Bayesian estimation of NIG-parameters by Markov chain Monte Carlo methods

Jostein Lillestøl
Institute of Finance and Management Science
The Norwegian School of Economics and Business Adminstration
Bergen, Norway

E-mail: jostein.lillestol@nhh.no

First draft September 16, 1998
Revision of January 15, 2001

Abstract

The Normal Inverse Gaussian (NIG) distribution recently introduced by Barndorff-Nielsen (1997) is a promising alternative for modelling financial data exhibiting skewness and fat tails. In this paper we explore the Bayesian estimation of NIG-parameters by Markov Chain Monte Carlo Methods.

KEY WORDS: Normal Inverse Gaussian distribution, Bayesian Analysis, Markov Chain Monte Carlo.

This work is partly done at Institut für Statistik und Ökonometrie, Humboldt Universität zu Berlin with support from the Ruhrgas Arena program.

1 Introduction

The empirical distributions of stock returns are typically skew and heavy tailed, and different families of distributions have been proposed to model returns. Most notably is the stable Paretian family with a long history, see McCulloch (1996) and Adler et.al. (1998). More recently the class of NIG-distributions, an acornym for Normal Inverted Gaussian, has been proposed by Barndorff-Nielsen (1997). He investigated its properties, the modelling of NIG-processes, the estimation of NIG-parameters and the fit to real financial data, see also Rydberg (1997). The NIG-framework has several desirable properties and opportunities for modeling. A number of problems in finance can be recasted within this framework, thus taking skewness and heavy tails into account, as demonstrated by Lillestøl (2000).

In this paper we focus on the estimation of NIG-parameters from the Bayesian viewpoint using Markov Chain Monte Carlo Methods. We will explore a convenience prior leading to simple updating formulas. This is a conjugate prior when an unobserved variate is included in the parameter set. We will give some examples on estimating real and simulated data, and make comparisons with the maximum likelihood estimates.

2 The NIG-distribution and a convenience prior

The NIG-distribution may be characterized by four parameters: $\mu, \delta, \alpha, \beta$ which relates mainly to location, scale, peakedness/heavy tail and skewness respectively. The moment-generating function of a $NIG(\alpha, \beta, \delta\gamma)$-variate X is given by

$$M_X(t) = E\exp(tX) = \exp(\mu t + \delta(\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta + t^2)}))$$

From this it is easily derived (let $\gamma = \sqrt{\alpha^2 - \beta^2}$)

$$EX = \mu + \delta \cdot \frac{\beta}{\gamma}$$

$$\mathrm{var}X = \delta\frac{\alpha^2}{\gamma^3}$$

$$\text{Skewness} = 3 \cdot \frac{\beta}{\alpha} \cdot \frac{1}{(\delta\gamma)^{1/2}}$$

$$\text{Kurtosis} = 3 \cdot (1 + 4(\frac{\beta}{\alpha})^2) \cdot \frac{1}{\delta\gamma}$$

The density function of X is fairly complicated involving Bessel functions. However, the distribution has a simple characterization as the marginal distribution of (X, Z) where

$$\begin{aligned} X \mid Z = z &\sim N(\mu + \beta z, z) \\ Z &\sim IG(\delta, \sqrt{\alpha^2 - \beta^2}) \quad \text{where} \quad 0 \leq \mid \beta \mid < \alpha \end{aligned}$$

Here $IG(\delta, \gamma)$ is the well known Inverted Gaussian distribution (also named Wald distribution), see Johnson and Kotz (1995).

Barndorff-Nielsen has studied the estimation of NIG-parameters by maximum likelihood methods. Given the complicated density this leads to likelihood equations that are very complicated and requires extensive programming combined with numerical skills. The program "hyp" developed by Blæsild et. al. (1992) solves the task, but this may not be readily available. Another possilibility is to use an EM type algorithm for maximum likelihood. This is more easily programmable with less challenging numerics, as long as the computer environment supports easy calculation of Bessel functions, as demonstrated by Karlis (2000). Still another possibility is to use the simple characterization of the distribution above and the fact that IG-variates are easy to simulate, see Michael et.al. (1976), and explore simulation-based approaches to the estimation problem within a Bayesian framework. The Bayesian approach to estimation of NIG parameters is concurrently explored by Karlis (2001), who also considers the case of covariates.

