

Testing for two components in a switching regression model

Jörn Dannemann

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

Hajo Holzmann

Fakultät für Mathematik und Informatik, Phillips-Universität Marburg, Germany.

Abstract

We consider switching regression models with independent or Markov-dependent regime. Based on the modified likelihood ratio test (LRT) statistic by Chen, Chen and Kalbfleisch (2004, JRSSB) we propose a test for two against more states of the underlying regime, and derive its asymptotic distribution in case when there is a single switching parameter. We show that its asymptotic distribution is robust when the regime is no longer independent but rather Markov-dependent. In a simulation study we investigate the finite-sample behavior of the test. Finally, we apply the methodology to data of a dental health trial. Here, the model selection criteria AIC and BIC favor distinct binomial regression models with switching intercept (AIC three states, BIC two states). The modified LRT allows us to reject the hypothesis of two states in favor of three states.

Key words: decayed, missing and filled teeth index, hypothesis testing, logistic regression, switching regression, Poisson regression, Markov regime

Running head: mixture regression models

1 Introduction

Mixture models and their extensions are extensively used for describing populations with unobserved heterogeneity, for comprehensive treatments see Frühwirth-Schnatter (2006); Böhning (1999); McLachlan and Peel (2000). Often, in addition to population heterogeneity covariates should be taken into account, in which case one speaks of switching regression models. For a Gaussian response, i.e. linear switching regression, these were introduced by Quandt and Ramsey (1978); Kiefer (1978). Further, switching regression models are also extensively used for count data, in particular Poisson switching regression as e.g. in Le et al. (1992); Wang and Puterman (2001, 1999), or for binary or binomial responses, i.e. switching logistic regression as e.g. in Wang and Puterman (1998). These models allow to incorporate overdispersion relative to the corresponding GLM, and can often be nicely interpreted.

In this paper we shall propose a penalized likelihood ratio test (LRTs) for two against more states in switching regression models with independent and also with Markov-dependent

regime. The proposed methodology is applied to the data on dental health analyzed in Böhning et al. (1999) and Skrandal and Rabe-Hesketh (2004) in order to decide between a switching binomial regression model with either two-states (favored by the BIC) or three states (favored by the AIC). Our proposed test will allow us to reject the hypothesis of two states in favor of three states at a 5% level.

Following Zhu and Zhang (2004), we shall introduce switching regression models in a longitudinal setup. Suppose we observe data from n units, and within each unit i we observe n_i measurements. More precisely, we observe $\mathbf{Y}_i = (Y_{i,1}, \dots, Y_{i,n_i})^T$, and covariates $\mathbf{X}_{i,j} \in \mathbb{R}^q$ for observation $Y_{i,j}$. Write $\mathbf{X}_i = (\mathbf{X}_{i,1}^T, \dots, \mathbf{X}_{i,n_i}^T)^T$. First suppose that for different i , the observations are independent, and that $f_i(\mathbf{y}_i, \mathbf{x}_i; \beta, \theta) \cdot g(\mathbf{x}_i)$ is a parametric family of densities for $(\mathbf{Y}_i, \mathbf{X}_i)$, where the parameters $\beta \in \mathcal{B} \subset \mathbb{R}^p$, $\theta \in \Theta \subset \mathbb{R}^l$, do not depend on i , and $\mathcal{B}$ and Θ are assumed to compact.

Adopting the notation of Chen, Chen and Kalbfleisch (2004) let

$$\mathfrak{M}_m = \left\{ G(\theta) = \sum_{k=1}^m \pi_k I_{\{\theta_k \leq \theta\}} : \theta_1 \leq \dots \leq \theta_m, \sum_{k=1}^m \pi_k = 1, \pi_k > 0 \right\}$$

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 $(\pi_1, \dots, \pi_m)$ and $(\theta_1, \dots, \theta_m)$ we let $f_{i, \text{switch}}(\mathbf{y}_i, \mathbf{x}_i; \beta, G)$ denote the m -density of the observations $(\mathbf{Y}_i, \mathbf{X}_i)$

$$f_{i, \text{switch}}(\mathbf{y}_i, \mathbf{x}_i; \beta, G) = (\pi_1 f_i(\mathbf{y}_i, \mathbf{x}_i; \beta, \theta_1) + \dots + \pi_m f_i(\mathbf{y}_i, \mathbf{x}_i; \beta, \theta_m)) g(\mathbf{x}_i). \quad (1)$$

The joint density of $(\mathbf{Y}_1, \mathbf{X}_1, \dots, \mathbf{Y}_n, \mathbf{X}_n)$ is then the corresponding product of the densities $f_{i, \text{switch}}$. For $G \in \mathfrak{M}_2$ this model corresponds to the model considered by Zhu and Zhang (2004, Eq.3).

Example 1 (*switching logistic regression*). Let U_i be independent copies of a latent variable with values in Θ and distribution function G . If $(Y_{i,1}, \mathbf{X}_{i,1}), \dots, (Y_{i,n_i}, \mathbf{X}_{i,n_i})$ are conditionally independent given U_i and satisfy

$$\text{logit } P(Y_{i,j} = 1 | \mathbf{X}_{i,j} = \mathbf{x}_{i,j}, U_i = \theta_k) = \mathbf{x}_{i,j}^T \beta + \mathbf{w}_{i,j}^T \theta_k,$$

where $\mathbf{w}_{i,j}$ is an l -dimensional vector of covariates, then the model is called a m -component switching logistic regression model. Note that the weights π_k of the components do not depend on the regression parameters, although the model could easily be extended in this direction. For $\mathbf{y}_i = (y_{i,1}, \dots, y_{i,n_i})^T$ we then have that

$$\begin{aligned} f_{i, \text{switch}}(\mathbf{y}_i, \mathbf{x}_i, \beta, G) &= \sum_{k=1}^m \pi_k \left(\prod_{j: y_{i,j}=1} \text{logit}^{-1}(\mathbf{x}_{i,j}^T \beta + \mathbf{w}_{i,j}^T \theta_k) \right. \\ &\quad \left. \prod_{j: y_{i,j}=0} (1 - \text{logit}^{-1}(\mathbf{x}_{i,j}^T \beta + \mathbf{w}_{i,j}^T \theta_k)) \right) g(\mathbf{x}_i). \end{aligned}$$

This model can be extended in a straightforward fashion to switching binomial regression models, in which case we denote the number of successes by $\mathbf{n}$.

