European Society of Computational Methods in Sciences and Engineering (ESCMSE)

Journal of Numerical Analysis, Industrial and Applied Mathematics (JNAIAM)
vol. 11, no. 1-2, 2017, pp. 1-10
ISSN 1790–8140

Gaussian Scale Mixtures

Miguel Martins Felgueiras[1]*, João Paulo Martins, Rui Filipe Santos****

* CEAUL Lisbon and ESTG, CIGS, Polytechnic Institute of Leiria, Portugal

** CEAUL Lisbon and ESTG, Polytechnic Institute of Leiria, Portugal

Abstract: In this paper we present a parsimonious approximation of a Gaussian mixture when its components share a common mean value, i.e. a scale mixture. We show that a shifted and scaled Student's t-distribution can be approximated to this type of mixture, and use the result to develop a hypothesis test for the equality of the components mean value. A simulation study to check the quality of the approximation is also provided, together with an application to logarithmic daily returns.

Keywords: Gaussian scale mixture, Student's t-distribution, log-returns, simulation.

Mathematics Subject Classification: 60E05; 62P20.

1 Introduction

A random variable (rv) X is called a Gaussian convex mixture of $N \in \mathbb{N}$ Gaussian subpopulations or components if its probability density function f_X is given by

$$f_X(x) = \sum_{j=1}^{N} w_j \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_j}{\sigma_j}\right)^2\right\}, \quad \sigma_j > 0, w_j > 0, \sum_{j=1}^{N} w_j = 1, \tag{1}$$

where μ_j, σ_j and w_j are respectively the mean value, the standard deviation and the weight of the j-th subpopulation, for $j = 1, ..., N$. Applications can be found in Economics [1, 5, 11], Biology [8, 15] and Astronomy [12], among others [7]. Parameters are usually estimated with the Expectation Maximization (EM) algorithm, a variation of the maximum likelihood method [3, 14]. However, the EM algorithm may not lead to good estimates, even when the initial estimates are equal to the real parameters values [7, 17]. The EM algorithm only guarantees that a local maximum is found and, therefore, the obtained estimates must be carefully analysed.

When in (1) $\mu_j = \mu$ for $j = 1, ..., N$ we have a Gaussian scale mixture. Since the derivative of f, denoted by f', is given by

$$f'_X(x) = -(x-\mu) \sum_{j=1}^{N} w_j \frac{1}{\sqrt{2\pi}\sigma_j^3} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma_j}\right)^2\right\}, \tag{2}$$

[1]Corresponding author. E-mail: mfelg@ipleiria.pt

the mixture is unimodal with mode $x = \mu$. Therefore, multimodal data should be a sufficient condition to reject the equal mean value hypothesis.

The X cumulant generating function can be developed as

$$\ln[\varphi_X(-it)] = \mu t + \left[\sum_{j=1}^{N} w_j \sigma_j^2\right] \frac{t^2}{2!} + \left[3\sum_{j=1}^{N} w_j \sigma_j^4 - 3\left(\sum_{j=1}^{N} w_j \sigma_j^2\right)^2\right] \frac{t^4}{4!} + O\left(t^5\right),$$

where $\varphi_X(t) = E\left[e^{-itX}\right]$ denotes the characteristic function. Previous equation leads to mean, denoted by μ_1', and centered moments, denoted by $\mu_{(k)}$,

$$\mu_1' = \mu, \qquad \mu_{(2)} = \sum_{j=1}^{N} w_j \sigma_j^2, \qquad \mu_{(3)} = 0, \qquad \mu_{(4)} = 3\sum_{j=1}^{N} w_j \sigma_j^4.$$

Thus, the skewness and the kurtosis coefficients, are given respectively by

$$\beta_1 = \frac{\mu_{(3)}}{\mu_{(2)}^{3/2}} = 0, \qquad \beta_2 = \frac{\mu_{(4)}}{\mu_{(2)}^2} = \frac{3\sum\limits_{j=1}^{N} w_j \sigma_j^4}{\left(\sum\limits_{j=1}^{N} w_j \sigma_j^2\right)^2}, \tag{3}$$

and therefore the Gaussian scale mixtures are always symmetric around μ (symmetry could also be derived from noting that $\varphi_{X-\mu}(t)$ is an even function).

