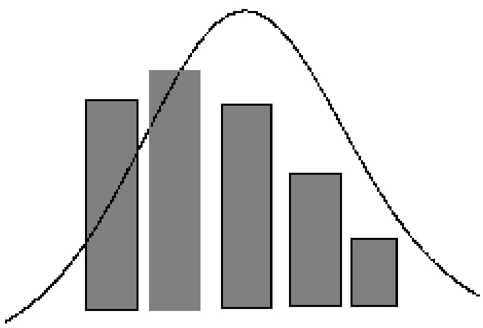
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Maximum likelihood estimation for the multivariate Normal Inverse Gaussian model

Dimitris Karlis and Argiro Papadimitriou
Department of Statistics
Athens University of Economics and Business

Technical Report No 202, September 2003



DEPARTMENT OF STATISTICS

TECHNICAL REPORT

Maximum likelihood estimation for the multivariate Normal Inverse Gaussian model

Dimitris Karlis and Argiro Papadimitriou

Department of Statistics

Athens University of Economics & Business

76, Patission Str., 10434, Athens, Greece,

September 2003

email: karlis@hermes.aueb.gr

Abstract

Multivariate Normal Inverse Gaussian model is obtained as mean-variance mixture of multivariate Normal distribution. The resulting distribution, while having heavier tails, it also accommodates skewness something not common to many multivariate distributions. In the present paper we propose Maximum likelihood estimation for the multivariate Normal Inverse Gaussian model through an EM algorithm making use of the mixture setting that generates the distribution. The algorithm is easily programmable to any statistical package. Properties of the distribution are also discussed. A financial application is also given to illustrate the proposed methodology

Keywords: mean-variance mixtures; EM algorithm; Athens Stock Exchange

1 Introduction

Most of the existing literature for multivariate models is based on the multivariate normal distribution. This implies a distribution with normal marginal distributions. To relax the normal assumption elliptically contoured multivariate distributions have been also proposed. They are generalizations of the multivariate normal distribution having elliptical contours and, thus, symmetric marginal distributions. More details for multivariate symmetric models can be found in the book of Fang *et al*. (1990). See also the interesting family of Kotz type distributions (see, e.g. Nadarajah, 2003) which is a particular case of the broad family of elliptically contoured distributions.

The class of elliptical distributions contain also scale mixtures of the multivariate normal distribution (see, e.g. Cambanis and Fotopoulos, 2000). According to the choice of the mixing distribution certain well known families arise like the multivariate t distribution or the multivariate modified Bessel distribution (Thabane and Drekic, 2003). Normal scale mixtures have heavier tails than the multivariate normal and they are always platycyrtic with respect the multivariate normal model. So, they can fit data with heavier tails and non-normal marginals.

Recently, there is an increasing interest in constructing models that allow for non-symmetric marginal distributions. Note, in particular, that the notion of symmetry in multivariate distributions can have a variety of forms (see chapter 1 of Fang *et al.* , 1990). We refer to the symmetry with respect to the marginal distributions. Azzalini and Dalla Valle (1996) introduced the multivariate skew-normal distribution (see also Azzalini and Capitanio, 1999 for applications and Genton and Liu, 2001). The marginal distributions are skewed-normal ones and thus this density is quite useful for practical applications.

In a similar fashion Gupta (2003) proposed multivariate skew t distribution, that allows heavy tails in addition to skewness. The same name has been used for a different multivariate distributions with marginals that are skewed and heavy tailed generalization of the well known t- distribution by Jones (2001). Azzalini and Capitanio (2003) described skew elliptical distribution extending the above mentioned classes (see also Fang, 2003).

We deal with a multivariate distribution that accomodates heavy tails and skewness at the same time, being a mean-variance mixture of the multivariate normal distribution. Namely we will discuss the multivariate normal-inverse gaussian (MNIG) distribution (see, Barndorff-Nielsen, 1997, Lillestol, 2001) which arises from a multivariate normal density mixed by the Inverse Gaussian distribution. We propose ML estimation of the density in the general multivariate form, through an EM type algorithm based on the mixture derivation of the density. The algorithm is easily programmable to any statistical package and thus it allows for easy application. This is quite important since many of the above mentioned model, while theoretically treated, present computational problems in order to be applied in real data problems.

The remaining of the paper proceeds as follows. In section 2 we introduce the MNIG distribution and discuss its properties. In section 3 we develop the EM algorithm, while in section 4 one can find real data applications. Finally in section 5, one can find concluding remarks.