Example 2 (*switching Poisson regression*). Again let U_i be independent copies of a latent variable with values in Θ and distribution function G . If $(Y_{i,1}, \mathbf{X}_{i,1}), \dots, (Y_{i,n_i}, \mathbf{X}_{i,n_i})$ are conditionally independent given U_i and satisfy

$$P(Y_{i,j} = y_{i,j} | \mathbf{X}_{i,j} = \mathbf{x}_{i,j}, U_i = \theta_k) = \frac{1}{y_{i,j}!} \lambda_{i,j;k}^{y_{i,j}} \exp(-\lambda_{i,j;k}),$$

where $\lambda_{i,j;k} = \exp(\mathbf{x}_{i,j}^T \beta + \mathbf{w}_{i,j}^T \theta_k)$, then the model is called an switching m component Poisson regression model. For $\mathbf{y}_i = (y_{i,1}, \dots, y_{i,n_i})^T$ we then have that

$$f_{i, \text{switch}}(\mathbf{y}_i, \mathbf{x}_i, \beta, G) = \sum_{k=1}^m \pi_k \left(\prod_{j=1}^{n_i} \frac{1}{y_{i,j}!} \lambda_{i,j;k}^{y_{i,j}} \exp(-\lambda_{i,j;k}) \right) g(\mathbf{x}_i)$$

Example 3 (*Linear switching regression*). Here,

$$Y_{i,j} = \mathbf{x}_{i,j}^T \beta + \mathbf{w}_{i,j}^T U_i + \epsilon_{i,j},$$

where the $\epsilon_{i,j}$ are independently distributed with $E\epsilon_{i,j} = 0$ and $\text{Var } \epsilon_{i,j} = \sigma^2 < \infty$. A special case is $\epsilon_{i,j} \sim N(0, \sigma^2)$, alternatively, one could also use t -distributed errors (with more than 3 degrees of freedom).

The prevalent method of estimation in switching regression models is maximum likelihood, as discussed e.g. in Zhu and Zhang (2004). Concerning identifiability of the parameters, Hennig (2000) gives interesting results for the linear case in which he shows that identifiability of finite mixtures of the error distribution together with a full-rank design matrix does not suffice in general to achieve identification. The area of application of these models is very wide and includes epileptic seizure modeling (Wang and Puterman, 2001), evolutionary biology (Wang and Puterman, 1998), economics (Quandt and Ramsey, 1978; Kiefer, 1978), Segregation Analysis (Zhang et al., 2003).

Model selection for switching regression models has been developed in two directions. First, the appropriate number of components of the latent distribution has to be determined. Second, the relevant covariates, possibly distinct sets for the distinct components, have to be chosen. These two selection problems can be addressed by model selection methods, either separately as in Khalili and Chen (2007) for a method addressing the first problem and in Chen and Khalili (2008) for the second, or jointly as in Naik et al. (2007).

Alternatively, one may apply formal hypothesis tests. For tests concerning the coefficients of the covariates, simple likelihood ratio tests (LRTs) can be applied. However, when testing for the number of components via the LRT, one encounters the same difficulties as for i.i.d. mixtures, the LRT is asymptotically distributed as the square of the supremum of a truncated Gaussian process (cf. Zhu and Zhang, 2004). Following Chen et al. (2001) for i.i.d. mixtures, Zhu and Zhang (2004) developed asymptotics for a modified LRT in case of testing one against two states for switching regressions.

Sometimes, when population heterogeneity is evident or has already been established by appropriate tests, it is further of interest to test whether the latent distribution only has two states or more than two states. Indeed, two states often correspond to contrasts (e.g. high and low volatility of a financial market, bull and bear market of an economy, low and high wind speed etc.), whereas more than two states express a whole range of possibilities. Therefore, following Chen et al. (2004) for i.i.d. mixtures in this paper we propose a penalized likelihood ratio test for two against more states in switching regression models.

If the observations have an additional times series structure in i , this can be incorporated by a Markov-dependent regime. For such models to be dependent, the underlying Markov chain has to have at least two states. Thus, it is also a natural problem to test for two against more states in switching regression models with Markov dependent regime. Following Dannemann and Holzmann (2008), we shall show that the proposed testing method can also be applied in this case without change, the asymptotic distribution remains the same.

We apply the test to the dental data set studied in Böhning et al. (1999). These consist of discrete data with covariates. Böhning et al. (1999) use zero-inflated Poisson regression to model these data. In Skrondal and Rabe-Hesketh (2004), a variety of Poisson and also binomial regression models with switching intercept are fitted to the data. It turns out that the best model in terms of AIC is the switching binomial regression model with three states, whereas in terms of the BIC it is the switching binomial with two states. Our test for two against more states allows to clearly reject the two-states model in favor of three states.

The structure of the paper is as follows. In Section 2 we propose the tests for two against more states, both for independent as well as for Markov dependent regime. Section 3 contains a simulation study. In Section 4 we apply the test to the dental data set.

Outlines of the proofs are deferred to an appendix.

2 Testing for two components

We shall suppose that $l = 1$, i.e. the dimension of the switching parameter θ is one. Often, this will be the intercept. For more than a single switching parameter, the asymptotic distribution becomes intractable, even for testing one against two states (cf. Zhu and Zhang (2004)), and bootstrap methods have to be applied.

2.1 Independent regime

Suppose that for different i , the observations $(\mathbf{Y}_i, \mathbf{X}_i)$ are independent. Let G_0 be the true switching distribution. Following Chen et al. (2004) we shall propose a test for

$$H : G_0 \in \mathfrak{M}_2 \quad \text{against} \quad K : G_0 \in \mathfrak{M} \setminus \mathfrak{M}_2. \quad (2)$$

Throughout we shall assume H . We indicate the true parameter value with a subindex “0” and denote the two-component switching distribution $G_0(\theta) = \pi_0 I_{\{\theta_1^0 \leq \theta\}} + (1 - \pi_0) I_{\{\theta_2^0 \leq \theta\}}$, 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 density of the true model is $f_{i, \text{switch}}(\mathbf{y}, \mathbf{x}; \beta_0, G_0)$.

For each $G(\theta) \in \mathfrak{M}_m$, the modified likelihood function as proposed by Chen et al. (2004) is defined as

$$\tilde{L}_n^{(m)}(\beta, G) = \sum_{i=1}^n \log f_{i, \text{switch}}(\mathbf{Y}_i, \mathbf{X}_i; \beta, G) + C_m \sum_{k=1}^m \log(\pi_k), \quad (3)$$

where $C_m > 0$ is a constant, a suitable choice of which is discussed in Chen et al. (2004). The estimate $(\hat{\beta}^{(m)}, \hat{G}^{(m)})$, or more explicitly

$$(\hat{\beta}^{(m)}, \hat{\pi}_1^{(m)}, \dots, \hat{\pi}_m^{(m)}, \hat{\theta}_1^{(m)}, \dots, \hat{\theta}_m^{(m)}),$$

resulting from maximization of $\tilde{L}_n^{(m)}(\cdot)$, is called modified maximum likelihood estimate. For a suitably large choice of m , the modified LRT for two components is based on the statistic

$$T_n^{\text{mod}} = 2(L_n^{(m)}(\hat{\beta}^{(m)}, \hat{G}^{(m)}) - L_n^{(2)}(\hat{\beta}^{(2)}, \hat{G}^{(2)})), \quad (4)$$

where $L_n^{(m)}$ is the ordinary likelihood function, i.e.