2 Gaussian mixtures and Student's *t*-distribution

Using the Pearson system of distributions [9], it is possible to show that the distribution of a Gaussian scale mixture can be approximated by a location-scale Student's t-distribution.

Theorem 1
Let X be a Gaussian mixture as defined in (1) where all subpopulations have equal mean value, i.e. $\mu_1 = \mu_2 = ... = \mu_N = \mu$. Then X can be approximated by Y,

$$Y \sim t_{(\mu,\sigma,\nu)}, \quad \sigma > 0, \nu > 0$$

where (μ, σ, ν) denotes the location, scale and freedom degrees of the Student's t-distribution as defined in [9], i.e. $\sigma = \sqrt{\sum\limits_{j=1}^{N} w_j \sigma_j}$ and υ is selected in order to fulfil $\beta_{2,X} = \beta_{2,Y}$.

Proof
Student's t-distribution has polynomial tails, heavier than the negative square exponential Gaussian tails, and therefore Student's t-distribution kurtosis is always greater than 3. Since the mixture is symmetric $(\beta_1 = 0)$, it can be approximated by a location-scale Student's t-distribution, i.e. a Pearson type VII distribution if $\beta_2 > 3$. Therefore, it is sufficient to establish that the rv X has kurtosis greater than 3. From the kurtosis coefficient defined in (3),

$$\frac{3\sum\limits_{j=1}^{N} w_j \sigma_j^4}{\left(\sum\limits_{j=1}^{N} w_j \sigma_j^2\right)^2} > 3 \iff \sum_{j=1}^{N} w_j \sigma_j^4 > \left(\sum_{j=1}^{N} w_j \sigma_j^2\right)^2.$$

Cauchy-Schwarz inequality,

$$\left(\sum_{j=1}^{N} x_j^2\right)\left(\sum_{j=1}^{N} y_j^2\right) \geq \left(\sum_{j=1}^{N} x_j y_j\right)^2,$$

can be applied considering $x_j = \sqrt{w_j}\sigma_j^2$ and $y_j = \sqrt{w_j}$, leading to

$$\left(\sum_{j=1}^{N} w_j\sigma_j^4\right)\left(\sum_{j=1}^{N} w_j\right) \quad \geq \quad \left(\sum_{j=1}^{N} w_j\sigma_j^2\right)^2 \Longleftrightarrow$$

$$\Longleftrightarrow \quad \sum_{j=1}^{N} w_j\sigma_j^4 \geq \left(\sum_{j=1}^{N} w_j\sigma_j^2\right)^2,$$

and therefore the mixture can be approximated by a location-scale Student's t-distribution. In fact, we are performing a location-scale transformation $Y = \mu + \sqrt{\sum_{j=1}^{N} w_j\sigma_j t_v}$ matching X and Y moments.

Theorem 1 allows to perform a test for the equality of the mean values,

$$H_0 : \mu_1 = \mu_2 = ... = \mu_N = \mu. \tag{4}$$

For $N = 2$, the condition $\beta_1 = 0$ on the Gaussian mixture implies that $\mu_1 = \mu_2$ or $w = 0.5$ and $\sigma_1^2 = \sigma_2^2$ [4]. For the second condition, the kurtosis of the mixture becomes

$$\beta_2 = \frac{-0.125\,(\mu_1 - \mu_2)^4}{\left(\sum\limits_{j=1}^{N} w_j\sigma_j^2\right)^2} + 3,$$

