

# Assessing the number of components in a normal mixture: an alternative approach

Katarzyna Maciejowska\*  
Wroclaw University of Technology

## Abstract

In this article, a new approach for model specification is proposed. The method allows to choose the correct order of a mixture model by testing if a components is significant. The hypothesis are set in a new way, in order to avoid regularity problems, which are typical for mixture models. If some of the parameters are know, the distribution of the LR statistic is  $\chi^2$ , with the degrees of freedom depending on the number of components and the number of parameters in each component. The advantage of the new approach is its simplicity and computational feasibility.

*JEL classification:* C13, C33, C43

*Keywords:* Gaussian mixture; Likelihood ratio test; Order selection

## 1 Motivation

There is a strong interest in finite mixture models in the last few decades. Applications of mixture models can be found in many disciplines such as biology, medicine and engineering, among others. In economics, they have been successfully used in marketing (Jedidi et al. (1997)), finance (Liesenfeld (1998), Liesenfeld (2001), Perez-Quiros and Timmermann (2001)) and macroeconomics (Lanne (2006)). The development of computers and computation techniques enables development of Maximum Likelihood (ML) estimation approaches (Day (1969)). However, it was the Expectation Maximization (EM) algorithm described by Dempster et al. (1977) that significantly simplified the estimation procedure and therefore helped to popularize the mixture models. Recently, mixture models have been expanded in various ways in order to allow for the conditional heteroscedasticity (Zhang et al. (2006)) or time varying mixing proportions (Wong and Li (2001)).

Mixtures, with normal components, belong to a very flexible family of distributions, which can approximate any distribution to an arbitrary degree of freedom (McLachlan and Peel (2000)). Therefore, they are suitable for modeling

---

\*Katarzyna Maciejowska, ul. Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland, katarzyna.maciejowska@pwr.wroc.pl

processes that are bimodal, asymmetric or characterized by heavy tails. They are also used for modeling of process with several distinguishable patterns of behavior, such as business cycle or fluctuations of financial assets characterized by periods of high and low volatility. It should be noticed that mixture models are also special cases of Markov switching models, which are extensively used in econometrics and macroeconomic analysis (Kim and Nelson (1999), Sims and Zha (2006), Smith et al. (2006)), especially in business cycle analysis (Hamilton (1989), Goodwin (1993), Diebold and Rudebusch (1996), Kim and Nelson (1998)).

An open problem that needs to be solved is a way to determine the number of components in the mixture model. It is an important but very difficult issue. There are a few reasons, why tests based on Maximum Likelihood (ML) estimation method do not behave well and have unknown distribution. As discussed by McLachlan and Peel (2000) and Chen and Li (2009), the problem arises because of: (i) lack of identifiability under the null, (ii) unbounded likelihood function, (iii) infinite Fisher information, (iv) ML estimators of the homogenous model being one of the local maxima of the likelihood of mixture model. Additionally, as shown by Hartigan (1985) and Liu and Shao (2004), the likelihood ratio (LR) statistic diverge to infinity as sample size increases.

A few different approaches were proposed in the literature. Starting with Wolfe (1971), who suggested that when the components have common variance, the distribution of  $LR$  statistic can be approximated with a  $\chi_2^2$  distribution. When the variances are allowed to differ, the distribution can be approximated with  $\chi_4^2$ . On the contrary, McLachlan (1987) suggested, that in a case of mixtures with different means and variances, the  $\chi_6^2$  distribution worked well for the sample size  $N = 100$ .

In order to derive a correct distribution, some separation conditions were imposed by Ghosh and Sen (1985). The restrictions were later relaxed by Garel (2001) and Garel (2005). The approach requires heavy computation and simulation of critical values. An alternative method was proposed by Chen and Li (2009), King and Chen (2010) and Chen et al. (2012), who used an  $EM$  test. Their test is based on the modified likelihood function, with a penalty component added to ensure identifiability. Although the asymptotic distribution of the proposed test is fairly standard, computation of the EM statistic is complicate and its properties depends on an arbitrary choices of additional parameters: number of EM iterations and parameters of penalty functions. Next, Lo et al. (2001) introduced a test based on Kullback-Leibler information criterion. However, Jeffrues (2003) showed that the regularity conditions used by Lo et al. (2001) are not met under the null of homogeneity. Finally, a resampling approach was proposed by McLachlan (1987), who present a way to compute the  $p$ -values of the LR test based on the bootstrap method.

In this article, an alternative approach to testing for the number of components is presented. It is based on a new hypotheses specification, which allows to overcome the identification problem. The hypotheses use only these parameters, which can be consistently estimated with the ML method, without imposing any separation conditions or adding penalty to the log likelihood function. A like-

likelihood ratio test statistic is proposed, which is fairly simple to compute and bounded under the null.

