ghyp: A package on generalized hyperbolic distributions

Preliminary draft

Wolfgang Breymann
David Lüthi

Institute of Data Analysis and Process Design
www.idp.zhwin.ch

Email: david.luethi@zhwin.ch

February 27, 2007

Abstract

In this document the `R` package `ghyp` is described in detail. Basically, the density functions of the generalized hyperbolic distribution and its special cases and the fitting procedure. Some code chunks indicate how the package `ghyp` can be used.

Contents

1 Introduction

The origin of this package goes back to the first authors' years at RiskLab, when he worked together with Alexander McNeil to develop an algorithm for fitting multivariate generalized hyperbolic distributions. Accordingly, the functionality of this package largely overlaps McNeil's S-Plus library QRMlib [2]. However, there are quite some differences in the implementation. From the user's point of view, one of the most important may be that one can choose between different parametrizations. In addition, with the present library it is possible to fit multivariate as well as univariate generalized hypberbolic distributions and not only the special cases.

2 Definition

Facts about generalized hyperbolic (GH) distributions are cited according to [1] chapter 3.2. The random vector $\mathbf{X}$ is said to have a multivariate GH distribution if

$$\mathbf{X} := \mu + W\gamma + \sqrt{W} A \mathbf{Z} \tag{2.1}$$

where

(i) $\mathbf{Z} \sim N_k(\mathbf{0}, I_k)$

(ii) $A \in \mathbb{R}^{d \times k}$

(iii) $\mu, \gamma \in \mathbb{R}^d$

(iv) $W \geq 0$ is a scalar-valued random variable which is independent of $\mathbf{Z}$ and has a Generalized Inverse Gauss distribution (see appendix B).

2.1 Expected value and variance

The expected value and the variance are given by

$$\begin{aligned} \mathrm{E}(\mathbf{X}) &= \mu + \mathrm{E}(W)\gamma && (2.2) \\ \mathrm{cov}(\mathbf{X}) &= \mathrm{E}(\mathrm{cov}(\mathbf{X}|W)) + \mathrm{cov}(\mathrm{E}(\mathbf{X}|W)) && (2.3) \\ &= \mathrm{var}(W)\gamma\gamma' + \mathrm{E}(W)\Sigma \end{aligned}$$

where $\Sigma = AA'$.

2.2 Linear transformations

The GH class is closed under linear operations:
If $\mathbf{X} \sim \mathrm{GH}_d(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$ and $\mathbf{Y} = B\mathbf{X} + \mathbf{b}$, where $B \in \mathbb{R}^{k \times d}$ and $\mathbf{b} \in \mathbb{R}^k$, then $Y \sim \mathrm{GH}_k(\lambda, \chi, \psi, B\mu + \mathbf{b}, B\Sigma B', B\gamma)$.

2.3 Density

Since the conditional distribution of $\mathbf{X}$ given W is gaussian with mean $\mu + W\gamma$ and variance $W\Sigma$ the GH density can be found in the following way.

$$\begin{aligned} f_{\mathbf{X}}(x) &= \int_0^\infty f_{\mathbf{X}|W}(x|w)\, f_W(w)\, \mathrm{d}w \qquad (2.4) \\ &= \int_0^\infty \frac{e^{(\mathbf{x}-\mu)'\Sigma^{-1}\gamma}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}} w^{\frac{d}{2}}} \exp\left\{-\frac{\mathrm{Q}(\mathbf{x})}{2w} - \frac{\gamma'\Sigma^{-1}\gamma}{2/w}\right\} f_W(w) dw \\ &= \frac{(\sqrt{\psi/\chi})^\lambda (\psi + \gamma'\Sigma^{-1}\gamma)^{\frac{d}{2}-\lambda}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}} \mathrm{K}_\lambda(\sqrt{\chi\psi})} \times \frac{\mathrm{K}_{\lambda-\frac{d}{2}}(\sqrt{(\chi + \mathrm{Q}(\mathbf{x}))(\psi + \gamma'\Sigma^{-1}\gamma)}\; e^{(\mathbf{x}-\mu)'\Sigma^{-1}\gamma}}{(\sqrt{(\chi + \mathrm{Q}(\mathbf{x}))(\psi + \gamma'\Sigma^{-1}\gamma)})^{\frac{d}{2}-\lambda}} \end{aligned}$$

where $\mathrm{Q}(\mathbf{x})$ denotes the mahalanobis distance and the relation (A.2) of the modified bessel function of the third kind $\mathrm{K}_\lambda(\cdot)$ (A.1) is used. The constraints of the parameters λ, χ and ψ were described in appendix B.

3 Parametrization

There are several alternative parametrizations for the GH distribution. In this package the user can choose between two of them, the $(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$-parametrization and the $(\lambda, \overline{\alpha}, \mu, \Sigma, \gamma)$-parametrization. Have a look at appendix F.1 to see how both of these parametrizations can be used.

3.1 $(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$-Parametrization

The $\lambda, \chi, \psi, \mu, \Sigma, \gamma$-parametrization is straight forward but has a drawback of an identification problem. Indeed, the distributions $\mathrm{GH}_d(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$ and $\mathrm{GH}_d(\lambda, \chi/k, k\psi, \mu, k\Sigma, k\gamma)$ are identical for any $k > 0$. Therefore, an identifying problem occurs when we start to fit the parameters of a `ghyp` distribution. This problem can be solved by introducing a suitable contraint. One possibility is to require the determinant of the covariance matrix to be 1.

