ORIGINAL PAPER

A likelihood ratio test for bimodality in two-component mixtures with application to regional income distribution in the EU

Hajo Holzmann · Sebastian Vollmer

Abstract We propose a parametric test for bimodality based on the likelihood principle by using two-component mixtures. The test uses explicit characterizations of the modal structure of such mixtures in terms of their parameters. Examples include the univariate and multivariate normal distributions and the von Mises distribution. We present the asymptotic distribution of the proposed test and analyze its finite sample performance in a simulation study. To illustrate our method, we use mixtures to investigate the modal structure of the cross-sectional distribution of per capita log GDP across EU regions from 1977 to 1993. Although these mixtures clearly have two components over the whole time period, the resulting distributions evolve from bimodality toward unimodality at the end of the 1970s.

1 Introduction

Analyzing the modality of a random sample's distribution is an important problem, especially for proper graphical visualization of the data. In particular, we must decide whether modes which are present in a certain fit are merely sampling artifacts or whether they are actual features of the underlying density.

H. Holzmann (✉)
Institut für Stochastik, Universität Karlsruhe, Englerstr. 2, 76128 Karlsruhe, Germany
e-mail: holzmann@stoch.uni-karlsruhe.de

S. Vollmer
Ibero-Amerika Institut für ökonomische Forschung und Institut für Statistik und Ökonometrie,
Universität Göttingen, Platz der Göttinger Sieben 3, 37073 Göttingen, Germany

Most testing procedures for multimodality suggested in the literature are nonparametric in nature. The arguably most popular method, which is based on kernel estimates with the normal kernel, was suggested by Silverman (1981). He observed that for fixed observations the number of modes in such an estimate is a monotonically decreasing function of the bandwidth. Using this fact Silverman (1981) defined the k-critical bandwidth h_k as the minimal bandwidth for which the kernel estimate still just has k modes. If h_k exceeds a critical value, which is constructed from a bootstrap procedure, then the hypothesis for k modes of the underlying density is rejected. See also Mammen et al. (1992), Fisher et al. (1994) and Hall and York (2001). A test for unimodality against multimodality—called the dip test—is based on measuring the distance between the empirical distribution function and the class of unimodal distribution functions (Hartigan and Hartigan 1985). Müller and Sawitzki (1991) used the so-called excess mass functional to construct a test for k-modality. For $k = 1$ their test is equivalent to the dip test. See also Fisher and Marron (2001).

The notion of multimodality of a population's distribution is closely related to the notion of population heterogeneity. A popular way to model population heterogeneity parametrically is via mixture models. In particular, the likelihood ratio test for homogeneity in two-component mixtures has been extensively studied in recent years, cf. e.g. Chen et al. (2001). However, mixtures with two distinct components need not be bimodal, and two component mixtures of unimodal component densities can have more than two modes. Therefore, there is no immediate connection between the number of components in a mixture and the number of modes of the resulting density. Nevertheless, the modal structure of two-component mixtures of certain parametric families, notably the normal distribution (Robertson and Fryer 1969) and the von Mises distribution (Mardia and Sutton 1975), is completely known in terms of the parameters of the mixture. For two-component mixtures, for which such an explicit characterization of the modal structure is available, we construct a likelihood ratio (LR) test for unimodality against bimodality. The asymptotic distribution of the LR test for bimodality, though not a standard χ^2-distribution, can be deduced from existing results on the behavior of LR statistics on the boundary of the parameter space, cf. Chernoff (1954) and Self and Liang (1987).

When compared to the nonparametric methods mentioned above, the LR test has certain merits as well as certain limitations. Concerning the advantages, the LR test is more powerful than competing nonparametric methods if the distributional assumptions are satisfied. Furthermore, using von Mises mixtures, the LR test can easily be applied to circular data. Note that for circular data, only a few methods are available, notably the tests by Fisher and Marron (2001) and by Basu and Jammalamadaka (2002). Moreover, using recent results by Ray and Lindsay (2005) on the modal structure of multivariate normal mixtures, it can be extended to the multivariate setting where no methods seem to be available yet. Concerning limitations, the LR test can only test for unimodality against bimodality and not for k against more than k modes, since there are no parametric descriptions for these cases. Furthermore, it loses power if the mixture component densities are not normally distributed but have heavier tails (like the t-distribution).