Now let $Y = (X, Z)$ with distribution as in the characterization above, i.e. probability density $f(y) = f(x, z) = f(z) \cdot f(x \mid z)$ where

$$\begin{aligned} f(x \mid z) &= (2\pi)^{-1/2} z^{-1/2} e^{-\frac{1}{2z}(x-(\mu+\beta z))^2} \\ f(z) &= (2\pi)^{-1/2} \delta e^{\delta\gamma} \cdot z^{-3/2} e^{-\frac{1}{2}(\delta^2 z^{-1} + \gamma^2 z)} \end{aligned}$$

that is

$$f(x, z) \propto \delta e^{\delta\gamma - \beta\mu} z^{-2} e^{\beta x + \mu\frac{x}{z} - \frac{1}{2}(\beta^2+\gamma^2)z - \frac{1}{2}(\mu^2+\delta^2)z^{-1}}$$

If we let $\theta = (\mu, \delta, \beta, \gamma)$ we see that the joint density is within the exponential family

$$f(y) = g(\theta) h(y) e^{\sum_{j=1}^{4} \phi_j(\theta) t_j(y)}$$

where

$$\begin{aligned}\phi_1(\theta) &= \beta \\ \phi_2(\theta) &= \mu \\ \phi_3(\theta) &= -(\beta^2 + \gamma^2) \\ \phi_4(\theta) &= -(\mu^2 + \delta^2)\end{aligned}$$

and

$$\begin{aligned}t_1(y) &= x \\ t_2(y) &= xz^{-1} \\ t_3(y) &= \frac{1}{2}z \\ t_4(y) &= \frac{1}{2}z^{-1}\end{aligned}$$

$$g(\theta) = \delta e^{\delta\gamma - \beta\mu}$$

This means that the conjugate prior distribution of θ is then of the form

$$p(\theta) \propto g(\theta)^{a_0} e^{\sum_{j=1}^{4} \phi_j(\theta) a_j}$$

where $(a_0, a_1, a_2, a_3, a_4)$ is a vector of superparameters, where a_0, a_3 and $a_4 > 0$. Given n independent realizations $y_1, y_2, ..., y_n$ of the variate Y, the posterior of θ is of the same form with

$$a_0' = a_0 + n \quad \text{and} \quad a_j' = a_j + \sum_{i=1}^{n} t_j(y_i)$$

This is an augmented posterior in the sense that we are really interested in a distribution of θ given the observable X, and the Z in $Y = (X, Z)$ is treated as unobservable or missing. However, we can get at the desired posterior by using Markov chain Monte Carlo ideas.

A closer look at the chosen distribution for $\theta = (\mu, \delta, \beta, \gamma)$ shows that (μ, β) is independent of (δ, γ) and that (μ, β) is bivariate normal with correlation

$$\rho = -\frac{1}{2}\frac{a_0}{\sqrt{a_3 a_4}}$$

The other parameters are

$$\begin{aligned}
\bar{\mu} &= \frac{1}{2(1-\rho^2)a_4}(a_2 - \frac{a_0}{2a_3} \cdot a_1) \\
\bar{\beta} &= \frac{1}{2(1-\rho^2)a_3}(a_1 - \frac{a_0}{2a_4} \cdot a_2) \\
\sigma_\mu^2 &= \frac{1}{2(1-\rho^2)a_4} \\
\sigma_\beta^2 &= \frac{1}{2(1-\rho^2)a_3}
\end{aligned}$$

Note therefore the expressions for $\bar{\mu}/\sigma_\mu^2$ and $\bar{\beta}/\sigma_\beta^2$ given by the two parentheses respectively. Note also the binding equation $\sigma_\beta^2/\sigma_\mu^2 = a_4/a_3$. By the reparametrization (μ, β) to (τ, β) where $\tau = \mu + \frac{a_0}{2a_4} \cdot \beta$ we get τ independent of β.

The distribution of (δ, γ) is

$$p(\delta, \gamma) \propto \delta^{a_0} e^{-a_3\gamma^2 + a_0\delta\gamma - a_4\delta^2} \quad \text{for} \quad \delta \geq 0, \gamma \geq 0.$$

By a linear transformation to get rid of the cross-term it is seen that

$$\delta^2 \sim Gamma(\frac{a_0+1}{2}, a_4 - \frac{a_0^2}{4a_3})$$

$$\gamma \mid \delta \sim Normal(\frac{a_0}{2a_3}\delta, \frac{1}{2a_3}) \quad \text{truncated at zero}$$

