Testing for two states in a hidden Markov model[1]

Jörn Dannemann

Institut für Mathematische Stochastik, Georg-August-Universität Göttingen, Germany.

Hajo Holzmann[2]

Institut für Stochastik, Universität Karlsruhe (TH), Germany.

Abstract

We consider hidden Markov models (HMMs) with finite-valued latent process and state-dependent distributions from a general one-parameter family. A test for $m = 2$ against $m \geq 3$ states of the underlying Markov chain is proposed. So far, no satisfactory methods for this problem are available. Our test is an extension to HMMs of the modified likelihood ratio test (LRT) for two-states in a finite mixture, as introduced by Chen, Chen and Kalbfleisch (J. R. Stat. Soc. Ser. B 66, 2004, 95–115). We develop its asymptotic distribution theory under the null hypothesis of two states, and investigate its finite sample properties in a simulation study. The test is based on inference for the marginal mixture distribution of the HMM. In order to illustrate the additional difficulties due to the dependence structure of the HMM, we also show how to test general regular hypotheses on the marginal mixture of HMMs via a quasi LRT. Two empirical illustrations conclude the paper.

1 Introduction

A hidden Markov model (HMM) is a bivariate process $(X_k, Y_k)_{k \geq 0}$, where $(X_k)_{k \geq 0}$ is an unobservable, finite-state Markov chain and $(Y_k)_{k \geq 0}$ is the observable process with values in a Borel-measurable subset $\mathcal{Y} \subset \mathbb{R}^k$, which are related as follows. Given $(X_k)_{k \geq 0}$, the

---

[1] Preprint submitted to the Canadian Journal of Statistics (6 Febr 2008).

[2] Address for correspondence: Hajo Holzmann, Institut für Stochastik, Englerstraße 2, 76128 Karlsruhe, Germany, Email: Hajo.Holzmann@stoch.uni-karlsruhe.de

$(Y_k)_{k \geq 0}$ are conditionally independent, and for each $j \geq 0$, the conditional distribution of Y_j depends on X_j only. The unobservable Markov chain is also called the regime or the latent process of the HMM. We shall assume that (X_k) is stationary and ergodic with state space $\mathcal{M} = \{1, \ldots, m\}$, so that the stationary distribution $\pi = (\pi_1, \ldots, \pi_m)$ of the associated transition matrix $\alpha_{ab} = P(X_{k+1} = b | X_k = a)$, $a, b \in \mathcal{M}$ is uniquely determined. The conditional distributions of Y_j given X_j are called the state-dependent distributions, we shall assume that they belong to a parametric family $\{f(y; \theta) | \theta \in \Theta\}$ of densities w.r.t. a σ-finite measure ν on $\mathcal{Y}$.

HMMs provide a flexible and very widely used class of models for dependent data, in particular in the presence of overdispersion (for series of count data) or unobserved heterogeneity. For independent data, these phenomena are often modeled by finite mixtures, and indeed, the marginal distribution of an HMM is a finite mixture in the state-dependent distributions, thus HMMs are sometimes also called Markov-dependent mixtures. Areas of applications of HMMs include speech recognition (Rabiner 1989), biological sequence analysis (Durbin *et al.* 1998), the modeling of animal movements (Leroux & Puterman 1992), series of overdispersed count data arising in medicine (Albert 1991, who analyzed series of epileptic seizures) and financial time series (Rydén *et al.* 1998). See the monographs by MacDonald & Zucchini (1997) for further examples of applications, and by Cappé *et al.* (2005) for a state-of-the-art overview of theoretical developments for HMMs. Related model classes include switching autoregressive models (Douc *et al.* 2004), switching regression (Zhang *et al.* 2003) and switching-regime GARCH models (Francq & Zakoïan 2005).

In statistical applications of HMMs, selection of the number of states m of the latent process is a task of major importance. To this end, model selection criteria are often used, which are either based on the full-model log-likelihood (McDonald & Zucchini 1997; Rydén 1995; Gassiat & Boucheron 2003; MacKay 2002), or on reducing the problem to selecting the number of components in the marginal mixture distribution (Poskitt & Zhang 2005). However, hypothesis tests for the number of components are also useful, since in contrast to model selection criteria they allow decisions with a formal significance level, and the likelihood ratio test (LRT) is then a canonical choice. For testing $m = 1$ against $m \geq 2$ for an HMM, Gassiat & Keribin (2000) show that the LRT statistic diverges to ∞. Note that for $m = 1$, the (Y_k) are simply an i.i.d. sequence from $f(y; \theta)$, thus the simplest non-trivial (i.e. dependent) HMM has to have at least two states. Therefore, testing for $m = 2$ versus $m \geq 3$ states for an HMM is the problem of primary practical interest. Short of any asymptotic theory, Rydén *et al.* (1998) used a bootstrap version of the LRT for this problem. However, bootstrapping in this context is computationally extremely demanding, since it requires repeated maximization of the full log-likelihood function of an HMM for more than two states. Further, the results by Gassiat & Keribin (2000) significantly reduce the hope to develop asymptotic distribution theory for the even more difficult problem of testing for $m = 2$ via the LRT.

In this paper, we shall propose a test for $m = 2$ against $m \geq 3$ in an HMM by extending the modified LRT of Chen, Chen and Kalbfleisch (2004) to HMMs. The test has a relatively simple limit theory, and is computationally easy to handle, since it does neither require

2

bootstrapping nor evaluation of the full log-likelihood function of the HMM. In fact, the test is based on quasi likelihood inference for the marginal mixture distribution of the HMM (cf. Lindgren 1987). Therefore, as an illustration in Section 2 we first discuss how to test regular hypotheses on parameters of the marginal mixture distribution of the HMM via a quasi LRT. It turns out that this test statistic is not asymptotically χ^2-distributed in general, but rather requires an adjustment for the dependence structure of the HMM. Surprisingly, as shown in Section 3, the modified LRT for $m = 2$ against $m \geq 3$ in an HMM does not require such an adjustment, the limit distribution is the same as for independent mixtures. This makes its use for HMMs particularly simple and attractive.

Section 4 contains the results of a Monte Carlo study. We first illustrate in Section 4.1 that even for simple regular hypotheses, the distribution of the quasi LRT for the marginal mixture can be quite far from the expected χ^2-distribution. Section 4.2 contains results of extensive simulation experiments for the modified LRT for two components in an HMM, both under the hypothesis of two states as well as under various alternative models, and with distinct forms of the transition matrix. It turns out that as expected from the asymptotic distribution theory, the performance of the modified LRT is hardly influenced by the form of the transition matrix, as long as its stationary distribution remains the same. The test has satisfactory finite-sample properties, both for keeping the nominal level under the hypothesis as well as in terms of power. In Section 5 we give two empirical illustrations, one for the series of fetal lamb movements analyzed in Leroux & Puterman (1992), and the other to the series of log-returns of the S&P 500 (cf. Rydén *et al.* 1998). Some formal assumptions and proofs are given in the Appendix.

