Statistical Modelling 2014; **14**(**4**): 293–313

Segmented mixed models with random changepoints: a maximum likelihood approach with application to treatment for depression study

Vito MR Muggeo[1], David C Atkins[2], Robert J Gallop[3] and Sona Dimidjian[4]
[1]Dipartimento di Scienze Statistiche e Matematiche "S. Vianelli", Universita di Palermo, Italy
[2]Dept. of Psychiatry and Behavioural Science, University of Washington - Seattle, WA, USA
[3]Dept. of Mathematics, Applied Statistics Program, West Chester University, West Chester, PA, USA
[4]Dept. of Psychology and Neuroscience, University of Colorado at Boulder, CO, USA

Abstract: We present a simple and effective iterative procedure to estimate segmented mixed models in a likelihood based framework. Random effects and covariates are allowed for each model parameter, including the changepoint. The method is practical and avoids the computational burdens related to estimation of nonlinear mixed effects models. A conventional linear mixed model with proper covariates that account for the changepoints is the key to our estimating algorithm. We illustrate the method via simulations and using data from a randomized clinical trial focused on change in depressive symptoms over time which characteristically show two separate phases of change.

Key words: changepoint; mixed segmented regression; nonlinear mixed models; random changepoints; psychiatric longitudinal data

Received June 2011; **revised** April 2013; **accepted** August 2013

1 Introduction

Depression is a serious psychiatric illness involving significantly reduced mood and loss of interest or pleasure. The World Health Organization has estimated that depression affects 121 million people worldwide; it is among the leading causes of disability and in its most severe forms, depression can lead to suicide. Fortunately, clinical research has shown that both anti-depressant medication and psychotherapy can be effective treatments for clinical depression. Treatment research with depression is now focusing on the pattern of recovery during treatment: typically sudden, short-term improvements and multiple phases of improvement are observed (Ilardi and Craighead, 1994). However, studying such patterns is challenging as individuals have different recovery slopes and different transition points between early and later phases of improvement. The potential importance of this regime

Address for correspondence: Vito MR Muggeo, Dipartimento di Scienze Statistiche e Matematiche "Vianelli", viale delle Scienze, edificio 19, 90121 Palermo – Italy. E-mail: vito.muggeo@unipa.it

 10.1177/1471082X13504721

switching feature for guiding practice and understanding the course and mechanisms of psychotherapeutic change, makes it imperative to investigate the phenomenon with appropriate statistical methods.

The segmented or piecewise linear mixed model constitutes a valuable framework for an in-depth investigation of these multi-phase profiles. These models may be useful when the individual response profiles exhibit a mixture of two or more phases of growth or decline with one or more unknown transition points usually referred as breakpoints or changepoints. The changepoints, which are parameters with substantive interest represent the main feature and appeal of this model. Beyond treatment research on depression, segmented mixed models could have broad applicability in Psychology and Medicine. In Psychology many developmental processes exhibit distinct phases; for instance, in learning experiments subjects show fast improvement in an initial learning period and a levelling off in improvement later (e.g., Cudeck and Harring, 2007). Studies on cognitive decline and dementia have described a two-stage pattern for the individual profiles: Hall *et al.* (2000, 2003) have dealt with the analysis of pre-dementia phases, Dominicus *et al.* (2008) have focussed on transitions in old age and Jacqmin-Gadda *et al.* (2006) have employed a joint modelling for cognitive decline and risk of dementia; Muniz Terrera *et al.* (2011) have focussed on presence of missing data in modelling cognitive decline within a Bayesian framework. Moreover, Lange *et al.* (1992), Kiuchi *et al.* (1995) and Ghosh and Vaida (2007) have applied random changepoints in the analysis of longitudinal biomarkers (CD4 T-cells or T4 counts) for HIV infected subjects. Morrell *et al.* (1995) have studied the development of prostate-specific antigen levels as a marker for prostate cancer and Scott *et al.* (2004) have modelled growth and decline in lung function in Duchenne's muscolar dystrophy using a mixture model for the changepoint.

Within the statistical literature, random changepoint models have been mostly investigated in a Bayesian perspective and few authors have faced the problem from a frequentist standpoint; difficulties in dealing with such nonstandard models are well-known and discussed, for instance, in Naumova *et al.* (2001) and Hall *et al.* (2003), who have warned about the difficulty of accounting for random changepoints in a likelihood framework. In fact the few papers presenting non-Bayesian methods have assumed smooth transitions between phases, linear-exponential (Morrell *et al.*, 1995) or linear-cubic (Jacqmin-Gadda *et al.*, 2006) transitions. Treatment of piecewise linear modelling in longitudinal analysis may be also found in the relatively recent book by Fitzmaurice *et al.* (2004), but discussion is confined to fixed changepoints only. When the changepoint is known, the model simplifies considerably as it becomes a conventional linear mixed model (LMM) and standard software may be used (e.g., Naumova *et al.*, 2001); even if the breakpoint is unknown but not varying among the subjects, the problem may be addressed rather straightforwardly by profiling the log likelihood and using a grid search to seek the best value, see Hall *et al.* (2003); random effects in the changepoint parameter represent the actual complication for segmented modelling in a likelihood framework. As an alternative to segmented modelling, a nonparametric approach could be followed leading to additive

mixed models (e.g., Durban *et al.*, 2005; Fitzmaurice *et al.*, 2004). These models enable to model the individual profiles via smooth but otherwise unspecified functions and therefore they are more flexible than the segmented models. While the fit might be greatly improved, the additive model does not provide interpretable parameters, such as slopes and changepoints. We will not discuss additive mixed models and their differences with the segmented framework in the present article.

The aim of this paper is to present a general approach for the estimation of linear segmented mixed models with random changepoints in a likelihood framework. As will be discussed in the next section, the proposed framework is simple as the nonlinear segmented mixed model is converted to a standard mixed model to be fitted iteratively. The article is structured as follows. In Section 2 we introduce the segmented mixed model and derive the estimating algorithm that is presented and discussed in detail. Section 3 presents results from a small simulation study and an illustration of the methods using data from a depression treatment study is reported in Section 4; final discussion and concluding remarks are included in the last section.

2 Methods

2.1 The model

Let y_{ij} be the measurement $j = 1, 2, \ldots, n_i$ for subject $i = 1, 2, \ldots, n$ and let t_{ij} be the time when y_{ij} occurs. Typically in longitudinal data analysis the focus is on time trajectories, namely to investigate how the outcome y_{ij} evolves in time and to assess possible differences due to covariates and/or heterogeneity. Roughly speaking, the mixed model framework postulates that $y_{ij} = s(t_{ij}, \phi_i) + \epsilon_{ij}$, where $s(t_{ij}, \cdot)$ represents the time trajectory for subject i controlled by the parameter vector ϕ_i and ϵ_{ij} is the usual stochastic term. The conventional linear mixed model assumes that $s(t_{ij}, \phi_i)$ is linear and also usually $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$. However, as discussed above, linearity may be too restrictive in many applications. As sketched in the Introduction, in Psychology it has been often observed that the outcome patterns obey two regimes, leading to a 'segmented' or 'piecewise' linear mixed model,

$$y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \delta_i (t_{ij} - \psi_i)_+ + \varepsilon_{ij}, \tag{2.1}$$

where the linear spline function $(t - \psi)_+ = (t - \psi) I(t > \psi)$ and $I(\cdot)$ is the indicator function equaling one when its argument is true and zero otherwise. For each subject i, the parameter β_0 is the mean response value at time $t = 0$, and ψ is the changepoint or breakpoint which determines the change in the trajectory. More specifically β_1 is the 'left' slope i.e., when $t \leq \psi$, and δ is the difference-in-slopes; therefore $\beta_1 + \delta$ is the 'right' slope, namely the slope when $t > \psi$. Note the regression function is continuous with discontinuous first derivative at the changepoint.