leading to $\beta_2 < 3$, and thus the Student's t-distribution approximation is only valid when the condition $\mu_1 = \mu_2$ is fulfilled. Nevertheless, for $N \geq 3$ it is possible to have $\beta_1 = 0$ and $\beta_2 > 3$ without $\mu_1 = \mu_2 = ... = \mu_N = \mu$ being fulfilled. Even though that situation is very unlikely since it only happens if the third cumulant is equal to zero, that is, if $\sum_{j=1}^{N} w_j \left(\mu_j^3 + 3\mu_j\sigma_j^2\right) + 2\mu^3 - 3\mu\sum_{j=1}^{N} w_j \left(\mu_j^2 + \sigma_j^2\right) = 0$ [6]. Thus, $\beta_1 = 0$ and $\beta_2 > 3$ may not entail $\mu_1 = \mu_2 = ... = \mu_N$, even theoretically. On the other hand, the conditions $\beta_1 \neq 0$ or $\beta_2 \leq 3$ or multimodal data imply that at least one of the mean values is different from the others. Hence, and assuming an underlying unimodal model, Table 1 summarizes the main conclusions.

Table 1: Main conclusions for equality of the components mean values test, where $\vee$ represents OR and $\wedge$ represents AND conditions.

		Decision	
H_0 : $\mu_1 = ... = \mu_N$		Reject H_0 ($\beta_1 \neq 0 \vee \beta_2 < 3$)	Don't reject H_0 ($\beta_1 = 0 \wedge \beta_2 > 3$)
	$N = 2$	implies $\mu_1 \neq \mu_2$	implies $\mu_1 = \mu_2$
	$N > 2$	implies $\exists\,(i,j) : \mu_i \neq \mu_j,\ i \neq j$	likely $\mu_1 = \mu_2 = ... = \mu_N$

Under the previous considerations, the Kolmogorov-Smirnov or other distribution fitting test [16] may be used to check the Student's t-distribution approximation under H_0 and therefore to perform the test for the equality of mean values. As a side result, note that if the Student's t-distribution approximation is valid, then the underlying distribution can be considered as symmetric and more heavy tailed than the Gaussian one.

3 A simulation study for the $\mu_1 = \mu_2$ test

To evaluate the proposed test accuracy, a simulation study was performed. Note that there are many parameters combinations, even when considering only two subpopulations, and thence any simulation study only represents a small subset of all the possible combinations of parameters. We believe that the conclusions within this section can be generalized for others subsets, however different parameters subsets require different simulation studies [6].

To set the parameters region for the simulation study, let us consider the model of daily log-returns, an often studied problem in finances. Log-returns are defined as

$$x_t = \ln(X_t) - \ln(X_{t-1}),$$

where X_t represents the close index value of the t-th day.

Several models have been considered for this kind of data, such as Gaussian mixtures, Student's t-distribution with location, scale or even skewness, stable Paretian models, generalized hyperbolic or generalized logF, among others [1, 5, 11, 13]. Unlike the often used stable Paretian models with $\alpha < 2$, the Gaussian mixtures have all order moments. Besides, they have a very flexible shape. Gaussian mixtures are recommended in [1, 11], preferentially with a small number of components to avoid over-fitting.

For the financial data problem, it is usually assumed that the model kurtosis should be higher than the Gaussian kurtosis. Besides, the data mean varies near zero, and skewness is sometimes present [11].

3.1 Fitting a two component Gaussian mixture to a data set

Let us consider the data set that corresponds to S&P 500 stock index daily returns, available on `http://finance.yahoo.com/q/hp?s=\%5Egspc`, between 01-08-1984 and 31-07-2014 (20 years of data) for a total of 5036 observations. For this sample, descriptive statistics are presented in Table 2.