2 The LRT under independence assumption

The marginal distribution of the observations (Y_k) is given by the finite mixture

$$f_{\text{mix}}\big(y; \pi_1, \ldots, \pi_m, \theta_1, \ldots, \theta_m\big) = \pi_1 f(y; \theta_1) + \ldots + \pi_m f(y; \theta_m). \tag{1}$$

Here we assume that the parameters $\big(\pi_1, \ldots, \pi_m, \theta_1, \ldots, \theta_m\big)$ of f_{mix} depend on a parameter $\omega \in \Omega \subset \mathbb{R}^p$, and we write $\pi_a(\omega)$ and $\theta_a(\omega)$, where $a \in \mathcal{M}$, and $f_{\text{mix}}(y; \omega) = f_{\text{mix}}(y; \pi_1(\omega), \ldots, \pi_m(\omega), \theta_1(\omega), \ldots, \theta_m(\omega))$. Lindgren (1978) proposed estimation of ω by maximizing the log-likelihood function under independence assumption

$$L_n^I(\omega) = \sum_{k=1}^{n} \log\, f_{\text{mix}}(Y_k; \omega), \tag{2}$$

and this approach can also be used to test hypothesis about ω via a LRT under independence assumption (LRTI).

In this section we briefly discuss the LRTI for regular hypotheses in order to illustrate that its asymptotic distribution is not given by a simple χ^2-distribution but is in general significantly influenced by the dependence structure of the HMM. Note that since π is uniquely determined by the transition matrix $\big(\alpha_{a,b}\big)$, $a, b, \in \mathcal{M}$, hypotheses on ω can in principle

be reformulated into hypotheses on the original parameters of the HMM, and hence be tested by the usual LRT for HMMs (cf. Giudici *et al.* 2000). However, the expression of π in terms of the entries of the transition matrix is highly nonlinear for $m \geq 3$, and thus the ordinary LRT becomes intractable in such situations. Hence, for $m \geq 3$ the LRTI is also an attractive procedure to test hypotheses on the stationary distribution of an HMM, e.g. testing for $\pi_a = \pi_b$ for $a, b \in \mathcal{M}$, or $\pi_1 = \ldots = \pi_m = 1/m$.

We shall assume that finite mixtures of $f(\cdot, \theta)$ are identifiable, as for example finite mixtures of Gaussian, Gamma or Poisson distributions, and that the components of the state dependent distribution θ_a are all distinct, and moreover that the parametrization in ω is such that there exists a unique true ω_0 which corresponds to the true law P_0 of the (Y_k). Let $\hat{\omega}$ be the maximizer of $L_n^I(\omega)$. A slight generalization of the argument of Lindgren (1978) shows that $\hat{\omega}$ is consistent and asymptotically normally distributed. More precisely, one shows that under assumptions similar to the assumption of Theorem 3.4 in Lindgren (1978), for the score under independence assumption we have that

$$\frac{D_\omega L_n^I(\omega)}{\sqrt{n}} \quad \xrightarrow{\mathcal{L}} \quad N(0, \mathrm{Cov}_0),$$

$$\mathrm{Cov}_0 \quad = \quad \Sigma_0 + \sum_{j \geq 2} E\Big(h(Y_1; \omega_0)\, h(Y_j; \omega_0)^T + h(Y_j; \omega_0)\, h(Y_1; \omega_0)^T\Big),$$

where $h(y; \omega) = \big(D_\omega \log f_{\mathrm{mix}}(y; \omega)\big)^T$, and

$$\frac{D_\omega\, D_\omega^T\, L_n^I(\omega)}{n} \to \Sigma_0 \quad \text{a.s.,} \qquad \Sigma_0 = E h(Y_1; \omega_0)\, h(Y_1; \omega_0)^T.$$

Hence, if Σ_0 is nonsingular, the usual argument gives $\sqrt{n}\big(\hat{\omega} - \omega_0\big) \xrightarrow{\mathcal{L}} N\big(0, \Sigma_0^{-1} \mathrm{Cov}_0 \Sigma_0^{-1}\big)$. The matrices Σ_0 and Cov_0 will typically differ, which causes the non-standard behavior of the LRTI.

More precisely, suppose that we want to test a regular r-dimensional restriction

$$H_s: \; s(\omega_0) = 0 \qquad \text{against} \quad K_s: \; s(\omega_0) \neq 0,$$

where $s : \mathbb{R}^p \to \mathbb{R}^r$, $r \leq p$, is a differentiable map with Jacobian $D_\omega s(\omega_0)$ of full rank r at ω_0. Let

$$\lambda_n^I = \frac{\sup_{\omega \in \Omega : s(\omega) = 0} L_n^I(\omega)}{\sup_{\omega \in \Omega} L_n^I(\omega)}$$

be the LRTI statistic. In order to derive the asymptotic distribution of λ_n^I, reparametrize H_s (at least locally around ω_0) as the image of a differentiable mapping $\varphi : \mathbb{R}^{p-r} \supset U \to \mathbb{R}^p$, i.e. $s(\varphi(t)) = 0$, and these are the only solutions locally around ω_0. Let $\Phi_0 = D_t \varphi(t_0)$, where $\varphi(t_0) = \omega_0$. Then, under H_s and non-singularity of Σ_0, by following the argument for the i.i.d. case (cf. e.g. Pruscha 2000, p. 251-256) one can show that

$$-2 \log \lambda_n^I \xrightarrow{\mathcal{L}} Z^T \mathrm{Cov}_0^{1/2}\Big(\Sigma_0^{-1} - \Phi_0(\Phi_0^T \Sigma_0 \Phi_0)^{-1}\Phi_0^T\Big)\mathrm{Cov}_0^{1/2} Z, \qquad (3)$$

where $Z \sim N(0, I_p)$. The quadratic form which occurs as asymptotic distribution in (3) is a linear combination of independent χ_1^2 distributed variables, where the weights are given by the eigenvalues of the matrix $\text{Cov}_0^{1/2}\left(\Sigma_0^{-1} - \Phi_0(\Phi_0^T \Sigma_0 \Phi_0)^{-1}\Phi_0^T\right)\text{Cov}_0^{1/2}$. This matrix is not an orthogonal projection in general since, as indicated above, Σ_0 and Cov_0 differ due to the dependence structure of an HMM. Hence, the asymptotic distribution of the LRTI will in general not be a simple χ^2-distribution.

For an application of (3), these eigenvalues have to be estimated, by first consistently estimating all component matrices Cov_0, Σ_0 and Φ_0 and using the fact that the eigenvalues depend continuously on the entries. Alternatively one can also use a Wald-type statistic as follows. Suppose that Σ_0 and Cov_0 are non-singular, and let Σ_n and $\widehat{\text{Cov}}_n$ be consistent estimates of Σ_0 and Cov_0, respectively. Then, under H_s and non-singularity of Σ_0, one shows by using the δ-method that

$$W_n^I = ns(\hat{\omega})^T \left(D_\omega s(\hat{\omega})\, \Sigma_n^{-1}\widehat{\text{Cov}}_n\Sigma_n^{-1}\left(D_\omega s(\hat{\omega})\right)^T\right)^{-1} s(\hat{\omega}) \xrightarrow{\mathcal{L}} \chi_r^2.$$

3   The modified LRT for two states in an HMM

Now we propose the modified LRT for $m = 2$ against $m \geq 3$ states in an HMM. The test is in fact designed for testing for two components of the marginal mixture distribution of an HMM. However, since we assume that the state-dependent distributions are all distinct, this is equivalent to testing for two states of the underlying latent process. In this section we assume that the parameter θ of the family $\{f(y; \theta) | \theta \in \Theta\}$ is univariate: $\Theta \subset \mathbb{R}$.