Section 2 describes the asymptotic distribution of the LR test for bimodality in two-component mixtures and gives two examples. Section 3 investigates the performance of the LR test via a simulation study. As an illustration, Sect. 4—following

Pittau and Zelli ([2005](), [2006]())—analyzes the cross-sectional distribution of per capita log GDP across EU regions via mixtures. After excluding the mere urban areas, it turns out that a two-component mixture model with equal variances for the two components adequately describes the data for all years. We further investigate whether the distribution is actually bimodal, both by using Silverman's test as well as via the LR test for bimodality. Silverman's test can never reject the hypothesis of unimodality. In contrast, for the years 1977–1979, the LRT rejects unimodality with level 5%, while in the following years it can no longer reject this hypothesis with increasing p-values. Thus, while the cross-sectional distribution of per capita log GDP in the EU regions under investigation remains heterogeneous in the sense of being well-modeled by a two-component mixture of normal distributions, these components only significantly result in a bimodal distribution in the years 1977–1979, while in the following years the two components start to merge and form a unimodal distribution.

2 The likelihood ratio test for bimodality

Let $f(x; \theta)$, $\theta \in \Theta \subset \mathbb{R}^d$, $x \in \mathbb{R}^q$, be a parametric family of q-dimensional densities, and consider the two-component mixture family

$$f(x; \theta_1, \theta_2, p) = p f(x; \theta_1) + (1 - p) f(x; \theta_2),$$

where

$$(\theta_1, \theta_2, p) \in \Theta \times \Theta \times [0, 1] = \Theta_{\text{mix}} \subset \mathbb{R}^{2d+1}.$$

In order to allow for possible joint parameters of the component densities (e.g. equal variances), we consider a subset $E_{\text{mix}} \subset \Theta_{\text{mix}}$, where $E_{\text{mix}} \subset \mathbb{R}^q$ for a minimal $q \leq 2d + 1$. Suppose that the mixture density is at most bimodal, so that we can split the set E_{mix} disjointly into $E_{\text{mix}} = E_{\text{unim}} \cup E_{\text{bim}}$, the unimodal part E_{unim} and the bimodal part E_{bim}. We will denote the boundary between E_{bim} and E_{unim} by ∂E_{unim}, i.e. $\partial E_{\text{unim}} = E_{\text{unim}} \cap \overline{E_{\text{bim}}}$, where $\overline{E_{\text{bim}}}$ denotes the closure of E_{bim}. Given observations $X_1, \ldots, X_n$ from the mixture density, we consider the log-likelihood function

$$\mathcal{L}_n(\theta_1, \theta_2, p) = \sum_{k=1}^{n} \log f(X_k; \theta_1, \theta_2, p).$$

Assumption 1 The partial derivatives of $\log f(x; \theta_1, \theta_2, p)$ of order 3 with respect to θ_1, θ_2 and p exist a.s., at least in a neighborhood N of the true value $(\theta_1^0, \theta_2^0, p^0)$.

Assumption 2 For $(\theta_1, \theta_2, p) \in N$, the first- and second-order partial derivatives of $f(x; \theta_1, \theta_2, p)$ are uniformly bounded in absolute value by a function $F(x) \in L_1(\mathbb{R})$, and the third-order partial derivatives of $\log f(x; \theta_1, \theta_2, p)$ are uniformly bounded in absolute value by a function $H(x)$ with $E H(X_1) < \infty$.

Assumption 3 The expectation of the matrix of second-order partial derivatives of $\log f(x; \theta_1, \theta_2, p)$ is finite and positive definite for $(\theta_1, \theta_2, p) \in N$.

Note that Assumption 3 will not be satisfied in a neighborhood of a single-component density (i.e. if $p = 0$ or $p = 1$ or $\theta_1 = \theta_2$, see e.g. Goffinet et al. 1992). However, for unimodal component densities such as normal or von Mises densities (see the following examples), a density on the boundary ∂E_{unim} will be a proper two-component mixture, so that Assumption 3 is satisfied.

Theorem 1 *Suppose that the true parameter vector* $(\theta_1^0, \theta_2^0, p^0)$ *of the mixture density lies on the boundary* ∂E_{unim}, *and that locally around* $(\theta_1^0, \theta_2^0, p^0)$, ∂E_{unim} *is a smooth* $(q-1)$-*dimensional surface in* $\mathbb{R}^q$. *If furthermore Assumptions 1–3 hold true, then we have that*

$$R_n := 2\Big(\sup_{(\theta_1, \theta_2, p) \in E_{mix}} \mathcal{L}_n(\theta_1, \theta_2, p) - \sup_{(\theta_1, \theta_2, p) \in E_{unim}} \mathcal{L}_n(\theta_1, \theta_2, p) \Big)$$

$$\xrightarrow{\mathcal{D}} \big(\chi_0^2 + \chi_1^2\big)/2, \tag{1}$$

where χ_0^2 *is the measure with mass one at* $x = 0$ *and* χ_1^2 *is the chi-square distribution with* 1 *degree of freedom.*

This result follows from the theory of the likelihood ratio test for parameter vectors which lie on the boundary of the parameter space, cf. Chernoff (1954) and Self and Liang (1987). Note that if the true parameter vector lies in the interior of E_{unim}, then due to consistency, the unrestricted maximum likelihood estimator will asymptotically lie in a neighborhood U of θ_0 with $U \subset E_{unim}$, so that $R_n \to 0$ in probability. Therefore the test will also asymptotically keep the level in this case.