2 The Multivariate Normal-Inverse Gaussian distribution

The derivation of the MNIG is based on a multivariate extension of mean variance normal mixtures (see Barndorff-Nielsen *et al.* , 1983). Namely, suppose that $\boldsymbol{x}$ is a random vector, which, conditional on z follows and m-variate normal distribution with mean vector $\mu + z\beta\Omega$ and covariance matrix $z\Omega$, where z is a scalar, β and μ are vectors with m elements and Ω is a variance covariance matrix. Assume further that z is a random variable follows an Inverse Gaussian distribution with parameters γ and δ and probability density function given by

$$f(z) = \frac{\delta}{\sqrt{2\pi}} exp(\delta\gamma) z^{-3/2} exp\left(-\frac{1}{2}\left(\frac{\delta^2}{z} + \gamma^2 z\right)\right). \tag{2.1}$$

denoted as $IG(\gamma, \delta)$. The mean and the variance of the $IG(\gamma, \delta)$ distribution are $E(Z) = \delta/\gamma$ and $Var(Z) = \delta/\gamma^3$ respectively. Note that different parameterizations of the Inverse Gaussian distribution has been also used (see, Seshardi, 1993).

For identifiability reasons we assume further that $\det \Omega = 1$. Alternatively one may assume that $E(Z) = 1$ but this reduces the IG distribution to a one-parameter family which is rather restrictive. Then, unconditionally, $\boldsymbol{x}$ follows an MNIG distribution.

The probability density function of the MNIG distribution of the d-dimensional column vector X has the following form

$$f_X(x) = \frac{\delta}{2^{\frac{d-1}{2}}} \left[\frac{a}{\pi \widetilde{q}(x)}\right]^{\frac{d+1}{2}} \exp\left(\widetilde{p}(x)\right) K_{\frac{d+1}{2}}\left(a\widetilde{q}(x)\right) \tag{2.2}$$

where

$$\widetilde{p}(x) = \delta\sqrt{a^2 - \beta^T \Delta \beta} + \beta^T (x - \mu) \tag{2.3}$$

and

$$\widetilde{q}(x) = \sqrt{\delta^2 + \left[(x-\mu)^T \Delta^{-1} (x-\mu)\right]} \tag{2.4}$$

One can see that β appears only on the term $exp(\widetilde{p}(x))$ and thus it adds asymmetry to the distribution. It is clear to see that if $\beta = 0$ the distribution is symmetric, it is merely a scale mixture of multivariate normal distribution. Parameter γ of the Inverse Gaussian distribution is related to the above parameters via $\gamma = \sqrt{a^2 - \beta^T \Delta \beta}$.

Changing the Inverse Gaussian distribution with its generalization, the Generalized Inverse Gaussian distribution (see, Jorgensen, 1982) one obtains the family of multivariate hyperbolic distributions (see, Blaesild and Jensen, 1985). We will discuss this family in a later section.

As it can be seen from the probability density function the MNIG distribution has two scalar parameters α,δ, which are assumed to be positive numbers, two d-dimensional vector parameters $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$ and one $d \times d$ symmetric and positive definite matrix $\boldsymbol{\Delta}$ which is not the covariance matrix.

The shape of the MNIG distribution is specified by the values of its parameters. The α parameter controls the "steepness" of the density. When α increases the steepness of the density increases monotonically too.

The parameter α affects also the tails of the density, in the sense that large values of α imply light tails, while smaller values correspond to heavier tails. The sign of β parameter controls the skewness of the distribution. Especially, for $\beta > \mathbf{0}$ the distribution is skew to the right, while for $\beta < \mathbf{0}$ is skew to the left. Parameter δ is a scale parameter and μ is a location parameter. Finally, the Δ is a semidefinite symmetric matrix with unity determinant, which controls the intercorrelations between the components of the vector $\boldsymbol{X}$. The distribution is symmetric if and only if $\beta = \mathbf{0}$ and $\Delta = I$, whereas if $\beta = \mathbf{0}$ and $\Delta \neq I$ then MNIG is semi-symmetric.