$$L_n^{(m)}(\beta, G) := \sum_{i=1}^n \log f_{i, \text{switch}}(\mathbf{Y}_i, \mathbf{X}_i; \beta, G). \quad (5)$$

Theorem 1. *Suppose that Assumptions 1 - 5 hold and that m in the definition of T_n^{mod} in (4) satisfies*

$$m \geq m^* := \max([1.5/\pi_1^0], [1.5/\pi_2^0], 4). \quad (6)$$

Then under H , the modified likelihood ratio test statistic T_n^{mod} converges in distribution to a mixture of χ^2 -distributions,

$$T_n^{\text{mod}} \xrightarrow{\mathcal{L}} \left(\frac{1}{2} - p\right) \chi_0^2 + \frac{1}{2} \chi_1^2 + p \chi_2^2, \quad (7)$$

where $p = (\cos^{-1} \rho)/(2\pi)$ and ρ is the correlation coefficient in the covariance matrix $\tilde{B}_{22}$ as defined below in (8).

In order to define $\tilde{B}_{22}$, we introduce several quantities which will also be used in the proof of Theorem 1. Set

$$\begin{aligned} \Delta_i(\beta) &= (f_i(\mathbf{Y}_i, \mathbf{X}_i; \beta, \theta_1^0) - f_i(\mathbf{Y}_i, \mathbf{X}_i; \beta, \theta_2^0)) / f_{i, \text{switch}}(\mathbf{Y}_i, \mathbf{X}_i; \beta_0, G_0), \\ Z_i^1(\beta, \theta) &= \frac{d}{d\beta} f_i(\mathbf{Y}_i, \mathbf{X}_i; \beta, \theta) / f_{i, \text{switch}}(\mathbf{Y}_i, \mathbf{X}_i; \beta_0, G_0), \\ Z_i^2(\beta, \theta) &= \frac{d}{d\theta} f_i(\mathbf{Y}_i, \mathbf{X}_i; \beta, \theta) / f_{i, \text{switch}}(\mathbf{Y}_i, \mathbf{X}_i; \beta_0, G_0), \\ Z_i^3(\beta, \theta) &= \frac{d^2}{d\theta^2} f_i(\mathbf{Y}_i, \mathbf{X}_i; \beta, \theta) / f_{i, \text{switch}}(\mathbf{Y}_i, \mathbf{X}_i; \beta_0, G_0). \end{aligned}$$

Further, let

$$b_i = (\Delta(\beta_0), Z_i^1(\beta_0, G_0), Z_i^2(\beta_0, \theta_1^0), Z_i^2(\beta_0, \theta_2^0), Z_i^3(\beta_0, \theta_1^0), Z_i^3(\beta_0, \theta_2^0))^T$$

and $\mathbf{b} = \sum_i b_i \in \mathbb{R}^{p+5}$ with $b_i^T = (b_{1i}^T, b_{2i}^T)$, $\mathbf{b}^T = (\mathbf{b}_1^T, \mathbf{b}_2^T)$ and b_{i2} , $\mathbf{b}_2 \in \mathbb{R}^2$. Let

$$\mathbf{B} = \lim_n \frac{1}{n} \sum_{i=1}^n E[b_i b_i^T] = \left(\begin{array}{c|c} B_{11} & B_{12} \\ \hline B_{21} & B_{22} \end{array} \right), \quad B_{22} \in \mathbb{R}^{2 \times 2}.$$

Finally, set

$$\tilde{B}_{22} = B_{22} - B_{21} (B_{11})^{-1} B_{12}, \quad (8)$$

the covariance of

$$\tilde{\mathbf{b}}_2 = \mathbf{b}_2 - B_{21} (B_{11})^{-1} \mathbf{b}_1. \quad (9)$$

2.2 Testing in case of a Markov dependent regime

An interesting extension of the model is to allow some dependence structure in the underlying latent variables. We shall consider the case of a Markov-dependent regime. Since in this case, a longitudinal set-up is no longer natural, we only consider the non-longitudinal case in the following.

More precisely, let the regime $U_1, \dots, U_n$ be a stationary, ergodic Markov chain on a finite set $\{\theta_1, \dots, \theta_m\} \subset \Theta$ with transition matrix P and stationary distribution G . Suppose that conditioned on the U_i the observations $(\mathbf{X}_i, \mathbf{Y}_i)$ are independent for different i , the distribution of $(\mathbf{X}_i, \mathbf{Y}_i)$ depends on $(U_j)_j$ through U_i only and given $U_i = \theta_k \in \{\theta_1, \dots, \theta_m\}$ the $(\mathbf{X}_i, \mathbf{Y}_i)$ have density $f(\mathbf{y}_i, \mathbf{x}_i; \beta, \theta_k) \cdot g(\mathbf{x}_i)$. The marginal density of the $(\mathbf{X}_i, \mathbf{Y}_i)$ is of the form (1) (but does not depend on i), where the weights $\pi_1, \dots, \pi_m$ are given by G , the stationary distribution of the Markov chain $(U_i)_i$. If the initial distribution of $(U_i)_i$ coincides with G then (X_i, Y_i) is a stationary process. We shall call this model Markov-switching regression model. It can be seen both as an extension of the independent switching regression model as well as of the hidden Markov model (Markov-switching model without covariates).

The testing problem (2) simply translates to testing whether the number of states m equals two,

$$H : G_0 \in \mathfrak{M}_2 \quad \text{against} \quad K : G_0 \in \mathfrak{M} \setminus \mathfrak{M}_2, \quad (10)$$

where G_0 is the stationary distribution of the Markov chain. The likelihood functions (3) and (5) neglect the introduced dependence structure, but they can still be used to estimate the parameters π, β and $\theta_1, \dots, \theta_m$, and to form the test statistic (4) in the present situation. However, one should expect that the asymptotic distribution (7) in Theorem 1 must be modified due to the dependence structure of the regime. Surprisingly, the asymptotic distribution remains the same for the switching regression model with independent regime. A similar effect was observed for the comparable testing problem for Hidden Markov Models (Danne-mann and Holzmann, 2008). To see this, we need to examine the asymptotic behavior of $\tilde{\mathbf{b}}_2$. One needs to show that the asymptotic covariance of $\tilde{\mathbf{b}}_2$ (see (9)) remains the same as for independent switching. First observe

$$\begin{aligned} \lim_{n \rightarrow \infty} \text{Cov}\left(\frac{1}{\sqrt{n}} \tilde{\mathbf{b}}_2\right) &= \lim_{n \rightarrow \infty} \frac{1}{n} E \left[\tilde{\mathbf{b}}_2 \tilde{\mathbf{b}}_2^T \right] \\ &= E \left[\tilde{b}_{21} \tilde{b}_{21}^T \right] + \sum_{i=2}^{\infty} E \left[\tilde{b}_{21} \tilde{b}_{2i}^T + \tilde{b}_{2i} \tilde{b}_{21}^T \right], \end{aligned} \quad (11)$$

with $\tilde{b}_{2i} = b_{2i} - B_{21} (B_{11})^{-1} b_{1i}$, where the first equality holds, since $\mathbf{b}$ and hence $\tilde{\mathbf{b}}_2$ has mean zero and the second equality follows by stationarity.