For independent mixtures, the LRT for homogeneity (i.e. one against two or more states) and related tests have been intensively investigated in the literature in recent years, cf. e.g. Dacunha-Castelle & Gassiat (1999) or Chen $et\ al.$ (2001). However, as indicated in the introduction, the main testing problem for HMMs is to test for $m = 2$ states, since for a dependent HMM, at least two states are required. Therefore, in this section we show how the modified LRT of Chen $et\ al.$ (2004) for testing for two components in a mixture can be extended to the marginal distribution of an HMM. Surprisingly, it turns out that in contrast to general LRTIs as considered in Section 2, the modified LRT for two components does not require a correction for the dependence structure for an HMM, which makes its use for HMMs particularly attractive. Let

$$\mathfrak{M}_m = \Big\{G(\theta) = \sum_{j=1}^m \pi_j I_{\{\theta_j \leq \theta\}} :\ \theta_1 \leq \ldots \leq \theta_m,\ \sum_{j=1}^m \pi_j = 1, \pi_j > 0\Big\}$$

denote the set of all m-point distributions on Θ, and let $\mathfrak{M} = \cup_{m \geq 2}\mathfrak{M}_m$. For $G \in \mathfrak{M}_m$ with parameters $(\pi_1, \ldots, \pi_m)$ and $(\theta_1, \ldots, \theta_m)$ we let $f_{\text{mix}}(y; G)$ denote the mixing density (1). Further, let G_0 be the true mixing distribution of the marginal distribution. We shall propose a test for

$$H :\ G_0 \in \mathfrak{M}_2 \qquad \text{against} \quad K :\ G_0 \in \mathfrak{M} \setminus \mathfrak{M}_2.$$

5

Throughout we shall assume H, and denote the true two-component mixing distribution of the marginal mixture as $G_0(\theta) = \pi_0 I_{\{\theta_1^0 \leq \theta\}} + (1 - \pi_0) I_{\{\theta_2^0 \leq \theta\}}$, where $(\pi_0, \theta_1^0, \theta_2^0) \in (0, 1) \times \text{Interior}(\Theta)^2$ with $\theta_1^0 < \theta_2^0$. Hence the true marginal density is $f_{\text{mix}}(y; G_0)$.

For each $G(\theta) \in \mathfrak{M}_m$, the modified likelihood function under independence assumption is defined as

$$\tilde{L}_n^{I(m)}(G) = \sum_{k=1}^n \log f_{\text{mix}}(Y_k; G) + C_m \sum_{j=1}^m \log(\pi_j),$$

where $C_m > 0$ is a constant, a suitable choice of which is discussed in Chen *et al.* (2004). The estimate $\hat{G}^{(m)}$, or more explicitly $(\hat{\pi}_1^{(m)}, \ldots, \hat{\pi}_m^{(m)}, \hat{\theta}_1^{(m)}, \ldots, \hat{\theta}_m^{(m)})$, resulting from maximization of $\tilde{L}_n^{I(m)}(\cdot)$, is called modified maximum likelihood estimate under independence assumption. For a suitably large choice of m, the modified LRTI for two components is based on the statistic

$$T_n^{\text{mod}} = 2\big(L_n^{I(m)}(\hat{G}^{(m)}) - L_n^{I(2)}(\hat{G}^{(2)})\big), \tag{4}$$

where $L_n^{I(m)}$ is the ordinary likelihood function under independence assumption as defined in (2).

In order to investigate the asymptotic distribution of T_n^{mod} in the HMM setting, following Chen *et al.* (2004) we introduce the following quantities

$$\begin{aligned}
\Delta_k &= \big(f(Y_k; \theta_1^0) - f(Y_k; \theta_2^0)\big)/f_{\text{mix}}(Y_k; G_0), \\
Z_k'(\theta) &= f'(Y_k; \theta)/f_{\text{mix}}(Y_k; G_0), \\
Z_k''(\theta) &= f''(Y_k; \theta)/f_{\text{mix}}(Y_k; G_0).
\end{aligned}$$

Set

$$b_{1k} = \big(\Delta_k, Z_k'(\theta_1^0), Z_k'(\theta_2^0)\big)^T, \quad b_{2k} = \big(Z_k''(\theta_1^0), Z_k''(\theta_2^0)\big)^T, \quad b_k^T = (b_{1k}^T, b_{2k}^T),$$

and further $\tilde{b}_{2k} = b_{2k} - B_{21}^1 (B_{11}^1)^{-1} b_{1k}$, $\tilde{\mathbf{b}}_2 = \sum_{k=1}^n \tilde{b}_{2k}$ and $\tilde{\Sigma} = B_{22}^1 - B_{21}^1 (B_{11}^1)^{-1} B_{12}^1$, where

$$B^1 = E(b_1 b_1^T) = \left(\begin{array}{c|c} B_{11}^1 & B_{12}^1 \\ \hline B_{21}^1 & B_{22}^1 \end{array}\right), \qquad B_{11}^1 \in \mathbb{R}^{3 \times 3}.$$

Following the arguments in Chen *et al.* (2004), under Assumptions 1 – 5, given in the Appendix, one obtains the following quadratic approximation to T_n^{mod}

$$T_n^{\text{mod}} = \sup_{t_2 \in \mathbb{R}_+^2} \left(2(\tilde{\mathbf{b}}_2^T/\sqrt{n}) t_2 - t_2^T \tilde{\Sigma} t_2\right) + o_p(1).$$

The final step in the argument of Chen *et al.* (2004) (and likelihood ratio tests in general) requires that $\tilde{\Sigma}$ is the covariance matrix in the asymptotic normal distribution of $n^{-1/2} \tilde{\mathbf{b}}_2$. However, due to the dependence structure, under Assumptions 1 – 5 we have that $n^{-1/2} \tilde{\mathbf{b}}_2$ is asymptotically normally distributed with mean zero and covariance matrix

$$\widetilde{\text{Cov}} = \tilde{\Sigma} + \sum_{k=2}^\infty E\big(\tilde{b}_{21} \tilde{b}_{2k}^T + \tilde{b}_{2k} \tilde{b}_{21}^T\big).$$

6

Surprisingly, for the asymptotic distribution of $n^{-1/2}\tilde{\mathbf{b}}_2$ we indeed have $\tilde{\Sigma} = \widetilde{\text{Cov}}$, as stated in the next proposition, the proof of which is given in the appendix.