3.2 $(\lambda, \overline{\alpha}, \mu, \Sigma, \gamma)$-Parametrization

There is a more elegant way to eliminate the degree of freedom. We simply constrain the expected value of the mixing variable W to be 1. This makes the interpretation of the skewness parameters γ easier and in addition, the fitting procedure becomes faster (see 4.1).
We define [3]

$$\mathrm{E}(W) = \sqrt{\frac{\chi}{\psi}} \frac{\mathrm{K}_{\lambda+1}(\sqrt{\chi\psi})}{\mathrm{K}_{\lambda}(\sqrt{\chi\psi})} = 1. \tag{3.1}$$

and set

$$\overline{\alpha} = \sqrt{\chi\psi}. \tag{3.2}$$

It follows that

$$\psi = \overline{\alpha}\ \frac{\mathrm{K}_{\lambda+1}(\overline{\alpha})}{\mathrm{K}_{\lambda}(\overline{\alpha})} \ \text{ and } \ \chi = \frac{\overline{\alpha}^2}{\psi} = \overline{\alpha}\ \frac{\mathrm{K}_{\lambda}(\overline{\alpha})}{\mathrm{K}_{\lambda+1}(\overline{\alpha})}. \tag{3.3}$$

The drawback of the $(\lambda, \overline{\alpha}, \mu, \Sigma, \gamma)$-parametrization is that it does not exist in the case $\overline{\alpha} = 0$ and $\lambda \in [-1, 0]$. This is the case of a student-t distribution with non-existing variance. Note that the $(\lambda, \overline{\alpha}, \mu, \Sigma, \gamma)$-parametrization yields to a slightly different student-t parametrization. See section (C.1) for details.

4 Fitting generalized hyperbolic distributions to data

Numerical optimizers can be used to fit univariate GH distributions to data by means of maximum likelihood estimation. Multivariate GH distributions can be fitted with algorithms based on the expectation-maximazion (EM) scheme.

4.1 EM-Scheme

Assume we have iid data $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and parameters represented by $\Theta = (\lambda, \overline{\alpha}, \mu, \Sigma, \gamma)$. The problem is to maximize

$$\ln L(\Theta; \mathbf{x}_1, \ldots, \mathbf{x}_n) = \sum_{i=1}^{n} \ln f_{\mathbf{X}}(\mathbf{x}_i; \Theta). \tag{4.1}$$

This problem is not easy to solve due to the number of parameters and necessity of maximizing over covariance matrices. We can proceed by introducing an augmented likelihood function

$$\ln \tilde{L}(\Theta; \mathbf{x}_1, \ldots, \mathbf{x}_n, w_1, \ldots, w_n) = \sum_{i=1}^{n} \ln f_{\mathbf{X}|W}(\mathbf{x}_i | w_i; \mu, \Sigma, \gamma) + \sum_{i=1}^{n} \ln f_W(w_i; \lambda, \overline{\alpha}) \tag{4.2}$$

and spend the effort on the estimation of the latent mixing variables w_i coming from the mixture representation of (2.1). This is where the EM algorithm comes into play.

E-step: Calculate the conditional expectation of the likelihood function (4.2) given the data $\mathbf{x}_1, \ldots, \mathbf{x}_n$ and the current estimates of parameters $\Theta^{[k]}$. This results in the objective function

$$Q(\Theta; \Theta^{[k]}) = \mathrm{E}(\ln \tilde{L}(\Theta; \mathbf{x}_1, \ldots, \mathbf{x}_n, w_1, \ldots, w_n) | \mathbf{x}_1, \ldots, \mathbf{x}_n; \Theta^{[k]}). \tag{4.3}$$

M-step: Maximize the objective function with respect to Θ to obtain the next set of estimates $\Theta^{[k+1]}$.

Alternating between these steps yields to the maximum likelihood estimation of the parameter set Θ.

In practice, performing the E-Step means maximizing the second summand of (4.2) numerically. The log density of the GIG distribution (see B.1) is

$$\ln f_W(w) = \frac{\lambda}{2} \ln(\psi/\chi) - \ln(2\mathrm{K}_\lambda \sqrt{\chi\psi}) + (\lambda - 1) \ln w - \frac{\chi}{2} \frac{1}{w} - \frac{\psi}{2} w. \tag{4.4}$$

When using the $(\lambda, \overline{\alpha})$-parametrization this problem is of dimension two instead of three and consequently increases the performance.
Since the w_i's are latent one has to replace w, $1/w$ and $\ln w$ with expected values in order to maximize the log likelihood function. Let

$$\eta_i^{[k]} := \mathrm{E}(w_i \,|\, \mathbf{x}_i; \Theta^{[k]}), \;\; \delta_i^{[k]} := \mathrm{E}(w_i^{-1} \,|\, \mathbf{x}_i; \Theta^{[k]}), \;\; \xi_i^{[k]} := \mathrm{E}(\ln w_i \,|\, \mathbf{x}_i; \Theta^{[k]}). \tag{4.5}$$

We have to find the conditional density of w_i given $\mathbf{x}_i$ to be able to calculate these quantities (see (D.1)).

4.2 MCECM estimation

In the `R` implementation we employ a modified EM scheme which is called multi-cycle, expectation, conditional estimation (MCECM) algorithm ([1], [2]). The different steps of the MCECM algorithm are sketched as follows:

(1) Select reasonable starting values for $\Theta^{[k]}$. For example $\lambda = 1$, $\overline{\alpha} = 1$, μ is set to the sample mean, Σ to the sample covariance matrix and γ to a zero skewness vector.

(2) Calculate $\chi^{[k]}$ and $\psi^{[k]}$ as a function of $\overline{\alpha}^{[k]}$ using (3.3).

(3) Use (4.5), (B.2) and (D.1) to calculate the weights $\eta_i^{[k]}$ and $\delta_i^{[k]}$. Average the weights to get

$$\bar{\eta}^{[k]} = \frac{1}{n}\sum_{i=1}^{n} \eta_i^{[k]} \text{ and } \bar{\delta}^{[k]} = \frac{1}{n}\sum_{i=1}^{n} \delta_i^{[k]}. \tag{4.6}$$

(4) If an asymmetric model is to be fitted set γ to $\mathbf{0}$, else set

$$\gamma^{[k+1]} = \frac{1}{n}\frac{\sum_{i=1}^{n} \delta_i^{[k]}(\bar{\mathbf{x}} - \mathbf{x}_i)}{\bar{\eta}^{[k]}\bar{\delta}^{[k]} - 1}. \tag{4.7}$$

(5) Update μ and Σ:

$$\mu^{[k+1]} = \frac{1}{n}\frac{\sum_{i=1}^{n} \delta_i^{[k]}(\mathbf{x}_i - \gamma^{[k+1]})}{\bar{\delta}^{[k]}} \tag{4.8}$$

$$\Sigma^{[k+1]} = \frac{1}{n}\sum_{i=1}^{n} \delta_i^{[k]}(\mathbf{x}_i - \mu^{[k+1]})(\mathbf{x}_i - \mu^{[k+1]})' - \bar{\eta}^{[k]}\gamma^{[k+1]}\gamma^{[k+1]\,\prime}. \tag{4.9}$$

(6) Set $\Theta^{[k,2]} = (\lambda^{[k]}, \overline{\alpha}^{[k]}, \mu^{[k+1]}, \Sigma^{[k+1]}, \gamma^{[k+1]})$ and calculate weights $\eta_i^{[k,2]}$, $\delta_i^{[k,2]}$ and $\xi_i^{[k,2]}$ using (4.5), (B.3) and (B.2).