Proposition 1. *Suppose that for a Markov-switching regression model Assumptions 1 - 5 hold true. Then, under the hypothesis H of a two-state Markov regime we have*

$$E \left[\tilde{b}_{21} \tilde{b}_{2i}^T + \tilde{b}_{2i} \tilde{b}_{21}^T \right] = 0 \quad \text{for all } i \geq 2.$$

Proposition 1 and (11) imply that the asymptotic distribution (7) remains true for Markov-switching regression models, where the weight p is determined as in Theorem 1 from the covariance matrix $\tilde{B}_{22}$ given in (8). We state this as a corollary.

Corollary 2. *Suppose that for a Markov-switching regression model Assumptions 1 - 5 hold and that m in the definition of the test statistic T_n^{mod} (in (4)) satisfies (6). Then under the hypothesis H of a two-state Markov regime the asymptotic distribution of T_n^{mod} is as in (7).*

Note that for the more general longitudinal setup Corollary 2 only holds true under additional assumptions on the sequence $(n_i)_i$, e.g. if these are chosen at random according to a bounded, stationary process.

One may also think about relaxing the i.i.d. assumption on the regressors. Our simulations indicate that the effect of dependent regressors, for example if $\mathbf{X}_i$ are univariate and follow an AR(1) process, on the asymptotic covariance matrices and hence on the asymptotic distribution of the test statistic under the hypothesis is small. So we may conclude that the described testing procedure is quite robust against violations of the independence assumptions.

3 Simulations

In this section we examine the finite sample behavior of the testing procedures for both independent and Markov dependent regime.

3.1 Independent regime

We consider two specific switching Poisson regression models and two switching Binomial regression models. To apply the testing procedure we use $C_2 = C_m = 1$ and choose m , the number of components in the alternative, by m^* , where we replace π_k^0 , $k = 1, 2$, by their estimates under the hypothesis.

The switching Poisson regression models are specified by $n_i = 1, q = 2, l = 1$ with covariates $\mathbf{X}_i = (1, X_i)^T$, X_i independent r.v.s uniformly distributed on the unit interval. For the first model P1 we choose $W_i = 1$ which leads to the Poisson regression model with switching intercept:

$$P(Y_i = y_i | X_i = x_i, U_i = \theta_k) = \frac{1}{y_i!} \lambda_{i;k}^{y_i} \exp(-\lambda_{i;k}),$$

with $\lambda_{i;k} = \exp(x_i \beta + \theta_k)$. Choosing $W_i = X_i$ leads to the Poisson regression model with switching regression coefficient (P2) where Poisson intensities is given by $\lambda_{i;k} = \exp(x_i \theta_k + \beta)$. The two Binomial regression models with switching intercept (B1, B2) with $n_i = 1$ are given by

$$P(Y_i = y_i | X_i = x_i, U_i = \theta_k) = \binom{\mathbf{n}}{y_i} p_{i;k}^{y_i} (1 - p_{i;k})^{\mathbf{n} - y_i}$$

with

$$\text{logit } p_{i;k} = x_i^T \beta + \theta_k$$

For both models we specify $\mathbf{n} = 8$ and for B1 we consider the covariates as in P1, i.e., $q = 2, l = 1$, $\mathbf{X}_i = (1, X_i)^T$, X_i independent r.v.s uniformly distributed on the unit interval and $W_i = 1$. For B2 the covariates are categorical variables as in the application we discuss below, i.e., $q = 9, l = 1$, $\mathbf{X}_i = (1, X_i)^T$, $X_i \in \{0, 1\}^8$ eight independent copies of Bernoulli r.v.s with success probability 1/2 and $W_i = 1$. The specific parameter combinations for models of P1, P2, B1, B2 under the hypothesis ($m = 2$) and under the alternative ($m = 3$) are given in Table 1. For B2 β is given by $(-0.8, -0.5, -0.3, -0.4, -0.2, -0.1, 0.1, 0.2)^T$.

Table 1: Parameter values of the switching Poisson regression models and the switching Binomial regression models under the hypothesis (P1.H, P2.H, B1.H, B2.H) and the alternative (P1.A, P2.A, B1.A, B2.A). For the model B2 the parameter β is given in the text.

	β	θ_1	θ_2	θ_3	π_1	π_2	π_3
P1.H	0.5	1	2		0.6	0.4	
P1.A	0.5	0	1	2	0.33	0.33	0.33
P2.H	1	1	2		0.6	0.4	
P2.A	1	0	1	2	0.33	0.33	0.33
B1.H	-2	-2	0		0.6	0.4	
B1.A	-2	-2	0	1	0.33	0.33	0.33
B2.H		-2	0		0.6	0.4	
B2.A		-2	0	1	0.33	0.33	0.33

Table 2: Simulated rejection rates of the modified LRT for the models under the hypothesis P1.H, P2.H, B1.H and B2.H in Table 1 for various sample sizes with $N = 5000$ replications.

P1.H (Poisson)					B1.H (Binomial)				
Level	$n = 50$	100	200	500	Level	$n = 100$	200	500	1000
0.025	0.017	0.018	0.022	0.027	0.025	0.010	0.013	0.020	0.019
0.05	0.031	0.038	0.040	0.047	0.05	0.018	0.028	0.039	0.039
0.1	0.057	0.068	0.073	0.085	0.1	0.040	0.060	0.073	0.081
P2.H (Poisson)					B2.H (Binomial)				
Level	$n = 50$	100	200	500	Level	$n = 100$	200	500	1000
0.025	0.011	0.015	0.023	0.027	0.025	0.011	0.014	0.017	0.016
0.05	0.021	0.030	0.046	0.052	0.05	0.017	0.025	0.034	0.033
0.1	0.044	0.060	0.083	0.091	0.1	0.031	0.048	0.062	0.066

In general, the simulated rejection rates correspond to the specified levels under the hypothesis in a satisfactory manner (see Table 2). For small sample size and for the model B2.H, the test is somewhat conservative. Note, that for small sample sizes the estimation of ρ as correlation coefficient might fail (if the empirical version of B_{11} is not invertible), in this case one may use $p = 0.5$ leading to a conservative test decision.