Table 2: Descriptive statistics for the S&P 500 stock index daily returns

n	$\widehat{\mu}$	$\widehat{\mu}_{(2)}$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
5036	-0.00028	0.00015	-0.24586	11.17124

The unknown parameters of the Gaussian mixture, $(\mu_1, \mu_2, \sigma_1, \sigma_2, \omega)$, where the parameter ω corresponds to the first component weight, were estimated by the EM algorithm using *Matlab R2012b* software [7] and are displayed in Table 3.

Table 3: Estimated parameters for the two components Gaussian mixture

$\widehat{w}$	$\widehat{\mu}_1$	$\widehat{\mu}_2$	$\widehat{\sigma}_1$	$\widehat{\sigma}_2$
0.75823	0.00085	-0.00148	0.00712	0.02122

Thus, we investigated the performance of the test for the equality of the mean values, considering parameter values near the ones estimated for the S&P 500 stock index daily returns data.

3.2 Simulations results

Let us consider the hypotheses $H_0 : \mu_1 = \mu_2$ and $H_1 : \mu_1 \neq \mu_2$. As usual, the test power is defined as $1-\beta = P(\text{reject } H_0 \text{ when } H_0 \text{ is false})$. All the simulations were performed considering 1000 runs of samples with 5000 observations each. For each parameter combination, the test power, $1-\beta$, is estimated by the number of location-scale Student's t-distribution approximations rejected by the Kolmogorov-Smirnov test, at 5% significance level, over 1000. Lilliefors test would be preferable, since the Student's t-distribution parameters are estimated from the sample, but this is a less powerful test [16] and as far as we know it is not yet implemented in any software for the location-scale Student's t-distribution. Matching the theoretical moments from the Gaussian mixture with the theoretical moments of the Student's t-distribution we can find out the parameters needed for the Kolmogorov-Smirnov test.

If $(\mu_2, \sigma_2) = (-0.0015, 0.0212)$, the test power as a function of μ_1 is plotted in Figure 1, where $\sigma_1 = 0.005$, $\sigma_1 = 0.007$ and $\sigma_1 = 0.010$ for the dotted, dashed and thick curves, respectively, and $w = 0.7$, $w = 0.75$ and $w = 0.8$ for the left, center and right figures, respectively. We considered 0.0002 increases of μ_1 (from -0.0015 to 0.0065) and the points were interpolated by a third order spline. Figure 2 is similar, but here we have $w = 0.7$, $w = 0.75$ and $w = 0.8$ for the dotted, dashed and thick curves, respectively, and $\sigma_1 = 0.005$, $\sigma_1 = 0.007$ and $\sigma_1 = 0.010$ for the left, center and right figures, respectively.

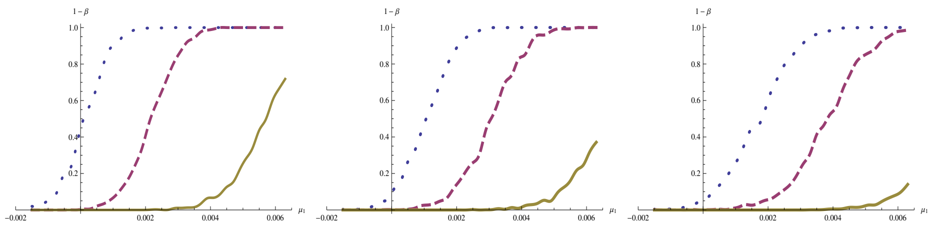

Figure 1: Test power for $w = 0.7$, $w = 0.75$ and $w = 0.8$ as a function of μ_1 and σ_1.

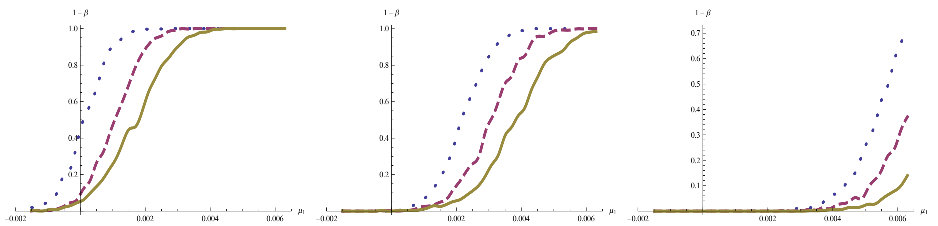

Figure 2: Test power for $\sigma_1 = 0.005$, $\sigma_1 = 0.007$ and $\sigma_1 = 0.010$ as a function of μ_1 and w.