(7) Maximise the second summand of (4.2) with respect to λ, χ and ψ to complete the calculation of $\Theta^{[k,2]}$ and go back to step (2). Note that the objective function must calculate χ and ψ in dependence of λ and $\overline{\alpha}$ using the relation (3.3).

5 Special cases of the generalized hyperbolic distribution

The GH distribution contains several special cases known under special names [1].

- If $\lambda = \frac{d+1}{2}$ the name generalized is dropped and we have a multivariate hyperbolic distribution. The univariate margins are still GH distributed. Inversely, when $\lambda = 1$ we get a multivariate GH distribution with hyperbolic margins.
- If $\lambda = -\frac{1}{2}$ the distribution is called Normal Inverse Gauss (NIG).
- If $\overline{\alpha} = 0$ and $\lambda > 0$ one gets a limiting case which is known amongst others as Variance Gamma (VG) distribution.

- If $\overline{\alpha} = 0$ and $\lambda < -2$ one gets a limiting case which is known as a skewed student-t distribution.

All the necessary formulas to fit the special cases can be found in the appendix.

A Modified Bessel function of the third kind

The modified bessel function of the third kind appears in the GIG density (B.1). This function is defined as

$$K_\lambda(x) := \frac{1}{2}\int_0^\infty w^{\lambda-1}\exp\left\{-\frac{1}{2}x\left(w+w^{-1}\right)\right\}\mathrm{d}w\,, \qquad x>0. \tag{A.1}$$

By means of the following relation the GH density (2.1) can be written in the closed form.

$$\int_0^\infty w^{\lambda-1}\exp\left\{-\frac{1}{2}\left(\frac{\chi}{w}+w\psi\right)\right\}\mathrm{d}w = 2\left(\frac{\chi}{\psi}\right)^{\frac{\lambda}{2}} K_\lambda(\sqrt{\chi\psi}) \tag{A.2}$$

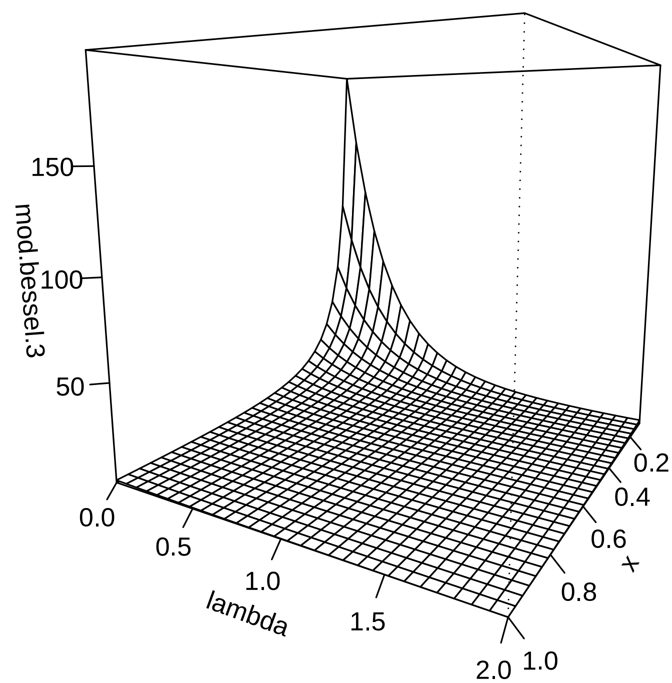
When calculating the densities of the special cases of the GH density we can use the asymtotic relations

$$\mathrm{K}_\lambda(x) \sim \Gamma(\lambda)\,2^{\lambda-1}x^{-\lambda} \quad \text{as} \quad x\to 0+ \quad \text{and} \quad \lambda>0 \tag{A.3}$$

and

$$\mathrm{K}_\lambda(x) \sim \Gamma(-\lambda)\,2^{-\lambda-1}x^{\lambda} \quad \text{as} \quad x\to 0+ \quad \text{and} \quad \lambda<0. \tag{A.4}$$

(A.4) follows from (A.3) and the observation that the Bessel function is symmetric with



respect to λ.

B Generalized Inverse Gaussian distribution

The density of a Generalized Inverse Gaussian (GIG) distribution is given as

$$f_{GIG}(w) = \left(\frac{\psi}{\chi}\right)^{\frac{\lambda}{2}} \frac{w^{\lambda-1}}{2K_\lambda(\sqrt{\chi\psi})} \exp\left\{-\frac{1}{2}\left(\frac{\chi}{w} + \psi w\right)\right\}, \tag{B.1}$$

with parameters satisfying

$\chi > 0, \psi \geq 0, \quad \lambda < 0$
$\chi > 0, \psi > 0, \quad \lambda = 0$
$\chi \geq 0, \psi > 0, \quad \lambda > 0$.

The GIG density contains the Gamma (Γ) and Inverse Gamma ($\mathrm{I}\Gamma$) densities as limiting cases. If $\chi = 0$ and $\lambda > 0$ then X is gamma distributed with parameters λ and $\frac{1}{2}\psi$ ($\Gamma(\lambda, \frac{1}{2}\psi)$).
If $\psi = 0$ and $\lambda < 0$ then X has an inverse gamma distribution with parameters $-\lambda$ and $\frac{1}{2}\chi$ ($\mathrm{I}\Gamma(-\lambda, \frac{1}{2}\chi)$).

The $n-th$ moment of a GIG distributed random variable can be found to be

$$\mathrm{E}(X^n) = \left(\frac{\chi}{\psi}\right)^{\frac{n}{2}} \frac{\mathrm{K}_{\lambda+n}(\sqrt{\chi\psi})}{\mathrm{K}_{\lambda}(\sqrt{\chi\psi})}. \tag{B.2}$$

Furthermore

$$\mathrm{E}(\ln X) = \frac{\mathrm{dE}(X^\alpha)}{\mathrm{d}\alpha}\bigg|_{\alpha=0}. \tag{B.3}$$

Numerical calculations may be performed with the integral representation as well. In the `R` package `ghyp` the derivative is implemented.