In a two-stage/hierarchical formulation, we specify each parameter as a sum of fixed and random effects (Fitzmaurice *et al.*, 2004, pp. 200–206). For the intercept, the left slope and the difference-in-slope parameters of equation (2.1) it is possible to write $\beta_{0i} = \beta_{00} + \boldsymbol{x}_{0i}^T\boldsymbol{\beta}_0 + b_{0i}$, $\beta_{1i} = \beta_{10} + \boldsymbol{x}_{1i}^T\boldsymbol{\beta}_1 + b_{1i}$, and $\delta_i = \delta_0 + \boldsymbol{x}_i^T\boldsymbol{\delta} + d_i$ respectively. Each covariate vector represents the factors affecting the relevant subject-specific parameter via the corresponding fixed effects and b_{0i}, b_{1i} and d_i are zero-mean random effects accounting for heterogeneity. To set-up a two-level model for the changepoint, we re-parameterize the changepoint into $\psi = \psi(\kappa_0) = (a_1 + a_2 e^{\kappa_0})/(1 + e^{\kappa_0})$ where a_1 and a_2 are respectively the minimum and the maximum of the observed times t_{ij}; note that κ_0 is unbounded and $\psi \in (a_1, a_2)$. Now, the logistic function readily allows us to set up a full model for the changepoint,

$$\psi_i = \psi(\boldsymbol{\kappa}, k_i; \boldsymbol{z}_i) = \frac{a_1 + a_2 e^{\kappa_0 + \boldsymbol{z}_i^T\boldsymbol{\kappa} + k_i}}{1 + e^{\kappa_0 + \boldsymbol{z}_i^T\boldsymbol{\kappa} + k_i}}. \tag{2.2}$$

Similarly to the other linear parameters, $\boldsymbol{z}_i$ represents the covariates affecting the location of changepoint via the parameter $\boldsymbol{\kappa}$, κ_0 is the changepoint (on the logit scale) when $\boldsymbol{z} = \boldsymbol{0}$, and the k_is are the random effects; we also write equation (2.2) as $\psi_i = \dfrac{a_1 + a_2 e^{\eta_i}}{1 + e^{\eta_i}}$, namely in terms of the linear predictor $\eta_i = \kappa_{0i} + \boldsymbol{z}_i^T\boldsymbol{\kappa}$ where the random term is absorbed in the intercept, i.e., $\kappa_{0i} = \kappa_0 + k_i$. As in conventional mixed models, the covariate vectors do not need to be distinct, i.e., the same variable may simultaneously affect each parameter of the segmented mixed model (2.1). To complete specification of the model we assume $(b_{0i}, b_{1i}, d_i, k_i)^T \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R})$, where the covariance matrix $\boldsymbol{R}$ is allowed to take any structure, such as the usual unstructured, compound symmetry or diagonal.

2.2 Linear approximation of the segmented function

As mentioned in the Introduction, estimation of model (2.1) is a difficult task, especially in a frequentist framework. Essentially the random effects enter the model nonlinearly and they cannot be integrated out of the likelihood in closed form; more importantly, the log-likelihood is not differentiable everywhere making ineffective the usual gradient-based optimization methods. We are aware of no previous method to estimate a linear segmented model with random changepoints in a frequentist-likelihood paradigm. We generalize the idea of Muggeo (2003) to estimate regression models with fixed-effects changepoints and extend the algorithm to account for covariates and random effects in the determination of the changepoints themselves. We discuss how by transplanting the nonlinear segmented mixed model into a linear framework, estimation proceeds in a rather straightforward manner by fitting at

each step a conventional LMM via 'simple' or 'restricted' maximum likelihood. The linear predictor of this 'working' LMM depends on specifications of stage 2 models for the changepoint and the difference-in-slopes parameters that are discussed in the following subsections.

2.2.1 Covariates in the changepoint

Let us ignore the subscripts and write $f = f(\psi(\kappa)) = (t - \psi(\kappa))_+$ where $\psi(\kappa)$ is the logistic function (2.2) for the changepoint and $\boldsymbol{\kappa}$ also includes the intercept κ_0. A first-order Taylor expansion around a known $\tilde{\boldsymbol{\kappa}}$ yields

$$f(\psi(\boldsymbol{\kappa})) \approx f(\psi(\tilde{\boldsymbol{\kappa}})) + (\boldsymbol{\kappa} - \tilde{\boldsymbol{\kappa}})^T \left.\frac{\partial f}{\partial \boldsymbol{\kappa}}\right|_{\boldsymbol{\kappa}=\tilde{\boldsymbol{\kappa}}} \tag{2.3}$$

where $f(\psi(\tilde{\boldsymbol{\kappa}}))$ is just the segmented function $(t - \psi(\boldsymbol{\kappa}))_+ = U$ evaluated at $\tilde{\boldsymbol{\kappa}}$, and $\left.\frac{\partial f}{\partial \boldsymbol{\kappa}}\right|_{\boldsymbol{\kappa}=\tilde{\boldsymbol{\kappa}}}$ means the gradient vector of f with respect to $\boldsymbol{\kappa}$, again evaluated at $\tilde{\boldsymbol{\kappa}}$. The hth component of the vector $\frac{\partial f}{\partial \boldsymbol{\kappa}}$ is the hth partial derivative which, using the chain rule, is

$$\frac{\partial f}{\partial \kappa_h} = \frac{\partial f}{\partial \psi}\frac{\partial \psi}{\partial \eta}\frac{\partial \eta}{\partial \kappa_h} = -I(t > \psi)\frac{e^{\eta}(a_2 - a_1)}{(1 + e^{\eta})^2} z_h = VDz_h,$$

where z_h refers to the hth explanatory variable for the changepoint model with $z_0 \equiv 1$. In the above expression we have set $V = -I(t > \psi)$ and $D = \frac{e^{\eta}(a_2 - a_1)}{(1 + e^{\eta})^2}$ where η is the linear predictor of ψ which is given by (2.2). Then, using the Taylor approximation derived above and using tilde to denote approximate values, it is possible to write

$$\delta_i(t_{ij} - \psi_i)_+ \approx \delta_i \widetilde{U}_{ij} + \delta_i(\kappa_{0i} - \tilde{\kappa}_{0i})\widetilde{V}_{ij}\widetilde{D}_{ij} + \delta_i(\kappa_1 - \tilde{\kappa}_1)\widetilde{V}_{ij}\widetilde{D}_{ij}z_{ij} \tag{2.4}$$

where we have purposely assumed for simplicity a single covariate z in the changepoint model (2.2). However, for reasons that will be clearer later, we can express (2.4) in an alternative way; assuming to know also an approximate value of δ_i, $\tilde{\delta}_i$ say, it turns out

$$\begin{aligned}\delta_i(t_{ij} - \psi_i)_+ &\approx \delta_i \widetilde{U}_{ij} + (\kappa_{0i} - \tilde{\kappa}_{0i})\{\tilde{\delta}_i \widetilde{V}_{ij}\widetilde{D}_{ij}\} + (\kappa_1 - \tilde{\kappa}_1)\{\tilde{\delta}_i \widetilde{V}_{ij}\widetilde{D}_{ij}z_{ij}\} \\ &= \delta_i \widetilde{U}_{ij} + \kappa_{0i}\widetilde{G}_{0ij} + \kappa_1 \widetilde{G}_{1ij} + \{-\tilde{\kappa}_{0i}\widetilde{G}_{0ij} - \tilde{\kappa}_1 \widetilde{G}_{1ij}\},\end{aligned} \tag{2.5}$$

where the subscript i means the parameter is given by the sum of fixed and random terms, and $\{-\tilde{\kappa}_{0i}\widetilde{G}_{0ij} - \tilde{\kappa}_1\widetilde{G}_{1ij}\}$ is a known term, i.e., an 'offset' in the generalized linear model terminology. Equation (2.4), or equivalently (2.5), emphasize the approximate linear formulation of the nonlinear segmented parameterization. Note that the nonlinear parameters κs of the second-level model (2.2) are 'converted' in (2.5) into linear parameters of the variables $\widetilde{G}_{0i}$ and $\widetilde{G}_{1i}$ that can be considered in the same way of standard explanatory variables entering the model linearly. This linearity is enabled by the first-order expansion (2.3); indeed, additional terms in the Taylor expansion would improve the accuracy of the approximation but these would also waste the derived linear structures (2.4) and (2.5).