The article is structured as follows. In Section 2 a new set of hypotheses is presented and its equivalence to the traditional hypotheses formulation is proved. Next, Section 3 introduces a simple likelihood ratio test and discusses its basic properties. Finally, in Section 4, the asymptotic distribution for models, in which some parameters are known, are presented and validated with a simulation studies. Section 5 ends the article with some discussion and conclusions.

## 2 Alternative hypothesis formulation

Let  $f(y; \beta, \sigma^2)$  be the normal density function with a mean  $\beta$  and a variance  $\sigma^2$ . Lets summarize the parameters with a vector  $\theta = [\beta', \sigma^2]$ ,  $\theta \in \Theta$ . The finite normal mixture model of order  $m$ , has the following density function

$$f_m(y; \theta^{(m)}, p^{(m)}) = \sum_{i=1}^m p_i f(y; \theta_i) \quad (1)$$

where  $\theta^{(m)} = [\theta_1, \dots, \theta_m]$ ,  $p^{(m)} = [p_1, \dots, p_m]$  and  $p_i$  denotes the mixing proportions with the properties:  $p_i \geq 0$  and  $\sum_{i=1}^m p_i = 1$ .

### 2.1 Testing normality against a mixture of two normal densities

Suppose, we want to test, if the underlying process is normal versus the alternative that it follows a mixture of two normal distributions. Under the null,  $y_t \sim N(\beta_0, \sigma_0^2)$ , whereas under the alternative

$$y_t \sim pN(\beta_1, \sigma_1^2) + (1 - p)N(\beta_2, \sigma_2^2)$$

Lets denote  $\theta_0 = [\beta_0, \sigma_0^2]$ ,  $\theta_1 = [\beta_1, \sigma_1^2]$  and  $\theta_2 = [\beta_2, \sigma_2^2]$ .

In the traditional setup, hypothesis are stated in the form of restrictions on the parameters of the mixture model. As mentioned by Garel (2001) and Garel (2007), the null hypothesis is represented by three curves:

- $p = 0$  and  $\theta_2 = \theta_0$
- $p = 1$  and  $\theta_1 = \theta_0$
- $\theta_1 = \theta_2 = \theta_0$

Because, in the mixture model, the ordering of the components doesn't affect the distribution function, we can assume, without loss of generality, that the first component is significant, meaning that  $p \geq 0.5$ . Then the null is reduced

to two curves:  $p = 1$ ,  $\theta_1 = \theta_0$  and  $\theta_1 = \theta_2 = \theta_0$ . The hypothesis can be stated as follows (Chen and Li (2009)):

$$H_0 : (1 - p)(\theta_1 - \theta_2) = 0$$

$$H_1 : (1 - p)(\theta_1 - \theta_2) \neq 0$$

Under the null, some of the parameters are not identifiable. For example, when  $p = 1$  then the parameter vector  $\theta_2$  can take any value from  $\Theta$  without changing the overall density function (1). Similarly, when  $\theta_1 = \theta_2$  then the mixing proportion  $p$  can't be identified. It leads to some serious problems, when the mixture model is estimated and the null hypothesis is tested.

At the same time, it can be noticed that when the true model is normal, then adding insignificant components doesn't change the parameters of the significant one. Hence, regardless of the number of redundant components, there should be  $\theta_0 = \theta_1$ . We will use this property to develop a new set of hypothesis.

First, let's denote  $\tilde{\beta}$  and  $\tilde{\sigma}^2$  the first two central moments of  $y$  :  $\tilde{\beta} = E(y)$  and  $\tilde{\sigma}^2 = Var(y)$ , respectively. Then

$$\tilde{\beta} = p\beta_1 + (1 - p)\beta_2 \tag{2}$$

$$\tilde{\sigma}^2 = p\sigma_1^2 + (1 - p)\sigma_2^2 + p(1 - p)(\beta_1 - \beta_2)^2 \tag{3}$$

Under the null, either  $p = 1$  and  $\theta_1 = \theta_0$  or  $\theta_1 = \theta_2 = \theta_0$ . In both cases  $\theta_0 = \theta_1$ . Moreover, the moments are equal to the parameters of the normal distribution and  $\tilde{\theta} = \theta_0 = \theta_1$ . Under the alternative, the second component is significant. It means that both  $p \neq 1$  and  $\theta_1 \neq \theta_2$ . From equations (2)-(3), it can be easily shown that also  $\tilde{\theta} \neq \theta_1$ . Therefore, we can write the hypothesis in the alternative way

$$H_0^* : \tilde{\theta} = \theta_1$$

$$H_1^* : \tilde{\theta} \neq \theta_1$$

Both sets of hypotheses  $(H_0, H_1)$  and  $(H_0^*, H_1^*)$  are equivalent.