The type of $\widetilde{p}(x)$implies that the equality $a^2 > \beta^T \Delta \beta$ must be satisfied, in order the MNIG distribution to exist. Figures 2 and 2 present contour plots for selected sets of parameters. Since δ and mu are scale and location we have kept them equal to 1 and 0 respectively. For the first plot $\Delta = \begin{bmatrix} \sqrt{2} & 1 \\ 1 & \sqrt{2} \end{bmatrix}$. We have used 4 different values for β, namely $\beta = (0,0), (0,0.5), (0.5,0)$ and $(0,-0.5)$. It is clear from figure how β regulates the skewness of the density. In figure 2 the same β's have been used but now $\Delta = \boldsymbol{I}$. Recall that this does not imply uncorrelated variables. Again the effect of β is obvious. Note also that the contours are not ellipses and that they can have quite different shapes.

2.1 Properties of the MNIG distribution

At this section we present some of the most important properties of the Multivariate Normal-Inverse Gaussian distribution, starting with the mean and variance. The mean

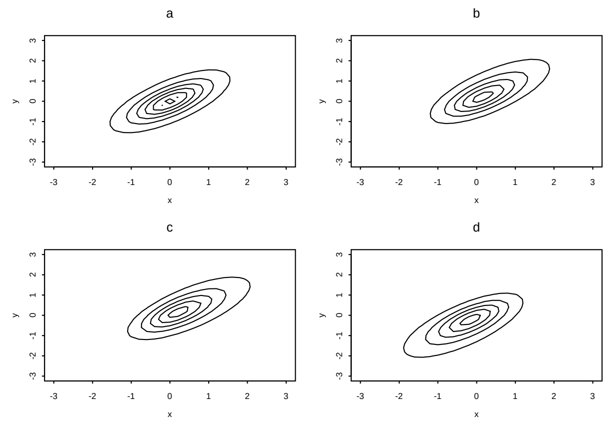


Figure 1: Contour plots for different values of β.

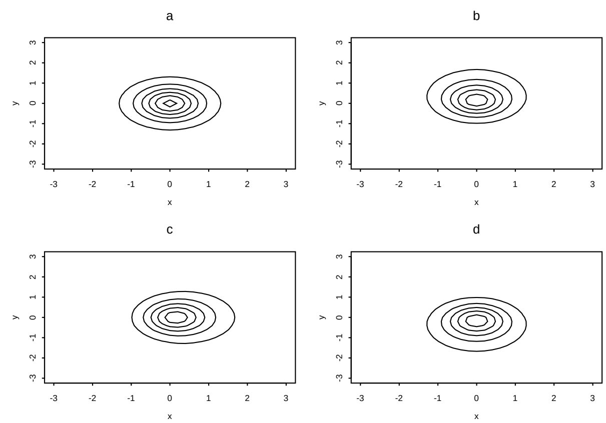


Figure 2: Contour plots for different values of β and $\Delta = \boldsymbol{I}$.

vector is

$$E(X) = \mu + \frac{\delta \Delta \beta}{\gamma}$$

while the covariance matrix of the vector X is given by

$$\Sigma = \delta \left(a^2 - \beta^T \Delta \beta\right)^{-1/2} \left[\Delta + \left(a^2 - \beta^T \Delta \beta\right)^{-1} \Delta \beta \beta^T \Delta^T\right] \tag{2.5}$$

which can be written in the better form

$$\Sigma = \frac{\delta}{\gamma^3} \left(\gamma^2 \Delta + \Delta \beta \beta^T \Delta^T\right)$$

This allows to see that the covariance is split in two parts one due to the mixing and one due to intrinsic covariance. It is interesting to see that it does not suffice the fact that Δ is diagonal in order to obtain independent variates. This is due to the fact that the share the same factor, the latent variable Z which induces correlation. Interesting special cases are when $\mu = \beta = 0$, then $\Sigma = \delta \Delta / a$.

A very attractive property of the MNIG distribution is that is closed under convolution. Particularly, if $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n$ are n independent MNIG variables with common α, β and Δ parameters and different location parameters $\mu_1, \mu_2, \ldots, \mu_n$ as well as different scale parameters $\delta_1, \delta_2, \ldots, \delta_n$ then the variable $\boldsymbol{Y} = \sum_{i=1}^{n} \boldsymbol{X}_i$ is also MNIG distributed with parameters $\alpha, \beta, \mu_{tot}, \delta_{tot}$, where $\mu_{tot} = \sum_{i=1}^{n} \mu_i$ and $\delta_{tot} = \sum_{i=1}^{n} \delta_i$.