B.1 Gamma distribution

When $\chi = 0$ and $\lambda > 0$ the GIG distribution reduces to the gamma distribution defined as

$$f_W(w) = \frac{\beta^\alpha}{\Gamma(\alpha)} w^{\alpha-1} \exp\{-\beta w\}.$$

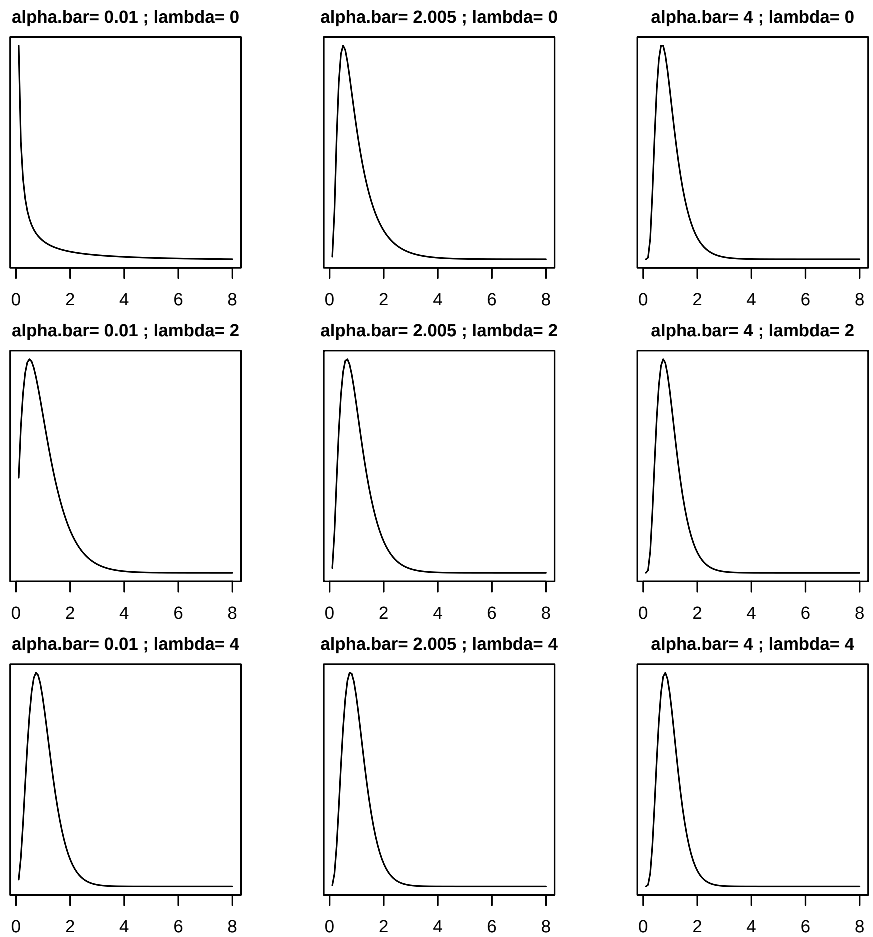
The expected value and the variance are $\mathrm{E}(X) = \beta/\alpha$ and $\mathrm{var}(X) = \alpha/\beta^2$, respectively. The expected value of the logarithm is $\mathrm{E}(\ln X) = \psi(\alpha) - \ln(\beta)$ where $\psi(\cdot)$ is the digamma function. We will see that this value is not needed to fit a multivariate variance gamma distribution (see D.3).

B.2 Inverse gamma distribution

When $\psi = 0$ and $\lambda < 0$ the GIG distribution reduces to the gamma distribution defined as

$$f_W(w) = \frac{\beta^\alpha}{\Gamma(\alpha)} w^{-\alpha-1} \exp\left\{-\frac{\beta}{w}\right\}.$$

The expected value and the variance are $\mathrm{E}(X) = \beta/(\alpha-1)$ and $\mathrm{var}(X) = \beta^2/((\alpha-1)^2(\alpha-2))$, and exist provided that $\alpha > 1$ and $\alpha > 2$ respectively. The expected value of the logarithm is $\mathrm{E}(\ln X) = \ln(\beta) - \psi(\alpha)$. This value is required in order to fit a symmetric multivariate student-t distribution by means of the MCECM algorithm (see D.2).



C Densities of the special cases of the generalized hyperbolic distribution

As mentioned in section 5 the GH distribution contains several special cases. In what follows the densities of the special cases are derived. In the case of a hyperbolic or normal inverse gaussian distribution we simply fix the parameter λ either to $(d+1)/2$ or -0.5.

C.1 Student-t distribution

With relation (A.4) it can be easily shown that when $\psi \to 0$ and $\lambda < 0$ the density results in

$$f_t(\mathbf{x}) = \frac{\chi^{-\lambda}(\gamma'\Sigma^{-1}\gamma)^{\frac{d}{2}-\lambda}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}\Gamma(-\lambda)2^{-\lambda-1}} \times \frac{\mathrm{K}_{\lambda-\frac{d}{2}}(\sqrt{(\chi+\mathrm{Q}(\mathbf{x}))\gamma'\Sigma^{-1}\gamma}e^{(\mathbf{x}-\mu)'\Sigma^{-1}\gamma}}{(\sqrt{(\chi+\mathrm{Q}(\mathbf{x}))\gamma'\Sigma^{-1}\gamma})^{\frac{d}{2}-\lambda}}. \tag{C.1}$$

We use the common student-t parametrization and set the degree of freedom $\nu = -2\lambda$ [1]. Because $\psi = 0$ the transformation of $\overline{\alpha}$ to χ and ψ (see 3.3) reduces to

$$\chi = \overline{\alpha}\ \frac{\mathrm{K}_\lambda(\overline{\alpha})}{\mathrm{K}_{\lambda+1}(\overline{\alpha})} \overset{\overline{\alpha}\to 0}{\longrightarrow} 2\ (-\lambda-1) = \nu - 2. \tag{C.2}$$