2.2.2 Covariates in the difference-in-slopes

When the difference-in-slopes parameter also depends on covariates, the linear approximation of the segmented function has to be slightly modified. For the sake of simplicity, we assume that δ_i depends on a single covariate x via the parameter δ_1, namely $\delta_i = \delta_{0i} + \delta_1 x_{ij}$, where δ_{0i} also includes the relevant random effect d_i, i.e., $\delta_{0i} = \delta_0 + d_i$; therefore, for individual i with covariate value x_{ij}, the difference-in-slopes is δ_{0i} when $x_{ij} = 0$ and $\delta_{0i} + \delta_1 x_{ij}$ if $x_{ij} \neq 0$. The relevant term is $(\delta_{0i} + \delta_1 x_{ij})(t_{ij} - \psi_i)_+$, and using the Taylor expansion for $(t_{ij} - \psi_i)_+$, it is possible to write

$$\begin{aligned}\delta_{0i}(t_{ij} - \psi_i)_+ + \delta_1(t_{ij} - \psi_i)_+ x_{ij} &\approx \delta_{0i}\widetilde{U}_{ij} + \delta_1\widetilde{U}_{ij}x_{ij} \\ &+ (\kappa_{0i} - \tilde{\kappa}_{0i})\delta_{0i}\widetilde{V}_{ij}\widetilde{D}_{ij} + (\kappa_{0i} - \tilde{\kappa}_{0i})\delta_1\widetilde{V}_{ij}\widetilde{D}_{ij}x_{ij} \\ &+ (\kappa_1 - \tilde{\kappa}_1)\delta_{0i}\widetilde{V}_{ij}\widetilde{D}_{ij}z_{ij} + (\kappa_1 - \tilde{\kappa}_1)\delta_1\widetilde{V}_{ij}\widetilde{D}_{ij}z_{ij}x_{ij}.\end{aligned}$$

Now this expression is apparently useless, since the same parameter refers to different pseudo-covariates preventing the terms to be managed as in (2.4) to obtain a linear predictor. However this problem is bypassed by using the same rationale justifying the shift from (2.4) to (2.5), namely by assuming to know the approximate values of δ_{0i} and δ_1, respectively $\tilde{\delta}_{0i}$ and $\tilde{\delta}_1$. Thus, it is possible to re-arrange the right hand side terms and to write

$$\begin{aligned}&\delta_{0i}(t_{ij} - \psi_i)_+ + \delta_1(t_{ij} - \psi_i)_+ x_{ij} \approx \delta_{0i}\widetilde{U}_{ij} + \delta_1\{\widetilde{U}_{ij}x_{ij}\} \\ &+ (\kappa_{0i} - \tilde{\kappa}_{0i})\{\tilde{\delta}_{0i}\widetilde{V}_{ij}\widetilde{D}_{ij} + \tilde{\delta}_1\widetilde{V}_{ij}\widetilde{D}_{ij}x_{ij}\} + (\kappa_1 - \tilde{\kappa}_1)\{\tilde{\delta}_{0i}\widetilde{V}_{ij}\widetilde{D}_{ij}z_{ij} + \tilde{\delta}_1\widetilde{V}_{ij}\widetilde{D}_{ij}z_{ij}x_{ij}\} \\ &= \delta_{0i}\widetilde{U}_{ij} + \delta_1\{\widetilde{U}_{ij}x_{ij}\} + \kappa_{0i}\widetilde{G}_{0ij} + \kappa_1\widetilde{G}_{1ij} + \{-\tilde{\kappa}_{0i}\widetilde{G}_{0ij} - \tilde{\kappa}_1\widetilde{G}_{1ij}\} \qquad (2.6)\end{aligned}$$

which has a linear structure and the same feature of expression (2.5), but somewhat more complicated expressions for the pseudo-covariates $\widetilde{G}_0$ and $\widetilde{G}_1$. We emphasize that expressions (2.5) or (2.6) are linear approximations of the nonlinear segmented term for a given value of the changepoint $\tilde{\psi}$ which in turn depends on $\tilde{\kappa}$ and which affects the values of the pseudo-covariates $\widetilde{U}$, $\widetilde{G}_0$, and $\widetilde{G}_1$. Of course when multiple explanatory variables are requested in equation (2.2), additional G-type pseudo-covariates are easily constructed and included. With $z_1, \quad , z_H$ explanatory variables in the changepoint model, and $x_1, \quad , x_S$ explanatory variables for the difference-in-slopes parameter, the h^{th} ($h = 0, 1, \quad , H$) generic auxiliary covariate for the working LMM is

$$\widetilde{G}_h = \tilde{\delta}_{0i} VDz_h + \tilde{\delta}_1 VDz_h x_1 + \ldots + \tilde{\delta}_S VDz_h x_S$$

where $z_0 \equiv 1$, and $\tilde{\delta}_0, \tilde{\delta}_1, \ldots, \tilde{\delta}_S$ are the (approximate) coefficients of variables $\widetilde{U}, \widetilde{U}x_1, \quad , \widetilde{U}x_S$. In the next section we detail how these pseudo-covariates $\widetilde{G}_h$ coming from the foregoing linear parameterizations may be used to estimate the segmented mixed model (2.1) with fixed and random effects for each model parameter.

2.3 Model estimation

The working LMM with original response y_i and linear predictor (2.5) or (2.6) allows estimation of model (2.1). We illustrate how an efficient algorithm may be implemented. We assume to know the $\tilde{\psi}_i$s, an approximation of changepoints; for instance, at the first step we can set $\tilde{\psi}_i = \tilde{\psi}$ for each subject $i = 1, 2, \quad , n$ with all elements of κ in equation (2.2) equal to zero. Of course $\tilde{\kappa}_0 = \log\left(\frac{\tilde{\psi} - a_1}{a_2 - \tilde{\psi}}\right)$ and $\tilde{\psi}$ may be obtained by means of visual inspection of the scatter-plot, or by means of substantive issues related to the data being analyzed; alternatively we can use the nadir estimate from a preliminary simplified fit based on quadratic polynomial profiles. To make ideas clearer we re-write the regression equation of the working LMM,

$$\begin{aligned}\mu_{ij} = {} & \beta_0 + \beta_1 t_{ij} + \delta_0 \widetilde{U}_{ij} + \delta_1 \widetilde{U}_{ij} x_{ij} + \kappa_0 \widetilde{G}_{0ij} + \kappa_1 \widetilde{G}_{1ij} \\ & + b_{0i} + b_{1i} t_{ij} + d_i \widetilde{U}_{ij} + k_i \widetilde{G}_{0ij} + \widetilde{O}_{ij} \end{aligned} \tag{2.7}$$

where Greek symbols refer to fixed effects and Italic to random effects. Of course $\delta_0 + d_i = \delta_{0i}$, $\kappa_0 + k_i = \kappa_{0i}$, and

$$\widetilde{O}_{ij} = -\sum_{h=0}^{H} \tilde{\kappa}_h \widetilde{G}_{hij} - \tilde{k}_i \widetilde{G}_{0ij} \tag{2.8}$$

is the offset term as in (2.5) and (2.6).