In the case that the conjugate prior is too restrictive to represent our data, it may be worthwhile to introduce another layer of superparameters. Most convenient is to take a_i's to be independent and

$$\begin{aligned}
a_i &\sim Normal(c_i, b_i) \quad \text{for} \quad i = 1, 2 \\
a_i &\sim Exponential(b_i) \quad \text{for} \quad i = 0, 3, 4.
\end{aligned}$$

If we let $a = (a_0, a_1, a_2, a_3, a_4)$, the posterior of a given θ is determined as

$$\begin{aligned}
a_1 &\sim Normal(c_1 + \beta b_1, b_1) \\
a_2 &\sim Normal(c_2 + \mu b_2, b_2) \\
a_3 &\sim Exponential(b_3 + \gamma^2 + \beta^2) \\
a_4 &\sim Exponential(b_4 + \mu^2 + \delta^2) \\
a_0 &\sim Exponential(b_0 + \beta\mu - \delta\gamma - log(\delta)).
\end{aligned}$$

We see that the estimation problem essentially splits in two parts, that of (μ, β) and that of (δ, γ), and that the conjugate prior imposes cross restrictions. These may be unrealistic and not remedied by introducing superparameters. A possibility would be to model the two parts separately, which leads to estimation of heteroscedastic normal regression and estimation of inverse Gaussian parameters, as suggested by Karlis (2001). There exists various partametrizations of the IG-distribution that may lead to different solutions to this Bayesian estimation problem.

3 A MCMC scheme for the posterior

In order to get at the posterior $p(\theta \mid x)$ we use the MCMC scheme using full conditional.

Let $w = (x, z, \theta)$ where x is observed, z unobserved, θ parameters. Let w_s be a subset of the components of w and w_{-s} the complementary set.

$$p(w_s \mid w_{-s}) \propto p(w)$$

where only the factors involving components of w_s in any product formula need to be retained. A special case of this is $p(\theta, z \mid x)$, where the marginal $p(\theta \mid x)$ is our interest. Various schemes for sampling from the posterior is given Robert and Cassela (1999). The NIG-model fits under the heading of data augmentation, which is a special case of the Gibbs sampler.

The full conditionals sufficient for the current problem are given by

1. $p(\theta, z \mid x) \propto p(x, z \mid \theta)p(\theta)$
2. $p(\theta_s \mid x, z, \theta_{-s}) \propto p(x, z \mid \theta)p(\theta_s \mid \theta_{-s})$
3. $p(z \mid x, \theta) \propto p(x \mid z, \theta)p(z \mid \theta)$

If this is written out in our case, we will see that formula 3 leads to

$$Z_i \quad i.i.d. \quad IG^*(\delta', \gamma')$$

where $IG^*(\delta, \gamma)$ denotes the distribution with density similar to IG, but with z^{-2} as multiplicative factor replacing $z^{-3/2}$. Both distributions are member of the family of generalized inverse Gaussian distributions $GIG(\lambda, \delta, \gamma)$ with $\lambda = -1/2$ and $\lambda = -1$ respectively.

The parameters in our case are given by

$$\delta' = (\delta^2 + (X_i - \mu)^2)^{1/2}$$
$$\gamma' = \sqrt{\gamma^2 + \beta^2} = \alpha$$

The Z_i's can be simulated using the ideas in Michael et.al. (1976), or by rejection sampling methods, as described in some detail in an appendix.

Formula 2 for the complete parameter set θ as well as for (μ, β) and (γ, δ) separately just reiterates the result of our choice of conjugate prior for the augmented (X, Z), where the superparameters are updated according to

$$a_0' = a_0 + n$$
$$a_1' = a_1 + \sum X_i$$
$$a_2' = a_2 + \sum Z_i^{-1} X_i$$
$$a_3' = a_3 + \frac{1}{2} \sum Z_i$$
$$a_4' = a_4 + \frac{1}{2} \sum Z_i^{-1}$$

New θ can then be simulated according to the distributional structure given above. That is, (μ, β) bivariate normal and δ^2 Gamma and γ computed from δ and a simulated Normal variate.

In the case of another layer of parameters we get another set to visit in each round, determined by the posterior given at the end of the previous section.