Example 1 (Normal distribution) For the normal density $f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{(-\frac{(x-\mu)^2}{2\sigma^2})}$ we obtain the general two-component mixture as follows:

$$f(x; p, \mu_1, \mu_2, \sigma_1, \sigma_2) = p f(x; \mu_1, \sigma_1) + (1 - p) f(x; \mu_2, \sigma_2), \tag{2}$$

where $0 \le p \le 1$ and without loss of generality, $\mu_2 \ge \mu_1$. Introducing the parameters $r = \sigma_1/\sigma_2$ and $d = (\mu_2 - \mu_1)/(2\sqrt{\sigma_1 \sigma_2})$, one easily sees that $f(x; p, \mu_1, \mu_2, \sigma_1, \sigma_2)$ has the same number of modes as $f(x; p, -d, d, r, 1/r)$, cf. Behboodian (1970). Thus, bimodality solely depends on the three parameters (p, d, r). Let us mention that such an argument can be made for general location-scale families (e.g. also for the t-distribution with fixed degrees of freedom). In case of equal variances, $\sigma_1 = \sigma_2 = \sigma$, one obtains $r = 1$, and the conditions read as follows: $f(x; p, \mu_1, \mu_2, \sigma)$ is unimodal if and only if $d \le 1$ or if

$$d > 1 \quad \text{and} \quad \big|\log(1 - p) - \log p\big| \ge 2\log\big(d - \sqrt{d^2 - 1}\big) + 2d\sqrt{d^2 - 1}, \tag{3}$$

otherwise, it is bimodal. In Fig. 1, the region of bimodality is depicted, where for $r \ne 1$ we used the characterization in Robertson and Fryer (1969).

If we assume that the variances $\sigma_1 = \sigma_2 = \sigma$ are equal (though possibly unknown) and that p is known and fixed, then the smoothness assumption on the boundary of the unimodal parameter domain is satisfied everywhere, and thus Theorem 1 holds true.

Fig. 1 Regions of bimodality of the normal distribution, where $d = (\mu_2 - \mu_1)/(2\sqrt{\sigma_1\sigma_2})$ and $r = \sigma_1/\sigma_2$. On the *right side* of the curves are those parameter constellations of p and d for which, given fixed r, the resulting mixture is bimodal
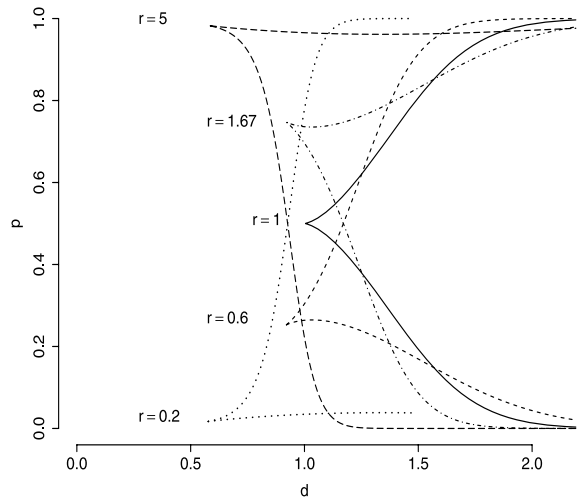
This is obvious, e.g., for the case $p = 1/2$, in which case the mixture is unimodal with mode at $(\mu_1 + \mu_2)/2$ if and only if $\mu_2 - \mu_1 \leq 2\sigma$. For other values of p it follows from (3). However, in case of variable p and equal variances, the boundary has a singularity for $d = 1$ and $p = 1/2$, which follows from (3) by taking equality there and the limit $d \to 1$.

If $d = 1$ and $p = 1/2$, the likelihood ratio statistic will be asymptotically stochastically smaller than the limit distribution in (1). This is because the angle between the tangents to the bimodal region at these points is less than π, so that the unrestricted ML estimator will in more than 50% of all cases fall into the unimodal region, and the LR statistic will be zero. See also Fig. 1. In summary, a test based on the critical value of the $1/2\,(\chi_0^2 + \chi_1^2)$ distribution will asymptotically keep the level everywhere in the unimodal parameter space. Extensions to the characterization of the number of modes of higher dimensional normal distributions were obtained by Ray and Lindsay (2005).