Additionally, a linear transformation of a MNIG random variable is also a MNIG random variable. In other words, if $\boldsymbol{Y} = AX + B$ and $\boldsymbol{X} \sim MNIG(\alpha, \delta, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\Delta})$, then the transformed variable $\boldsymbol{Y} \sim MNIG(\alpha, \delta, \boldsymbol{\beta}, \boldsymbol{\mu}, \Delta)$ where

$$\begin{aligned}
\alpha' &= \alpha \left\|A\right\|^{-1/d} \\
\boldsymbol{\beta'} &= A^{-T} \beta \\
\delta' &= \delta \left\|A\right\|^{1/d} \\
\boldsymbol{\mu}' &= A\boldsymbol{\mu} + B \\
\Delta' &= A^T \Delta A \left\|A\right\|^{2/d}
\end{aligned}$$

In the above equations $\|A\|$denotes the magnitude of the determinant of the matrix $\boldsymbol{A}$.

Another basic property of the MNIG distribution is its limiting behavior. When $\delta \to \infty$ and $a \to \infty$ then $X \sim N_d\left(\mu + \sigma^2 \Delta \beta, \sigma^2 \Delta\right)$ where $\delta / \alpha = \sigma^2$, thus the Multivariate Normal distribution is a limiting distribution for the Multivariate Normal-Inverse Gaussian distribution.

An interesting point related to the MNIG distribution is that the contours are not necessarily hyperellipsoids and they are not symmetric in all cases. Thus the MNIG is quite flexible and can model a variety of different situations.

An interesting result is the following:

Lemma 1 *The conditional density of $z \mid \boldsymbol{x}$ is a GIG* $\left(-\frac{m+1}{2}, \delta\sqrt{\phi(\mathbf{x})}, \alpha\right)$ *distribution.*

Define as $\phi(\mathbf{x}) = 1 + \delta^{-1}(\mathbf{x}-\mu)\Delta^{-1}(\mathbf{x}-\mu)^T$ and $\gamma = \sqrt{\alpha^2 - \beta\Delta\beta^T}$. To see this, note that, ignoring all terms not involving z one obtains that

$$\begin{aligned}
f(z \mid x) &\propto z^{-m/2} exp\left(-\frac{1}{2z}(x-\mu-z\beta\Delta)^T\Delta^{-1}(x-\mu-z\beta\Delta)\right) z^{-3/2} exp\left(-\frac{1}{2}(\delta^2 z^{-1} + \gamma^2 z)\right) \\
&\propto z^{-m/2} exp\left\{-\frac{1}{2z}\left((x-\mu)\Delta^{-1}(x-\mu)^T - 2z\beta(x-\mu)^T + z^2\beta\Delta\beta^T\right)\right\} z^{-3/2} \times \\
&\times \; exp\left(-\frac{1}{2}(\delta^2 z^{-1} + \gamma^2 z)\right) \\
&\propto z^{-(m+3)/2} exp\left(-\frac{(x-\mu)\Delta^{-1}(x-\mu)^T}{2z} + \beta(x-\mu)^T - \frac{z\beta\Delta\beta^T}{2}\right) exp\left(-\frac{1}{2}(\delta^2 z^{-1} + \gamma^2 z)\right) \\
&\propto z^{-(m+3)/2} exp\left(-\frac{1}{2}\left\{(\delta^2 + (x-\mu)\Delta^{-1}(x-\mu)^T)z^{-1} + (\gamma^2 + \beta\Delta\beta^T)z\right\}\right)
\end{aligned}$$

and hence we obtain that the posterior is a $GIG\left(-\frac{m+1}{2}, \delta\sqrt{\phi(\mathbf{x})}, \alpha\right)$. Note that the posterior looks quite similar with the one derived for the univariate case.

This property will be quite helpful for deriving the EM algorithm in the next section.

3 Maximum Likelihood Estimation through an EM algorithm

Let us assume that we have a sample of vector observations $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$. Standard ML estimation would proceed by maximizing the loglikehood. Closed form expressions do not exist and the derivatives of the likelihood are quite cumbersome. Numerical maximization is also an option. We will provide an EM type algorithm. The algorithm makes use of the mixture derivation of the distribution.

The Expectation-Maximization algorithm (formally introduced by Dempster *et al.* , 1975) is a widely used method, which simplifies effectively the iterative computation of the maximum likelihood estimates. The algorithm is applicable in situations where the Newton-Raphson method is more complicated.