Putting it all together the density is calculated to be

$$f_t(\mathbf{x}) = \frac{(\nu-2)^{\frac{\nu}{2}}(\gamma'\Sigma^{-1}\gamma)^{\frac{\nu+d}{2}}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}\Gamma(\frac{\nu}{2})2^{\frac{\nu}{2}-1}} \times \frac{\mathrm{K}_{\frac{\nu+d}{2}}(\sqrt{(\nu-2+\mathrm{Q}(\mathbf{x}))\gamma'\Sigma^{-1}\gamma})\ e^{(\mathbf{x}-\mu)'\Sigma^{-1}\gamma}}{(\sqrt{(\nu-2+\mathrm{Q}(\mathbf{x}))\gamma'\Sigma^{-1}\gamma})^{\frac{\nu+d}{2}}}. \tag{C.3}$$

When $\gamma \to 0$ we observe the symmetric multivariate t distribution

$$f_t(\mathbf{x})) = \frac{(\nu-2)^{\frac{\nu}{2}}\Gamma(\frac{\nu+d}{2})}{\pi^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}\Gamma(\frac{\nu}{2})(\nu-2+\mathrm{Q}(\mathbf{x}))^{\frac{\nu+d}{2}}}. \tag{C.4}$$

C.2 Variance gamma distribution

The relation (A.4) can be used again to show that when $\psi \to 0$ and $\lambda > 0$ the density of the GH distribution results in

$$f_t(\mathbf{x}) = \frac{\psi^\lambda(\psi+\gamma'\Sigma^{-1}\gamma)^{\frac{d}{2}-\lambda}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}\Gamma(\lambda)2^{\lambda-1}} \times \frac{\mathrm{K}_{\lambda-\frac{d}{2}}(\sqrt{\mathrm{Q}(\mathbf{x})(\psi+\gamma'\Sigma^{-1}\gamma)})\ e^{(\mathbf{x}-\mu)'\Sigma^{-1}\gamma}}{(\sqrt{\mathrm{Q}(\mathbf{x})(\psi+\gamma'\Sigma^{-1}\gamma)})^{\frac{d}{2}-\lambda}}. \tag{C.5}$$

In the case of a variance gamma distribution the transformation of $\overline{\alpha}$ to χ and ψ (see 3.3) reduces to

$$\psi = \overline{\alpha}\ \frac{\mathrm{K}_{\lambda+1}(\overline{\alpha})}{\mathrm{K}_\lambda(\overline{\alpha})} = 2\lambda \tag{C.6}$$

[1]Note that the $(\lambda, \overline{\alpha}, \mu, \Sigma, \gamma)$ parametrization yields to a slightly different student-t parametrization: In this package the parameter Σ denotes the variance in the multivariate case and the standard deviation in the univariate case. Thus, set $\sigma = \sqrt{\nu/(\nu-2)}$ in the univariate case to get the same results as with the standard R implementation of the student-t distribution.

Thus, the density is

$$f_t(\mathbf{x}) = \frac{2\,\lambda^{\lambda}(2\lambda+\gamma'\Sigma^{-1}\gamma)^{\frac{d}{2}-\lambda}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}\Gamma(\lambda)} \times \frac{\mathrm{K}_{\lambda-\frac{d}{2}}(\sqrt{\mathrm{Q}(\mathbf{x})(2\lambda+\gamma'\Sigma^{-1}\gamma)})\; e^{(\mathbf{x}-\mu)'\Sigma^{-1}\gamma}}{(\sqrt{\mathrm{Q}(\mathbf{x})(2\lambda+\gamma'\Sigma^{-1}\gamma)})^{\frac{d}{2}-\lambda}}. \tag{C.7}$$

D Conditional density of the mixing variable W

Performing the E-Step of the MCECM algorithm requires the calculation of the conditional expectation of w_i given $\mathbf{x}_i$. In this section the conditional density is derived.

D.1 Generalized hyperbolic, hyperbolic and NIG distribution

The mixing term w is GIG distributed. By using (2.4) and (B.1) the density of w_i given $\mathbf{x}_i$ can be calculated to be again the GIG density with parameters $(\lambda-\frac{d}{2}, \mathrm{Q}(\mathbf{x})+\chi, \psi+\gamma'\Sigma^{-1}\gamma)$.

$$\begin{aligned} f_{w|\mathbf{x}}(w) &= \frac{f_{\mathbf{X},W}(\mathbf{x},w)}{f_{\mathbf{X}}(\mathbf{x})} \\ &= \frac{f_{\mathbf{X}|W}(\mathbf{x})f_{GIG}(w)}{\int_0^\infty f_{\mathbf{X}|W}(\mathbf{x})f_{GIG}(w)\mathrm{d}w} \\ &= \left(\frac{\gamma'\Sigma^{-1}\gamma+\psi}{\mathrm{Q}(\mathbf{x})+\chi}\right)^{0.5(\lambda-\frac{d}{2})} \times \\ &\quad \frac{w^{\lambda-\frac{d}{2}-1}\exp\left\{-\frac{1}{2}\left(\frac{\mathrm{Q}(\mathbf{x})+\chi}{w}+w\,(\gamma'\Sigma^{-1}\gamma+\psi)\right)\right\}}{2\,K_{\lambda-\frac{d}{2}}(\sqrt{(\mathrm{Q}(\mathbf{x})+\chi)\,(\gamma'\Sigma^{-1}\gamma+\psi)})} \end{aligned} \tag{D.1}$$

D.2 Student-t distribution

The mixing term w is IΓ distributed. Again the conditional density of w_i given $\mathbf{x}_i$ results in the GIG density. The equations (2.4) and (B.4) were used. The parameters of the GIG density are $(\lambda-\frac{d}{2}, \mathrm{Q}(\mathbf{x})+\chi, \gamma'\Sigma^{-1}\gamma)$. When γ becomes 0 the conditional density reduces to the IΓ density with parameters $(\frac{d}{2} - \lambda, \frac{\mathrm{Q}(\mathbf{x})+\chi}{2})$.