Proposition 1. *Suppose that Assumptions 1 – 5 hold. Under the hypothesis H of a two-component marginal mixture, we have*

$$E\big(\tilde{b}_{21}\tilde{b}_{2k}^T\big) = E\big(\tilde{b}_{2k}\tilde{b}_{21}^T\big) = 0 \qquad \text{for all } k \geq 2.$$

This is much in contrast to the relation of the matrices Σ_0 and Cov_0 introduced in Section 2, as we shall illustrate in the simulation study in Section 4.1. Proposition 1 implies that T_n^{mod} will have the same limit distribution as for independent mixtures. In particular, analogously to Theorem 2 in Chen *et al.* (2004) we have

Theorem 1. *Suppose that Assumptions 1 – 5 hold and that the true marginal distribution of (Y_k) is a two-component finite mixture. Further assume that m in the definition of T_n^{mod} in (4) satisfies $m \geq m^* := \max\left\{\lfloor 1.5/\pi_1^0 \rfloor, \lfloor 1.5/\pi_2^0 \rfloor, 4\right\}$. Then*

$$T_n^{mod} \xrightarrow{\mathcal{L}} \left(\frac{1}{2} - p\right)\chi_0^2 + \frac{1}{2}\chi_1^2 + p\,\chi_2^2, \tag{5}$$

where $p = \big(\cos^{-1}\rho\big)/(2\pi)$ and ρ is the correlation coefficient in the covariance matrix $\tilde{\Sigma}$.

For an application of Theorem 1, Chen *et al.* (2004) discuss a data-driven choice of m as well as estimation of the correlation coefficient ρ. Since these methods can be applied here without change, we refer to their paper for the details.

4 Simulations

Here we present some of the results of an extensive simulation study of the tests proposed in the two previous sections. For the maximization of the log-likelihood function (under independence assumption) we use direct maximization via a Newton-type algorithm, as advocated by McDonald & Zucchini (1997).

4.1 The LRT under independence assumption

In this section we shall illustrate two aspects about the LRT under independence assumption. First, the difference between Σ_0 and Cov_0 can be quite large and the distribution of λ_n^I can be quite far from a χ^2-distribution, even in a simple setting. Second, we show that (at least in a particular example), the LRTI and the Wald test under independence assumption have little loss in power when compared to the LRT based on full-model MLEs. Thus, ignoring the dependence structure in the test statistic need not result in a significant loss of power.

We start by suggesting estimators for the matrices Σ_0 and Cov_0, where Σ_0 is estimated by

$\Sigma_n = \frac{1}{n} \sum_{k=1}^{n} h(Y_k; \hat{\omega}) h(Y_k; \hat{\omega})^T$, and Cov_0 by

$$\widehat{\text{Cov}}_n = \Sigma_n + \sum_{j=1}^{J} \frac{n-j}{n} \Sigma_{n,j}, \quad \Sigma_{n,j} = \frac{1}{n-j} \sum_{k=1}^{n-j} \left(h(Y_k; \hat{\omega}) h(Y_{k+j}; \hat{\omega})^T + h(Y_{k+j}; \hat{\omega}) h(Y_k; \hat{\omega})^T \right),$$

where J is small compared to n. Typically, the covariances decrease exponentially fast, so a small number for J will suffice. In practice one can simply check for each j whether the entries of $\Sigma_{n,j}$ are small compared to Σ_n.

We simulate from a stationary three-state Poisson HMM, where the means of the state-dependent Poisson distributions are given by $\theta_1 = 1$, $\theta_2 = 5$ and $\theta_3 = 9$, and the transition matrix of the underlying Markov chain is of the form

$$\begin{pmatrix} 1 - \alpha & \alpha & 0 \\ \beta & 1 - \beta - \gamma & \gamma \\ 0 & \delta & 1 - \delta \end{pmatrix}. \tag{6}$$

We examine testing the hypothesis $H : \pi_1 = \pi_3$. Under H, we choose the entries in (6) as $\alpha = 0.4$, $\beta = 0.2$, $\gamma = 0.3$ and $\delta = 0.6$, yielding for the stationary distribution $\pi_1 = \pi_3 = 0.25$, $\pi_2 = 0.5$.

In the following, for simplicity we fix the θ's at their true values, and estimate the parameter $\omega = (\pi_1, \pi_3)$ only. First, we generate estimates of Σ_0 and Cov_0 from a single sample of size 10^6, yielding for $J = 8$

$$\Sigma_n = \begin{pmatrix} 3.56 & 0.16 \\ 0.16 & 2.12 \end{pmatrix}, \qquad \widehat{\text{Cov}}_n = \begin{pmatrix} 8.13 & -1.61 \\ -1.61 & 2.74 \end{pmatrix}, \qquad P_n = \begin{pmatrix} 1.34 & -1.18 \\ -1.18 & 1.03 \end{pmatrix}$$

where $P_n = \widehat{\text{Cov}}_n^{1/2} \left(\Sigma_n^{-1} - \Phi_0 (\Phi_0^T \Sigma_n \Phi_0)^{-1} \Phi_0^T \right) \widehat{\text{Cov}}_n^{1/2}$ is an estimate of the matrix in the quadratic form in (3) (here, Φ_0 does not depend on ω). Thus, the matrices Σ_0 and Cov_0 apparently differ significantly. The matrix P_n is singular, its non-zero eigenvalue is equal to 2.38. Hence, the asymptotic distribution of the LRTI is a scaled χ_1^2–distribution with scaling factor 2.38.

The distribution of the LRTI-Statistic and the Wald-Statistic was investigated for sample size $n = 500$ with $N = 10000$ replications. Figure 1 shows the empirical cumulative distribution functions. In both cases one can hardly visually distinguish between the sample and the asymptotic distribution functions. However, one clearly observes that the distribution of LRTI differs strongly from the standard χ_1^2-distribution.

Finally, we conduct a power comparison between the LRTI, the Wald test under independence assumption and the LRT based on the full model MLEs. We again test the hypothesis $H : \pi_1 = \pi_3$, and the parameters in (6) are taken as $\alpha_1 = \alpha_0 + \kappa$, $\beta_1 = \beta_0 + \kappa$, $\gamma_1 = \gamma_0 - \kappa$ and $\delta_1 = \delta_0 - \kappa$, where $\alpha_0, \ldots, \delta_0$ are chosen as above, and for κ we use $\kappa = 0, 0.05, 0.1, 0.15, 0.25$. For all tests, the asymptotic critical values are employed (in case of the LRTI the critical value is estimated for each sample). The sample size was taken as $n = 500$, and $N = 10000$ samples were used to estimate the power in each setting. The results are displayed in Table 1. It turns out that at least in this specific scenario, there is little loss in power when using the tests based on the MLEI.
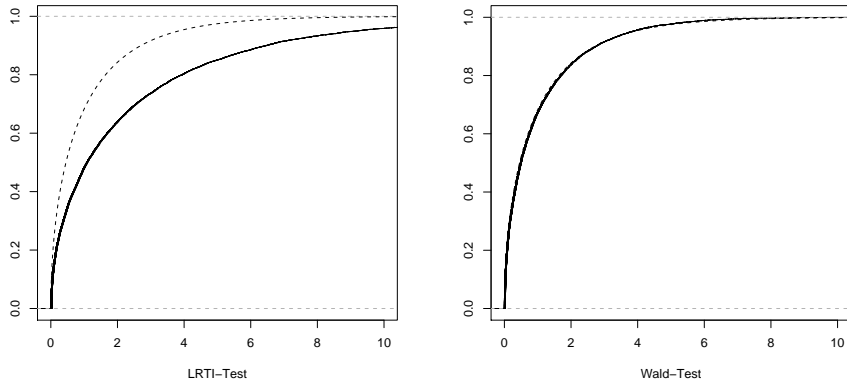
Figure 1: Distribution of the LRTI-Statistic and of the Wald-Statistic (solid), the dotted line (hardly visible) indicates the asymptotic distribution of the LRTI-Statistic and the dashed line the χ_1^2–distribution.