The general purpose of the algorithm is to calculate the maximum likelihood estimates in *incomplete-data* problems. All these situations do not only include the problems with obvious incompleteness, such as missing values or truncated distributions, but also problems where the incompleteness is not natural or so clear. Therefore, even if there is no incomplete-data problem -with any possible form-, it is usually helpful to express the given problem as an incomplete-data problem and work it out using the EM algorithm. In our case the mixing operation is considered as producing missing data.

The algorithm is very easy to implement and it is numerically stable. By choosing proper initials values we can achieve an appropriate convergence for the algorithm. Concerns about whether the obtained maximum is a local or a global one can be handled by starting from different initial values to see whether different solution are found. More details on the algorithm can be found in McLachlan and Krishnan (1997).

The proposed algorithm is an extension of the algorithm provided in Karlis (2002).

In our case the complete data would contain observation $\boldsymbol{y}_i = (\boldsymbol{x}_i, z_i)$, where $\boldsymbol{x}_i$ denotes the observable part (the observed data that are considered as incomplete) and z_i the unobservable part of the data for each data point, which corresponds to the mixing variate.

The EM algorithm consists of two steps· the E-step and the M-step. The E-step computes the expectation of the unobservable part given the current values of the parameters and the M-step maximizes the complete data likelihood and updates the parameters using the expectations of the E-step. If we work with a member of the exponential family, as as in our case the inverse gaussian distribution, the calculations are simplified a lot, as at the E-step we have to calculate the conditional expectation of the sufficient statistics for the Inverse Gaussian distribution. These are $\sum z_i$and $\sum z_i^{-1}$. Starting with suitable initial parameters the two steps are repeated until convergence.

As mentioned in Lemma 1, the conditional density of $z \mid x$ according to our representation of the MNIG distribution is a GIG distribution, and thus we need only some moments of this distribution. More details about the GIG distribution can be found in the Appendix.

Now we can present the EM algorithm for the MNIG model. We assume that Δ is known and equal to the scaled covariance matrix so as the determinant to be equal to 1. At the E-step we have to calculate the conditional expectations of the first order moment and the inverse first moment order. Since the conditional densities are Generalized Inverse Gaussian distributions the moments are given as

$$
\begin{aligned}
s_i &= E\left(z_i \mid x_i\right) = \frac{\widetilde{q}(x_i)}{a} \frac{K_{(d-1)/2}\left[a\widetilde{q}(x_i)\right]}{K_{(d+1)/2}\left[a\widetilde{q}(x_i)\right]} \\
\varphi_i &= E\left(z_i^{-1} \mid x_i\right) = \frac{a}{\widetilde{q}(x_i)} \frac{K_{(d+3)/2}\left[a\widetilde{q}(x_i)\right]}{K_{(d+1)/2}\left[a\widetilde{q}(x_i)\right]}
\end{aligned}
$$

for i=1,...,n. The quantities $\widetilde{q}(x_i)$ are given by the equation 4.

The parameters are updated as follows:

$$
\hat{\mu} = \frac{n^{-1}\sum_{i=1}^{n} x_i\phi_i - \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} s_i}}{n^{-1}\sum_{i=1}^{n} \phi_i - \frac{n}{\sum_{i=1}^{n} s_i}}, \quad \hat{\beta} = \frac{\sum_{i=}^{n}(x_i - \hat{\mu})}{\sum_{i=1}^{n} s_i}
$$

$$
\begin{aligned}
\hat{\Sigma} &= n^{-1}\sum_{i=1}^{n}\Big[(x_i-\hat{\mu})'(x_i-\hat{\mu})\phi_i - \hat{\beta}(x_i-\hat{\mu}) \\
&\quad -(x_i-\hat{\mu})'\hat{\beta} + \beta'\beta s_i\Big] \\
\hat{\gamma} &= \frac{n}{\sum_{i=1}^{n} s_i}
\end{aligned}
$$

and then transform back to the true parameters by

$$
\mu^{new} = \hat{\mu}, \delta^{new} = |\hat{\Sigma}|^{1/2d}, \gamma^{new} = \hat{\gamma}/\delta^{new}
$$

$$
\beta^{new} = \hat{\beta}\hat{\Sigma}^{-1} \;\text{ and }\; \Delta^{new} = |\hat{\Sigma}|^{-1/d}\hat{\Sigma}
$$

The algorithm is iterated between these two steps until a convergence criterion is been satisfied. The criterion for the algorithm termination is based on the changes of the likelihood between two iterations. Specifically, we stop iterating when the relative change in the loglikelihood is smaller than a small positive value, i.e. when $\left|\frac{L^k - L^{k+1}}{L^k}\right| < 10^{-8}$, where L^k is the log-likelihood after the k-th iteration.