It should be noticed that when the new pair of hypotheses is considered, no distinction between the two curves  $p = 1$  and  $\theta_1 = \theta_2$  is needed. Moreover, under the null, parameters  $\tilde{\theta}$ ,  $\theta_0$  and  $\theta_1$  are unique, even when the remaining parameters of the mixture model are not identifiable.

## 2.2 Testing normality against a mixture of two normal densities with equal variances

The hypotheses can be simplified, when equal component variances are assumed. Let's consider a case, when under the null  $y_t \sim N(\beta_0, \sigma_0^2)$ , whereas under the alternative there is

$$y_t \sim pN(\beta_1, \sigma^2) + (1 - p)N(\beta_2, \sigma^2)$$

In the traditional setup, the hypotheses are

$$H_0 : (1 - p)(\beta_1 - \beta_2) = 0$$

$$H_1 : (1 - p)(\beta_1 - \beta_2) \neq 0$$

Under the null, some of the parameters of the mixture model are not identifiable (either  $\beta_2$  or  $p$ ). In order to solve the problem, the second, central moment of the sample can be used. Similarly to the previous case, the variance,  $\tilde{\sigma}^2$ , of the variable  $y$  is equal to (Chen and Li (2009))

$$\tilde{\sigma}^2 = \sigma^2 + p(1 - p)(\beta_1 - \beta_2)^2 \quad (4)$$

It is straightforward that the two conditions  $(1 - p)(\beta_1 - \beta_2) = 0$  and  $\tilde{\sigma}^2 = \sigma^2$  are equivalent. Hence, the hypotheses can be reformulated as follows

$$H_0^* : \tilde{\sigma}^2 = \sigma^2$$

$$H_1^* : \tilde{\sigma}^2 \neq \sigma^2$$

Moreover, under the null the variance of the variable  $y$  is equal to the variance of the normal distribution,  $\tilde{\sigma}^2 = \sigma_0^2$  and therefore also  $\sigma_0^2 = \sigma^2$ .

The new pair of hypotheses could be also based on the comparison of the mean of the significant component and the mean of the process. When the mean is a more complex phenomena, for example in regression models, then this approach requires more degrees of freedom than the variance comparison. Therefore, I found the hypotheses based on variances simpler and more suitable, especially when some prior information about  $\sigma^2$  is available.

### 2.3 Testing $m$ -component against $(m+1)$ -component mixture

The approach can be easily extended to the case, in which the significance of the last component in the  $(m + 1)$ -component mixture is tested. Lets assume that under the null, the data is described by  $m$ -component mixture.

$$y_t \sim \sum_{i=1}^m p_i f(y; \theta_i) \quad (5)$$

whereas under the alternative, there are  $m + 1$  significant components

$$y_t \sim \sum_{i=1}^{m+1} q_i f(y; \theta_i) \quad (6)$$

We can say, that the last component is insignificant, if either its probability is zero,  $q_{m+1} = 0$  or its parameters are equal to parameters of some other

component,  $\exists i \in \{1, \dots, m\} : \theta_{m+1} = \theta_i$ . These conditions are typically described in the form of restrictions on the parameters of the  $(m+1)$ -component mixture.

$$H_0 : q_{m+1} \prod_{i=1}^m (\theta_{m+1} - \theta_i) = 0$$

$$H_1 : q_{m+1} \prod_{i=1}^m (\theta_{m+1} - \theta_i) \neq 0$$

Similarly to the previous case, in which  $m = 1$ , some of the parameters of the model (6) are not identifiable under the null. When  $q_{m+1} = 0$  then  $\theta_{m+1}$  may take any value from  $\Theta$ . On the other hand, when there exists  $i \in \{1, \dots, m\}$  such that  $\theta_{m+1} = \theta_i$ , then both  $q_i$  and  $q_{m+1}$  are not identifiable.