$$\begin{aligned}
f_{w|\mathbf{x}}(w) &= \frac{f_{\mathbf{X},W}(\mathbf{x},w)}{f_{\mathbf{X}}(\mathbf{x})} \\
&= \frac{f_{\mathbf{X}|W}(\mathbf{x}) f_{\mathrm{I}\Gamma}(w)}{\int_0^\infty f_{\mathbf{X}|W}(\mathbf{x}) f_{\mathrm{I}\Gamma}(w) \mathrm{d}w} \\
&= \left(\frac{\gamma'\Sigma^{-1}\gamma}{\mathrm{Q}(\mathbf{x})+\chi}\right)^{0.5(\lambda-\frac{d}{2})} \times \frac{w^{\lambda-\frac{d}{2}-1} \exp\left\{-\frac{1}{2}\left(\frac{\mathrm{Q}(\mathbf{x})+\chi}{w} + w\,\gamma'\Sigma^{-1}\gamma\right)\right\}}{2\,K_{\lambda-\frac{d}{2}}(\sqrt{(\mathrm{Q}(\mathbf{x})+\chi)\,\gamma'\Sigma^{-1}\gamma})} \qquad \text{(D.2)}
\end{aligned}$$

D.3 Variance gamma distribution

The mixing term w is Γ distributed. By using (2.4) and (B.4) the density of w_i given $\mathbf{x}_i$ can be calculated to be again the GIG density with parameters $(\lambda - \frac{d}{2}, \mathrm{Q}(\mathbf{x}), \psi + \gamma'\Sigma^{-1}\gamma)$.

$$\begin{aligned}
f_{w|\mathbf{x}}(w) &= \frac{f_{\mathbf{X},W}(\mathbf{x},w)}{f_{\mathbf{X}}(\mathbf{x})} \\
&= \frac{f_{\mathbf{X}|W}(\mathbf{x}) f_{\Gamma}(w)}{\int_0^\infty f_{\mathbf{X}|W}(\mathbf{x}) f_{\Gamma}(w) \mathrm{d}w} \\
&= \left(\frac{\gamma'\Sigma^{-1}\gamma+\psi}{\mathrm{Q}(\mathbf{x})}\right)^{0.5(\lambda-\frac{d}{2})} \times \qquad \text{(D.3)} \\
& \quad \frac{w^{\lambda-\frac{d}{2}-1} \exp\left\{-\frac{1}{2}\left(\frac{\mathrm{Q}(\mathbf{x})}{w} + w\,(\gamma'\Sigma^{-1}\gamma+\psi)\right)\right\}}{2\,K_{\lambda-\frac{d}{2}}(\sqrt{\mathrm{Q}(\mathbf{x})\,(\gamma'\Sigma^{-1}\gamma+\psi)})} \qquad \text{(D.4)}
\end{aligned}$$

E Distribution objects

In the package `ghyp` we follow an object oriented programming approach and introduce distribution objects. There are mainly four reasons for that:

1. Unlike most distributions the GH distribution has quite a few parameters which have to fulfill some consistency requirements. Consistency checks can be performed uniquely when an object is initialized.

2. Once initialized the common functions belonging to a distribution can be called conveniently by passing the distribution object. A repeated input of the parameters is avoided.

3. Distributions returned from fitting procedures can be directly passed to, e.g., the density function since fitted distribution objects add information to the distribution object and consequently inherit from the class of the distribution object.

4. Generic method dispatching can be used to provide a uniform interface to, e.g., simulate random variates of a specific distribution like `rand(n, distribution.object)`. Additionally, one can take advantage of generic programming since R provides virtual classes and some forms of polymorphism.

See appendix F for several examples and F.2 for particular examples concerning the object oriented approach.

F Examples

This section provides examples of distribution objects and the object oriented approach as well as fitting to data and portfolio optimization.

F.1 Initializing distribution object

This example shows how GH distribution objects can be initialized by either using the $(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$ or the $(\lambda, \overline{\alpha}, \mu, \Sigma, \gamma)$-parametrization.

```
> library(ghyp)
> ghyp(lambda = -2, alpha.bar = 0.5, mu = 10, sigma = 5, gamma = 1)

Object of class 'ghypuv' (Univariate Generalized Hyperbolic)

Model:
Asymmetric Generalized Hyperbolic

Parameters:
    lambda  alpha.bar         mu      sigma      gamma
-2.0000000  0.5000000 10.0000000  5.0000000  1.0000000


Slot 'data' is NULL.
```

```
> ghyp(lambda = -2, chi = 5, psi = 0.1, mu = 10:11, sigma = diag(5:6),
+     gamma = -1:0)

Object of class 'ghypmv' (Multivariate Generalized Hyperbolic)

Model:
Asymmetric Generalized Hyperbolic

Mixing parameters (lambda / alpha.bar):
   lambda  alpha.bar
-2.000000   0.707107

mu:
[1] 10 11

sigma:
     [,1] [,2]
[1,]    5    0
[2,]    0    6

gamma:
[1] -1  0


Slot 'data' is NULL.
```