3.2 Dependent regime

Next we investigate the behavior of the proposed test when the regime is Markov-dependent. As discussed in Section 2.2 if the switching process is a Markov chain rather than an i.i.d. process, the asymptotic behavior of the test statistic remains the same. In addition, we also investigate the case of dependent covariates. Although we have no formal theory, our simulation indicates that also in this case there is no a significant change in the asymptotic behavior of the test statistic.

Table 3: Simulated rejection rates of the modified LRT for the models under the alternative P1.A, P2.A, B1.A and B2.A in Table 1 for various sample sizes with $N = 5000$ replications.

P1.A (Poisson)					B1.A (Binomial)				
Level	$n = 50$	100	200	500	Level	$n = 100$	200	500	1000
0.025	0.113	0.248	0.502	0.909	0.025	0.097	0.208	0.507	0.813
0.05	0.180	0.356	0.622	0.947	0.05	0.161	0.292	0.630	0.883
0.1	0.281	0.485	0.743	0.976	0.1	0.256	0.419	0.749	0.937

P2.A (Poisson)					B2.A (Binomial)				
Level	$n = 50$	100	200	500	Level	$n = 100$	200	500	1000
0.025	0.085	0.220	0.450	0.864	0.025	0.058	0.137	0.394	0.763
0.05	0.144	0.320	0.564	0.917	0.05	0.094	0.202	0.510	0.838
0.1	0.233	0.440	0.686	0.956	0.1	0.141	0.301	0.636	0.908

For our simulations we consider the models P1.H and B1.H (see Table 1) and choose the switching process as a Markov chain with transition matrix and P and stationary distribution π specified as

$$P = \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix} \quad \text{and} \quad \pi = \begin{pmatrix} 0.6 \\ 0.4 \end{pmatrix}.$$

In addition we construct the covariate $\mathbf{X}_i = (1, X_i)^T$ based on an autoregressive process

$$\tilde{X}_i = \rho_0 \tilde{X}_{i-1} + \varepsilon_i$$

with $\varepsilon_i \sim N(0, \sigma^2)$ i.i.d., with $\sigma^2 = 1$. Based on $\tilde{X}_i$ we construct a process with uniform marginals by

$$X_i = \phi^{-1} \left(\tilde{X}_i / \sqrt{\sigma^2 / (1 - \rho_0^2)} \right).$$

The regression models with Markov-switching and independent covariates are denoted by P1.MC.H and B1.MC.H, whereas the models with Markov-switching and dependent covariates (with $\rho_0 = 0.5$) are denoted by P1.MCAR.H and B1.MCAR.H. The results displayed in Table 4 confirm the small effect of the dependency structure on the asymptotic distribution of the test statistic.

4 Application to dental health trial

We discuss an application of the switching regression model to the dental data set analyzed by Böhning et al. (1999). In a dental health trial 797 children were exposed to different treatments for the improvement of their dental health. This was measured by the number of decayed, missing or filled teeth (DMFT - Index). The index provides counting data, which cannot exceed 8 in our case, since only eight molars were under examination in the trial. As covariates there are the six different treatment groups, sex and three ethnic groups.

Table 4: Simulated rejection rates of the modified LRT for the models with dependency structure over time P1.MC.H, P1.MCAR.H, B1.MC.H and B1.MCAR.H under the hypothesis for various sample sizes with $N = 5000$ replications.

P1.MC.H				B1.MC.H			
Level	$n = 100$	200	500	Level	$n = 200$	500	1000
0.025	0.018	0.024	0.024	0.025	0.014	0.018	0.019
0.05	0.032	0.041	0.046	0.05	0.026	0.037	0.040
0.1	0.064	0.074	0.081	0.1	0.059	0.078	0.085

P1.MCAR.H				B1.MCAR.H			
Level	$n = 100$	200	500	Level	$n = 200$	500	1000
0.025	0.023	0.025	0.027	0.025	0.013	0.018	0.027
0.05	0.043	0.047	0.045	0.05	0.027	0.036	0.049
0.1	0.071	0.085	0.079	0.1	0.060	0.070	0.086

Skrondal and Rabe-Hesketh (2004, p. 349) fit several latent Poisson and binomial regression models to the data. They consider zero-inflation, two - and three component switching regression models with switching intercept as well as models with a normally-distributed intercept (normal intercept models).

In the notation of the introduction, we have $n = 797$, $q = 9$, $l = 1$ and $n_i = 1$ for $i = 1, \dots, n$ with covariate $\mathbf{X}_i = (1, X_i)^T$, $X_i \in \{0, 1\}^8$ and $W_i = 1$. The zero-inflated models are defined by two switching components, where one component is a point mass at zero. Further, in the normal intercept models, the latent variable U_i is distributed $N(\mu, \sigma^2)$. Specifically, the density of $Y_i | \mathbf{X}_i = \mathbf{x}_i, U_i = \theta_k$ is, in the Poisson case,

$$f(y_i, \mathbf{x}_i, \beta, \theta_k) = \exp(-\exp(\mathbf{x}_i^T \beta + \theta_k)) \frac{\exp(y_i (\mathbf{x}_i^T \beta + \theta_k))}{y!}$$

and in the binomial case with $\mathbf{n} = 8$,

$$f(y_i, \mathbf{x}_i, \beta, \theta_k) = \binom{8}{y_i} (\text{logit}^{-1}(\mathbf{x}_i^T \beta + \theta_k))^{y_i} (1 - \text{logit}^{-1}(\mathbf{x}_i^T \beta + \theta_k))^{8-y_i}.$$

A point mass at zero arises as a limiting case for $\theta_1 = -\infty$, thus, the zero-inflated model can be thought of as a sub-model of the two-component switching model (with one parameter less).

Table 5 contains the results of the model selection criteria AIC and BIC for the above mentioned latent regression models (we omit the simple Poisson and binomial regression model without latent variable). We note that except for the zero-inflated variant, the models based on the binomial distribution perform better than the corresponding model based on the Poisson distribution. Further, the overall best model in terms of AIC is the two-state binomial model, and in terms of BIC the three-state binomial model. Thus, in this problem model selection criteria give a clear indication of population heterogeneity, but do not allow to decide between a two- and a three component model. Therefore, we test for two against three states

	Poisson				Binomial			
	ZI	m=2	m=3	NI	ZI	m=2	m=3	NI
Log-likelihood	-1410	-1406	-1406	-1433	-1431	-1400	-1397	-1409
AIC	2841	2834	2838	2886	2882	2822	2821	2838
BIC	2887	2886	2899	2932	2929	2874	2882	2885

Table 5: Log-Likelihood, AIC and BIC for the latent regression models. ZI: zero-inflated model, NI: normal intercept model

by using the above methodology. Note that since m^* is at least 4, the test will be asymptotically conservative. Table 6 provides the penalized maximum likelihood estimators for the two- and three-component binomial switching regression model. When fitting a model with four potential components, two components (i.e. parameters of the binomial components) were equal, thus, it reduced to the three-component model.