4 Application to real data

MNIG models have been proposed as appropriate models for financial data. The benchmark theory of mathematical finance is the Black-Scholes-Merton theory, based on

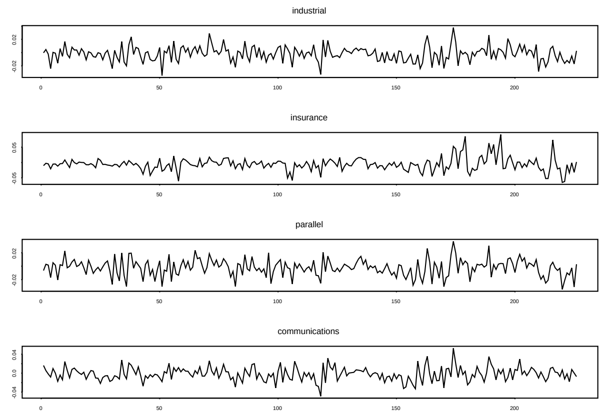


Figure 3: Tsplots for the four indices

Brownian motion as the driving noise process for asset prices. According to this model the distributions of returns of the assets in a portfolio are multivariate normal. The two most obvious limitations here concern symmetry and thin tails, neither being consistent with real data. MNIG models can have both skewness and heavy tails, thus they are quite useful as model for financial data

At this section the EM algorithm was applied to a multivariate 4-dimensional problem, which concerns the log returns of the Athens Stock Exchange for the year 2002. In total $n = 226$ observations were used.

The data were selected as follows. Let Y_t denotes the value of the index for the day t.The log-returns are obtained as

$$X_t = \ln \frac{Y_t}{Y_{t-1}}$$

For our application we used four indices, namely those related to the industrial, the insurance, the parallel market and the communications respectively.

At figure 3 one can see the series plot for all the index used. It is clear from the figure that the series are correlated. Scattermatrix presented in figure 4 shows the shapes of all pairs of variables. It is evident that they are not symmetric. Histograms also reveal heavy tails. Concluding, the MNIG distribution seems to be the appropriate model for fitting the data instead of the four dimensional Gaussian distribution.

Therefore we will use the EM-algorithm to estimate the parameters of the MNIG distribution, as the direct maximization of the log-likelihood would be an extremely painful approach.

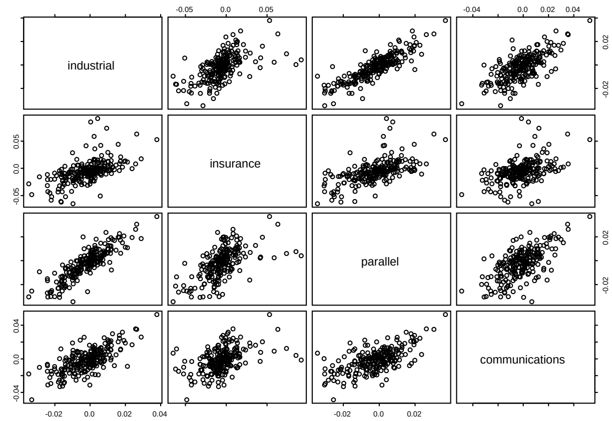


Figure 4: Matrix plot for the four variables

Starting with arbitrary, initials values $\alpha = 1, \delta = 2, \beta = (0,0,0,0)', \mu = (0,0,0,0)'$ we took the following values of the estimates of the parameters: $\hat{a} = 11.08764$, $\hat{\beta} = \begin{bmatrix} -3.041931 \\ -1.199446 \\ -5.562134 \\ 4.037525 \end{bmatrix}$, $\hat{\mu} = \begin{bmatrix} -0.0006977 \\ -0.00328029 \\ -0.00081249 \\ -0.00100034 \end{bmatrix}$ and $\hat{\delta} = 0.00731856$

while the values for the matrix Δ is

$$\Delta = \begin{bmatrix} 1.148994 & 1.293078 & 1.103333 & 1.014983 \\ 1.293078 & 4.304846 & 1.474563 & 1.276138 \\ 1.103333 & 1.474563 & 1.415006 & 1.035096 \\ 1.014983 & 1.276138 & 1.035096 & 1.817839 \end{bmatrix}$$

Standard errors can be derived by inverting the matrix with second derivatives. However since this involves a lot of Bessel functions, one can avoid it by using bootstrap standard errors. Simulation from the MNIG is straightforward based on the derivation of the model. In addition, since good initial values are available, the EM converges after a few iterations.