Example 2 (Von Mises distribution) The von Mises distribution is given by the density

$$f(x; \mu, \kappa) = \frac{1}{2\pi\, I_0(\kappa)} \exp\big(\kappa \cos(x - \mu)\big), \quad 0 \leq x < 2\pi,$$

where $\mu \in [0, 2\pi)$, $\kappa > 0$ and $I_0(\kappa)$ is a norming factor given by the modified Bessel function of the second kind. The two-component mixture of von Mises distributions will be denoted by

$$f(x; p, \mu_1, \mu_2, \kappa_1, \kappa_2) = pf(x; \mu_1, \kappa_1) + (1 - p)f(x; \mu_2, \kappa_2),$$

where w.l.o.g. $\mu_2 - \mu_1 =: d \in [0, \pi]$ (since the maximal distance on the circle of two points along the arc is π). Mardia and Sutton (1975) gave precise conditions for bimodality for general parameters constellations. Here we review only their results

for the case of equal concentration parameters, i.e. $\kappa_1 = \kappa_2 = \kappa$. In this case, the mixture is unimodal if and only if either $d = 0$ or

- $d = \pi$ and $0 \le p \le (1 + \exp(2\kappa))^{-1}$ or $(1 + \exp(-2\kappa))^{-1} \le p \le 1$ or
- $0 < d < \pi$ and either
  - $\sin d < 2\kappa \sin^3(d/2)$, $0 \le p \le 1$, or
  - $\sin d \ge 2\kappa \sin^3(d/2)$ and $0 \le p \le -t(\delta)/(1 - t(\delta))$ or $1/(1 - t(\delta)) \le p \le 1$, where δ is the solution of

$$2\kappa \cos^3 \delta - \kappa(1 + \cos d)\cos \delta - 2\cos(d/2) = 0, \quad 0 < \delta < d/2, \qquad (4)$$

  and

$$t(\delta) = -\frac{\sin(d/2 + \delta)}{\sin(d/2 - \delta)} \exp\left(-2\kappa \sin(d/2)\sin \delta\right)$$

The conditions for von Mises mixtures are more complicated than those for normal mixtures, since it is not a simple location-scale family (this notion is not defined for circular distributions). Still, although one has to distinguish several cases, these cases merge continuously. For example, for $\sin d = 2\kappa \sin^3(d/2)$, the mixture is still unimodal for all $0 \le p \le 1$. Furthermore, for $d \to \pi$, one has that $-t(\delta)/(1 - t(\delta)) \to (1 + \exp(2\kappa))^{-1}$ (and similarly on the other side). In Fig. 2, the region of bimodality is displayed.

For fixed p, Theorem 1 is again generally applicable. However, for p variable and equal concentration parameters, there occurs a singularity on the boundary of the set of unimodal parameter constellations for $p = 0.5$ if $\sin d = 2\kappa \sin^3(d/2)$. Again, the test will nevertheless asymptotically keep the critical value.

Fig. 2 Regions of bimodality of the von Mises distribution, where $\mu = \mu_2 - \mu_1$ and $\kappa_1 = \kappa_2 = \kappa$. On the *right side* of the curves are those parameter constellations of p and μ for which, given fixed κ, the resulting mixture is bimodal
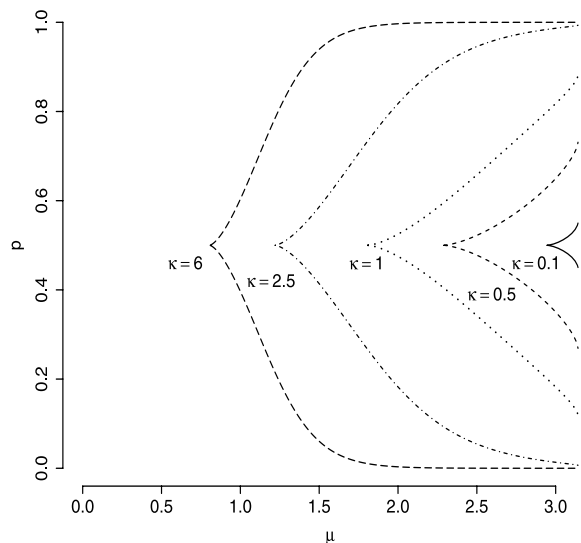
3 Simulations

In this section we conduct a simulation study in order to analyze the practical feasibility of the LR test for bimodality.

First let us investigate the quality of the approximation by the asymptotic distribution as given in Theorem 1. To this end, for certain parameter constellations on the boundary of the unimodal region we simulate the actual level of the test when using asymptotic critical values. Here we use 10^4 samples of various sizes. Furthermore, we employ direct numerical maximization of the log-likelihood function and, for the constrained estimate, we reparametrize the problem in order to use unconstrained maximization. The results for the normal distribution are displayed in Table 1; simulations for the von Mises distribution led to similar results. It turns out that the test keeps the nominal level quite well even for moderate sample sizes, both for normal and von Mises mixtures, as long as either equal variances (or concentration parameters) are employed or if the variances are assumed to be known. However, further simulations indicated that if both variances are allowed to vary, the approximation is rather inaccurate and should not be used.

