \frac{\sqrt{2}|y_t|}{2\exp\left(f_t/2\right)} - \frac{1}{2} \\ -\mathrm{E}_{t-1}\left[\nabla_t \nabla_t'\right]^{-1} &= 4 \\ f_{t+1} &= \omega + A\left(\frac{2\sqrt{2}}{\exp\left(f_t/2\right)} |y_t| - 2\right) + B f_t\end{aligned}$$

If we add the current value of y_t as an explanatory variable multiplied by a parameter γ to this last expression, we obtain an expression similar to the EGARCH(1,1) model of Nelson (1991). To make the models exactly equivalent, one needs to specify the observation density as an asymmetric Laplace distribution; see, Creal, Koopman, and Lucas (2012) for details. A multivariate Laplace distribution is implicitly part of the multivariate generalized hyperbolic distribution developed by Zhang, Creal, Koopman, and Lucas (2012).

3 Generalized Hyperbolic GAS models (PRELIMINARY)

The generalized hyperbolic (GH) distribution is useful because for different limits of its parameters it contains as special cases the Gaussian, Student's t, and Laplace distributions as well as the skewed Student's t, skewed Laplace, normal gamma (NG), normal inverse Gaussian (NIG), and the normal reciprocal inverse Gaussian (NRIG) distributions. By deriving the GAS model for the GH distribution, one automatically derives the models for the sub-classes. The main drawback of the distribution is that it can lead to complicated expressions for the score and expressions for the information matrix that may not have closed-form solutions. Prause (1999) provides a detailed discussion of the GH distribution and its properties while Barndorff-Nielsen and Shephard (2012) discuss its relationship to Lévy processes.

3.1 Generalized hyperbolic distribution

The generalized hyperbolic (GH) distribution was introduced by Barndorff-Nielsen (1977) as a mean-variance mixture of normal distributions. It is constructed as

$$
\begin{aligned}
y &= \mu + \beta\zeta + \sqrt{\zeta} Z \qquad\qquad Z \sim \mathrm{N}(0,1) \\
\zeta &\sim \mathrm{GIG}(\lambda, \chi, \psi)
\end{aligned}
$$

where the variance ζ is a generalized inverse Gaussian (GIG) random variable with probability density function

$$
p(\zeta|\lambda, \chi, \psi) = \left(\sqrt{\frac{\psi}{\chi}}\right)^{\lambda} \frac{1}{2K_\lambda\left(\sqrt{\chi\psi}\right)} \zeta^{\lambda-1} \exp\left(-\frac{1}{2}\left[\chi\zeta^{-1} + \psi\zeta\right]\right)
$$

and $K_\nu(x)$ is the modified Bessel function of the second kind. The GH distribution has five parameters $\mathrm{GH}(\mu, \beta, \lambda, \chi, \psi)$. The parameter μ controls the location and γ determines the skewness. We provide a brief listing of the distributions that are special cases of the GH distribution

$$
\begin{aligned}
\mathrm{GH}\left(\mu, \beta, -\frac{1}{2}, \chi, \psi\right) &= \text{Normal inverse Gaussian (NIG)} \\
\mathrm{GH}\left(\mu, \beta, \frac{1}{2}, \chi, \psi\right) &= \text{Normal reciprocal inverse Gaussian (NRIG)} \\
\mathrm{GH}(\mu, \beta, 1, \chi, \psi) &= \text{Hyperbolic} \\
\mathrm{GH}(\mu, \beta, 0, \chi, \psi) &= \text{Hyperboloid} \\
\mathrm{GH}\left(\mu, \beta, \frac{1}{\gamma}, 0, \frac{2}{\gamma}\right) &= \text{Normal gamma (variance gamma)} \\
\mathrm{GH}\left(\mu, \beta, -\frac{\nu}{2}, \nu, 0\right) &= \text{skewed Student's } t
\end{aligned}
$$

where we omit the details of the parameterizations.[†] Note that the normal gamma distribution is a generalization of the Laplace distribution and is sometimes titled the generalized asymmetric

[†] In a future version of this document, we will add these.

Laplace distribution; see, e.g. Kotz, Kozubowski, and Podgórski (2001). For further discussion of the GH distribution and how these distributions are derived see Prause (1999) and Barndorff-Nielsen and Shephard (2012).