F.2 Object oriented approach

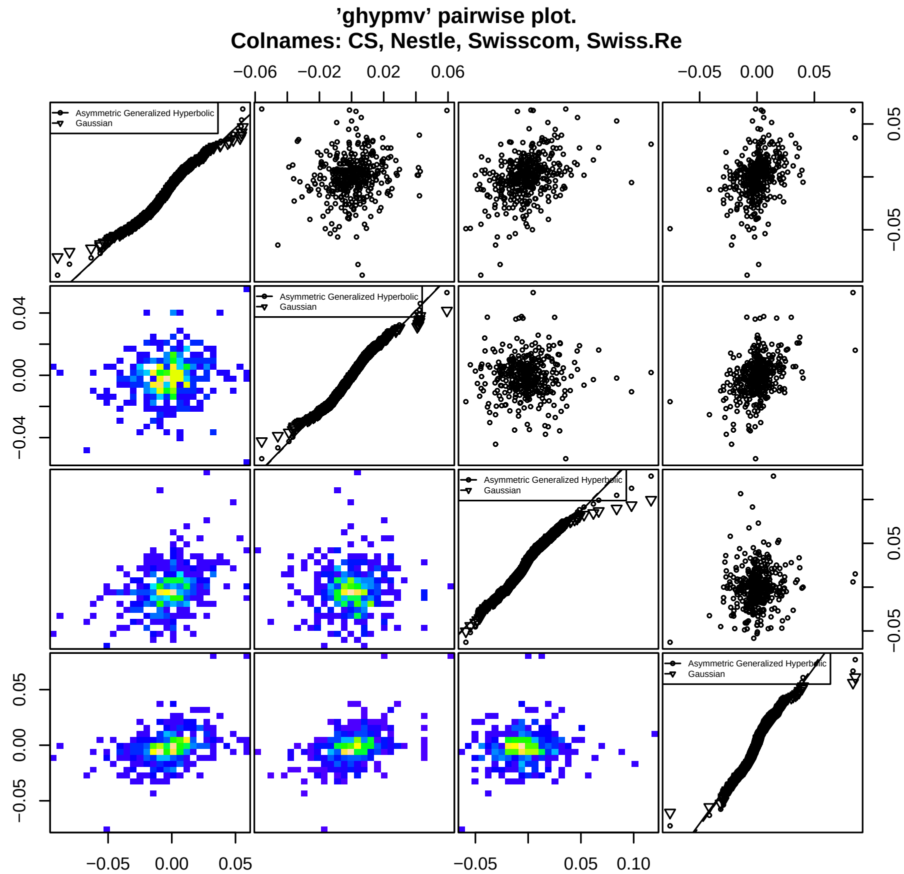
First of all a GH distribution object is initialized and a consistency check takes place. The second command shows how the initialized distribution object is passed to the density function. Then a student-t distribution is fitted to the daily log-returns of the company Novartis. The fitted distribution object is passed to the quantile function. Since the fitted distribution object inherits from the distribution object this constitutes no problem. The generic methods *hist*, *mean* and *vcov* are defined for distribution objects inheriting from classes ”ghypuv” and ”ghypbase”, respectively.

```
> data(smi.stocks)
> univariate.ghyp.object <- ghyp(lambda = -2, alpha.bar = 0.5,
+     mu = 10, sigma = 5, gamma = 1)
> dghyp(10:14, univariate.ghyp.object)

[1] 0.09967129 0.09923090 0.09078867 0.07723714 0.06207470

> fitted.ghyp.object <- fit.tuv(smi.stocks[, "Novartis"], silent = T)
> qghyp(c(0.01, 0.05), fitted.ghyp.object)

[1] -0.03622398 -0.01980706

> hist(fitted.ghyp.object, legend.cex = 0.7)
> mean(fitted.ghyp.object)

[1] 0.0001696209

> vcov(univariate.ghyp.object)

[1] 26.54375
```

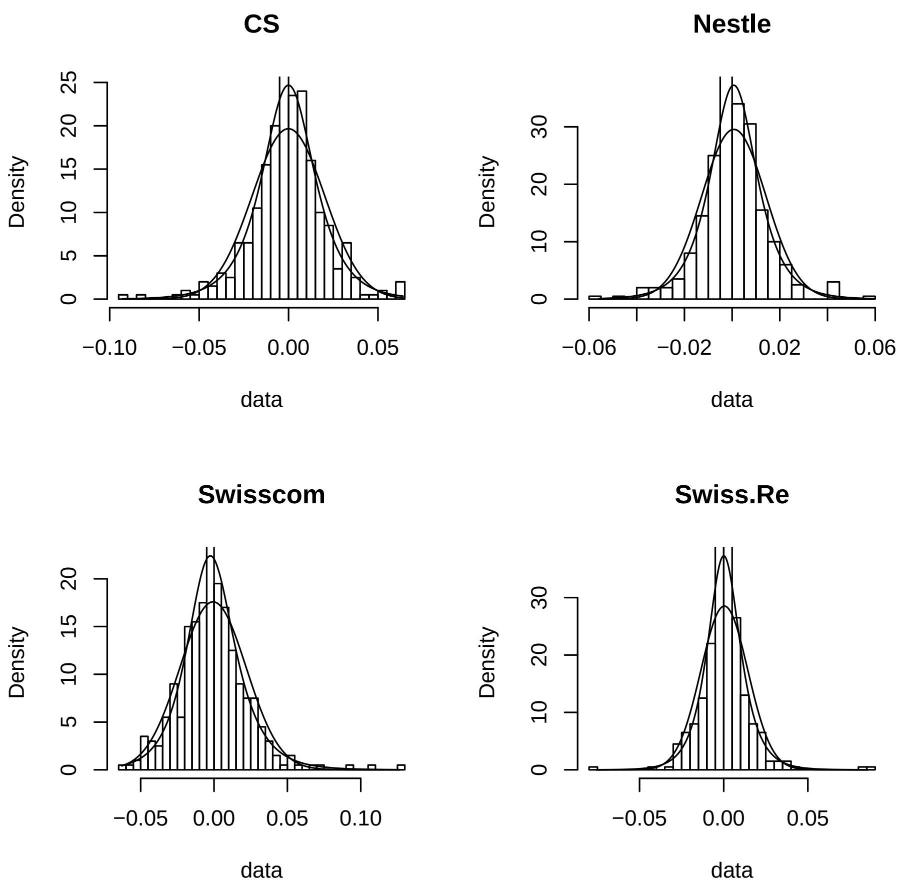
F.3 Fitting generalized hyperbolic distributions to data

A multivariate GH distribution is fitted to the daily returns of four swiss blue chips Credit Suisse, Nestle, Swisscom and Swiss Re. A *pairs* plot and four histogramms are plotted in order to do some graphical diagnostics of the accuracy of the fit.

```
> fitted.stocks <- fit.ghypmv(data = smi.stocks[1:400, c("CS",
+     "Nestle", "Swisscom", "Swiss.Re")], silent = TRUE)
> pairs(fitted.stocks, cex = 0.5, legend.cex = 0.5)
```




```
> par(mfrow = c(2, 2))
> for (i in 1:4) {
+     hist(redim(fitted.stocks, i), legend = FALSE,
+          main = colnames(ghyp.data(fitted.stocks))[i])
+ }
```