After specification of the covariance matrix $\boldsymbol{R}$ for $(b_{0i}, b_{1i}, d_i, k_i)^T$, the LMM (2.7) may be fitted via (restricted) maximum likelihood to obtain parameter estimates and empirical Bayes predictions of the random effects; see for instance Pinheiro and Bates (2000, pp. 57–81) or Fitzmaurice *et al.* (2004, pp. 88–102, 206–210) for the relevant formulas. Given the new estimates/predictions of the changepoints and difference in slopes parameters, the covariates U, G_0, and G_1 are re-computed along with the offset term O, and model (2.7) is re-fitted to obtain the updated estimates; the process is repeated up to convergence when the changepoint estimates, fixed and subject-specific, are available from the last fit. Assuming a single covariate for the changepoint and difference-in-slopes parameters, the algorithm may be schematized as follows.

Initialize: Fix a starting value for the changepoints, for instance $\tilde{\psi}_i = \tilde{\psi}$ for each i, and compute the covariates $\widetilde{U}$ and $\widetilde{Ux}$. Fit a LMM

$$\text{fixed} = 1 + t + \widetilde{U} + \widetilde{Ux}, \ \text{random} = 1 + t + \widetilde{U}$$

to get preliminary estimates of the difference-in-slopes parameters $\tilde{\delta}_{0i} = \tilde{\delta}_0 + \tilde{d}_i$ and $\tilde{\delta}_1$, in addition to $\tilde{\kappa}_{0i} = \text{logit}(\tilde{\kappa})$, $\tilde{\kappa}_1 = 0$ and $k_i = 0$;

Step 1: With approximate values $\tilde{\kappa}_{0i}$, $\tilde{\kappa}_1$, $\tilde{\psi}_i$, $\tilde{\delta}_{0i}$, $\tilde{\delta}_1$, compute the covariates $\widetilde{U}$, $\widetilde{Ux}$, $\widetilde{G}_0$, $\widetilde{G}_1$, and the offset term $\widetilde{O}$;

Step 2: Fit the working LMM (2.7), i.e.,

$$\text{fixed} = 1 + t + \widetilde{U} + \widetilde{Ux} + \widetilde{G}_0 + \widetilde{G}_1, \ \text{random} = 1 + t + \widetilde{U} + \widetilde{G}_0$$

Step 3: From the fitted LMM extract the estimates $\hat{\kappa}_0$, $\hat{\kappa}_1$ and the predictions $\hat{k}_i$ to update $\hat{\eta}_i = \hat{\kappa}_0 + \hat{\kappa}_1 z_i + \hat{k}_i$ and the changepoint values $\hat{\psi}_i = \dfrac{a_1 + a_2 \exp \hat{\eta}_i}{1 + \exp \hat{\eta}_i}$. Also extract $\hat{\delta}_0$, $\hat{\delta}_1$ and $\hat{\delta}_i$;

Step 4: Repeat steps 1 to 3, up to the relative difference between the log likelihoods of the last two fits is small.

At convergence the last fitted working LMM provides estimates for the initial segmented mixed model: fixed effects and predictions for the d_is and the k_is from which subject-specific changepoints can be computed via (2.2). Variances and covariances estimates for the random effects can be extracted from the estimated matrix $\boldsymbol{R}$: for instance, entries corresponding to covariate G_0 in $\boldsymbol{R}$ represent covariances between the changepoint random effects and random effects of the remaining terms.

2.3.1 Some remarks

The choice of initial values to start the algorithm, which is generally non trivial for nonlinear mixed models (NLMMs), becomes a relatively minor issue in our approach since they have to be supplied only for the changepoint. Of course for quite noisy data or bland segmented relationships, different starting values may affect the final results. However, in these circumstances we have found the bootstrap restarting to be very effective to escape local optima of the log likelihood: here the starting values are obtained by fitting the model to different bootstrap samples. More specifically, following Wood (2001), the bootstrap restarting proceeds as follows: *(i)* fit the model to the observed data $\boldsymbol{y}$ using starting values $\tilde{\boldsymbol{\kappa}}$ and save the changepoint parameter estimates $\hat{\boldsymbol{\kappa}}$; *(ii)* fit the model to a nonparametric bootstrap sample $\boldsymbol{y}^*$ using the 'previous' estimates $\hat{\boldsymbol{\kappa}}$ as starting values and save the relevant estimates $\hat{\boldsymbol{\kappa}}^*$; *(iii)* fit the model to observed data $\boldsymbol{y}$ using $\hat{\boldsymbol{\kappa}}^*$ as starting values, and save the relevant estimates $\hat{\hat{\boldsymbol{\kappa}}}$; *(iv)* if $\hat{\hat{\boldsymbol{\kappa}}}$ improve the log likelihood, then consider them as starting values in (i), otherwise discard them; *(v)* repeat steps (i) to (iv) for several bootstrap samples. Since bootstrap objective functions share the same large scale but different small scale structure of the original objective, bootstrap samples should provide starting values from which maximization of the original objective function is likely to escape local optima.

The algorithm has been illustrated for a complete model with intercept and left slope and a single covariate in the difference-in-slopes and in the changepoint submodels. However intercepts and left slope may be deleted if requested and for each model parameter multiple explanatory variables may be included straightforwardly leaving unchanged the algorithm. Also, whether a second changepoint is required, the Taylor expansion described above would produce additional G_0-type and G_1-type variables (depending on the values of the second changepoint) to be included in the working LMM (2.7).

As in standard mixed modelling, the structure of the covariance matrix $\boldsymbol{R}$ of the random effects $(b_{0i}, b_{1i}, d_i, k_i)^T$ has to be specified. While the algorithm does not depend on it, the structure of $\boldsymbol{R}$ may affect the final predictions. As in the usual LMMs there are not in general rules of thumb, although more complex structures allow more flexible predictions but also make model estimation more demanding. Likelihood-based criteria can be employed to select the more appropriate $\boldsymbol{R}$, see Section 4 for an illustrative example with discussion.

As an alternative to logistic model (2.2), a linear formulation for the changepoint would be possible, such as $\psi_i = \kappa_0 + \boldsymbol{z}_i^T\boldsymbol{\kappa} + k_i$. In this linear case, the estimating algorithm remains unchanged, with the only difference that $\frac{\partial \psi}{\partial \eta} = D \equiv 1$. Parameter interpretation would benefit from a linear model, but it also would lead to some problems related to the range of admissible values for the changepoints which have to lie within the range of observed times t_{ij}. Whatever the model for the changepoint is, the algorithm assumes that the changepoint parameter exists. The parameter κ_0 in expression (2.2) has to be defined, i.e., the 'identified case' according to the terminology

of Feder (1975). Actually $\kappa_0 = 0$ does not imply the non-existence of the changepoint and when the changepoint does not exist the algorithm should fail to converge when all the observed profiles exhibit linear patterns. However, as reported in Application, the algorithm is still effective when there are some linear profiles in the sample.

Note the proposed algorithm is strictly related to the so-called zero-expansion methods commonly employed to estimate NLMM (Davidian and Giltinan, 1995), see also Wolfinger and Lin (1997) for a concise discussion. However the rationale is substantially different since, as opposed to our approach, the zero-expansion methods use the Taylor expansion to approximate the integrand of the log-likelihood from the original NLMM.