Clearly the test power increases when μ_1 is farther from $\mu_2 = -0.0015$, as expected. Figure 2 shows that when w increases, i.e. more unbalanced mixtures, then $1-\beta$ decreases, since the dot line dominates the dash line, and the dash line dominates the thick line in yy axis. On the other hand, Figure 1 shows that when σ_1 increases, i.e. the subpopulations are more mixed, then $1-\beta$

decreases, since the dot line dominates the dash line, and the dash line dominates the thick line in yy axis. Therefore, small values of σ_1 and balanced values of w and $1-w$ contribute to a higher test power.

Our estimated parameters in Table 3 are similar to the ones that originate the dash line in the center graph of both figures. Table 4 presents the obtained test powers for that situation, that is, when $(w, \mu_2, \sigma_1, \sigma_2) = (0.75, -0.0015, 0.007, 0.0212)$. Note that the test seems quite sensible to the mean values difference, since when $\mu_1 = 0.0031$, then $\mu_2 - \mu_1 = -0.0046$ and $1-\beta = 0.5130$, that is, it is likely to be detected a mean values difference that is approximately two thirds of σ_1 and one fifth of σ_2. However, for $\mu_1 = 0.0009$, the test power is small ($1-\beta = 0.0230$) and therefore mean values are expected to be considered as equal.

Table 4: Obtained test power $1-\beta$ as a function of μ_1 when $(w, \mu_2, \sigma_1, \sigma_2) = (0.75, -0.0015, 0.007, 0.0212)$

μ_1	-0.0015	-0.0009	-0.0003	0.0003	0.0009	0.0015	0.0021	0.0023	0.0025	0.0027
$1-\beta$	0	0	0.0010	0.0010	0.0230	0.0490	0.1620	0.2160	0.2590	0.3040
μ_1	0.0029	0.0031	0.0033	0.0035	0.0037	0.0039	0.0045	0.0051	0.0057	0.0063
$1-\beta$	0.4380	0.5130	0.6040	0.7010	0.7400	0.8250	0.9580	0.9920	0.9990	1

4 Comparing models

The comparative measures Akaike information criterion (AIC) and Bayesian information criterion (BIC) are two popular measures for comparing maximum likelihood models of non-nested models [2], such as Student's t-distribution and Gaussian mixtures models. Other alternatives are, for instance, the Bayes factor or the Schwarz criterion [10], which will not be used in this work. AIC and BIC statistics are defined as

$$AIC = -2\ln(likelihood) + 2p$$

and

$$BIC = -2\ln(likelihood) + p\ln(n)$$

where n is the sample size and p the number of estimated parameters. AIC and BIC can be viewed as measures that combine both fit and complexity. Fit is measured negatively by $-2\ln(likelihood)$; the larger the value, the worse the fit. Complexity is measured positively, either by $2p$ (AIC) or $p\ln(n)$ (BIC). Given two models fitting the same data, the model with the smallest value of the information criterion should be preferred. BIC measure leads to more parsimonious models than AIC measure. When the two criteria do not provide the same solution, a final decision should be taken according to the analyst experience.

5 Fitting Gaussian mixtures and Student's t-distribution to the data

5.1 The data set

The analysed data set corresponds to the one introduced in subsection 3.1. To assess the approximation in Theorem 1, we considered different time intervals – 1 year, 5 years, 10 years, 15 years and 20 years – counting backwards from 31-07-2014. Some descriptive statistics for data in each time interval are presented in Table 5.