In order to solve the identification problem, the hypothesis need to be reformulated. The new hypothesis are based on the observation that adding insignificant components doesn't change parameters of significant ones. Lets denote  $g(y; \tilde{\theta})$  a density function that is either normal or a mixture of two components, with the first two moments described as before by  $\tilde{\theta} = [\tilde{\beta}', \tilde{\sigma}^2]$ . Then

$$y_t \sim \sum_{i=1}^{m-1} \tilde{p}_i f(y; \tilde{\theta}_i) + \tilde{p}_m g(y; \tilde{\theta}_m)$$

Lets denote  $\theta_{1:m}^{(m+1)}$  the parameters of the first  $m$ , significant components of the model (6). Under the null, the  $g(y)$  component is normally distributed and the parameter vectors  $\theta_{1:m}^{(m+1)}$  and  $\tilde{\theta}^{(m)}$  are equal, up to the permutation of the components. Lets order the components in  $\theta^{(m)}$  and  $\theta_{1:m}^{(m+1)}$  in such a way that  $\beta_1 \leq \dots \leq \beta_m$  and  $\sigma_i^2 \leq \sigma_{i+1}^2$  for  $\beta_i = \beta_{i+1}$ . Then, under the null

$$H_0^* : \theta_{1:m}^{(m+1)} = \tilde{\theta}^{(m)} \tag{7}$$

On the other hand, it can be easily shown that when  $H_0^*$  holds then also  $H_0$  is true. It follows directly from the section .. and the fact that the finite mixture of normal distributions is uniquely identified up to the permutation of its components. Hence, the two null hypotheses,  $H_0$  and  $H_0^*$ , are equivalent.

Lets set the alternative hypothesis as follows

$$H_1^* : \theta_{1:m}^{(m+1)} \neq \tilde{\theta}^{(m)} \tag{8}$$

then  $(H_0, H_1)$  can be replaced by  $(H_0^*, H_1^*)$ . It should be noticed that under  $H_0^*$ , parameter vectors  $\theta_{1:m}^{(m+1)}$  and  $\tilde{\theta}^{(m)}$  are identifiable. This property simplified the testing procedure significantly.

### 3 Test for the significance of a component in a normal mixture model

Lets start the construction of the test by describing the log-likelihood function of the mixture of  $m$  normal distributions

$$l_{m,T}(\theta, p) = \sum_{t=1}^T \ln f_m(y_t; \theta, p) \quad (9)$$

where  $f_m(y_t; \theta, p)$  is defined by (1). In a traditional setup, the following likelihood ratio test is considered

$$LR = 2 \left( l_{m+1,T}(\hat{\theta}^{(m+1)}, \hat{p}^{(m+1)}) - l_{m,T}(\hat{\theta}^{(m)}, \hat{p}^{(m)}) \right)$$

with  $\hat{\theta}^{(m)}$ ,  $\hat{p}^{(m)}$ ,  $\hat{\theta}^{(m+1)}$  and  $\hat{p}^{(m+1)}$  being the ML estimators of parameters of models with  $m$  and  $(m + 1)$  components, respectively.

It has been discussed by many authors (Hartigan (1985), McLachlan and Peel (2000), Liu and Shao (2004), Chen and Li (2009) among others) that the  $LR$  statistic doesn't have a standard  $\chi^2$  distribution. Although the model seems relatively simple, the likelihood function and ML estimators suffer serious problems.

First, as discussed in the previous section, parameters of the mixture model are not identifiable under the null. It has important implications for the asymptotic distribution of the  $LR$  statistic. The problem has been discussed by Davies (1977), Andrews and Ploberger (1994), Hansen (1996). They showed that when there are nuisance parameters, which can't be identified under the null, then the critical values of the  $LR$  statistic can't be tabulated and need to be simulated for each application separately. Unfortunately, their approach requires some regularity conditions, which are not satisfied in the MN context. Therefore, their method should be used with much caution. There are a few other papers, which try to solve the identification problem. For example, King and Shively (1993) reparametrized the non-identifiable model using a transformation involving polar coordinates, Ghosh and Sen (1985) imposed some separation conditions on component parameters. In Chen and Li (2009) the identification is achieved by assuming known mixing proportion parameter. These approaches are difficult to generalize and sometimes hard to justify.

Second, the log likelihood function (9) is unbounded, and diverges to infinity when the variance of one component approaches 0, see Hathaway (1985) and Chen and Li (2009) for discussion. It means that the model parameters can't be consistently estimated with the ML method, even when the true underlying model is a mixture of distributions. The problem can be solved by imposing restrictions on the components relative variances, as proposed by Hathaway (1985), or adding penalty function to the log likelihood (Chen and Li (2009), Chen et al. (2012)). It was shown by Redner and Walker (1984) that when the parameter space is compact and contains the true parameter values, then the ML estimator is super consistent. Unfortunately, placing constraints on variances

doesn't ensure that the  $LR$  statistic is well behaved. As shown by Hartigan (1985) and confirmed by Liu and Shao (2004), when the mean parameters are not bounded, the  $LR$  statistic diverges to infinity even in a model with equal component variances.

Finally, the log likelihood is so irregular ( $\hat{\theta}^{(m)}$  is one of the local maxima of the  $l_{m+1,T}(\theta, p)$ , the Fisher information matrix is not well defined) that even under strong parameter constraints, the distribution of  $LR$  is nonstandard and very difficult to simulate (Chen and Li (2009)).