Now let us investigate the power properties of the LR test for bimodality. For simplicity we restrict ourselves to normal mixtures, and we compare the performance with Silverman's (1981) test and with the dip test by Hartigan and Hartigan (1985). When implementing Silverman's (1981) test we use 1000 bootstrap replications to estimate the critical value for the bandwidth. We also use the R library's "Diptest"

Table 1 Simulated level of R_n on the boundary for normal mixtures, using asymptotic critical values

Scenario	Free parameters	Sample size	Nominal level	Simulated level
$\mu_1 = 0, \mu_2 = 3$	μ_1, μ_2, p	100	0.10	0.12
$\sigma_1 = \sigma_2 = \sigma = 1.3$			0.05	0.059
$p = 0.442$			0.01	0.015
		250	0.10	0.11
			0.05	0.057
			0.01	0.012
	μ_1, μ_2, p, σ	250	0.10	0.14
			0.05	0.072
			0.01	0.016
		500	0.10	0.13
			0.05	0.066
			0.01	0.014
$\mu_1 = 0, \mu_2 = 2.5$	μ_1, μ_2, p	100	0.10	0.16
$\sigma_1 = 1.1, \sigma_2 = 0.8$			0.05	0.086
			0.01	0.019
		250	0.10	0.10
			0.05	0.052
			0.01	0.011

Fig. 3 Density of normal mixture $f_1(x) = f(x, 0.5, -1.5, 1.5, 1, 1)$ (*solid line*), together with unrestricted fit to sample of size 500 from f_1 (*dashed line*) and restricted unimodal fit (*dotted line*)
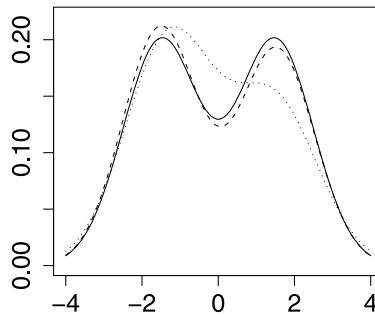

Table 2 Power under the first alternative, the normal mixture $f_1 = f(x, 0.5, -1.5, 1.5, 1, 1)$

Sample size	Nominal level	LR	Silverman's	Dip
200	0.10	0.89	0.77	0.30
	0.05	0.80	0.63	0.20
	0.01	0.53	0.35	0.06
500	0.10	0.99	0.97	0.60
	0.05	0.98	0.92	0.47
	0.01	0.93	0.75	0.24

for the dip test by Hartigan and Hartigan (1985). We consider several alternative scenarios. In each scenario 1000 samples of various sizes are generated.

(a) First alternative: a normal mixture $f_1(x) = f(x, 0.5, -1.5, 1.5, 1, 1)$. The density is symmetric and clearly bimodal, cf. Fig. 3. where also the unrestricted ML fit and the ML fit restricted to the unimodal region are displayed for a sample from f_1. Here we use only equal variances, and σ is allowed to vary. The LR test performs slightly superior to Silverman's test, and both outperform the dip test. See Table 2 for the simulation results.

(b) Second alternative: a normal mixture $f_2(x) = f(x, 0.3, -1.5, 1, 0.75, 0.75)$. The density is asymmetric but bimodal, cf. Fig. 4. Again we use only equal variances for the fit, and σ is allowed to vary. The results are displayed in Table 3. The dip test has no significant power exceeding the level for this hypothesis for sample sizes up to $n = 500$, and the LR test strongly outperforms Silverman's test.

(c) Third alternative: a normal mixture $f_3(x) = f(x, 0.6, -1.5, 1.4, 0.6, 1.4)$. The density is asymmetric but bimodal, cf. Fig. 4. Here we use distinct variances which are assumed to be known. Only the LR test has a reasonable power against this alternative, and therefore we do not present a table for this simulation setting. In fact, the comparison is not really fair since the LR test uses the exact values of the variances, which are hardly available in practice.