Table 1: Simulated rejection rates of the LRT based on the MLE, LRTI and Wald test based on MLEIs under the hypothesis ($\kappa = 0$) and under the alternative ($\kappa > 0$).

κ	0	0.05	0.1	0.15	0.25
π_1 (true value)	0.25	0.276	0.300	0.323	0.377
π_3 (true value)	0.25	0.226	0.200	0.169	0.078
Power LRT	0.052	0.135	0.381	0.717	0.994
Power LRTI	0.050	0.131	0.373	0.709	0.999
Power W_n^I	0.047	0.121	0.351	0.683	0.997

4.2 The modified LRT for two states

In the following we investigate the finite-sample performance of the modified LRT for $m = 2$ against $m \geq 3$ states as suggested in Section 3. We both consider the standard example of HMMs with state-dependent Poisson distributions, as well as with state-dependent zero-mean Gaussian distributions which are used to model financial times series (cf. Rydén *et al.* 1998; Robert *et al.* 2000).

First, we examine the empirical levels under the hypothesis and consider two-state HMMs with Gaussian state-dependent distributions (N1, N2) and Poisson state-dependent distributions (P1) and five different transition matrices T1 - T5. The specific parameter combinations of N1, N2 and P1 are displayed in Table 2 and transition matrices T1 - T5 are given in Table 3.

To perform the test we need to specify the number of states m for the evaluation of $L_n^{I(m)}(\hat{G}^{(m)})$ and the constants C_2, C_m. Under the hypothesis we choose the minimal $m = m^*$, i.e. $m = 4$ for T1-T4 and $m = 6$ for T5. Under the alternative we always take $m = 4$.

Table 2: Parameter values of the Gaussian state-dependent distributions under the hypothesis (N1, N2) and the alternative (A1, A2) as well as parameter values of the Poisson state-dependent distributions under the hypothesis (P1) and the alternative (A3, A4).

Gaussian ($\mu = 0$)				Poisson		
	σ_1^2	σ_2^2	σ_3^2		θ_1	θ_2	θ_3
N1	1	2.5		P1	3	12	
N2	1	4					
A1	1	2.5	4	A3	3	8	1
A2	1	3	6	A4	3	12	7

Table 3: Transition probabilities for models under the hypothesis.

	α_{12}	α_{21}	π_1
T1	0.50	0.50	0.50
T2	0.25	0.25	0.50
T3	0.75	0.75	0.50
T4	0.80	0.40	0.33
T5	0.90	0.30	0.25

We set $C_2 = C_4 = C_6 = 1$ and choose the starting values as suggested by Chen *et al.* (2004).

Tables 4 - 5 show the simulated rejection rates for sample sizes $n = 200$ and $n = 1000$ for different levels. Note that models N1 and N2 are often used for financial time series analysis where large data sets are available (Rydén *et al.* 1998).

In general, the simulated rejection rates correspond to the specified levels under the hypothesis in a satisfactory manner. Only for sample size $n = 200$ and for N1 and N2, the test is somewhat conservative. The simulations also show better results for N2, where the components differ clearly, than for N1. Note that as expected from the asymptotic theory, the different transition matrices T1-T5 do not seem to have much influence on the results. Indeed, the finite sample behavior for different transition matrices with equal stationary distribution hardly differs, at least as long as transitions are not made too rarely or too frequently (i.e. the diagonal entries are not too close to 0 or 1).

Second, we examine the power of the tests under alternative models. We consider three-state HMMs with Gaussian state-dependent distributions (A1, A2) and Poisson state-dependent distributions (A3, A4) and four different transition matrices T6 - T9, where T6, T7 and T8 are matrices of the form (6) and T9 the corresponding i.i.d. mixture model. The specific parameter combinations of A1 -A4 are displayed in Table 2 and transition probabilities for T6 - T9 are given in Table 6.

Table 4: Simulated rejection rates of the modified LRT for the models under the hypothesis N1, N2 and P1 in Table 2 with transition probabilities T1 - T5 given in Table 3 for sample size $n = 200$ with $N = 10000$ replications.

N1 (Gaussian), n = 200					
Level	T1	T2	T3	T4	T5
0.025	0.010	0.009	0.012	0.008	0.010
0.05	0.021	0.021	0.022	0.016	0.018
0.1	0.045	0.045	0.044	0.034	0.037

N2 (Gaussian), n = 200					
Level	T1	T2	T3	T4	T5
0.025	0.020	0.022	0.019	0.013	0.013
0.05	0.039	0.040	0.039	0.028	0.029
0.1	0.073	0.074	0.072	0.061	0.058

P1 (Poisson), n = 200					
Level	T1	T2	T3	T4	T5
0.025	0.032	0.032	0.030	0.031	0.032
0.05	0.056	0.056	0.056	0.054	0.060
0.1	0.101	0.101	0.098	0.098	0.109

The results for sample sizes $n = 200$ and $n = 500$ and additionally for $n = 1000$ for A1 and A2 are displayed in Tables 7 - 8.

Generally speaking, the simulations show that one should expect only a slight loss of power when introducing dependence. In fact, the influence of the different transition matrices on the resulting power is small. Only, for models where transitions are sparse as for the models with transition matrix T8 one observes a slight loss of power, as might be expected. Furthermore, one observes that the test is more powerful against A2 than against A1. Similarly, for the Poisson case there is a higher power against A4 than against A3. Note that Poisson-mixtures were also investigated in the simulations by Chen *et al.* (2004), our results are rather close to those obtained in that paper.

Table 5: Simulated rejection rates of the modified LRT for the models under the hypothesis N1, N2 and P1 in Table 2 with transition probabilities T1 - T5 given in Table 3 for sample size $n = 1000$ with $N = 10000$ replications.

N1 (Gaussian), n = 1000					
Level	T1	T2	T3	T4	T5
0.025	0.022	0.023	0.020	0.016	0.018
0.05	0.044	0.044	0.039	0.035	0.034
0.1	0.082	0.080	0.076	0.067	0.070

N2 (Gaussian), n = 1000					
	T1	T2	T3	T4	T5
0.025	0.033	0.031	0.030	0.027	0.027
0.05	0.063	0.057	0.055	0.053	0.054
0.1	0.110	0.104	0.103	0.103	0.107

P1 (Poisson), n = 1000					
Level	T1	T2	T3	T4	T5
0.025	0.030	0.030	0.034	0.034	0.036
0.05	0.055	0.055	0.059	0.061	0.063
0.1	0.096	0.102	0.104	0.111	0.116

Table 6: Transitions probabilities for models under the alternative. The transition matrix is of the form (6).

	α	β	γ	δ	π_1	π_2	π_3
T6	0.60	0.60	0.35	0.70	0.40	0.40	0.20
T7	0.10	0.10	0.20	0.40	0.40	0.40	0.20
T8	0.05	0.05	0.05	0.10	0.40	0.40	0.20
T9		i.i.d.			0.40	0.40	0.20

Table 7: Simulated rejection rates of the modified LRT for the models under the alternative A1 - A4 in Table 2 with transition probabilities T6 - T9 given in Table 6 for sample size $n = 200$ and $n = 500$ with $N = 10000$ replications.