4 The choice of superparameter values

We will now examine the choice of superparameters in order to reflect our (lack of) knowledge about parameters. It is of course convenient to have few superparameters to address as is the case with our conjugate prior. A drawback may be little flexibility, i.e. our choices affect the parameters jointly in a manner which may not be transparent. A basic restriction for the expressions of variances to stay positive is

$$a_3 \cdot a_4 > \frac{1}{4} a_0^2$$

From the five updating equations we see that more information is accumulated as a_0, a_3 and a_4 are increasing (since the z's are non-negative). The

choice of a small a_0 to reflect prior ignorance may go together with small a_i's, but with some prior knowledge, and choosing a larger a_0, this has to go along with larger a_3 and/or a_4 as well to match the restriction. The parameter ρ may then be helful for calibration purposes.

A possible consequence of few parameters is when we try to express ignorance in some sense, it may have unwanted and even contradictory implications. This is in fact the case here, where $a_0 = 0$ at first sight, is a natural choice for ignorance. This means that $\rho = 0$ and consequently

$$\begin{aligned}
\bar{\mu} &= \frac{a_2}{2a_4} \\
\bar{\beta} &= \frac{a_1}{2a_3} \\
\sigma_\mu^2 &= \frac{1}{2a_4} = E\gamma^2 \\
\sigma_\beta^2 &= \frac{1}{2a_3} = E\delta^2
\end{aligned}$$

i.e. we have implicitely assumed that the more certain you are about μ (resp. β, the smaller you expect δ (resp. γ) to be, and judged from the corresponding variances of δ^2 and γ^2 you are even more certain about that. So, $a_0 = 0$ is an ignorants choice to represent ignorance!

It is likely that opinions are initially formed by observed centers and shapes of empirical distributions. It seems therefore natural to first reflect on the parameters (μ, β) (stage 1), and then on (δ, γ) (stage 2), We may first ask to what extent our knowledge of either of these is affected by the other. In order to match the expression for the expected value $EX = \mu + \beta\delta\gamma^{-1}$, a large/small μ departing from its prior expected value has to be balanced by a small/large β compared to its expectation, i.e. μ and β have to be negatively correlated, as reflected by the conjugate prior. It is easily checked that the linear combination $\tau = \mu + t \cdot \beta$ with the smallest prior variance is given by

$$\tau = \mu + \frac{a_0}{2a_4} \cdot \beta$$

Note from section 2 that this τ is stochastically independent of β. EX is such a linear combination, and it is not unresonable to assume that we are more certain about EX than any other linear combination of of μ and β. This assumption as well as its immediate consequenses will be referred to later as A0. We see that the corresponding prior mean and variance are

given by

$$\text{A0}: \qquad \bar{\tau} = \frac{a_2}{2a_4} \qquad \sigma_\tau^2 = \frac{1}{2a_4}$$

Note that ρ now has disappeared and that the expressions are the same as for μ in the case of $\rho = 0$. If we go back to the original expression for EX and use the prior independence of (μ, β) and (δ, γ) we get the equation

$$\text{A0}: \qquad E(\frac{\delta}{\gamma}) = \frac{a_0}{2a_4} = -\rho\sqrt{\frac{a_3}{a_4}}$$

i.e. our assumption A0 has implications for the two other parameters as well. Note also that if the "ignorance" assumption $a_0 = 0$ is used in conjunction with A0, it implies that $\delta = 0$, which leads to the one-point distribution at μ, which is quite the opposite of ignorance!

Let assumption A1 be that the the prior mean equal to zero. We then get

$$\text{A0 + A1}: \qquad a_2 = 0$$

In the case of a non-zero prior mean, we could as well subtract the prior mean from all observations and start from there.

In order to determine a_1 we have to be more specific about μ or β. In the case of $a_2 = 0$ we see that

$$\bar{\mu}/\bar{\beta} = -\frac{a_0}{2a_4} \quad , \quad \bar{\mu}/\sigma_\mu^2 = -\frac{a_0}{2a_3} \cdot a_1 \quad , \quad \bar{\beta}/\sigma_\beta^2 = a_1.$$

Thus $\bar{\beta}$ determines the sign of a_1 and $\bar{\mu}$ (opposite) in this case, which holds in particular for A0 + A1.

Let us also look into assuming $\bar{\mu} = 0$ (assumption A2) and $\bar{\beta} = 0$ (assumption A3). This gives the following restrictions on the superparameters respectively:

$$\text{A2}: \qquad a_2 = \frac{a_0}{2a_3} \cdot a_1$$
$$\text{A3}: \qquad a_1 = \frac{a_0}{2a_4} \cdot a_2$$

with obvious inequalities for the prior expectations less than or greater than zero. Note that any two of A1, A2 and A3 taken together are equivalent and corresponds to $a_1 = a_2 = 0$, omitting the case of $\rho^2 = 1$ leading to infinite prior variances of μ, β and δ^2.