(d) We also investigated the behavior of the LR test for bimodality if the distributional assumption of normal mixtures is not satisfied, and briefly report the results. We simulated 1000 samples of sizes 200 and 500 from a mixture of two t-distributions with five degrees of freedom. One component has location parameter 0 and the other 3. Both have a unit scaling parameter, and a weight of

Fig. 4 Densities of normal mixtures $f_2(x) = f(x, 0.3, -1.5, 1, 0.75, 0.75)$ (*solid line*) and $f_3(x) = f(x, 0.6, -1.5, 1.4, 0.6, 1.4)$ (*dashed line*)
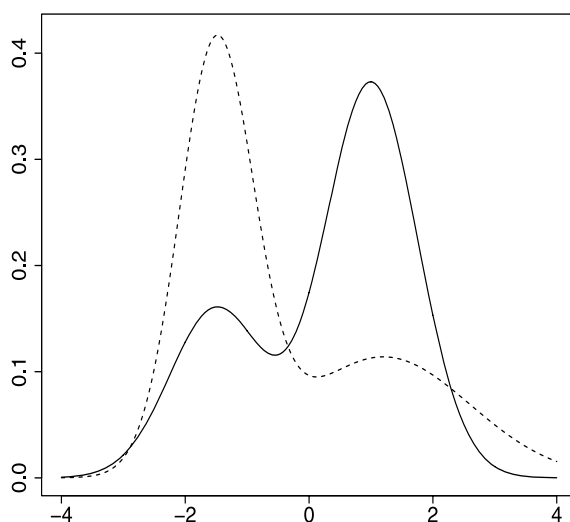

Table 3 Power under the second alternative, the normal mixture $f_2 = f(x, 0.3, -1.5, 1, 0.75, 0.75)$

Sample size	Nominal level	LR	Silverman's
200	0.10	0.80	0.57
	0.05	0.70	0.39
	0.01	0.45	0.11
500	0.10	0.99	0.97
	0.05	0.98	0.92
	0.01	0.93	0.75

$p = 0.4$ for the first component is used. The density is clearly bimodal. However, due to the heavy tails of the t-distribution, the variance in the normal fit is typically too large. Therefore the modes in the normal mixture are much less distinctive than in the t mixture, and the LR test loses power. In fact, the LR test loses a great deal of power in comparison to Silverman's test, while the LR test and the dip test perform similarly Silverman's test.

4 Application to the cross-sectional regional income distribution in the EU

The convergence hypothesis states that poorer economies are growing faster than richer ones and, hence, catching up so that eventually there will be no differences between real average per capita income across countries. This would imply a unimodal cross-national or even cross-regional distribution of income which should become constantly less dispersed. The literature distinguishes between two types of convergence, β-convergence and σ-convergence (Sala-i-Martin 1996). By definition, β-convergence occurs if the coefficient on initial income is negative when regressed

on the change of log real income or, in other words, if initially poorer economies grow on average faster than the initially rich. Moreover, σ-convergence is defined as the decrease of the dispersion of the entire income distribution measured by the standard deviation of log incomes. If there are no other control variables in the growth regression, we speak of absolute β-convergence, which would be a necessary but not a sufficient condition for σ-convergence. Thus, for the convergence hypothesis to hold we need absolute β-convergence and σ-convergence such that the income distribution converges to one common mode.

However, the extended Solow growth model, given its assumptions, only implies a restricted type of conditional β-convergence. Indeed, if two groups of countries are governed by different parameters, but display within-group homogeneity of parameters, it would imply a divergence of the two groups, but a within-group convergence of economies to their respective group steady state. Quah (1997) developed a theoretical and empirical framework for so-called club convergence from the viewpoint of income distribution dynamics, implying an emerging twin peaks phenomenon for the global cross-country income distribution. Bianchi (1997) found empirical evidence for a bimodal cross-country income distribution occurring in the 1970s for all subsequent years. For regions in the European Union, however, this picture is less clear (cf. Quah 1996; Le Gallo 2004).

The framework of EU regions is of special interest, since cohesion among EU regions has been a major priority of all EU treaties so far. The European Development Fund (EDF) has been in operation since the very beginning in 1959. Starting with 3.4 million Euro, it went up to 244.7 million Euro in 1977 and 1353.6 million Euro in 1993. Other relevant policy outcomes are the European Agricultural Guidance and Guarantee Fund (EAGGF), established in 1962, and the European Regional Development Fund (ERDF), created in 1975. The EAGGF started with 28.7 million Euro in 1975, increasing to 6587.1 million Euro in 1977 and 34935.8 million Euro in 1993. The EDF started with 150 million Euro and increased to 400 million Euro in 1977 and 5382.6 million Euro in 1993.[1] Hence, one should expect that policy interventions assimilate the parameters of the extended Solow growth model in the European Union over time, implying absolute convergence in the long run. Furthermore, Barro and Sala-i-Martin (1991) argued that convergence of incomes between regions is in general supported and accelerated by an economic environment without restrictions on the free movement of capital, labor and tradeable goods, which is the case in the European Union.