Alternatively, we can construct a test statistic, which will verify the hypothesis (7). If the true model is (5) then  $\theta_{1:m}^{(m+1)} = \hat{\theta}^{(m)}$  and  $\hat{\theta}^{(m)} = \theta^{(m)}$ . So under the null  $\theta_{1:m}^{(m+1)} = \theta^{(m)}$ . Therefore, in order to test whether the last component of the  $(m+1)$ -mixture is insignificant, we propose to compare the parameter vectors  $\theta_{1:m}^{(m+1)}$  and  $\theta^{(m)}$ . It can be done with the likelihood ratio test based on likelihood function of the  $m$ -component mixture model

$$LR^* = 2(l_{m,T}(\hat{\theta}^{(m)}) - l_{m,T}(\hat{\theta}_{1:m}^{(m+1)})) \quad (10)$$

Small values of the  $LR^*$  statistic indicates that  $\theta^{(m)}$  and  $\theta_{1:m}^{(m+1)}$  don't differ significantly. Hence, the last component in the  $(m+1)$ -mixture is insignificant. On the other hand, when the true underlying model is (6) then  $\theta^{(m)} \neq \theta_{1:m}^{(m+1)}$  and  $LR^*$  will rise with the sample size and will lead to the rejection of null.

Unlike the traditional setup, parameters  $\theta^{(m)}$  and  $\theta_{1:m}^{(m+1)}$  are identifiable under the null. When some constraints, similar to those proposed by Hathaway (1985), are imposed, then both parameter vectors can be consistently estimated with the ML method. The estimators  $\hat{\theta}^{(m)}$  and  $\hat{\theta}_{1:m}^{(m+1)}$  converges, with the same rate,  $\sqrt{T}$ , to the true vector of parameters  $\theta^{(m)}$ . Hence also

$$\sqrt{T}(\hat{\theta}^{(m)} - \hat{\theta}_{1:m}^{(m+1)}) = O_p(1)$$

Moreover, it can be shown that the test statistic  $LR^*$  is bounded in the probability, that is  $LR^* = O_p(1)$ . It follows from the fact that under the null, the matrix of second derivatives of the  $m$ -component mixture is well behaved and nonsingular. The same can be said about the variance-covariance matrix,  $\Sigma(\theta^{(m)})$ , of the estimator  $\hat{\theta}^{(m)}$ . Therefore,

$$LR^* \simeq T(\hat{\theta}^{(m)} - \hat{\theta}_{1:m}^{(m+1)})\Sigma^{-1}(\theta^{(m)})(\hat{\theta}^{(m)} - \hat{\theta}_{1:m}^{(m+1)}) = O_p(1)$$

Hence, the  $LR^*$  statistic doesn't diverge to infinity, both for  $m = 1$  and  $m > 1$ .

Unfortunately, the distribution of  $LR^*$  is nonstandard. Firstly, when no restrictions on parameter space are imposed, then the difference between the estimators is  $o_p(1)$  and it converges to zero, for  $\sigma_{m+1}^2 \rightarrow 0$ . Second, when some constraints are added, both vectors of parameters are still estimated and the variance-covariance matrix of their difference is different from  $\Sigma(\theta^{(m)})$ . Hence,  $LR^*$  doesn't have a  $\chi^2$  distribution. However, it can be shown that in those cases, when some of the parameters are known, the  $LR^*$  test statistic will have a standard distribution and will follow  $\chi^2$  with the number of degrees of freedom depending on the number of components under the null and the model setup.

## 4 Results

### 4.1 Known parameters of significant components

Suppose, we want to test if the data is normally distributed against the alternative that it follows a mixture of two normal densities. In the following section, it is assumed that parameters of the significant component are known and equal to:  $\beta_1 = 0$  and  $\sigma_1^2 = 1$ . Hence the hypothesis are

$$H_0 : y_t \sim N(\beta_0, \sigma_1^2)$$

$$H_1 : y_t \sim pN(0, 1) + (1 - p)N(\beta, \sigma^2)$$

This specification was considered by many authors, for example by Chen and Chen (2001), Liu and Shao (2004), Garel (2001) and Garel (2007). In some of the papers, additional restrictions on parameters are imposed, such as equal and known variances (Garel (2001), Liu and Shao (2004)). Although, the problem seems simple, it is very difficult to verify in the traditional setup. Under the null, the  $LR$  statistic has a nonstandard distribution and its critical values have to be simulated. Lets consider the new pair of hypotheses, which are based on the comparison of parameters of the significate component and the first two central moments of the sample. Under the null, the moments are known and the hypotheses are trivial

$$H_0^* : \theta_0 = \tilde{\theta} = [0, 1]$$

$$H_1^* : \tilde{\theta} \neq [0, 1]$$

The null can be easily tested with the  $LR^*$  statistic

$$LR^* = 2(l_{1,T}(\hat{\theta}_0) - l_{1,T}([0, 1]))$$

Under the null, the statistic has a  $\chi^2(2)$  distribution.