Table 5: Some descriptive statistics

	1 year	5 years	10 years	15 years	20 years
n	251	1257	2517	3773	5036
$\widehat{\mu}$	0.00049	0.00052	0.00022	0.00010	−0.00028
$\widehat{\mu}_{(2)}$	0.00004	0.00011	0.00017	0.00017	0.00015
$\widehat{\beta}_1$	−0.49510	−0.48858	−0.33722	−0.17784	−0.24586
$\widehat{\beta}_2$	4.30307	7.29288	14.1000	10.8356	11.1712

The data mean is always near 0. Besides, $\widehat{\mu}_{(2)}$ seems immune to the length of time period chosen for analysis (despite an initial increase form 0.00004 to 0.00011) and therefore models with infinite variance, as the stable Paretian model with $\alpha < 2$ [13] or the Student's t-distribution with $\nu < 2$ are not appropriate. Negative skewness, although not very sharp, is always present. Finally, kurtosis is clearly above 3, and thence Gaussian mixtures or location-scale Student's t-distribution can be suitable options if the skewness is considered as negligible.

5.2 Selecting models

Using *Matlab R2012b*, parameters for Gaussian mixtures with two, three and four components are estimated using the EM algorithm, selecting k observations from X at random as initial component mean values and considering equal mixing proportions and variances for all components in the first stage. As previously stated, authors advice Gaussian mixtures with a small number of components. Besides, in this study Gaussian mixtures with more than four components lead only to negligible improvements. Parameters for the Student's t-distribution are estimated using the maximum likelihood method. Next, the best model is selected for each period of time according with observed AIC and BIC statistics, provided in Table 6.

Considering BIC measure, the Student's t-distribution is always selected as the best model. However, for AIC measure the best model changes according to the period of time considered, between Gaussian mixtures with two, three and four components, although location-scale Student's t-distribution has presented similar results. Hence, if parsimonious is the priority for the researcher, location-scale Student's t-distribution is the advisable choice. Otherwise, Gaussian mixtures are a better option, although it is unclear how many components should be selected since their optimal number vary according to the time gap.

5.3 Testing the equality of the components mean values

The equality of the components mean values test, defined in (4), was carried out, leading to the results displayed in Table 7.

Note that we do not need to know the number of components that we are testing, because Student's t-distribution approximation has the advantage of being independent of the number of components. However, conclusions are different according to that number, as previously stated. The observed p-values are always "large" enough, thence the Student's t-distribution is never rejected. Naturally the p-value decreases with the increase of n, as it is usual in this kind of test. For the Gaussian mixture with two subpopulations, the preferable option according with [1], the decision implies that the mean values of the components can be considered the same.

Graphically, the location-scale Student's t-distribution seems to fit well to the data represented in the histogram (Figure 3) and the probability plot for the estimated Student's t-distribution (Figure 4) does not reveal relevant problems, although the extreme values present some lack of fit.

Table 6: AIC and BIC values for the fitted models

		AIC	BIC
1 year	location-scale Student's t	-1816.47	**-1811.42**
	2 Gaussian mixture	**-1817.50**	-1799.87
	3 Gaussian mixture	-1811.54	-1783.34
	4 Gaussian mixture	-1808.53	-1769.75
5 years	location-scale Student's t	-8154.54	**-8146.27**
	2 Gaussian mixture	-8153.15	-8127.47
	3 Gaussian mixture	**-8162.16**	-8121.06
	4 Gaussian mixture	-8158.25	-8101.75
10 years	location-scale Student's t	-15759.2	**-15749.5**
	2 Gaussian mixture	-15704.0	-15674.8
	3 Gaussian mixture	-15770.1	-15723.5
	4 Gaussian mixture	**-15776.4**	-15712.3
15 years	location-scale Student's t	-23055.8	**-23045.3**
	2 Gaussian mixture	-22992.6	-22961.4
	3 Gaussian mixture	-23079.0	-23029.1
	4 Gaussian mixture	**-23081.0**	-23012.4
20 years	location-scale Student's t	-31338.1	**-31327.1**
	2 Gaussian mixture	-31248.8	-31216.2
	3 Gaussian mixture	**-31361.6**	-31309.4
	4 Gaussian mixture	-31356.8	-31285.1