We use a data set on regional GDP in the European Union available from CRENoS,[2] covering the period from 1977 to 1993 and including administrative regions defined by the Nomenclature of the Territorial Units for Statistics (NUTS) established by Eurostat. The GDP figures are given in 1990 constant prices and are converted to Purchasing Power Standard (PPS). The data set includes regions from all EU-12 member countries at that time. Following Pittau and Zelli (2005, 2006) we use the territorial units as follows: NUTS-0 (countries) for Denmark, Luxembourg and Ireland; NUTS-1 for Belgium (3 Régions), West Germany (11 Länder),

---

[1]See European Commission (2000), The Community Budget: The Facts in Figures.

[2]Center for North South Economic Research, http://www.crenos.it.

Table 4 Model selection criteria for mixtures models fitted to the cross-sectional log-income distribution in of European regions

Comp.	Variances	No. param.	AIC 1977	BIC 1977	AIC 1990	BIC 1990
1	–	2	−131.70	−126.39	−133.96	−128.66
2	equal	4	−160.99	−150.38	− 145.47	−134.85
	distinct	5	−159.56	−146.29	−144.26	−130.97
3	equal	6	−156.99	−141.07	−141.47	−125.55
	distinct	8	−154.09	−132.86	−144.14	−122.91

the Netherlands (4 Landsdelen) and UK (9 Government Office Regions and 3 Countries); NUTS-2 for Italy (20 Regioni), France (22 Régions), Spain (17 Communitades Autonomas), Portugal (5 Comissaoes de Coordenacao Regional), and Greece (13 Development Regions). Though not equally sized, these regions are, due to administrative structure of the different countries, the best units available for comparisons below the national level. Urban regions usually have their own economic structure and are not comparable to regions covering both urban and rural parts. Therefore we decided to exclude the mainly urban regions from our analysis (Brüssel B3, Bremen D4, Hamburg D5, Ile de France F1 and Luxembourg LU).

Pittau and Zelli (2005, 2006) analyzed a similar data set (without the exclusion of the urban regions) by using finite mixtures of normal distributions. For our more homogeneous data set, in a first step we determine the number of components in the mixture as well as the structure of the mixture (in particular equal or unequal variances for the components of the mixture). To this end we use the model selection criteria AIC and BIC. Table 4 shows the results for the years 1977 and 1990. Both model selection criteria select the two-component mixture model with equal variances for the components. This is in fact true for all years from 1977–1993, thus, it is the model of choice for this period.

Pittau and Zelli (2005, 2006) also tested the number of components by using a bootstrap version of the likelihood ratio test. Here, in order to confirm that two components are indeed present in the data, we use the modified likelihood ratio test for homogeneity (cf. Chen et al. 2001). In contrast to the usual LRT for homogeneity, this test retains a comparatively simple limit theory, thus a parametric bootstrap (with the resulting loss in power) is not necessary. We test the hypothesis of a single normal distribution against a two-component mixture with equal (but unknown) variance, and to this end use a version of the modified LRT with a structural parameter as investigated by Chen and Kalbfleisch (2005). They showed that the χ_2^2-distribution is an asymptotic upper bound for the distribution of the modified LRT statistic in this case. Based on this bound we find p-values of less than 0.001 for all years in the period 1977–1993, thus, there is strong statistical evidence of two components in the distribution.

Nevertheless, as discussed in the introduction, this does not necessarily imply that the distribution is bimodal, so that the components are strongly pronounced. Therefore, we test for unimodality against bimodality, both by using the LRT for bimodality in a two-component normal mixture with equal variances, as well as using Silverman's test. The results are displayed in Table 5. While Silverman's test never rejects

Table 5 P-values for tests for unimodality for the distribution of log GDP PPS per capita in European regions

Year	p-val LRT	p-val Silverman
1977	0.006	0.14
1978	0.009	0.23
1979	0.042	0.29
1980	0.105	0.28
1985	0.222	0.98
1990	0.214	0.69
1993	0.347	0.54
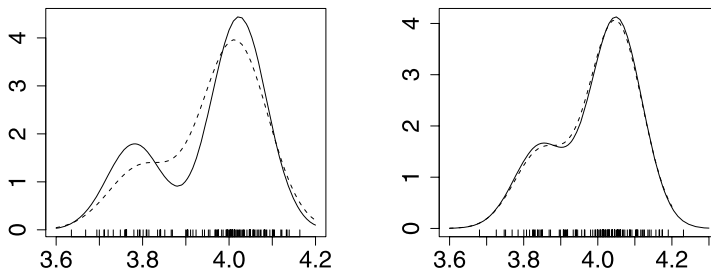
Fig. 5 Unrestricted (*solid lines*) and restricted unimodal (*dashed lines*) fits to the cross-sectional distribution of log GDP PPS per capita for European regions in the years 1977 (*left*) and 1993 (*right*)

the hypothesis of a single mode, the LRT rejects in favor of bimodality in 1977–1979. Afterward, the hypothesis can no longer be rejected, indicating that the two groups, though still present, start to merge. Figure 5 shows the restricted unimodal and the unrestricted fit for the years 1977 and 1993, respectively.