2.3.2 Standard errors and confidence intervals

The proposed iterative procedure makes maximization quite feasible, since standard LMM have to be fitted iteratively; however likelihood inference is not simplified accordingly. In standard LMM, inference on the fixed effect parameters relies on approximate distribution for the (restricted) maximum likelihood estimates which are asymptotically Normal with approximate covariance matrix depending on the inverse of the information matrix; these standard errors (SE) which are generally returned as model output, may be employed to build approximate Normal-based confidence intervals for the linear parameters. Notice that the working LMM (2.7), apparently returns standard errors also for the changepoint parameter estimates, $\hat{\kappa}_0$ and $\hat{\kappa}_1$. However simulations reported in the next section have shown that these SE underestimate the actual variability of the changepoint estimators and thus the resulting Normal-based confidence intervals are not recommended in general. However, since the sampling distributions appear to be Normal we suggest to use the bootstrap to estimate the standard errors. The 'subjects', rather than the measurements, are re-sampled and the model is estimated for each bootstrap sample (e.g., van der Leeden *et al.*, 2007); the empirical standard deviation of the bootstrap distributions is employed for the variance estimator. Thus for the generic changepoint parameter κ, we obtain $b = 1, \quad , B$ bootstrap estimates $\hat{\kappa}_b^*$, the bootstrap standard error $\mathrm{SE}^* = \left\{ \sum_b (\hat{\kappa}_b^* - \sum_b \hat{\kappa}_b^* / B)^2 / (B-1) \right\}^{1/2}$, leading to the 95%, say, confidence interval with endpoints $(\hat{\kappa} \pm 1.96\ \mathrm{SE}^*)$. The advantage of using the bootstrap to estimate SE* is that much less bootstrap samples are needed, usually even 25 or 50 suffice (Efron and Tibshirani, 1993, p. 52); instead, much more bootstrap replicates are needed to estimate the sampling distribution.

3 Simulation study

A simulation study was carried out to assess the performance of the proposed framework as regard to point and interval estimators of the model parameters, especially $\hat{\kappa}_0$ and $\hat{\kappa}_1$.

Data were generated from a segmented mixed model (2.1), with t_{ij} varying in $[0,1]$, $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon)$, $\beta_{0i} = 2 + b_{0i}$, $\beta_{1i} = -0.2 + b_{1i}$, $\delta_i = 0.3 + 0.5x_i + d_i$, and x_i a binary covariate with balanced design. Two regression equations for the changepoints with parameters κ_0, κ_1 and random effects k_i, were considered: linear $\psi_i = 0.4 + 0.2z_i + k_i$, and logistic $\psi_i = (1 + \exp\{-(-0.3 + 0.7z_i + k_i)\})^{-1}$ where z_i is an additional binary covariate with balanced categories. Furthermore $(b_{0i}, b_{1i}, d_i, k_i)^T \sim \mathcal{N}(\mathbf{0}, \mathrm{diag}(0.1^2, 0.1^2, 0.1^2, \sigma_k^2))$ where $\sigma_k \in \{0.10, 0.20\}$ if the ψ_is follow a logistic model and $\sigma_k \in \{0.05, 0.10\}$ when a linear model holds. Finally two different sample sizes $n = 50$ and $n = 100$ and two different values of the variance response $\sigma_\varepsilon \in \{0.03, 0.05\}$ were employed and for each subject the number of repeated measurements was assigned according to the values of a discrete uniform, i.e., $n_i \sim U(5,15)$.

Table 1 reports empirical means and variances for the point estimators of all the model parameters based on 500 replicates. At each replicate starting values for breakpoint have been obtained via the nadir estimate from a preliminary fitting of a quadratic polynomial and bootstrap restarting has been employed with 10 bootstrap samples.

The behaviour of the estimators coming from the proposed framework is satisfactory; each estimator appears to be approximately unbiased with variability decreasing at larger samples. As expected results get worse when the variances, both σ_k and σ_ε, increase, since the piecewise-linear relationships get flat and difficult to estimate. However absolute relative biases are always within the 3%, with the largest value obtained for $\hat{\kappa}_0$ in the most difficult situation at large values of σ_ε and σ_k.

We also examine the interval estimators for the changepoint parameters κ_0 and κ_1, by assessing the coverage levels (CL) and average width (AW) of the 95% confidence intervals (CI). As previously discussed, we exploit the nonparametric bootstrap by resampling 100 times the 'subjects' to obtain the standard error employed to build the confidence interval along with the gaussian quantiles.

Table 2 displays the CL of the aforementioned 95% CIs for the same scenarios reported in Table 1 based on 200 replicates.

Results get better when both the error and the random effect variance keep at lower values, although σ_k appears to have higher impact on the coverage levels and average widths as well. The effect of sample size is less noticeable and apparently vanishes when heterogeneity in breakpoint is large. Overall we can consider the bootstrap-based confidence intervals reasonable acceptable, but not fully satisfactory. This may not be surprising at all, as the bootstrap is not intended to be valid in every context; for instance it does not work for the median where the objective function is not smooth (Ghosh *et al.*, 1984). We conjecture that a similar problem related to the unsmoothness of the objective function affects the bootstrap paradigm for changepoint problems. However it should be emphasized that interval estimation for

Table 1 Mean and standard deviation (values × 10) of the sampling distributions for the parameter estimators in the segmented mixed model. True fixed effects: $\beta_0 = 2$, $\beta_1 = -0.2$, $\delta_0 = 0.3$, $\delta_1 = 0.5$; $\kappa_0 = 0.4$ and $\kappa_1 = 0.2$ for the linear changepoint model, while $\kappa_0 = -0.3$ and $\kappa_1 = 0.7$ for the logit changepoint model.

		Linear				Logit			
		$N = 50$		$N = 100$		$N = 50$		$N = 100$	
σ_ϵ		Mean	sd	Mean	sd	Mean	sd	Mean	sd
		$\sigma_k = 0.05$				$\sigma_k = 0.10$			
0.03	β_0	2.001	0.146	2.001	0.106	2.001	0.133	2.001	0.105
	β_1	−0.203	0.191	−0.200	0.143	−0.201	0.202	−0.201	0.127
	δ_0	0.303	0.317	0.300	0.225	0.301	0.319	0.302	0.222
	δ_1	0.497	0.412	0.499	0.304	0.503	0.416	0.497	0.285
	κ_0	0.398	0.187	0.400	0.137	−0.302	0.623	−0.301	0.442
	κ_1	0.200	0.255	0.199	0.192	0.698	0.862	0.688	0.604
		$\sigma_k = 0.10$				$\sigma_k = 0.20$			
	β_0	2.001	0.144	2.000	0.108	1.999	0.137	2.001	0.104
	β_1	−0.201	0.205	−0.199	0.140	−0.201	0.176	−0.202	0.136
	δ_0	0.299	0.341	0.297	0.233	0.301	0.328	0.302	0.233
	δ_1	0.500	0.462	0.497	0.310	0.499	0.432	0.498	0.300
	κ_0	0.401	0.288	0.399	0.205	−0.298	0.850	−0.297	0.536
	κ_1	0.199	0.403	0.201	0.296	0.682	1.118	0.674	0.758
		$\sigma_k = 0.05$				$\sigma_k = 0.10$			
0.05	β_0	2.001	0.145	2.002	0.110	2.000	0.166	2.001	0.104
	β_1	−0.206	0.264	−0.205	0.191	−0.204	0.237	−0.205	0.174
	δ_0	0.305	0.453	0.305	0.313	0.304	0.469	0.306	0.283
	δ_1	0.500	0.565	0.499	0.378	0.500	0.575	0.501	0.380
	κ_0	0.396	0.265	0.396	0.185	−0.304	0.951	−0.305	0.665
	κ_1	0.199	0.327	0.200	0.256	0.691	1.308	0.689	0.820
		$\sigma_k = 0.10$				$\sigma_k = 0.20$			
	β_0	2.004	0.157	2.003	0.112	2.000	0.154	2.001	0.109
	β_1	−0.210	0.309	−0.206	0.202	−0.207	0.272	−0.206	0.178
	δ_0	0.309	0.468	0.306	0.327	0.310	0.447	0.307	0.315
	δ_1	0.497	0.565	0.498	0.397	0.496	0.557	0.498	0.384
	κ_0	0.396	0.379	0.393	0.272	−0.310	1.172	−0.313	0.801
	κ_1	0.200	0.496	0.203	0.357	0.678	1.446	0.691	1.007

Source: Authors' own.

the changepoint is a quite hard and challenging problem; in fact no paper among those discussed in the Introduction and listed in literature deals with interval estimation for the changepoint and the other linear parameters of the piecewise model, except when the transition between segments is assumed to be smooth (Jacqmin-Gadda *et al.*, 2006). This comes out unsurprisingly since the model becomes somewhat more

Table 2 Coverage levels (CL) and average widths (AW) of the 95% CIs (using the standard normal quantiles and standard errors based on 100 nonparametric bootstrap resamples) for the changepoint parameters.