A1 (Gaussian), n = 200					A1 (Gaussian), n = 500			
Level	T6	T7	T8	T9	Level	T6	T7	T8	T9
0.025	0.048	0.049	0.040	0.046	0.025	0.146	0.145	0.137	0.149
0.05	0.090	0.086	0.072	0.083	0.05	0.227	0.219	0.213	0.227
0.1	0.157	0.153	0.129	0.155	0.1	0.343	0.333	0.315	0.342

A2 (Gaussian), n = 200					A2 (Gaussian), n = 500			
Level	T6	T7	T8	T9	Level	T6	T7	T8	T9
0.025	0.214	0.192	0.152	0.216	0.025	0.570	0.552	0.540	0.582
0.05	0.313	0.288	0.234	0.313	0.05	0.684	0.664	0.649	0.695
0.1	0.441	0.413	0.351	0.443	0.1	0.793	0.772	0.759	0.799

A3 (Poisson), n = 200					A3 (Poisson), n = 500			
Level	T6	T7	T8	T9	Level	T6	T7	T8	T9
0.025	0.327	0.293	0.239	0.326	0.025	0.700	0.680	0.622	0.714
0.05	0.437	0.399	0.334	0.446	0.05	0.791	0.774	0.713	0.807
0.1	0.567	0.529	0.451	0.573	0.1	0.878	0.857	0.804	0.881

A4 (Poisson), n = 200					A4 (Poisson), n = 500			
Level	T6	T7	T8	T9	Level	T6	T7	T8	T9
0.025	0.247	0.233	0.216	0.231	0.025	0.532	0.529	0.506	0.531
0.05	0.348	0.329	0.305	0.330	0.05	0.646	0.640	0.609	0.646
0.1	0.476	0.463	0.427	0.466	0.1	0.761	0.754	0.722	0.769

Table 8: Simulated rejection rates of the modified LRT for the models under the alternative A1 - A4 in Table 2 with transition probabilities T6 - T9 given in Table 6 for sample size $n = 1000$ with $N = 10000$ replications.

A1 (Gaussian), n = 1000				
Level	T6	T7	T8	T9
0.025	0.313	0.310	0.304	0.307
0.05	0.422	0.419	0.410	0.417
0.1	0.556	0.550	0.546	0.551

A2 (Gaussian), n = 1000				
Level	T6	T7	T8	T9
0.025	0.882	0.884	0.868	0.884
0.05	0.929	0.930	0.916	0.933
0.1	0.965	0.965	0.957	0.966

5 Empirical illustrations

5.1 Fetal lamb movements

As a first illustration, let us revisit the fetal movement data set which is analyzed in Leroux & Puterman (1992) and Chen *et al.* (2004). Leroux & Puterman (1992) fit both two- and three component independent Poisson mixtures as well as two- and three state Poisson HMMs. They find for these data that while independent mixtures are only marginally better than a negative binomial model, the fits provided by the HMMs are much superior and should be used. In fact, there is strong evidence for autocorrelation in these data (cf. Figure 2). For a two-state Poisson HMM, ordinary maximum likelihood yields the following estimates: $\hat{\alpha}_{12} = 0.011$, $\hat{\alpha}_{21} = 0.310$, $\hat{\theta}_1 = 0.256$ and $\hat{\theta}_2 = 3.115$. Assuming $m = 2$, an ordinary likelihood ratio test rejects the hypothesis of independence, i.e. $H: \alpha_{12} = 1 - \alpha_{21}$ with a p-value nearly zero. The comparison of the autocorrelation functions of the sample and the two-state Poisson HMM with parameters $\left(\hat{\alpha}_{12}, \hat{\alpha}_{21}, \hat{\theta}_1, \hat{\theta}_2\right)$ displayed in Figure 2 indicates that a two-state Poisson HMM is an appropriate model for the given data.

However, using formal model selection criteria one cannot decide between the two-state HMM (selected by BIC) and the three-state HMM (selected by the AIC). Using the modified LRT for two components in independent mixtures, Chen *et al.* (2004) test the hypothesis of two components which, yielding a p-value of 0.085, cannot be rejected. From Theorem 1, it follows that their analysis remains valid for the marginal mixture distribution even if the model of choice is an HMM.
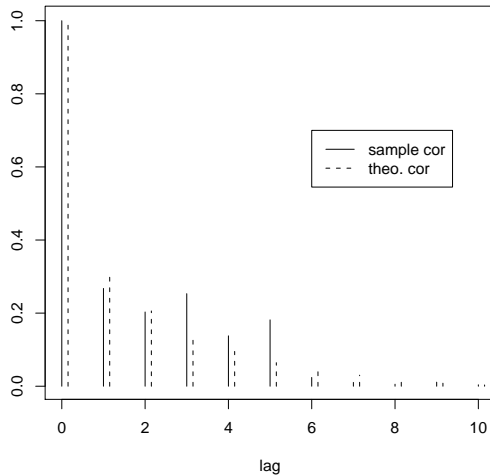
Figure 2: Autocorrelation function for the series of fetal lamb movements.

5.2 Series of log-returns of the S&P 500

Rydén *et al.* (1998) use HMMs with zero-mean Gaussian state-dependent distributions
to analyze the series of log-returns of daily values of the S&P 500 index (formerly called
S&P 90). Specifically, they consider the series of log-returns of ten subseries of length 1700
of the S&P 90/500 from 3 January 1928 to 30 April 1991. We shall examine the same
ten subseries A, ..., J, with outlier replacement and centering of each subseries being
conducted as in their paper.

In order to determine the number of states of the HMM, Rydén *et al.* (1998) use an M- out-
of N ($M = 800$) bootstrap for the full-model LRT for two against three components. This
procedure requires repeated maximization of the full log-likelihood function of an HMM
with three states. We found this procedure extremely computationally expensive, since
proper maximization also require the choice of several starting value combinations, and
we were not able to investigate the properties in an adequate simulation. In fact, in their
analysis Rydén *et al.* (1998) only used very small bootstrap samples for the distribution
of the LRT of size 50, and rejected the hypothesis if the LRT statistic from the first M
observations of the sample exceeded 48 (or more) values of the bootstrap distribution.
Also, the choice of M in the M- out-of N bootstrap is a somewhat subjective manner, and
may (at least in practice) significantly influence the results.

Therefore, we apply the modified LRT for two against more states to this data set, where
we use $m = m^*$ and set $C_m = 1$ for all m. As illustration, we present the estimates of the
fitted models $\hat{G}^{(2)}$ and $\hat{G}^{(m)}$ for the subseries H ($m = 4$), I ($m = 4$) and J ($m = 5$) in Table
9.

One observes that for H and I the mixture distribution $\hat{G}^{(m)}$ exhibits only three different
components, while for series J four distinct components are present. The values of the

Table 9: Estimates $\hat{G}^{(2)}$ and $\hat{G}^{(m)}$ for the subseries H, I and J of the series of log-returns of the S&P 500 index, each of length 1700.