From these values we conclude that, as the value of the estimation of α is too high, the distribution has light tails, while the significant difference of β from 0, implies that the distribution has an obvious skewness.

The criteria that should be achieved in order to stop the iterations is that the absolute relative difference between the values of the likelihood in two consecutive iterations should be smaller than 0.00001.

5 Concluding Remarks

The above described algorithm can be extended in several ways. In the case when $\boldsymbol{\beta} = 0$, i.e. when the distribution is symmetric, such an algorithm will be an extension of the one for multivariate t distribution described in McLachlan and Krishnan (1997). The algorithm also applies to the distribution examined in Thabane and Drekic (2003).

Another interesting extension might be the modelling of multivariate hyperbolic distributions. Note that the MNIG corresponds to the case when $\lambda = -1/2$. The key idea is that if λ is known the EM algorithm is applicable, since the conditional expectations due to the Lemma are easily obtained and hence the whole algorithm applies with small modifications. P However in this case the M-step has not closed form expressions in all cases and thus the EM algorithm becomes an ECM (Expectation conditional Maximization) algorithm. Note also that the loglikelihood might be unbounded as noted in Bleasild (1981).

Furthermore, the latent structure used to derive the EM algorithm can be the basis for Bayesian treatment as well. Some more details for the univariate case can be found in Karlis and Lillestol (2003).

Finally we must point out that in fact the presented model is in fact multivariate heteroscedastic regression model and thus it can find a large number of applications in certain other disciplines apart from modelling financial data. Note that the extension to allow for more covariates is easy since the first part of the M-step is just a fit of a regression model and thus it can be adjusted to allow for more covariates other than the latent variable z.

6 References

Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, **83**, 715-726.

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society*, **B 61**, 579-602.

Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the*

Royal Statistical Society, **B 65**, 367-389.

Barndorff-Nielsen, O.E., (1997). Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling. *Scandinavian Journal of Statistics,* **24,** 1-13.

Barndorff-Nielsen, O.E., Kent, J. and Sorensen, M., (1983). Normal Variance-Mean Mixtures and z-Distributions. *International Statistical Review,* **50,** 145-159.

Blaesild, P and Jensen, J.L. (1985) Multivariate distributions of hyperbolic type. In *Statistical Distribution in Scientific Work* (eds C. Talile, G.P. Patil and B.A. Baldessari), **4**, 45-66. Reidel, Dordecht.

Blaesild, P. (1981). The Two-Dimensional Hyperbolic Distribution and Related Distributions, with an Application to Johannsen's Bean Data. *Biometrika*, **68**, pp. 251-263.

Cambanis, S., Fotopoulos, S.B. and He, L. (2000). On the conditional variance of scale mixtures of normal distributions. *Journal of Multivariate Analysis*, **74**, 163-192.

Dempster, A.P., Laird, N.M. and Rubin, D., (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society,* **B 39,** 1-38.

Fang, B.Q. (2003). The skew elliptical distributions and their quadratic forms. *Journal of Multivariate Analysis, (forthcoming).*

Fang,K. T., Kotz, S. and Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions* Monographs on Statistics and Applied Probability, Vol.36, Chapman & Hall,London,New York.

Genton, M.G. He, L. and Liu, X. (2001). Moments of skew normal random vectors and their quadratic forms. *Statistics and Probability Letters*, **51**, 319-325.

Gupta, A.K. (2003). Multivariate skew-t distribution. ıStatistics, **37**, 359-363.

Jones, M.C. (2001). Multivariate t and beta distributions associated with the multivariate F distributions. *Metrika*, **54**, 215-231.

Jorgensen, B. (1982) The Generalized Inverse-Gaussian Distribution. *Springer-Verlag.*

Karlis, D. (2002) An EM type algorithm for maximum likelihood estimation for the Normal Inverse Gaussian distribution. *Statistics and Probability Letters*, 57, 43-52

Karlis D., (2002) Models based on the Normal-Inverse Gaussian Distribution, *Proceedings of the 17th International Workshop of Statistical Modelling, Chania, pp 375-383*.