The empirical size and a power of the test are evaluated on the basis of a Monte Carlo experiment. In order to asses the size of the test, 10000 normally distributed,  $N(0, 1)$  random samples are generated for different sample sizes,  $T = 50, 100, 500, 1000$ . The results are presented in the Table .... The empirical size is close to the nominal significance level. It is not surprising, since the testing procedure collapses to a well known statistical problem: verification of the hypothesis that a sample comes from a normal distribution with zero mean and an unit variance.

The observed power of the test was estimated on the basis of 10000 samples drawn from a two-component normal mixture. 15 different scenarios were considered, based on three values of mixing proportions  $p = 0.5, 0.7$  and  $0.9$  and different combinations of parameters  $(\beta_2, \sigma_2^2)$ . The empirical power of the test for the significance level  $\alpha = 0.05$  are presented in Table ... Firstly, it can be observed that the power of the test is larger than 50% for all  $T$  and the mixing proportion  $p = 0.5, 0.7$ . For all parameter specifications, it increases with  $T$  and reaches 100% for  $T = 1000$ . Moreover, the further are the components from

Table 1: Simulated size for testing a single normal versus a mixture of two normals with known parameters of the significant state; based on 10000 replications for each sample size

$T/\alpha$	0.1	0.05	0.01
50	0.1038	0.0526	0.0108
100	0.1033	0.0513	0.0100
500	0.1011	0.0503	0.0101
1000	0.0989	0.0501	0.0100

each other, the higher is the empirical power. For example, when the alternatives with parameters  $(0, 4)$  and  $(0, 9)$  are compared, the latter is detected more often, especially for short samples or large  $p$ . On the other hand, the mixing proportion,  $p$ , has an adverse effect on the power. Models with  $p$  close to 1 require much larger samples to detect the nonlinearity.

The results are in line with previous findings of Mendell et al. (1991) and Garel (2001), who showed, with much more demanding techniques, that the bigger is the distance between the components, the larger is the power of  $LR$  statistics. Moreover, values of  $p$  close to 1 required larger samples to reach the power level of 50% (Mendell et al. (1991)). When the traditional  $LR$  statistic is considered, our  $LR^*$  test exhibits larger power that, for example, the approximation used by Mendell et al. (1991). The results of Garel (2001) are hard to compare, because in his paper a one sided test is used, with  $\beta_2 > 0$  and  $\sigma_2^2 = 1$  under the alternative.

## 4.2 Unknown means, equal and known variances

Finally, lets consider a case of a null hypothesis that a variable is normally distributed and the alternative that is follows a mixture of two normal distributions with equal and known variances,  $\sigma^2 = 1$ . Then under the alternative there is

$$y_t \sim pN(\beta_1, 1) + (1 - p)N(\beta_2, 1)$$

This problem has been discussed by Ghosh and Sen (1985), Chen and Chen (2001), Garel (2001) and King and Chen (2010). Different approaches to the estimation of mixture model parameters and approximation of critical values were proposed. In this paper, a new pair of hypotheses, which follows from Section 2.2, is considered. They are formulated as follows:

$$H_0^* : \tilde{\sigma}^2 = 1$$

$$H_1^* : \tilde{\sigma}^2 \neq 1$$

with  $\tilde{\sigma}^2 = \sigma_0^2$  under the null.  $H_0^*$  can be tested with the  $LR^*$  statistic, which verifies if the linear model has a unit variance. It has a standard,  $\chi^2(1)$  distribution.

Table 2: Simulated powers for testing a single normal versus a mixture of two normals with known parameters of the significant state; based on 10000 replications under each alternative

$T$	$(\beta_2, \sigma_2^2)$				
	(1,1)	(2,1)	(0,4)	(0,9)	(1,4)
$p = 0.5$					
50	0.897	0.999	0.965	0.999	0.993
100	0.995	1.000	0.999	1.000	1.000
500	1.000	1.000	1.000	1.000	1.000
1000	1.000	1.000	1.000	1.000	1.000
$p = 0.7$					
50	0.521	0.985	0.765	0.982	0.891
100	0.819	1.000	0.951	1.000	0.992
500	1.000	1.000	1.000	1.000	1.000
1000	1.000	1.000	1.000	1.000	1.000
$p = 0.9$					
50	0.115	0.441	0.256	0.605	0.358
100	0.173	0.704	0.391	0.823	0.566
500	0.640	0.999	0.924	1.000	0.988
1000	0.910	1.000	0.997	1.000	1.000