Table 7: Obtained p-values for the equality of the components mean values test

	1 year	5 years	10 years	15 years	20 years
p-value	0.87742	0.59909	0.19535	0.17060	0.11653

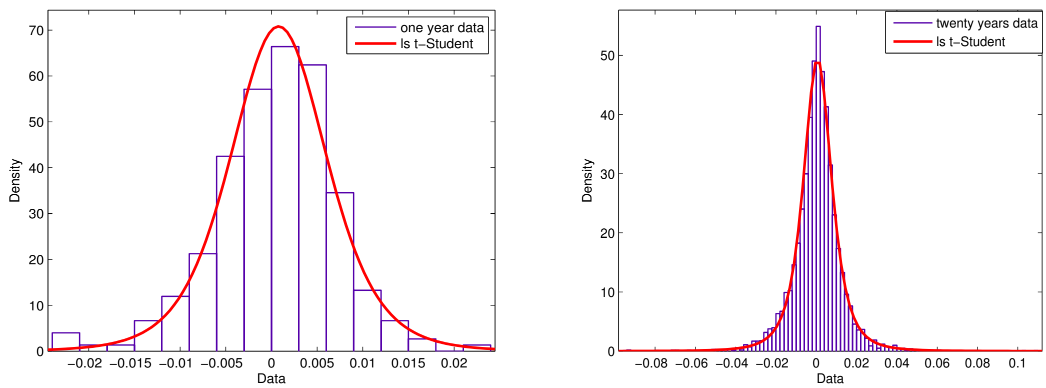

Figure 3: Fitted location-scale (ls) Student's t considering one year data (left) and twenty years data (right).

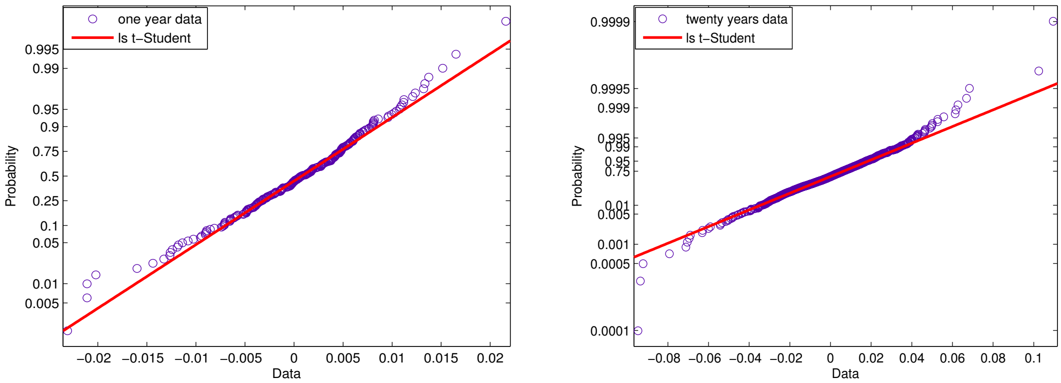

Figure 4: *pp*-plots for location-scale (ls) Student's t considering one year data (left) and twenty years data (right).

Therefore, even if previously obtained AIC results show that Gaussian mixtures are preferable, there is no evidence to support that the components mean values are different, and thus skewness can be considered as irrelevant as stated in Section 2.