		Linear				Logit			
		$N = 50$		$N = 100$		$N = 50$		$N = 100$	
σ_ϵ		CL	AW	CL	AW	CL	AW	CL	AW
		$\sigma_k = 0.05$				$\sigma_k = 0.10$			
0.03	κ_0	0.906	0.07	0.920	0.05	0.921	0.23	0.934	0.16
	κ_1	0.916	0.10	0.940	0.07	0.921	0.32	0.929	0.22
		$\sigma_k = 0.10$				$\sigma_k = 0.20$			
	κ_0	0.912	0.11	0.910	0.08	0.901	0.29	0.927	0.20
	κ_1	0.909	0.15	0.909	0.11	0.907	0.40	0.963	0.29
		$\sigma_k = 0.05$				$\sigma_k = 0.10$			
0.05	κ_0	0.910	0.11	0.894	0.07	0.938	0.37	0.933	0.25
	κ_1	0.921	0.13	0.930	0.09	0.928	0.47	0.954	0.32
		$\sigma_k = 0.10$				$\sigma_k = 0.20$			
	κ_0	0.917	0.14	0.905	0.10	0.907	0.41	0.898	0.29
	κ_1	0.915	0.19	0.894	0.13	0.938	0.53	0.918	0.38

Source: Authors' own.

regular with continuous first and second log likelihood derivatives and relatively more conventional inferential tools.

4 Application

We illustrate the aforementioned methodology with data from a randomized controlled trial on treatments for depression carried out in Seattle, Washington between 1998 and 2001. All participants met criteria for a major depressive episode based on the Diagnostic and Statistical Manual of Mental Disorders. Out of almost 400 eligible subjects, about 150 individuals declined treatment or were excluded due to factors that contra-indicated outpatient treatment for depression (e.g., imminent risk of suicide or psychotic symptoms). At beginning of study, individuals were randomized in four treatment arms: psychotherapy via cognitive therapy, psychotherapy via behavioural activation, pharmacotherapy via antidepressant medication and placebo. However, based on previous analyses that revealed no important difference between cognitive therapy and behavioural activation (Dimidjian *et al.*, 2006; Gallop *et al.*, 2011), in this article we focus only on three treatment arms: any psychotherapy (PST), pharmacotherapy (MED) and placebo (PLA). Moreover substantive issues lead us to confine our attention to severely depressed subjects.

The aim of the trial was to assess effectiveness in treatment of depression. After randomization, treatment lasted for a total of about 16 weeks, except for the placebo condition, which only lasted 8 weeks (at which point, these individuals were offered any of the active treatments). At each treatment session, patients completed a self-report questionnaire of depression symptoms, the Beck Depression Inventory, hereafter BDI. The BDI is a widely used self-report symptom checklist of depressive symptoms: Twenty-one items are separately rated on a 0–3 scale from 0 representing the absence of the symptom and 3 representing the most intense level of the symptom. The total BDI score can range from 0 to 63 and the higher the score, the more serious the depression. The BDI was administered by the therapist or psychiatrist at every session, providing a detailed description of the course of depressive symptoms during the follow-up period.

In addition to BDI representing the outcome variable, information was obtained on gender, age and race. Table 3 reports some descriptive statistics for the $n = 121$ severely depressed subjects analyzed in this article.

The interest focuses on assessing the BDI response (bdi) profiles with respect to time (weeks) by emphasizing possible differences due to the treatment. In analyses of data from psychiatric trials, longitudinal profiles have been typically modelled using either a log transformation or polynomials (e.g., Christensen *et al.*, 2004); however, segmented modelling may provide a useful framework, possibly leading to better fit when individual profiles show two or more clear-cut phases. It should be emphasized that while alternative nonlinear models may provide better fits, specifically the semiparametric additive mixed model (Durban *et al.*, 2005), the segmented parameterization could be preferred on the grounds of its simple and interpretable parameters, slopes and changepoints in particular. Our guess is that semiparametric mixed models are not direct competitors of the segmented

Table 3 Description of study participants considered in the present article according to the three treatment groups: PST (psychotherapy), MED (pharmacotherapy) and PLA (placebo).

	Treatment group		
	PST	MED	PLA
No. of subjects	48	48	25
Race			
white	41	40	21
other	7	8	4
Gender			
male	17	14	6
female	31	34	19
Age (years)			
median (range)	41 (16–60)	36 (20–59)	45 (22–56)
Follow-up (weeks)			
median (range)	15.1 (4.4–18.3)	15.0 (0–16.6)	7.0 (1.0–8.7)

Source: Authors' own.

mixed model and instead they should be used when detection and estimation of changepoints is not the primary goal.

We fit a segmented mixed model for the sqrt-transformed BDI score to normalize the response and guarantee positive fitted values with regression equation

$$\sqrt{(bdi+1)}_{ij} = \beta_{0i} + \beta_{1i}\,\mathsf{weeks}_{ij} + \delta_i(\mathsf{weeks}_{ij} - \psi_i)_+ + \varepsilon_{ij}, \tag{4.1}$$

where we assume $\varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$ and each parameter depending on the subject treatment, namely

$$\begin{aligned}
\beta_{0i} &= \beta_{00} + \beta_{0P}\mathsf{PST}_i + \beta_{0M}\mathsf{MED}_i + b_{0i}\\
\beta_{1i} &= \beta_{10} + \beta_{1P}\mathsf{PST}_i + \beta_{1M}\mathsf{MED}_i + b_{1i}\\
\delta_i &= \delta_0 + \delta_P\mathsf{PST}_i + \delta_M\mathsf{MED}_i + d_i\\
\log\left(\frac{\psi_i - a_1}{a_2 - \psi_i}\right) &= \kappa_0 + \kappa_P\mathsf{PST}_i + \kappa_M\mathsf{MED}_i + k_i,
\end{aligned}$$

where $a_1 = 0$, $a_2 = 18.3$ weeks, while PST and MED are the two dummies for the treatment variable having the placebo as baseline category: $\mathsf{PST}_i = 1$ if subject i experienced psychotherapy and zero otherwise and $\mathsf{MED}_i = 1$ when subject i experienced pharmacotherapy and zero otherwise. b_{0i}, b_{1i}, d_i and k_i are the zero-mean Gaussian random effects with covariance matrix $\boldsymbol{R}$. As previously discussed, specification of $\boldsymbol{R}$ does not affect the outline of the estimating algorithm, but covariance matrices with more parameters can guarantee better fits. To illustrate we fit two segmented with different covariance matrix $\boldsymbol{R}$: diagonal, i.e., independent random effects and block diagonal where only covariances between random effects of intercept and left slope ($\sigma_{b_0 b_1}$), and difference-in-slope and breakpoint (σ_{dk}) are estimated. The resulting two marginal log likelihoods are –1459.7 and –1452.1 suggesting that the two covariance parameters cannot be set to zero. Parameter estimates from the segmented linear mixed model with a block diagonal covariance matrix are reported in Table 4.