	$\hat{\pi}_1$	$\hat{\sigma}_1$	$\hat{\sigma}_2$
H	0.679	0.0064	0.0125
I	0.562	0.0062	0.0115
J	0.704	0.0063	0.0154

	$\hat{\pi}_1$	$\hat{\pi}_2$	$\hat{\pi}_3$	$\hat{\pi}_4$	$\hat{\pi}_5$	$\hat{\sigma}_1$	$\hat{\sigma}_2$	$\hat{\sigma}_3$	$\hat{\sigma}_4$	$\hat{\sigma}_5$
H	0.181	0.308	0.308	0.203		0.0043	0.0077	0.0077	0.0136	
I	0.232	0.287	0.287	0.193		0.0049	0.0084	0.0084	0.0131	
J	0.173	0.242	0.252	0.252	0.081	0.0032	0.0063	0.0101	0.0101	0.0210

LRT statistic, the estimated $\hat{p}$ in the limit distribution in (5) as well as the p-values of the test are displayed in Table 10.

Table 10: Test results of the hypothesis $m = 2$ for the subseries H, I and J of the series of log-returns of the S&P 500 index, each of length 1700.

	LRT	$\hat{p}$	p-value
H	2.68	0.09	0.074
I	2.16	0.08	0.099
J	21.72	0.12	0.000

While for the series J, the hypothesis of two states can be rejected at a level of $\alpha < 0.001$, for series H and I rejection is only possible at a nominal level of 0.1. Note that from the simulations in Section 4.2 we may expect that the test is somewhat conservative in such settings, so that a test decision on a nominal level of 0.1 appears to be reasonable.

6 Discussion

Inference for the parameters of the marginal mixture distribution of an HMM can be conducted, apart from maximum likelihood estimation in the full model, by using the likelihood function under independence assumption (Lindgren 1978)

$$L_n^I(\pi_1, \ldots, \pi_{m-1}, \theta_1, \ldots, \theta_m) = \prod_{i=1}^{m} \big(\pi_1 f(Y_i, \theta_1) + \ldots + \pi_m f(Y_i, \theta_m)\big).$$

Generally speaking, compared to full model log-likelihood inference one expects this strategy to result in a gain of simplicity, accompanied by a loss of efficiency. In Sections 2 and

16

4.1 we showed that when testing regular restrictions on the parameters of the marginal mixture via a LRT based on L_n^I (LRTI), the loss of efficiency can be quite small, however, the gain in simplicity is also limited by the fact that the distribution of LRTI is not simply a χ^2-distribution, but requires an adjustment for the dependence structure of the HMM. Model selection criteria for choosing the number of components in a finite mixture can also be extended via L_n^I to choosing the number of states in an HMM (cf. Poskitt & Zhang 2005). However, methods for model selection based on the full model likelihood are also available (Gassiat & Boucheron 2003), and the gain in simplicity is rather marginal since these full model based criteria only require a single maximization of the log-likelihood function of the HMM.

When testing for the number of states of an HMM, the situation is quite different. Regarding full-model likelihood inference, it turns out that the asymptotic distribution of the LRT is inaccessible (Gassiat & Keribin 2000), while a bootstrap test for $m = 2$ against $m \geq 3$ states (Rydén *et al.* 1998) is computationally extremely expensive so that proper bootstrap sample sizes cannot be realized. In contrast, we showed in this paper how to extend the modified LRT for $m = 2$ against $m \geq 3$ in a finite mixture by Chen *et al.* (2004) to HMMs, using a modified version of L_n^I. Quite surprisingly, in contrast to the ordinary LRT based on L_n^I, the modified LRT for two states in an HMM does not require any adjustment for its dependence structure, the asymptotic distribution under the hypothesis is the same as for independent mixtures. Judging the loss of efficiency of the modified LRT for HMMs is hard since it does not have any feasible full-model likelihood based (or in fact, any) competitors. A simulation study concerning its power properties turned out to be quite promising.

We believe that our approach to reduce the testing problem for the number of states in latent models with Markov-dependent regime to independent regimes provides a general principle which can be extended to other latent models such as switching autoregression (Douc *et al.* 2004).

Appendix

Assumption 1. The parameter space $\Theta \subset \mathbb{R}$ is compact. Let

$$E\left(|\log f_{\text{mix}}(Y; G_0)|\right) < \infty.$$

and there exists $\varepsilon > 0$ such that, for each G, $f_{\text{mix}}(y; G, \varepsilon) := 1 + \sup_{|Q-G| \leq \varepsilon} f_{\text{mix}}(y; Q)$ is measurable and

$$E\left(\log f_{\text{mix}}(Y; G, \varepsilon)\right) < \infty.$$

Assumption 2. The support of $f(y; \theta)$ does not depend on θ and $f(y; \theta)$ is three times continuously differentiable w.r.t. $\theta \subset \Theta$. The derivatives are jointly continuous in x and θ.

Assumption 3. The family $\{f(y; \theta) \mid \theta \in \Theta\}$ is strong identifiable, i.e. for $\theta_1 \neq \theta_2$

$$\sum_{j=1}^{2} \left(a_j f(y; \theta_j) + b_j f'(y; \theta_j) + c_j f''(y; \theta_j)\right) = 0$$

for all y implies $a_j + b_j + c_j = 0$ for $j = 1, 2$.

Assumption 4. There exists a $\varepsilon > 0$ such that for $j = 1, 2$

$$E\left(\sup_{\theta \in \Theta} \left| \frac{f(Y_1; \theta) - f(Y_1; \theta_j^0)}{f_{\text{mix}}(Y_1; G_0)} \right|^{4+\varepsilon} \right) < \infty;$$

and for $i = 1, 2, 3$

$$E\left(\sup_{\theta \in \Theta} \left| \frac{\frac{d^i}{d\theta^i} f(Y_1; \theta)}{f_{\text{mix}}(Y_1; G_0)} \right|^3 \right) < \infty.$$

Assumption 5. The processes

$$n^{-1/2} \sum_k \frac{f(Y_k; \theta) - f(Y_k; \theta_j^0)}{f_{\text{mix}}(Y_k; G_0)}$$

for $j = 1, 2$ and

$$n^{-1/2} \sum_k \frac{\frac{d^i}{d\theta^i} f(Y_k; \theta)}{f_{\text{mix}}(Y_k; G_0)}$$

and for $i = 1, 2, 3$ are tight.