We use $C_2 = C_4 = 1$ and obtain $T_n^{mod} = 5.71$, which yields a P-value of 1.4%. Thus, the test clearly rejects two components in favor of three components. Finally, in Table 7 we display the observed and expected frequencies under the fitted models (estimated without penalization). We also see here that the three component model provides quite a good fit to the data.

	π_1	π_2	π_3	θ_1	θ_2	θ_3	Log-likelihood	
m=2	0.49	0.51		-2.30	-0.28		-1400.19	
m=3	0.23	0.44	0.33	-3.47	-1.27	-0.01	-1397.33	

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
m=2	-0.42	-0.17	-0.46	-0.28	-0.74	0.17	0.13	-0.14
m=3	-0.41	-0.16	-0.50	-0.38	-0.81	0.16	0.13	-0.18

Table 6: Penalized ML-estimators for the two- and three-component model.

	obs	m=1	ZI	m=2	m=3
1	231	107.21	227.88	221.72	226.84
2	163	230.81	120.50	180.94	171.36
3	140	233.84	174.21	130.26	141.90
4	116	145.04	150.21	113.99	109.83
5	70	59.90	84.21	85.11	77.62
6	55	16.76	31.33	45.26	45.10
7	22	3.08	7.53	16.00	18.86
8	0	0.34	1.06	3.40	4.90
9	0	0.02	0.07	0.33	0.59

Table 7: Observed and expected frequencies under the fitted binomial models.

5 Conclusions

When fitting regression models with regime switches, apart from the relevant covariates one needs to choose the appropriate number of states of the underlying regime. Here, one is particularly interested in identifying a single state (i.e. homogeneity) and two states (corresponding to contrasts as opposed to a whole range of possibilities in case of more than two states). However, distinct model selection criteria such as AIC or BIC often yield distinct optimal models. Therefore, formal hypothesis are also useful since they allow to reject homogeneity or the hypothesis of two states with a specified significance level.

Using penalized likelihood ratio test statistics, one may develop tests with a feasible asymptotic distribution if (as we assume here) a single parameter is allowed to switch. In more general cases, however, these asymptotic distributions quickly become intractable. Therefore, it would be of interest to study in more detail bootstrap methods and their asymptotic properties for such testing problems.

6 Acknowledgements

Jörn Dannemann acknowledges financial support from the Georg Lichtenberg program "Applied Statistics & Empirical Methods". Hajo Holzmann gratefully acknowledges financial support from the Claussen-Simon Stiftung and from the Landesstiftung Baden-Württemberg, "Juniorprofessorenprogramm".

Appendix

Assumption 1. The modified maximum likelihood estimators $\hat{\beta}^{(m)}$ and $\hat{G}^{(m)}$ are consistent, i.e. $\hat{\beta}^{(m)} = \beta_0 + o_p(1)$ and $|\hat{G}_k - G_k| = \sup_{\theta} |\hat{G}_k(\theta) - G_k(\theta)| = o_p(1)$.

Indeed, for the i.i.d. model the standard Wald's integrability conditions ensure consistency (see Chen et al., 2004). In general, Zhu and Zhang (2003) illustrate that consistency basically follows from uniform laws of large numbers on $L_n^{(m)}$.

Assumption 2. The support of each function $f_i(\mathbf{y}_i, \mathbf{x}_i; \beta, \theta)$ does not depend on β, θ , and the derivatives

$$\frac{d^{i_1+i_2}}{d\theta^{i_1}d\beta^{i_2}} f_i(\mathbf{y}_i, \mathbf{x}_i; \beta, \theta)$$

with $i_1 = 1, 2, 3$, $i_2 = 1, 2$ exist and are jointly continuous in (x, y) and (β, θ) .

Assumption 3. For each i the family $\{f_i(\mathbf{y}_i, \mathbf{x}_i; \beta, \theta)\}$ is strong identifiable, in the sense that for $\theta_1 \neq \theta_2$

$$\sum_{j=1}^2 a_j f_i(\mathbf{y}_i, \mathbf{x}_i; \beta, \theta_j) + \sum_{l=1}^p b_{jl} \frac{d}{d\beta_l} f_i(\mathbf{y}_i, \mathbf{x}_i; \beta, \theta_j) + c_j \frac{d}{d\theta} f_i(\mathbf{y}_i, \mathbf{x}_i; \beta, \theta_j) + d_j \frac{d^2}{d\theta^2} f_i(\mathbf{y}_i, \mathbf{x}_i; \beta, \theta_j) = 0$$

for all $(\mathbf{y}_i, \mathbf{x}_i)$ implies $a_j = b_{j1} = \dots = b_{jp} = c_j = d_j = 0$ for $j = 1, 2$.

This assumption implies that the asymptotic covariance matrix $\mathbf{B}$ is positive definite.

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

$$E\left(\sup_{\beta \in \mathcal{B}, \theta \in \Theta} |(f_i(\mathbf{Y}_i, \mathbf{X}_i; \beta, \theta) - f_i(\mathbf{Y}_i, \mathbf{X}_i; \beta_0, \theta_j^0))/f_{i, \text{switch}}(\mathbf{Y}_i, \mathbf{X}_i; \beta_0, G_0)|^{4+\varepsilon}\right) < \infty,$$

and for $i_1 = 1, 2, 3, i_2 = 1, 2$

$$E\left(\sup_{\beta \in \mathcal{B}, \theta \in \Theta} \left|\frac{\frac{d^{i_1+i_2}}{d\theta^{i_1}d\beta^{i_2}} f_i(\mathbf{Y}_i, \mathbf{X}_i; \beta, \theta)}{f_{i, \text{switch}}(\mathbf{Y}_i, \mathbf{X}_i; \beta_0, G_0)}\right|^3\right) < \infty.$$

Assumption 5. The processes

$$n^{-1/2} \sum_{i=1}^n \frac{f_i(\mathbf{Y}_i, \mathbf{X}_i; \beta, \theta) - f_i(\mathbf{Y}_i, \mathbf{X}_i; \beta_0, \theta_j^0)}{f_{i, \text{switch}}(\mathbf{Y}_i, \mathbf{X}_i; \beta_0, G_0)}$$

for $j = 1, 2$ and

$$n^{-1/2} \sum_{i=1}^n \frac{\frac{d^{i_1+i_2}}{d\theta^{i_1}d\beta^{i_2}} f_i(\mathbf{Y}_i, \mathbf{X}_i; \beta, \theta)}{f_{i, \text{switch}}(\mathbf{Y}_i, \mathbf{X}_i; \beta_0, G_0)}$$

and for $i_1 = 1, 2, 3, i_2 = 1, 2$ are tight.