We may just want combine A0 and A2 to get

$$\text{A0 + A2}: \qquad E(\frac{\delta}{\gamma}) = \frac{a_2}{a_1}$$

Let us now turn to the prior of (δ, γ). It may be harder to have opinions about this than the prior of (μ, β), and it is likely that we want to express ignorance. We therefore have to choose parameters at the first stage carefully in order to give room for this. Note that the prior of (δ, γ). is determined by the parameters a_0, a_3 and a_4. Recall the binding resctriction above, where in fact the ratio of a_3 and a_4 is determined by the prior variances of μ and β. As said earlier this is an unwanted restriction, which we have to balance off according to where our knowledge is best. Note that

$$\delta^2 \sim Gamma(\frac{a_0 + 1}{2}, a_4(1 - \rho^2))$$

It will be of interest to see numerically how ad hoc assumptions representing various types of knowledge will work, when trying to balance or ignore the consequences for other parameters. Among these are ignorance assumptions taking $a_0 = 1$ or $a_0 = 2$ (assumption B1 and B2), and saying you are equally uncertain about μ and β, thus taking $a_3 = a_4$. Of particular interest is the case $a_0 = a_3 = a_4$, which means that $\rho = -1/2$.

5 Examples of estimation of NIG parameters

Reliable estimates of the parameters of a heavy tailed distiribution will require a minimum of observations in the tails, and small data sets of independent observations are not likely to give good results. Our experience so far with the NIG family based on simulated data suggests that difficulties may arise even for 100 observations and that about 400 observations are desirable, see Lillestøl (2000). We provide here two examples of NIG parameter fitting, one for simulated data and one for finance data. The first data set NIG2121 is 400 simulated independent NIG(2,1,2,1) variates. The second data set FTARET is the monthly nominal returns of the FT-Actuaries All-Share Index for the UK from January 1965 to December 1995 (372 observations). The empirical distributions are shown as histograms in Figure 1.

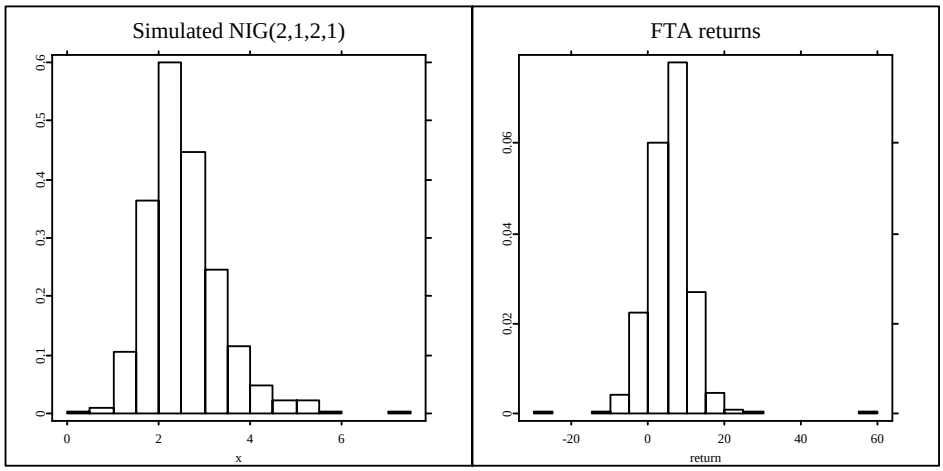

Figure 1: Histograms of data sets NIG2121 and FTARET

Descriptive measures are given in Table 1.

	Mean	StdDev	Skewness	Kurtosis
NIG2121	2.55	0.85	1.22	3.51
FTARET	5.53	6.05	1.12	14.43

Table 1: Descriptive measures

In Table 2 we give estimates by the MC-method with single layer of parameters and two layers of parameters (MC2) and compare this with the corresponding ML-estimate. The MC estimation with a single layer is based on a prior taking $a_0 = a_3 = a_4 = 1$ and $a_1 = a_2 = 0$ and iterating 10 rounds before a sample of α, β, μ and δ is taken. This is repeated 400 times and is the basis for calculating the posterior means and the smoothed histograms to represent posterior densities. The posterior means are estimated by the average of the sampled values. Estimates obtained by Rao-Blackwellization is also computed for comparison and as a check for convergence. For the two-layer estimation the values of the superparameters b_i's and c_i's are chosen to match the expected values of the a_i's for the single-layer specification above. The ML-estimates are obtained by the program 'hyp', and the results are confirmed by the EM-algorithm of Karlis mentioned above. We have also computed moment estimates (MM) obtained by inverting the expressions for the first four moments given in section 2.