3.2 Generalized hyperbolic GAS models

Zhang, Creal, Koopman, and Lucas (2012) construct multivariate GH volatility models for time-varying covariance matrices. Here, we will consider several univariate versions of their model. We start by specifying the model as

$$y_t = \sigma_t \varepsilon_t$$

where ε_t is a GH random variable standardized such that $\mathrm{E}[\varepsilon_t] = 0$ and $\mathrm{V}[\varepsilon_t] = 1$. These properties imply that σ_t^2 is the variance of y_t.

In order to ensure that ε_t is standardized we construct it as follows

$$\begin{aligned} \varepsilon_t &= \mu_\eta + \zeta_t \gamma \alpha + \sqrt{\zeta_t} \alpha \eta_t \qquad \eta_t \sim \mathrm{N}(0,1) \\ \zeta_t &\sim \mathrm{GIG}(\lambda, \chi, \psi) \end{aligned}$$

The parameter α is a scale parameter while μ_η determines the location. We will see that α and μ_η are not free parameters as they are functions of the parameters (λ, χ, ψ) of the mixing variable ζ_t. Let $\mathrm{E}[\zeta_t] = \mu_\zeta$ and $\mathrm{V}[\zeta_t] = \sigma_\zeta^2$ denote the mean and variance of the mixing variable ζ_t. To ensure that $\mathrm{E}[\varepsilon_t] = 0$ and $\mathrm{V}[\varepsilon_t] = 1$, we impose the restrictions

$$\begin{aligned} \mu_\eta &= -\mu_\zeta \gamma \alpha \\ \alpha^{-2} &= \mu_\zeta + \sigma_\zeta^2 \gamma^2 \end{aligned}$$

These restrictions ensure that σ_t^2 is the time-varying variance parameter. We note that in this formulation the parameter γ no longer controls the skewness alone as the parameters (χ, ψ) of the mixing variable ζ_t are also involved. However, it is still the case that $\gamma = 0$ is symmetric, $\gamma < 0$ is left skewed, and $\gamma > 0$ is right skewed.

Finally, we need to impose one additional restriction on the model in order to identify it.

The parameters χ and ψ cannot both be identified as they act as scale parameters controlling the level of volatility. We impose the condition that $\mathrm{E}\left[\zeta_t\right] = 1$ and we estimate the parameter $\kappa = \sqrt{\chi\psi}$. For a fixed value of κ, we can obtain χ and ψ by the equality

$$1 = \mu_\zeta = \frac{\sqrt{\chi\psi}K_{\lambda+1}\left(\sqrt{\chi\psi}\right)}{\psi K_\lambda\left(\sqrt{\chi\psi}\right)} \quad \Leftrightarrow \quad \psi = \frac{\kappa \cdot K_{\lambda+1}\left(\kappa\right)}{K_\lambda\left(\kappa\right)},$$

with $\chi = \kappa^2/\psi$.

From the results in the appendix, the GH density under this parameterization and with the identifying restriction $\mathrm{E}\left[\zeta_t\right] = \mu_\zeta = 1$ can be expressed as

$$p(y_t|f_t;\theta) = \frac{1}{\sqrt{2\pi\alpha^2\sigma_t^2}}\left(\sqrt{\frac{\gamma^2+\psi}{\left(\frac{(y_t-\sigma_t\mu_\eta)^2}{\sigma_t^2\alpha^2}+\chi\right)}}\right)^{\frac{1}{2}}\left(\sqrt{\frac{\psi\left(\frac{(y_t-\sigma_t\mu_\eta)^2}{\sigma_t^2\alpha^2}+\chi\right)}{\chi\left(\gamma^2+\psi\right)}}\right)^{\lambda}\exp\left(\frac{\gamma\left(y_t-\sigma_t\mu_\eta\right)}{\sqrt{\sigma_t^2\alpha^2}}\right)$$
$$\frac{K_{\lambda-\frac{1}{2}}\left(\sqrt{\left(\frac{(y_t-\sigma_t\mu_\eta)^2}{\sigma_t^2\alpha^2}+\chi\right)\left(\gamma^2+\psi\right)}\right)}{K_\lambda\left(\sqrt{\chi\psi}\right)}$$