Proof of Theorem 1. We shall use the notation introduced below Theorem 1. As in Chen et al. (2004) we decompose

$$T_n^{\text{mod}} = T_{1n}^{\text{mod}} - T_{0n}^{\text{mod}} = 2(L_n^{(m)}(\hat{\beta}^{(m)}, \hat{G}^{(m)}) - L_n^{(2)}(\beta_0, G_0)) - 2(L_n^{(2)}(\hat{\beta}^{(2)}, \hat{G}^{(2)}) - L_n^{(2)}(\beta_0, G_0))$$

and observe that $T_{1n}^{\text{mod}} = 2 \sum_i \log(1 + \delta_i)$ with

$$\delta_i = (f_{i, \text{switch}}(\mathbf{Y}_i, \mathbf{X}_i; \hat{\beta}^{(m)}, \hat{G}^{(m)}) - f_{i, \text{switch}}(\mathbf{Y}_i, \mathbf{X}_i; \beta_0, G_0))/f_{i, \text{switch}}(\mathbf{Y}_i, \mathbf{X}_i; \beta_0, G_0).$$

For the further analysis of T_{1n}^{mod} we omit the index m , i.e. $\hat{\beta} := \hat{\beta}^{(m)}$ and $\hat{G} := \hat{G}^{(m)}$. Following Chen et al. (2004) we define $\hat{\pi} = \hat{G}(\theta_0)$ for $\theta_0 := (\theta_1^0 + \theta_2^0)/2$ and $\hat{G}_k$ for $k = 1, 2$ such that $\hat{G}(\theta) = \hat{\pi} \hat{G}_1(\theta) + (1 - \hat{\pi}) \hat{G}_2(\theta)$ and observe

$$\begin{aligned} \delta_i &= (\hat{\pi} - \pi_0) \Delta_i(\hat{\beta}) \\ &+ (f_{i, \text{switch}}(\mathbf{Y}_i, \mathbf{X}_i; \hat{\beta}, G_0) - f_{i, \text{switch}}(\mathbf{Y}_i, \mathbf{X}_i; \beta_0, G_0))/f_{i, \text{switch}}(\mathbf{Y}_i, \mathbf{X}_i; \beta_0, G_0) \\ &+ \hat{\pi} (f_{i, \text{switch}}(\mathbf{Y}_i, \mathbf{X}_i; \hat{\beta}, \hat{G}_1) - f_{i, \text{switch}}(\mathbf{Y}_i, \mathbf{X}_i; \hat{\beta}, \theta_1^0))/f_{i, \text{switch}}(\mathbf{Y}_i, \mathbf{X}_i; \beta_0, G_0) \\ &+ (1 - \hat{\pi}) (f_{i, \text{switch}}(\mathbf{Y}_i, \mathbf{X}_i; \hat{\beta}, \hat{G}_2) - f_{i, \text{switch}}(\mathbf{Y}_i, \mathbf{X}_i; \hat{\beta}, \theta_2^0))/f_{i, \text{switch}}(\mathbf{Y}_i, \mathbf{X}_i; \beta_0, G_0). \end{aligned} \quad (12)$$

As in Chen et al. (2004) we define $\hat{m}_{jk} = \int (\theta - \theta_k^0)^j d\hat{G}_k(\theta)$ and using the same expansion argument one gets

$$\begin{aligned} \delta_i &= (\hat{\pi} - \pi_0) \Delta_i(\beta_0) + (\hat{\beta} - \beta_0) Z_i^1(\beta_0, G_0) + \hat{\pi} \hat{m}_{11} Z_i^2(\beta_0, \theta_1^0) + (1 - \hat{\pi}) \hat{m}_{12} Z_i^2(\beta_0, \theta_2^0) \\ &+ \hat{\pi} \hat{m}_{21}/2 Z_i^3(\beta_0, \theta_1^0) + (1 - \hat{\pi}) \hat{m}_{22}/2 Z_i^3(\beta_0, \theta_2^0) + \varepsilon_{in}, \end{aligned}$$

Chen et al. (2004) and Chen, Li and Fu (2008) show that the tightness condition (see Assumption 5) ensures that

$$\varepsilon_n := \sum_{i=1}^n \varepsilon_{in} = o_p(1).$$

Using $\log(1 + \delta_i) \leq \delta_i - \delta_i^2/2 + \delta_i^3/3$ and the fact that the remainder of the square and cubic part is at least of the same order as the linear part (cf. Chen et al., 2008), we obtain in terms of the notation introduced after Theorem 1

$$T_{1n}^{\text{mod}} = 2\mathbf{b}^T \hat{\mathbf{t}} - \hat{\mathbf{t}}^T \mathbf{B} \hat{\mathbf{t}} + o_p(1)$$

where

$$\hat{\mathbf{t}}^T = (\hat{\pi} - \pi_0, \hat{\beta} - \beta_0, \hat{\pi}\hat{m}_{11}, (1 - \hat{\pi})\hat{m}_{12}, \hat{\pi}\hat{m}_{21}/2, (1 - \hat{\pi})\hat{m}_{22}/2)^T.$$

Setting

$$\begin{aligned} \mathbf{t}^T &= \mathbf{t}(\beta, G)^T \\ &= (\pi(G) - \pi_0, \beta - \beta_0, \pi(G)m_{11}(G), (1 - \pi(G))m_{12}, \pi(G)m_{21}(G)/2, (1 - \pi(G))m_{22}(G)/2)^T \end{aligned}$$

one lets $\tilde{\mathbf{t}}_1 = \mathbf{t}_1 + (B_{11})^{-1} B_{12} \mathbf{t}_2$ with $\mathbf{t}^T = (\mathbf{t}_1^T, \mathbf{t}_2^T)$, $\mathbf{t}_2 \in \mathbb{R}^2$ and decomposes

$$2\mathbf{b}^T \mathbf{t} - \mathbf{t}^T \mathbf{B} \mathbf{t} = 2\mathbf{b}_1^T \tilde{\mathbf{t}}_1 - \tilde{\mathbf{t}}_1^T B_{11} \tilde{\mathbf{t}}_1 + 2\tilde{\mathbf{b}}_2^T \mathbf{t}_2 - \mathbf{t}_2^T \tilde{B}_{22} \mathbf{t}_2.$$

Since there exist (β, G) such that $\tilde{\mathbf{t}}_1(\beta, G) = B_{11}^{-1} \mathbf{b}_1$ and since $\mathbf{t}_2 \geq 0$ for all (β, G) we see that

$$T_{1n}^{\text{mod}} \leq \mathbf{b}_1^T B_{11}^{-1} \mathbf{b}_1 + \sup_{\mathbf{t}_2 \in \mathbb{R}_{\geq 0}^2} (2\tilde{\mathbf{b}}_2^T \mathbf{t}_2 - \mathbf{t}_2^T \tilde{B}_{22} \mathbf{t}_2) + o_p(1)$$

serves as an upper bound for T_{1n}^{mod} . Following the arguments in Chen et al. (2004, Lemma 2) one concludes that this upper bound can be attained as long as $m \geq m^*$. For the expansion of T_{0n}^{mod} we again follow Chen et al. (2004) and observe

$$T_{0n}^{\text{mod}} = \mathbf{b}_1^T B_{11}^{-1} \mathbf{b}_1 + o_p(1),$$

since under the hypothesis $m = 2$, $\hat{G}_k$ are single point distributions leading to $m_{k2} = m_{k1}^2$ for $k = 1, 2$.