	α	β	μ	δ
NIG2121 (MC)	1.360	0.868	2.205	0.407
NIG2121 (MC2)	1.669	0.965	2.142	0.579
NIG2121 (ML)	2.490	1.343	1.876	1.049
NIG2121 (MM)	2.445	1.396	1.872	0.972
FTARET (MC)	0.823	0.742	1.428	1.971
FTARET (MC2)	0.818	0.733	1.486	2.019
FTARET (ML)	0.174	-0.004	5.662	5.697
FTARET (MM)	0.083	0.015	5.001	2.872

Table 2: Parameter estimates

We see that the parameter estimates of the ML-method and the MC-method turned out somewhat different for both data sets. Let us first comment on the NIG2121 data.

The MC-method has overestimated μ while the ML-method has underestimated this parameter. The other three parameters are underestimated by the MC-method and overrstimated by the ML-method, except for δ which are about on target by the latter method. The low estimate for δ by the MC-method is somewhat disturbing. We see that adding the second parameter layer has lead to improvement for the MC-method. Now the estimate of both α and β are closer to their true values by this method than the ML-method, with β about on target. The estimate of δ is also improved somewhat, but is still at some distant from the true value. We see that the

MM-estimates are fairly close to the ML-estimates. It is also of interest to compare the first four moments by plugging in the estimated parameter values in the theoretical expressions for the moments in section 2 and compare with their true values. The main difference between the methods is that MC attributes more skewness and kurtosis than the true ones of the parent distribution, while ML attributes less skewness and kurtosis, and is closer to "target".

The smoothed histograms of sampled posterior densities for NIG2121 based on the MC2 simulations are given in Figure 2.

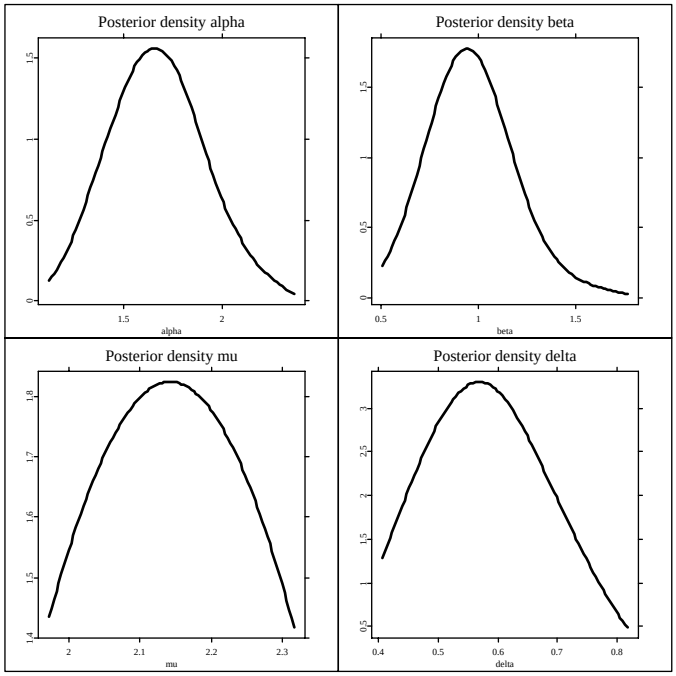

Figure 2: Smoothed histograms of sampled posterior densities for NIG2121

Looking at the posterior distributions for the simulated data we see that the true parameter values for β is at the central part of the distribution, while the true μ and α are out in the tails. For δ we have been rather unsuccessful indeed!

For the data set FTARET there are even more striking differences. We see that The ML-method has given a much higher estimate for μ and δ than the MC-method. On the other hand both α and β are higher with

the MC-method than the ML-method. We see that adding the second layer has not lead to appreciable changes. Her the MM-estimates differ more from the ML-estimates, but stay closer the the MC-estimates, except for the parameter δ.

The smoothed histograms of sampled posterior densities for FTARET based on the MC2 simulations are given in Figure 3.