Karlis, D. and Lilestol, J. (2003). Bayesian estimation for the Normal Inverse Gaussian distribution *submitted*.

Lillestol, J., (2000). Risk Analysis and the NIG Distribution. *Journal of Risk*, **2**, 41-44.

McLachlan, G. and Krishnan, N., (1997). *The EM Algorithm and Extensions*. Chichester: John Wiley & Sons.

Nadarajah, S. (2003). The Kotz-type distribution with applications. *Statistics*, **37**, 341-358.

Øigard T. A., Hanssen A. (2001), The Normal-Inverse Gaussian distribution as a flexible model for heavy-tailed processes, *paper presented in IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing, 2001*.

Øigard T. A., Hanssen A. (2002). The Multivariate Normal-Inverse Gaussian Heavy-Tailed Distribution: Simulation and Estimation. *Paper presented in 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP) May, 2002 Orlando, Florida, USA*

Seshardi, V. (1993). *The Inverse Gaussian Distribution*. Oxford Science Publications.

Thabane, L. and Drekic. S. (2003). Hypothesis testing for the generalized multivariate modified Bessel model. *Journal of Multivariate Analysis*, **86**, 360-374.

Appendix 1. The Modified Bessel Functions

As it can be noted form the equations 1 and 2, the probability density function of both the Univariate and the Multivariate Normal-Inverse Gaussian distribution depends on one specific order of the Modified Bessel function. At this section we will present some properties of the Bessel functions that will simplify our calculations.

First of all, a basic property of the Modified Bessel function of the third kind is that $K_{-n}(x) = K_n(x)$. Moreover, if we know the order of 0 and 1, we can easily calculate the Bessel function of order 2 using the form $K_2(x) = K_0(x)+(2/x)K_1(x)$. Generally, for n¿1 the Modified Bessel function of order n+1 is given by $K_{n+1}(x) = \frac{2n}{x}K_n(x) + K_{n-1}(x)$. Therefore, one needs only to evaluate $K_0(x)$ and $K_1(x)$ and the greater orders are given by the previous retrospective form.

Additionally, closed forms of the Bessel function $K_{(d+1)/2}$ exist when the vector $\mathbf{X}$ is of even dimensionality, d=2,4,6,.... This not only avoids the numerical evaluation of the Bessel function, but also simplifies the probability density function. For d=2, 4, 6 the corresponding Bessel function is

$$K_{3/2}(x) = \sqrt{\tfrac{\pi}{2}} \cdot e^{-x} \cdot x^{-3/2} \cdot (1 + x)$$
$$K_{5/2}(x) = \sqrt{\tfrac{\pi}{2}} \cdot e^{-x} \cdot x^{-5/2} \cdot (3 + 3x + x^2)$$
$$K_{7/2}(x) = \sqrt{\tfrac{\pi}{2}} \cdot e^{-x} \cdot x^{-7/2} \cdot (15 + 15x + 6x^2 + x^3)$$

The Bessel function of order 1/2 is given by the following equation

$$K_{1/2}(x) = \sqrt{\frac{\pi}{2}} \cdot e^{-x} \cdot x^{-1/2}$$

Appendix 2. The Generalized Inverse Gaussian distribution

The probability density function of the Generalized Inverse Gaussian distribution is given by

$$f(z; \lambda, \delta, \gamma) = \left(\frac{\gamma}{\delta}\right)^{\lambda} \frac{z^{\lambda-1}}{2K_{\lambda}(\delta\gamma)} exp\left(-\frac{1}{2}\left(\frac{\delta^2}{z} + \gamma^2 z\right)\right)$$

We will denote this distribution as $GIG(\lambda, \delta, \gamma)$. It will be quite helpful for the derivation of our algorithm in the sequel. The moments around the origin of the $GIG(\lambda, \delta, \gamma)$ distribution are given by

$$E(z^r) = \left(\frac{\delta}{\gamma}\right)^{r} \frac{K_{\lambda+r}(\delta\gamma)}{K_{\lambda}(\delta\gamma)} \tag{6.6}$$

and this formula holds for negative values of r , i.e. for inverse moments, too. The Inverse Gaussian distribution is a special case of the Generalized Inverse Gaussian distribution for $\lambda = -1/2$. The gamma distribution is also a special case of the GIG distribution. More details on the GIG distribution can be found in Jorgensen (1982).