As it would be expected from randomization design, no significant difference was found among the three treatment groups in the BDI values at the beginning of study, although patients with medical treatment exhibit higher values with respect to others. As a consequence of treatments, depression starts to decrease at beginning of study, but patterns of decay differ among the treatment groups and as a function of follow-up times. In the early follow-up all three treatment groups share significant drop in the BDI values; it turns out medical treatment guarantees the steepest decline, while psychotherapy and placebo exhibit almost identical decrease

Table 4 Parameter estimates from the segmented mixed model with a block diagonal covariance matrix for the random effects. The covariance estimates $\hat{\sigma}_{dk}$ and $\hat{\sigma}_{b_0b_1}$ are omitted, see text for details.

		Fixed		Random
Terms		Est.	SE	st.dev
Intercept	β_{00}	5.472	0.181	0.792
	β_{0P}	0.112	0.218	
	β_{0M}	0.415	0.224	
left slope	β_{10}	−0.239	0.047	0.095
	β_{1P}	0.065	0.051	
	β_{1M}	−0.193	0.063	
diff slope	δ_0	0.423	0.085	0.131
	δ_P	−0.359	0.089	
	δ_M	−0.057	0.097	
breakpoint	κ_0	−1.079	0.221	0.671
	κ_P	0.483	0.291	
	κ_M	−0.399	0.269	

Source: Authors' own.

rates. As previously discussed in the Introduction, patterns of recovery is two-phases and our data makes no exception: the estimated changepoints for the placebo, psychotherapy and pharmacotherapy groups are 4.64, 6.50 and 3.40 respectively: these are obtained via Equation (2.2) with $a_1 = 0$ and $a_2 = 18.3$. In the late follow-up, i.e., after the estimated breakpoints, the BDI values rebound but with apparent and important differences. Now the placebo and pharmacotherapy arms share the same behaviour and psychotherapy exhibits the lowest rebound.

To test for a different segmented curve among the three treatment groups, it is possible to test for nonzero contrast parameters using corresponding Wald statistics. Thus, concerning the left slope, the medical treatment is significatively different from the placebo ($\hat{\beta}_{1M} = -0.193$, p-value = 0.002), while no significant difference appears for the psychotherapy group ($\hat{\beta}_{1P} = 0.065$, p-value = 0.20). On the other hand, the difference-in-slopes parameter turns out to be different for the psychotherapy treatment arm ($\hat{\delta}_P = -0.359$, p-value < 0.0001), but not for medical treatment group ($\hat{\delta}_M = -0.057$, p-value = 0.56). Finally no important difference comes out among the changepoints since both the contrast parameter estimates are not significant ($\hat{\kappa}_P = 0.483$, p-value = 0.097) and ($\hat{\kappa}_M = -0.399$, p-value = 0.14). Figure 1 displays the fitted mean profiles for the three treatment groups.

Variances of the random effects appear to be non-negligible, especially for the difference-in-slopes ($\hat{\sigma}_d = 0.131$), suggesting higher heterogeneity at late follow-up. The smaller variance for the left slope ($\hat{\sigma}_{b_1} = 0.095$) emphasizes that most patients have quite similar slope at early follow-up. The two estimated correlations not

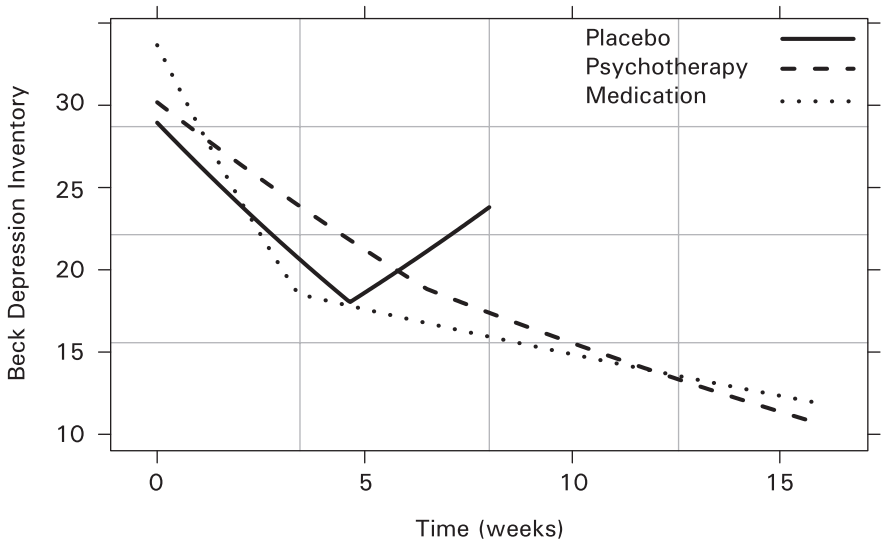

Figure 1 Fitted BDI score mean profiles (on the original scale) for the three treatment groups
Source: Authors' own.

reported in Table 4 are $r_{b_0 b_1} = 0.211$ and $r_{dk} = 0.697$; the latter correlation between the $\hat{d}_i$s and the $\hat{k}_i$s indicates that, regardless of the treatment group, subjects having later breakpoints also tend to experience more pronounced rebound. The model residual standard deviation is estimated 0.514.

Figure 2 summarizes for 12 selected individuals the observed data along with fitted values from the chosen segmented mixed model. Due to random effects in each model parameter, we note that the fitted segmented mixed model is able to capture quite different profiles in the observed data, including linear profiles without changepoints: this feature would allow to identify patients with a simple linear, rather than piecewise linear, decay in the response values. However this topic needs more research and conclusions should be taken with care. In this respect we remark that the structure of $\boldsymbol{R}$ which affects predictions of random effects even in standard LMM or NLMM, can become crucial when interest is in identifying subjects without changepoint. As an example Figure 3 compares individual predictions from the two aforementioned segmented mixed models for four subjects.

When the piecewise linear relationship is clear cut enough differences are negligible, see subject #196. However in other cases differences can be considerable, as in subject #131 where the simpler segmented model with independent random effects does not succeed in identifying the changepoint. Subjects #192 and #209 also show somewhat important differences among the fitted lines. Therefore in presence of high heterogeneity and flat segmented relationships, selection of structure of $\boldsymbol{R}$ should be done with care.

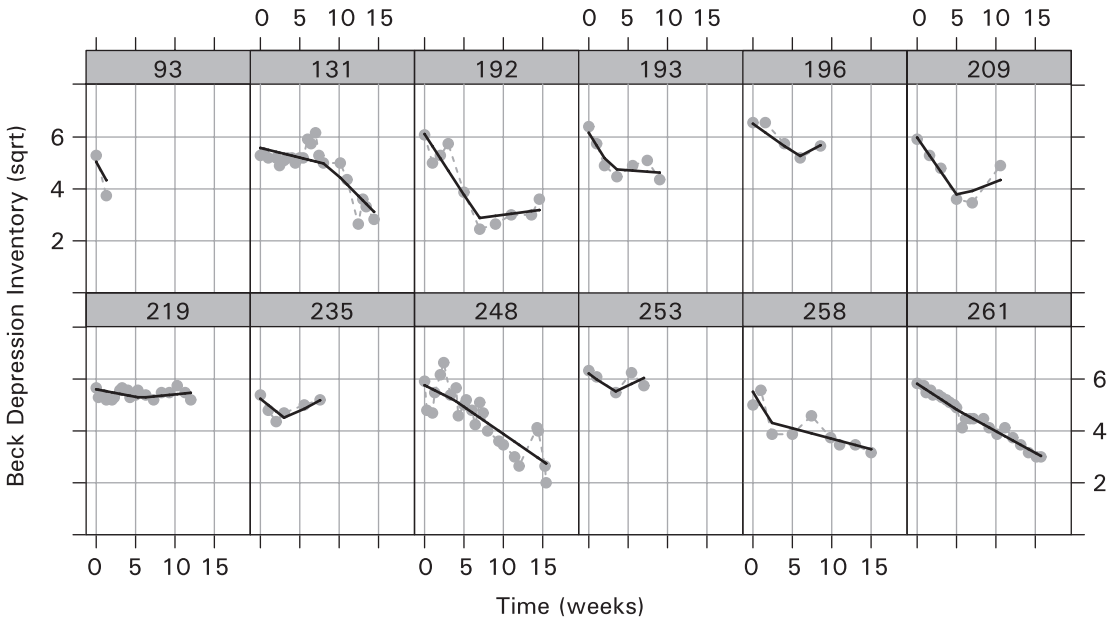

Figure 2 Observed data points for 12 selected individuals along with fitted piecewise linear profiles from the segmented mixed models with block diagonal random effects covariance matrix

Source: Authors' own.