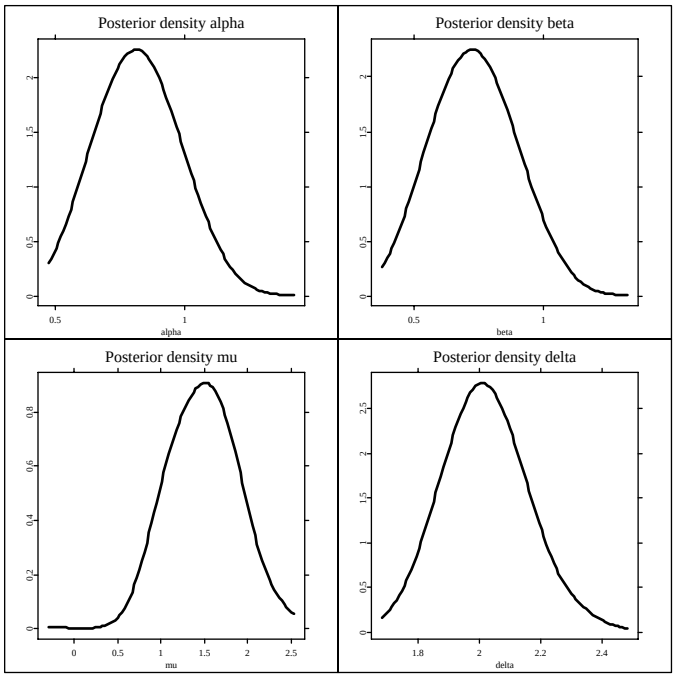

Figure 3: Smoothed histograms of sampled posterior densities for FTARET

It may be instructive to compare the fits in terms of QQ-plots. For the NIG2121 data we may compare with the true distribution. Exact computations require access to a Bessel function routine, and we may as well simulate the distribution. We have simulated n=10000 observations from the true distribution as well as from the distribution with the estimated parameters by the ML-method and MC-method respectively. The QQ-plot is given in Figure 4. We clearly see that the MC-fit is inferior to the ML-fit, and that the MC2-fit is an imrovement which makes the fit comparable to the success of the ML-fit, but that the two fits obviously have some distinct features that separates them.

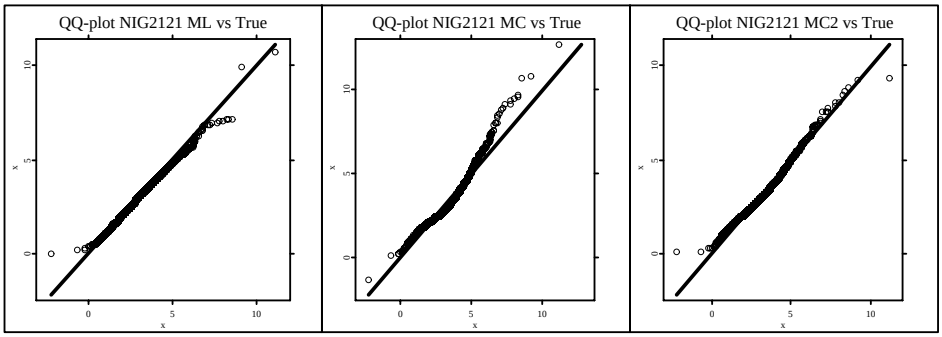

Figure 4: QQ-plot for distributions fitted by ML and MC vs True NIG2121

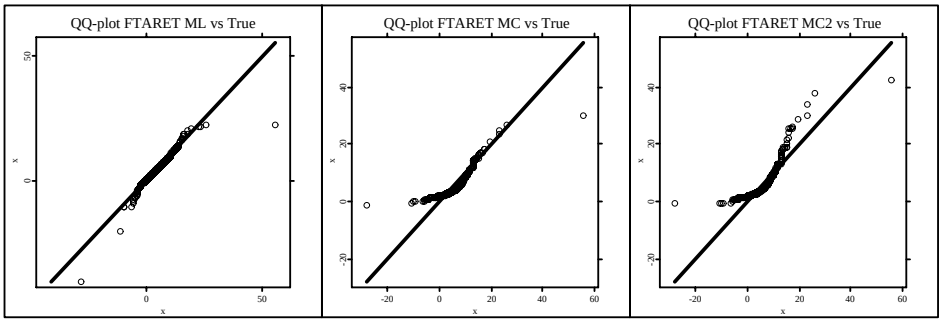

Figure 5: QQ-plot for distributions fitted by ML and MC vs Observed FTARET

For the FTARET data we compare the observed data with data simulated from the distributions fitted by the ML-method and MC-method respectively. We chose here to simulate data of the same size, namely n=372 observations. The resulting QQ-plot is given in Figure 5. It seems that the ML-fit is superior to both MC-fits.