which is unfortunately complicated due to the modified Bessel function. Zhang, Creal, Koopman, and Lucas (2012) provide the score for the general multivariate case with a time-varying covariance matrix. However, we focus here on several special cases. Indeed, there are two models for which the score does not involve the Bessel functions which will provide some intuition. We use the fact that the modified Bessel function of the second kind has simple expressions for any half-integer $K_{\pm\frac{1}{2}}\left(x\right) = \sqrt{\frac{\pi}{2x}}\exp\left(-x\right)$ and the recursion $K_{\lambda+1}(x) = K_{\lambda-1}(x) + \frac{2\lambda}{x}K_\lambda(x)$.

TO BE COMPLETED. We plan to derive each of these special cases in a future version of this document. The current version of the code has implemented several univariate versions of these models, but they should be viewed as preliminary. New updates to the Matlab code will be available shortly.

4 Matlab Code

There exists a set of Matlab code that implements the models listed above for two types of parameterizations $f_t = \sigma_t^2$ and $f_t = \log\left(\sigma_t^2\right)$. We will illustrate the models on returns for the S&P500 (GSPC) and J.P. Morgan (JPM). There are several key aspects of the code that users should consider before using it. This is because depending on the application users may want to change these.

1. For all models, the code uses the unconditional mean of the factor f_t as the initial condition: $f_0 = \omega\,(1 - B)^{-1}$.

2. For all models, the code estimates the unconditional mean of the factor f_t. This is denoted by ω in the output. Depending on how the factor is parameterized, this may or may not be the long-run mean of the variance.

3. For all models, the code only implements GAS(1,1) versions of the model but it should be easy to extend this to higher order processes.

4. For all models, the code outputs the score vector ∇_t, the variances σ_t^2, and the scaled scores $s_t = S_t \nabla_t$ for each date.

5. For all models, the code has a function that returns the value of the log-likelihood function to the optimizer. Since the optimizer is a minimizer, these functions return the (negative) of the log-likelihood function scaled by the total number of observations.

6. As the code is intended for illustration purposes, we do not model the conditional mean function of the data y_t. Currently, the code just demeans the data using the sample mean.

7. For the Gaussian, Laplace, and Student's t distributions, the code uses the Fisher information matrix as the scaling matrix S_t in the GAS factor recursion.

8. For the models built from the GH distribution, we have not managed to calculate the Fisher information matrix in closed form. The code currently uses the information matrix of the Student's t distribution. We hope to change this in a future version of the code.

9. The skewed Student's t distribution imposes that the degrees of freedom is greater than four, $\nu > 4$.

10. The standard errors are computed by numerically inverting the Hessian matrix at the ML estimates. We hope to implement robust standard errors in a future version of the code.

11. The code for the GH models is unfortunately slow as it is written in Matlab. We hope to provide routines implemented in C code soon.

5 Conclusion

This document was created as an overview for univariate GAS volatility models. It is intended to accompany publicly available Matlab code that implements the models. An interesting GAS volatility model that we have not discussed is a model built from the skewed Student's t distribution of Bauwens and Laurent (2005). Additional models that have not been developed but could be interesting are GAS versions of the ARCH-M model of Engle, Lilien, and Robins (1987), where the variance enters the conditional mean of returns.

References

Barndorff-Nielsen, O. E. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences 353*(1674), 401–419.

Barndorff-Nielsen, O. E. and N. Shephard (2012). Basics of Lévy processes. Working paper, Oxford-Man Institute, University of Oxford.

Bauwens, L. and S. Laurent (2005). A new class of multivariate skew densities, with application to generalized autoregressive conditional heteroscedasticity models. *Journal of Business and Economic Statistics 23*(3), 346–354.

Blasques, F., S. J. Koopman, and A. Lucas (2012). Stationarity and ergodicity of univariate generalized autoregressive score processes. Working paper, VU University, Amsterdam.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics 31*(3), 307–327.

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

Creal, D. D., S. J. Koopman, and A. Lucas (2012). Generalized autoregressive score models with applications. *Journal of Applied Econometrics*. forthcoming.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica 50*(4), 987–1007.