Conclusions can be drawn both from an economical and a statistical point of view. Economically the empirical results indicate that the two-component mixture describing the cross-regional income distribution in the European Union became less and less dispersed, meaning that well-separated clusters of poor and rich regions in the EU moved closer together and might tend to converge to a single group in the long run. However, further research is necessary to evaluate the long-run impact of EU cohesion policy on regional GDP. Of special interest would be an analysis of the distribution dynamics following the more recent EU enlargements. From a statistical point of view, we find that the LRT is able to detect a second mode in a real-data application, while Silverman's test is not able to do so.

References

Barro, R., Sala-i-Martin, X.: Convergence across states and regions. Brookings Pap. Econ. Act. **1991**(1), 107–182 (1991)

Basu, S., Jammalamadaka, S.R.: Unimodality in circular data: a Bayes test. In: Balakrishnan, N., et al. (eds.) Advances on Methodological and Applied Aspects of Probability and Statistics, pp. 141–153. Taylor and Francis, New York (2002)

Behboodian, J.: On the modes of a mixture of two normal distributions. Technometrics **12**, 131–139 (1970)

Bianchi, M.: Testing for convergence: Evidence from non-parametric multimodality tests. J. Appl. Econom. **12**, 393–409 (1997)

Chen, J., Kalbfleisch, J.D.: Modified likelihood ratio test in finite mixture models with a structural parameter. J. Stat. Plan. Inference **129**, 93–107 (2005)

Chen, J., Chen, H., Kalbfleisch, J.: A modified likelihood ratio test for homogeneity in finite mixture models. J. R. Stat. Soc. Ser. B **63**, 19–29 (2001)

Chernoff, H.: On the distribution of the likelihood ratio. Ann. Math. Stat. **25**, 573–578 (1954)

Fisher, N.I., Marron, J.S.: Mode testing via the excess mass estimate. Biometrika **88**, 499–517 (2001)

Fisher, N.I., Mammen, E., Marron, J.S.: Testing for multimodality. Comput. Stat. Data Anal. **18**, 499–512 (1994)

Goffinet, B., Loisel, P., Laurent, B.: Testing in normal mixture models when the proportions are known. Biometrika **79**, 842–846 (1992)

Hall, P., York, M.: On the calibration of Silverman's test for multimodality. Stat. Sin. **11**, 515–536 (2001)

Hartigan, J.A., Hartigan, P.M.: The dip test of unimodality. Ann. Stat. **13**, 70–84 (1985)

Le Gallo, J.: Space-time analysis of GDP disparities among European regions: a Markov chain approach. Int. Reg. Sci. Rev. **27**, 138–163 (2004)

Mammen, E., Marron, J.S., Fisher, N.I.: Some asymptotics for multimodality tests based on kernel density estimates. Probab. Theory Relat. Fields **91**, 115–132 (1992)

Mardia, K.V., Sutton, T.W.: On the modes of a mixture of two von Mises distributions. Biometrika **62**, 699–701 (1975)

Müller, D.W., Sawitzki, G.: Excess mass estimates and tests for multimodality. J. Am. Stat. Assoc. **86**, 738–746 (1991)

Pittau, M.G., Zelli, R.: Fitting regional income distributions in the European Union. Oxf. Bull. Econ. Stat. **67**, 135–161 (2005)

Pittau, M.G., Zelli, R.: Empirical evidence of income dynamics across EU regions. J. Appl. Econom. **21**, 605–628 (2006)

Quah, D.: Regional convergence clusters across Europe. Eur. Econ. Rev. **40**, 1353–1375 (1996)

Quah, D.: Empirics for growth and distribution: polarization, stratification, and convergence clubs. J. Econ. Growth **2**, 25–59 (1997)

Ray, S., Lindsay, B.G.: The topography of multivariate normal mixtures. Ann. Stat. **33**, 2042–2065 (2005)

Robertson, C.A., Fryer, J.G.: Some descriptive properties of normal mixtures. Skand. Aktuarietidskr. **69**, 137–146 (1969)

Sala-i-Martin, X.: The classical approach to convergence analysis. Econ. J. **106**, 1019–1036 (1996)

Self, S.G., Liang, K.-Y.: Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J. Am. Stat. Assoc. **82**, 605–610 (1987)

Silverman, B.W.: Using kernel density estimates to investigate multimodality. J. R. Stat. Soc. Ser. B **43**, 97–99 (1981)