Similarly as before, the empirical size and the power are estimated on the basis of 10000 randomly drawn samples. When the size is considered, the samples are normally distributed,  $N(0, 1)$ , whereas for power analysis six different scenarios are used. The scenarios are characterized by the mixing proportion parameter,  $p$ , which is either 0.5, 0.7 or 0.9 and the normalized difference between the mean parameters of two components  $D = (\beta_1 - \beta_2)/\sigma$ ;  $\beta_1$  is set to 0 and  $\sigma$  is set to 1. The results are presented in Table .. and Table.

Table 3: Simulated size for testing a single normal versus a mixture of two normals with equal and known variances; based on 10000 replications for each sample size

$T/\alpha$	0.1	0.05	0.01
50	0.1066	0.0536	0.0113
100	0.1030	0.0517	0.0110
500	0.1011	0.0506	0.0097
1000	0.0992	0.0495	0.0098

The empirical size of the test is close to the nominal significance levels, for all sample sizes. Moreover, the observed power of the test approaches 1, for long

Table 4: Simulated powers for testing a single normal versus a mixture of two normals with equal and known variances; based on 10000 replications for each sample size

D	1			2		
T/p	0.5	0.7	0.9	0.5	0.7	0.9
50	0.174	0.139	0.071	0.943	0.856	0.325
100	0.336	0.259	0.090	0.999	0.991	0.565
500	0.941	0.847	0.277	1.000	1.000	0.996
1000	0.998	0.987	0.489	1.000	1.000	1.000

samples, with  $T = 500$  and  $1000$ , and almost all parameter configuration. Only, when the mixing parameter is  $0.9$  and the components are not well separated ( $D = 1$ ), the power doesn't exceed  $0,5$  for  $T = 1000$ . It can be observed that our results are still superior to the outcomes of Mendell et al. (1991). Unfortunately, up to my knowledge, there are no other articles, which presents the power of the  $LR$  test for the above model setup. Most of the results discussed in the literature are theoretical and are presented without further verification.

## 5 Conclusions

In this article, an alternative approach to hypotheses formulation, when testing for the number of components in a mixture model, is presented. The new set of hypotheses doesn't involve non-identifiable parameters, such as parameters of insignificant component or mixing proportions. Hence, it overcomes one of the crucial problems associated with verification of mixture models. Moreover, a simple  $LR^*$  statistic is proposed, which is based on a well behaved likelihood function. When the parameter space is compact, then the ML parameter estimators converge to the true ones and the  $LR^*$  statistic, unlike the traditional  $LR$  statistic, is bounded. No further assumptions or separation conditions are needed.

The simulation studies suggest that the test work well for cases, when some of parameters are known. It is shown that the test has a correct size, when parameters of the significant component are known or when components have equal and known variances. These examples were previously analyzed by many authors, for example by Chen and Chen (2001), Garel (2001), Garel (2007) and King and Chen (2010), which used much more demanding approaches and simulation techniques to verify the null of normal distribution against the alternative of a mixture of two normal distributions. The simulation results indicates that the test has a good power, when compared with a traditional testing procedure and therefore seems promising.

It would be of interest to extend the simulation studies to models with unknown parameters, similarly to Lo et al. (2001) and Chen and Li (2009). This

requires deriving a consistent estimator of the variance-covariance matrix of the the difference between parameters from restricted and unrestricted models. It could be also checked, what are the losses of power and size, when the incorrect assumption of the  $\chi^2$  distribution of  $LR^*$  is used.

The proposed approach is very simple and relatively easy to apply. It has an intuitive interpretation and therefore could be extended to other cases and model setup. It would be interesting to further develop it, in order to cover more general model specification, such as linear regression models and multi-dimensional normal distribution models.

## References

- Andrews, D. W. K. and W. Ploberger (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica* 62(6), 1383–1414.
- Chen, H. and J. Chen (2001). Large sample distribution of the likelihood ratio test for normal mixtures. *Statistics & Probability Letters* 52, 125–133.
- Chen, H. and P. Li (2009). Hypothesis test for normal mixture models: the em approach. *The Annals of Statistics* 37(5).
- Chen, H., P. Li, and Y. Fu (2012). Inference on the order of a normal mixture. *Journal of the American Statistical Association* 107(499), 1096–1105.
- Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 64(2), 247–254.
- Day, N. E. (1969). Estimating the components of a mixture of normal distributions. *Biometrika* 56(3), 463–474.
- Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum-likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistics Society Ser. B (methodological)* 39, 1–38.
- Diebold, F. X. and G. D. Rudebusch (1996). Measuring business cycle: A modern perspective. *Review of Economics and Statistics* 78, 67–77.
- Garel, B. (2001). Likelihood ratio test for univariate gaussian mixture. *Journal of Statistical Planning and Inference* 96, 325–350.
- Garel, B. (2005). Asymptotic theory of the likelihood ratio test for the identification of a mixture. *Journal of Statistical Planning and Inference* 131, 271–196.
- Garel, B. (2007). Recent asymptotic results in testing for mixtures. *Computational Statistics & Data Analysis* 51, 5295–5304.