Now the expansions of T_{1n}^{mod} and T_{0n}^{mod} give

$$T_n^{\text{mod}} = \sup_{\mathbf{t}_2} (2\tilde{\mathbf{b}}_2^T \mathbf{t}_2 - \mathbf{t}_2^T \tilde{B}_{22} \mathbf{t}_2) + o_p(1)$$

with $\mathbf{t}_2 \in \mathbb{R}_{\geq 0}^2$. This leads to the mixture of χ^2 -distributions as described in Chen et al. (2004). $\square$

Proof of Proposition 1. We define for $i \geq 1$

$$B^i = E[b_1 b_i^T] = \left(\frac{B_{11}^i}{B_{21}^i} \middle| \frac{B_{12}^i}{B_{22}^i} \right), \quad B_{22}^i \in \mathbb{R}^{2 \times 2}.$$

We follow the calculations by Dannemann and Holzmam (2008) and observe firstly that $\lambda_1 = -C\lambda_2$ with $C > 0$ for $\lambda_k = E[b_1 | U_1 = k]$, $k = 1, 2$, since $E[b_1] = 0$.

Secondly,

$$\mathbf{B} \bar{\mathbf{1}} = B^1 \bar{\mathbf{1}} = \lambda_1 - \lambda_2 = (1 + C)\lambda_1 \tag{13}$$

where $\bar{\mathbf{1}} = (1, 0, \dots, 0)^T \in \mathbb{R}^{p+5}$, since

$$E[\Delta_1(\beta_0)b_1] = E[b_1 | U_1 = 1] - E[b_1 | U_1 = 2] = \lambda_1 - \lambda_2 = (1 + C)\lambda_1$$

Using the conditional independence of $(\mathbf{X}_i, \mathbf{Y}_i)_i$ given $(U_i)_i$ one shows

$$B^i = C_i \lambda_1 \lambda_1^T \quad (14)$$

with constants C_i depending on the transition matrix P . Similar calculations as in Danne-
mann and Holzmann (2008) show that for $i \geq 2$

$$E[\tilde{b}_{21} \tilde{b}_{2i}^T] = B_{22}^i - B_{21}^i (B_{11}^1)^{-1} B_{12}^1 - B_{21}^1 (B_{11}^1)^{-1} B_{12}^i + B_{21}^1 (B_{11}^1)^{-1} B_{11}^i (B_{11}^1)^{-1} B_{12}^1 = 0,$$

since all four summands are equal, which follows from (13) and (14). $\square$

References

- BÖHNING, D., DIETZ, E., SCHLATTMANN, P., MENDONCA, L. and KIRCHNER, U. (1999). The zero-inflated poisson model and the decayed, missing and filled teeth index in dental epidemiology. *J. R. Stat. Soc. Ser. A*, **162** 195–209.
- BÖHNING, D. (1999). *Computer-assisted analysis of mixtures and applications*, vol. 81 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, Boca Raton, FL. Meta-analysis, disease mapping and others.
- CHEN, H., CHEN, J. and KALBFLEISCH, J. D. (2001). A modified likelihood ratio test for homogeneity in finite mixture models. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **63** 19–29.
- CHEN, H., CHEN, J. and KALBFLEISCH, J. D. (2004). Testing for a finite mixture model with two components. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **66** 95–115.
- CHEN, J. and KHALILI, A. (2008). Order selection in finite mixture models with a non-smooth penalty. *J. American Statist. Assoc.* In press.
- CHEN, J., LI, P. and FU, Y. (2008). Testing homogeneity in a mixture of von Mises distributions with a structural parameter. *Canad. J. Statist.*, **36** 129–142.
- DANNEMANN, J. and HOLZMANN, H. (2008). Testing for two states in a hidden markov model. *Canad. J. Statist.*, **36** 505–520. In press.
- FRÜHWIRTH-SCHNATTER, S. (2006). *Finite mixture and Markov switching models*. Springer Series in Statistics, Springer, New York.
- HENNIG, C. (2000). Identifiability of models for clusterwise linear regression. *J. Classification*, **17** 273–296.
- KHALILI, A. and CHEN, J. (2007). Variable selection in finite mixture of regression models. *J. American Statist. Assoc.*, **102** 1025–1038.
- KIEFER, N. M. (1978). Discrete parameter variation: efficient estimation of a switching regression model. *Econometrica*, **46** 427–434.
- LE, N. D., LEROUX, B. G. and PUTERMAN, M. L. (1992). Exact likelihood evaluation in a markov mixture model for time series of seizure counts. *Biometrics*, **48** 317–323.

- McLACHLAN, G. and PEEL, D. (2000). *Finite mixture models*. Wiley Series in Probability and Statistics: Applied Probability and Statistics, Wiley-Interscience, New York.
- NAIK, P. A., SHI, P. and TSAI, C.-L. (2007). Extending the Akaike information criterion to mixture regression models. *J. Amer. Statist. Assoc.*, **102** 244–254.
- QUANDT, R. E. and RAMSEY, J. B. (1978). Estimating mixtures of normal distributions and switching regressions. *J. Amer. Statist. Assoc.*, **73** 730–752. With comments and a rejoinder by the authors.
- SKRONDAL, A. and RABE-HESKETH, S. (2004). *Generalized latent variable modeling*. Interdisciplinary Statistics, Chapman & Hall/CRC, Boca Raton, FL. Multilevel, longitudinal, and structural equation models.
- WANG, P. and PUTERMAN, M. L. (1998). Mixed logistic regression models. *J. Agric. Biol. Environ. Stat.*, **3** 175–200.
- WANG, P. and PUTERMAN, M. L. (1999). Markov Poisson regression models for discrete time series. I. Methodology. *J. Appl. Statist.*, **26** 855–869.
- WANG, P. and PUTERMAN, M. L. (2001). Analysis of longitudinal data of epileptic seizure counts—a two-state hidden Markov regression approach. *Biom. J.*, **43** 941–962.
- ZHANG, H., FENG, R. and ZHU, H. (2003). A latent variable model of segregation analysis for ordinal traits. *J. Amer. Statist. Assoc.*, **98** 1023–1034.
- ZHU, H.-T. and ZHANG, H. (2003). Hypothesis testing in mixture regression models (mathematical details). Tech. rep., Yale University School of Medicine, New Haven.
- ZHU, H.-T. and ZHANG, H. (2004). Hypothesis testing in mixture regression models. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **66** 3–16.