Proof of Proposition 1. Set $B^k = E\left(b_1 b_k^T\right)$, and partition B^k into

$$B^k = \left(\begin{array}{c|c} B_{11}^k & B_{12}^k \\ \hline B_{21}^k & B_{22}^k \end{array} \right), \qquad B_{11}^k \in \mathbb{R}^{3 \times 3}.$$

Let

$$\lambda_j = E\left(b_1 | X_1 = j\right) = \int b_1(y) f(y; \theta_j^0) \, dy \in \mathbb{R}^5.$$

From $E[b_1] = 0$ it easily follows that $\lambda_2 = c_1 \lambda_1$, where $c_1 = -\frac{\alpha_{21}}{\alpha_{12}} \neq 0$. Using this and $E(\Delta_1 b_1) = E(b_1 | X_1 = 1) - E(b_1 | X_1 = 2)$ we arrive at

$$B^1 \bar{1} = \lambda_1 - \lambda_2 = (1 - c_1)\lambda_1 \tag{7}$$

where $\bar{1} = (1, 0, 0, 0, 0)^T$. Further, using $\lambda_2 = c_1 \lambda_1$ and $E\left(b_1 b_k^T | X_1, X_k\right) = E\left(b_1 | X_1\right) E\left(b_k | X_k\right)^T$ one shows that

$$B^k = E\left(b_1 b_k^T\right) = c_k \lambda_1 \lambda_1^T, \qquad k \geq 2, \tag{8}$$

where $c_k = \frac{\alpha_{21}}{\alpha_{12}} \left(1 - \alpha_{12}^{(k-1)} - \alpha_{21}^{(k-1)}\right)$, and $\alpha_{ij}^{(k)} = P(X_{k+1} = j | X_1 = i)$ denotes the k-step transition probability. Note, that $c_k = 0$ for all k if and only if $a_{12} + a_{21} = 1$, which leads

to independence of the (Y_k). Furthermore, note that (8) implies the symmetry of B^k. In order to show $\tilde{B}^k = 0$ for $k \geq 2$, we compute

$$
\begin{aligned}
\tilde{B}^k &= E\left(\tilde{b}_{21}\tilde{b}_{2k}^T\right) \\
&= B_{22}^k - B_{21}^k\left(B_{11}^1\right)^{-1}B_{12}^1 - B_{21}^1\left(B_{11}^1\right)^{-1}B_{12}^k + B_{21}^1\left(B_{11}^1\right)^{-1}B_{11}^k\left(B_{11}^1\right)^{-1}B_{12}^1
\end{aligned}
$$

To establish our claim, we show that all four summands in this expansion coincide. From (7),

$$
B_{11}^1\,\bar{1} = (1-c_1)\left(\lambda_{11},\lambda_{12},\lambda_{13}\right)^T \quad \Longrightarrow \quad \left(B_{11}^1\right)^{-1}\left(\lambda_{11},\lambda_{12},\lambda_{13}\right)^T = \frac{1}{1-c_1}\,\bar{1},
$$

where $\bar{1} = (1,0,0)^T$ and λ_{1i} denotes the ith component of λ_1. Using this, (7) and (8) give

$$
\begin{aligned}
B_{21}^1\left(B_{11}^1\right)^{-1}B_{12}^k &= c_k B_{21}^1\left(B_{11}^1\right)^{-1}\left(\lambda_{11},\lambda_{12},\lambda_{13}\right)^T\left(\lambda_{14},\lambda_{15}\right) \\
&= \frac{c_k}{1-c_1}B_{21}^1\,\bar{1}\left(\lambda_{14},\lambda_{15}\right) = c_k\left(\lambda_{14},\lambda_{15}\right)^T\left(\lambda_{14},\lambda_{15}\right) = B_{22}^k.
\end{aligned}
$$

Since B^1 and B^k are symmetric, one also has $B_{21}^k\left(B_{11}^1\right)^{-1}B_{12}^1 = B_{22}^k$. The same argument applies to the last matrix

$$
\begin{aligned}
&B_{21}^1\left(B_{11}^1\right)^{-1}B_{11}^k\left(B_{11}^1\right)^{-1}B_{12}^1 \\
&= c_k B_{21}^1\left(B_{11}^1\right)^{-1}\left(\lambda_{11},\lambda_{12},\lambda_{13}\right)^T\left(\lambda_{11},\lambda_{12},\lambda_{13}\right)\left(B_{11}^1\right)^{-1}B_{12}^1 \\
&= \frac{c_k}{(1-c_1)^2}B_{21}^1\,\bar{1}\,\bar{1}^T B_{12}^1 = c_k\left(\lambda_{14},\lambda_{15}\right)^T\left(\lambda_{14},\lambda_{15}\right) = B_{22}^k.
\end{aligned}
$$

This concludes the proof. $\qquad\square$

7 References

P. S. Albert (1991). A two-state Markov mixture model for time series of epileptic seizure counts. *Biometrics* **47**, 1371–1381.

O. Cappé, E. Moulines and T. Rydén (2005). *Inference in hidden Markov models.* New York: Springer.

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

H. Chen, J. Chen and J. D. Kalbfleisch (2004). Testing for a finite mixture model with two components. *J. R. Stat. Soc. Ser. B*, **66**, 95–115.

D. Dacunha-Castelle and E. Gassiat (1999). Testing the order of a model using locally conic parametrization: population mixtures and stationary ARMA processes. *Ann. Statist.* **27**, 1178–1209

R. Durbin, S. R. Eddy, A. Krogh and G. Mitchison (1998). *Biological Sequence Analysis:*

Probabilistic Models of Proteins and Nucleic Acids. Cambridge: Cambridge University Press.

R. Douc, É. Moulines and T. Rydén (2004). Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *Ann. Statist.*, **32**, 2254–2304.

C. Francq, J.-M. Zakoïan (2005). The L^2-structures of standard and switching-regime GARCH models. *Stochastic Process. Appl.* **115**, 1557–1582.

E. Gassiat and S. Boucheron (2003). Optimal error exponents in hidden Markov models order estimation. *IEEE Trans. Inform. Theory* **49**, 964–980.

E. Gassiat and C. Keribin (2000). The likelihood ratio test for the number of components in a mixture with Markov regime. *ESAIM, Probab. Stat.* **4**, 25-52 (2000).

P. Giudici, T. Rydén and P. Vandekerkhove (2000). Likelihood-ratio tests for hidden Markov models, *Biometrics*, **56**(3), 742–747.

B. G. Leroux and M. L. Puterman (1992). Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models. *Biometrics*, **48**, 545–558.

G. Lindgren (1978). Markov regime models for mixed distributions and switching regressions. *Scand. J. Statist.*, **5**, 81–91.

I. L. MacDonald and W. Zucchini (1997). *Hidden Markov and other models for discrete-valued time series*, London: Chapman & Hall.

R. J. MacKay (2002). Estimating the order of a hidden Markov model. *Canad. J. Statist.* **30**, 573–589.

D. S. Poskitt and J. Zhang (2005). Estimating components in finite mixtures and hidden Markov models. *Aust. N. Z. J. Stat.* **47**, 269–286.

H. Pruscha (2000). *Vorlesungen über mathematische Statistik*, Stuttgart: B. G. Teubner. L. R. Rabiner (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, **77**, 257–286.

C. P. Robert, T. Rydén and D. M. Titterington (2000). Bayesian inference in hidden Markov models through the reversible jump Markov chain Monte Carlo method. *J. R. Stat. Soc. Ser. B* **62**, 57–75.

T. Rydén (1995). Estimating the order of hidden Markov models. *Statistics* **26**, 345–354.

T. Rydén, T. Terasvirta and S. Asbrink (1998). Stylized facts of daily return series and the hidden Markov model, *J. Appl. Econometrics*, **13**, 217–244.

H. Zhang, R. Feng and H. Zhu (2003). A latent variable model of segregation analysis for ordinal traits. *J. Amer. Statist. Assoc.* **98**, 1023–1034.