The comparisons above are based on limited experience so far, and more suitable choices of prior specifications may hopefully lead to further improvement. Admittedly, the results obtained with the current parametrization are not impressive, and alternative parametrizations should be compared before general conclusions can be drawn.

Acknowledgements

This research is supported from Deutsch-Norwegisches Stipendprogramm-Ruhrgas Stipendium for visiting Institut für Statistik und Ökonometrie, Humboldt Universität zu Berlin. I am grateful for this support, and to Wolfgang Härdle and his staff at Sonderforchungsbereich 373: Quantifikation und Simulation Ökonomische Prozesse, for providing an excellent environment for reaserch and the computing facilitites based on their program XploRe. I also have to thank Michael Kjær Sørensen and Dimitris Karlis for providing the ML-estimates based on my data, and Dimitris Karlis for valuable personal communications.

References

Adler, R.J., Feldman R.E. and Taqqu, M.S. (eds.) (1998) *A Practical Guide to Heavy Tails: Statistical Techniques and Applications.* Boston: Birkhäuser.

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

Blæsild, P. and Sørensen, M.K. (1992) 'Hyp' - A computer program for analyzing by means of the hyperbolic distribution. Research Report No.300, Department of Theoretical Statistics, University of Aarhus.

Damien, P. and Walker, S. (1997) Sampling nonstandard distributions via the Gibbs sampler. *Preprint.*

Johnson, N.L., Kotz, S. and Balakrishnan, N. (1995) *Continuous Univariate Distributions vol.2.* New York: Wiley.

Karlis, K. (2000) An EM type algorithm for maximum lilelihood estimation of the Normal-Inverse Gaussian distribution. *Research report Athens University of Economics and Business.*

Karlis, K. (2001) Bayesian estimation for the NIG distribution. *Personal communications.*

Lillestøl, J. (2000) Risk analysis and the NIG distribution. *Journal of Risk*, 2, 41-55.

McCulloch, J.H. (1996) Financial applications of stable distributions. Statistical Methods in Finance, Handbook of Statistics, vol.14 Maddala, G.S. and Rao, C.R. (eds.) New York: North Holland.

Michael, J.R., Schuhany, W.R. and Haas, R.W. (1976) Generating random variables using transformations with multiple roots. *The American Statistician*, 30, 88-89.

Robert, C.P. and Cassela, G. (1999) *Monte Carlo Statistical Methods.* New York: Springer Verlag.

Rydberg, T.H. (1997) The normal inverse Gaussian Levy process: Simulation and approximations. *Communications in Statistics-Stochastic Models* , 13, 887-910.

Appendix: Simulation of IG*-variates

Let V be a chisquare variate with 2 degrees of freedom and compute the roots with respect to Z of

$$V = \frac{(\gamma Z - \delta)^2}{Z}$$

They are given by

$$Z = \frac{\delta}{\gamma} + \frac{1}{2\gamma^2}(V \pm \sqrt{V^2 + 4\gamma\delta V})$$

Let Z_1 and Z_2 be the minus and plus root respectively, and note that $Z_2 = \delta^2/Z_1$. Let $\mu = \delta/\gamma$ and

$$\begin{aligned} Z = Z_1 \quad & \text{with probability} \quad \frac{\mu^2}{\mu^2 + Z_1^2} \\ = Z_2 \quad & \text{with probability} \quad \frac{Z_1^2}{\mu^2 + Z_1^2} = \frac{\mu^2}{\mu^2 + Z_2^2} \end{aligned}$$

Then Z is $IG^*(\delta, \gamma)$ An alternative way if simulating IG* variates is by

rejection sampling as follows

1. Generate Z by Z^{-1} being exponential with parameter $\lambda = \frac{1}{2}\delta^2$.
2. Compute $T = exp(-\frac{1}{2}\gamma^2)$
3. Generate U Uniform[0,1], if $T > U$ keep Z otherwise not.

This procedure follows from taking $q(z \mid \theta) = z^{-2}exp(-\frac{1}{2}\delta^2 z^{-1})$ as envelope. A problem with this procedure is that many observations are likely to be rejected. Some improvement are obtained by taking z^{-1} Gamma(k, λ) instead, and adjust k to the situation at hand. However the improvement is only slight.

A simulation method that works for any generalized inverse Gaussian distribution is proposed by Damien and Walker (1997), using a Gibbs sampler with auxiliary latent variables.