Engle, R. F., D. M. Lilien, and R. P. Robins (1987). Estimating time varying risk premia in the term structure: the ARCH-M model. *Econometrica 55*(2), 391–407.

Kotz, S., T. J. Kozubowski, and K. Podgórski (2001). *The Laplace Distribution and Generalizations*. Boston, MA: Birkhauser.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica 59*(2), 347–370.

Prause, K. (1999). The generalized hyperbolic model: estimation, financial derivatives, and risk measures.

Zhang, X., D. D. Creal, S. J. Koopman, and A. Lucas (2012). Modeling dynamic volatilities and correlations under skewness and fat tails. Working paper, University of Chicago, Booth School of Business.

A Appendix

$$
\begin{aligned}
p(y_t|f_t;\theta) &= \int_0^\infty \frac{1}{\sqrt{2\pi\alpha^2\sigma_t^2}}\zeta_t^{-\frac{1}{2}}\exp\left(-\frac{1}{2\sigma_t^2\alpha^2}\zeta_t^{-1}\left[y_t-\sigma_t\left\{\mu_\eta+\zeta_t\gamma\alpha\right\}\right]^2\right) \\
&\quad \left(\sqrt{\frac{\psi}{\chi}}\right)^\lambda \frac{1}{2K_\lambda\left(\sqrt{\chi\psi}\right)}\zeta^{\lambda-1}\exp\left(-\frac{1}{2}\left[\chi\zeta^{-1}+\psi\zeta\right]\right)d\zeta_t \\
&= \frac{1}{\sqrt{2\pi\alpha^2\sigma_t^2}}\left(\sqrt{\frac{\psi}{\chi}}\right)^\lambda \frac{1}{2K_\lambda\left(\sqrt{\chi\psi}\right)}\int_0^\infty \zeta_t^{\lambda-\frac{1}{2}-1}\exp\left(-\frac{1}{2\sigma_t^2\alpha^2}\zeta_t^{-1}\left[y_t-\sigma_t\left\{\mu_\eta+\zeta_t\gamma\alpha\right\}\right]^2\right) \\
&\quad \exp\left(-\frac{1}{2}\left[\chi\zeta^{-1}+\psi\zeta\right]\right)d\zeta_t \\
&= \frac{1}{\sqrt{2\pi\alpha^2\sigma_t^2}}\left(\sqrt{\frac{\psi}{\chi}}\right)^\lambda \frac{1}{2K_\lambda\left(\sqrt{\chi\psi}\right)}\exp\left(\frac{\gamma\left(y_t-\sigma_t\mu_\eta\right)}{\sqrt{\sigma_t^2\alpha^2}}\right) \\
&\quad \int_0^\infty \zeta_t^{\lambda-\frac{1}{2}-1}\exp\left(-\frac{1}{2}\left[\left\{\frac{\left(y_t-\sigma_t\mu_\eta\right)^2}{\sigma_t^2\alpha^2}+\chi\right\}\zeta^{-1}+\left(\gamma^2+\psi\right)\zeta\right]\right)d\zeta_t \\
&= \frac{1}{\sqrt{2\pi\alpha^2\sigma_t^2}}\left(\sqrt{\frac{\gamma^2+\psi}{\left(\frac{(y_t-\sigma_t\mu_\eta)^2}{\sigma_t^2\alpha^2}+\chi\right)}}\right)^{\frac{1}{2}}\left(\sqrt{\frac{\psi\left(\frac{(y_t-\sigma_t\mu_\eta)^2}{\sigma_t^2\alpha^2}+\chi\right)}{\chi\left(\gamma^2+\psi\right)}}\right)^\lambda \exp\left(\frac{\gamma\left(y_t-\sigma_t\mu_\eta\right)}{\sqrt{\sigma_t^2\alpha^2}}\right) \\
&\quad \frac{K_{\lambda-\frac{1}{2}}\left(\sqrt{\left(\frac{(y_t-\sigma_t\mu_\eta)^2}{\sigma_t^2\alpha^2}+\chi\right)\left(\gamma^2+\psi\right)}\right)}{K_\lambda\left(\sqrt{\chi\psi}\right)}
\end{aligned}
$$

where $\alpha^2=\left(1+\sigma_\zeta^2\gamma\right)$.