- Ghosh, J. K. and P. K. Sen (1985). On the asymptotic performance of the log-likelihood ratio statistic for the mixture model and related results. In L. LeCam and R. A. Olshen (Eds.), *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, Volume 2, pp. 789–806. Wadsworth, Monterey, CA.
- Goodwin, T. H. (1993). Business-cycle analysis with a markov-switching model. *Journal of Business & Economic Statistics* 11(3), 331–339.
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57, 357–384.
- Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64(2), 413–430.
- Hartigan, J. A. (1985). Statistical theory in clustering. *Journal of Classification* 2.
- Hathaway, R. J. (1985). A constrained formulation of maximum-likelihood estimation for normal mixture distribution. *The Annals of Statistics* 18(2), 795–800.
- Jedidi, K., H. S. Jagpal, and W. S. Desarbo (1997). Finite mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science* 16, 35–59.
- Jeffries, N. O. (2003). A note on testing the number of components in a normal mixture. *Biometrika* 90, 991–994.
- Kim, C.-J. and C. R. Nelson (1998). Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime switching. *The Review of Economics and Statistics* 80(2), 188–201.
- Kim, C.-J. and C. R. Nelson (1999). Has the U.S. economy become more stable? a bayesian approach based on a markov-switching model of the business cycle. *The Review of Economics and Statistics* 81(4), 608–616.
- King, M. L. and J. Chen (2010). Testing the order of a finite mixture. *Journal of the American Statistical Association* 105(491), 1084–1092.
- King, M. L. and T. S. Shively (1993). Locally optimal testing when a nuisance parameter is presented only under the laternative. *The Review of Economics and Statistics* 75(1), 1–7.
- Lanne, M. (2006). Nonlinear dynamics of interest rate and inflation. *Journal of Applied Econometrics* 21(8), 1157–1168.
- Liesenfeld, R. (1998). Dynamic bivariate mixture models: modeling the behavior of prices and trading volume. *Journal of Business and Economic Statistics* 16, 101–109.

- Liesenfeld, R. (2001). A generalized bivariate mixture model for stock price volatility and trading volume. *Journal of Econometrics* 104, 141–178.
- Liu, H. and Y. Shao (2004). Asymptotics for the likelihood ratio test in a two-component normal mixture model. *Journal of Statistical Planning and Inference* 123, 61–81.
- Lo, Y., N. R. Mendell, and D. B. Rubin (2001). Testing the number of components in a normal mixture. *Biometrika* 88(3), 767–778.
- McLachlan, G. J. (1987). On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Applied Statistics* 36, 318–324.
- McLachlan, G. J. and D. Peel (2000). *Finite Mixture Models*. New York: Wiley.
- Mendell, N. R., H. C. Thode, and S. J. Finch (1991). The likelihood ratio test for the two-component normal mixture problem: Power and sample size analysis. *Biometrics* 47, 1143–1148.
- Perez-Quiros, G. and A. Timmermann (2001). Business cycle asymmetries in stock returns: Evidence from higher order moments and conditional densities. *Journal of Econometrics* 103, 259–306.
- Redner, R. A. and H. F. Walker (1984). Mixture densities, maximum likelihood and the EM algorithm. *Society for Industrial and Applied Mathematics* 26, 195–239.
- Sims, C. A. and T. Zha (2006). Were there regime switches in U.S. monetary policy? *American Economic Review* 96, 54–81.
- SMith, A., P. A. Naik, and C.-L. Tsai (2006). Markov-switching model selection using kullbackleibler divergence. *Journal of Econometrics* 134, 553–577.
- Wolfe, J. H. (1971). A monte carlo study of the sampling distribution of the likelihood ratio for mixtures of multinormal distributions. technical Bulletin STB 72-2, San Diego, US.
- Wong, C. S. and W. K. Li (2001). On a logistic mixture autoregressive model. *Biometrika* 88(3), 883–846.
- Zhang, Z., W. K. Li, and K. C. Yuen (2006). On a mixture garch time-series model. *Journal of Time Series Analysis* 27(4), 577–597.