6 Conclusion

Gaussian mixtures are indeed a very powerful tool to fit a model to a data set, although caution is advised in their application, especially whenever used in conjunction with the EM algorithm. First, there is no straightforward way to select the number of components as was illustrated in stock log-returns example. Besides, EM algorithm may not lead to accurate estimates, and it is sensitive to the arbitrary initial estimates [7, 17]. Finally, the obtained model has a large number of unknown parameters to estimate. For instance, a three component mixture has eight parameters. Kon [11] analysed several unimodal stock log return indices, some much more skewed than the S&P 500 analysed in this paper. In that situation, Student's t-distribution approximation does not provide a good fit, and Gaussian mixtures or other skewed models should be applied, despite of the possible problems discussed above. For unimodal and slightly skewed data, a Student's t-distribution may be applied with possible good results. The simulation study shows that the equality of mean values test is quite sensible for detecting mean values differences, considering two subpopulations and parameters near the estimated ones for the data. For the S&P 500 stock index log-returns, Student's t-distribution always outperforms Gaussian mixtures, considering the BIC measure, although, for the AIC statistic, the results are inconclusive. The Student's t-distribution approximation was never rejected by the Kolmogorov-Smirnov test and, therefore, no evidence to deny that $\mu_1 = ... = \mu_N$ was found. Summarizing, Student's t-distribution seems to be an advisable choice for unimodal and slightly skewed data, specially when parsimonious models are preferred by the user.

Acknowledgment

Research partially funded by FCT, Portugal, through the project Pest-OE/MAT/UI0006/2014.

References

[1] Behr, A. and Pötter, U. (2009). Alternatives to the normal model of stock returns: Gaussian mixture, generalized logF and generalized hyperbolic models. *Ann. Finance,* **5**, 49-68.

[2] Burnham, K. and Anderson, D. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* New York: Springer-Verlag.

[3] Dempster, A., Laird, N. and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc.*, **B 39**, 1-37.

[4] Everitt, B. S. and Hand D. J. (1981). *Finite Mixture Distributions.* London: Chapman & Hall.

[5] Fama, E. (1965). The behavior of stock-market prices. *J. Bus.,* **38**, 34-105.

[6] Felgueiras, M. (2009). *Modelação Estatística com Misturas e Pseudo-Misturas.* PhD Thesis. Lisbon: FCUL.

[7] Frühwirth-Schnatter, S. (2006). *Finite Mixture and Markov Switching Models.* New York: Springer-Verlag.

[8] Grantyn, R., Shapovalov, A. and Shiriaev, B. (1984). Relation between structural and release parameters at frog sensory-motor synapse. *J. Physiol.*, **349**, 459-474.

[9] Johnson, N., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions.* Volume I. New York: Wiley.

[10] Kass, R. and Raftery, A. (1995). Bayes factors, *J. Am. Stat. Assoc.*, **90**, 773-795.

[11] Kon, S. (1984). Models of stock returns — a comparison. *J. Finance,* **39,** 1, 147-165.

[12] Murtagh, F., Starck, J. and Bijaoui, A. (1995). Image restoration with noise suppression using a multiresolution support. *Astron. Astrophys. Supp. Ser.*, **112**, 179-189.

[13] Rachev, S. and Mittnik, S. (2000). *Stable Paretian Models in Finance.* New York: Wiley.

[14] Redner, R. and Walker, H. (1984). Mixture Densities, Maximum Likelihood and the EM Algorithm *SIAM Rev.*, **26**, 2, 195-239.

[15] Shapovalov, A. and Shiriaev, B. (1980). Dual mode of junctional transmission at synapses between single primary afferent fibres and motoneurones in the amphibian. *The J. Physiol.*, **306**, 1-15.

[16] Sheskin, D. (2002). *Handbook of Parametric and Nonparametric Statistical Procedures.* Boca Raton: Chapman & Hall/CRC.

[17] Verbeek, V., Vlassis, N. and Krose, B. (2003). Efficient greedy learning of Gaussian mixture models. *Neural Comput.*, **15**, 2, 469-485.