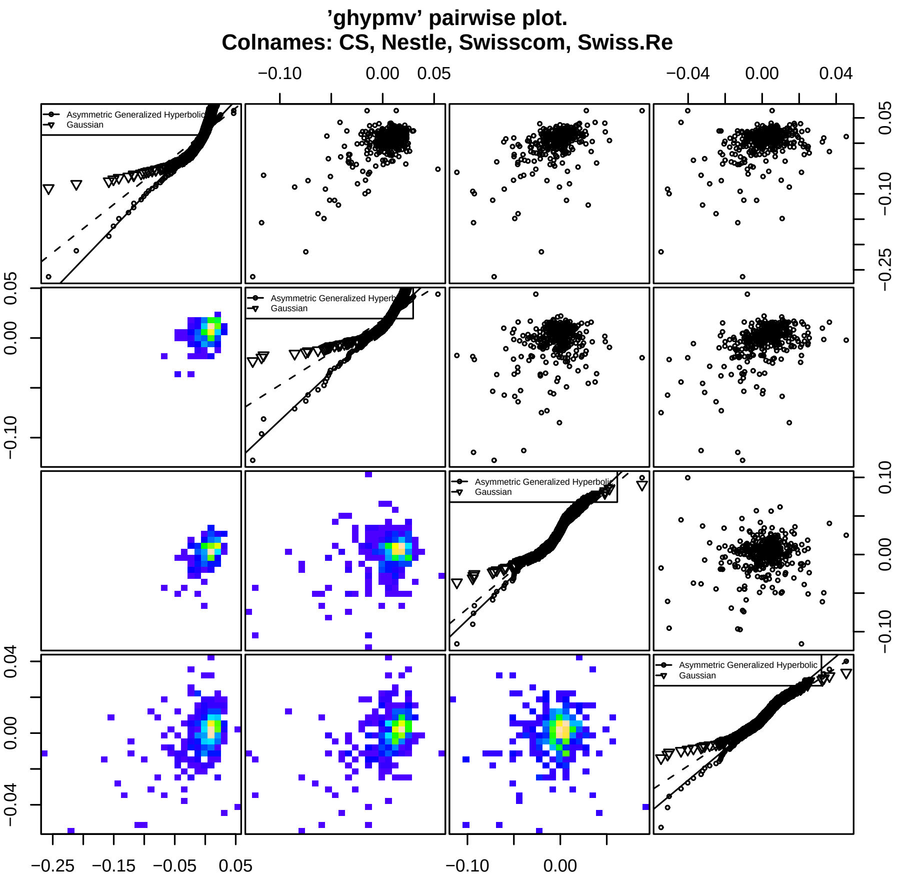


F.4 Portfolio optimization

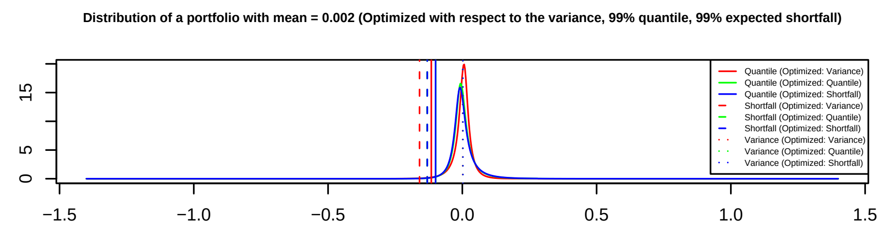
In the following a portfolio is created and optimized with respect to the variance, quantile and expected-shortfall. In the case of a normal distribution the result is exactly the same and does not depend on the method. When the returns are skewed or leptocurtic this procedure gives quite different results for each of the optimization criterion. Hence we build a synthetic skewed and leptocurtic multivariate GH distribution and see what happens. The *pairs* plot emphasizes the non-normality of the return distribution.

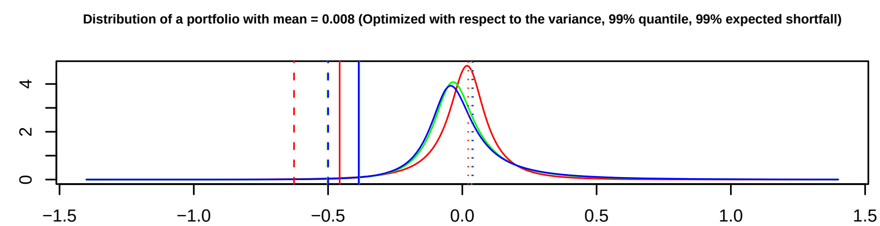
```
> smi.stock.params <- ghyp.params(fitted.stocks, type = "alpha.bar")
> smi.stock.params$alpha.bar <- 0.35
> smi.stock.params$gamma <- smi.stock.params$gamma - 0.005 * (4:1)
> smi.stock.params$mu <- smi.stock.params$mu + 0.005 * (4:1)
```

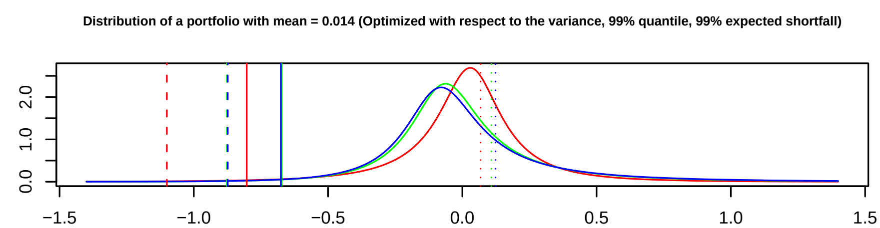
```
> skewed.returns <- do.call("ghyp", smi.stock.params)
> pairs(skewed.returns, rghyp(400, skewed.returns), cex = 0.5,
+     legend.cex = 0.5)
```



The following figure shows the distribution of the optimized portfolio with respect to the variance, quantile and expected shortfall. One can see that the skewness is negative when optimizing with respect to the variance whereas the skewness is positive when optimizing with respect to the quantile or expected shortfall. Since market risks are mostly measured in terms of value at risk (VaR) or expected shortfall one could mitigate the risk by optimizing portfolios with respect to the qunatile or expected shortfall. Additionally, the variance and the magnitude of "upward" fluctuation increases. This results in a higher expected growth when modelling log-returns.







References

[1] Alexander J. McNeil, Rüdiger Frey, Paul Embrechts, Quantitative Risk Management, Concepts, Techniques and Tools, 2005

[2] Alexander J. McNeil, S-Plus Library for Quantitative Risk Management *QRMlib*, 2005 , http://www.math.ethz.ch/~mcneil/book/QRMlib.html

[3] Wolfgang Breymann, One-dimensional hyperbolic distributions, unpublished, 2006,