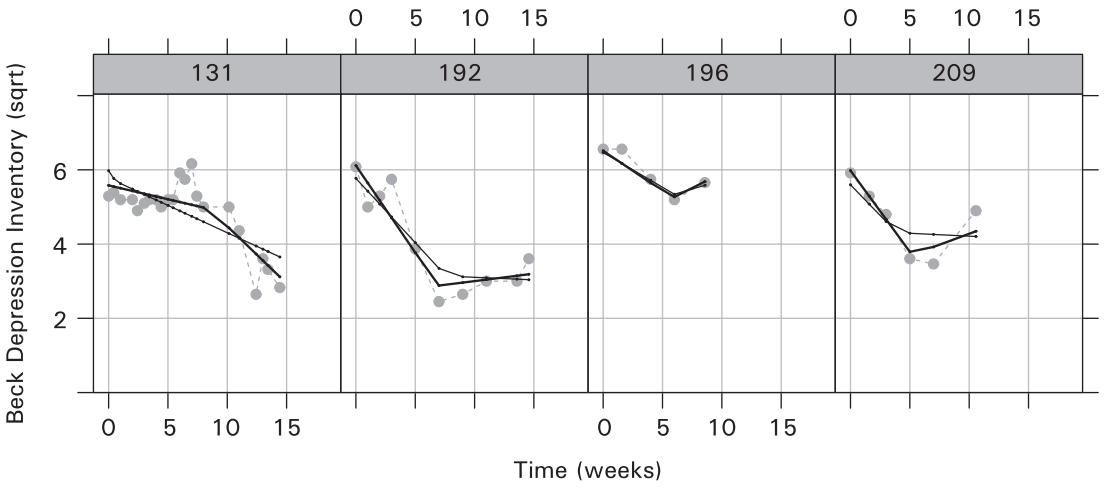

Figure 3 Contrasting individual predictions from two segmented mixed models for four subjects. The thick lines come from the model with block diagonal ***R*** and thin lines from the model with diagonal ***R***

Source: Authors' own.

5 Conclusion

The piecewise linear mixed model with random changepoints has been employed in several experimental contexts in biological, education and behavioural sciences; sometimes it may be preferred over more flexible approaches, e.g., additive mixed models, due to its comprehensible and interpretable parameters, slopes and changepoints. However its wider application has been probably limited by the computational burden which is not negligible in practice. We have presented a simple and effective estimation approach which represents a possible and useful alternative to the Bayesian approach. The proposed method may be helpful to researchers who are not familiar, or are not willing to work within the Bayesian framework. The main appeal of the proposed method lies in its relative simplicity, as the problem is transplanted into an LMM framework where the computational burden is considerably reduced. The alternative approach based on explicit estimation of a nonlinear LMM is far from being simple and effective. When the segmented relationships are bland, estimation can become problematic due to presence of multiple maxima of the objective function. To escape them, we have employed the bootstrap restarting algorithm which appears to be quite effective in practice.

Some authors discuss the distribution of the random effects for changepoint is likely to be skewed (e.g., Scott *et al.*, 2004; Ghosh and Vaida, 2007); although we still assume Gaussian distribution, our guess is that logistic transformation can deal with asymmetric distributions for the changepoints.

Although model estimation is feasible, some inference procedures are open research questions. Beyond interval estimation discussed in the simulation section, additional inferential issues relate to testing for heterogeneity among the changepoints and testing for the existence of a (fixed effect) changepoint. As previously sketched, the inferential concerns of the random changepoint models have not been discussed in literature and currently represent challenging research points. Another point worth investigating is the generalization to changepoints with smooth transitions, such as linear-quadratic segmented where the linear spline terms is replaced by $(t_{ij} - \psi)_+^2$.

From a practical viewpoint, the method is likely to be favoured as it may be easily implemented with existing statistical language having facilities for estimation of linear mixed models. R code is currently available from the first author and a set of functions is planned to be included in the R package segmented which currently handles only fixed effects segmented regression (Muggeo, 2008).

Acknowledgements

We would like to thank the associate editor and two referees whose comments have greatly improved the manuscript. The treatments for depression randomized trial was supported by National Institute of Mental Health Grant MH55502.

References

Christensen A, Atkins DC, Berns SB, Wheeler J, Baucom DH and Simpson L (2004) Integrative versus traditional behavioural couple therapy for moderately and severely distressed couples. *Journal of Consulting and Clinical Psychology*, **72**, 176–91.

Cudeck R and Harring JR (2007) Analysis of nonlinear patterns of change with random coefficient models. *Annual Review of Psychology*, **58**, 615–37.

Davidian M and Giltinan D (1995) *Nonlinear models for repeated measurement data*. New York: Chapman & Hall.

Dimidjian S, Hollon S, Dobson K, Schmaling K Kohlenberg R, Addis M, Gallop R, McGlinchey J, Markley D, Gollan JK, Atkins D and Dunner D (2006) Behavioural activation, cognitive therapy, and antidepressant medication in the acute treatment of major depression. *Journal of Consulting and Clinical Psychology*, **74**, 658–70.

Dominicus A, Ripatti S, Pedersen NL and Palmgren J (2008) A random change point model for assessing variability in repeated measures of cognitive function. *Statistics Medicine*, **27**, 5786–98.

Durban M, Harezlak J, Wand MP and Carroll RJ (2005) Simple fitting of subject-specific curves for longitudinal data. *Statistics in Medicine*, **24**, 1153–67.

Efron B and Tibshirani RJ (1993) *An Introduction to bootstrap*. New York: Chapman & Hall.

Feder P (1975) On asymptotic distribution theory in segmented regression problems—identified case. *Annals of Statistics*, **3**, 49–83.

Fitzmaurice GM, Laird NM and Ware JH (2004) *Applied Longitudinal Analysis*. New York: Wiley.

Gallop RJ, Dimidjian S, Atkins DC and Muggeo V (2011) Quantifying treatment effects when flexibly modeling individual change in a nonlinear mixed effects model. *Journal of Data Science*, **9**, 243–59.

Ghosh M, Parr WC, Singh K and Babu GJ (1984) A note on bootstrapping the sample median. *Annals of Statistics*, **12**, 1130–35.

Ghosh P and Vaida F (2007) Random changepoint modelling of HIV immunological responses. *Statistics in Medicine*, **26**, 2074–87.

Hall CB, Lipton RB, Sliwinski M and Stewart WF (2000) A change point model for estimating the onset of cognitive decline in preclinical Alzheimer's disease. *Statistics in Medicine*, **19**, 1555–66.

Hall CB, Ying J, Kuo L and Lipton RB (2003) Bayesian and profile likelihood change point methods for modeling cognitive function over time. *Computational Statistics & Data Analysis*, **42**, 91–109.

Ilardi SS and Craighead WE (1994) The role of nonspecific factors in cognitive-behaviour therapy for depression. *Clinical Psychology: Science and Practice*, **1**, 138–56.

Jacqmin-Gadda H, Commenges D and Dartigues J-F (2006) Random changepoint model for joint modeling of cognitive decline and dementia. *Biometrics*, **62**, 254–60.

Kiuchi AS, Hartigan JA, Holford TR, *et al.* (1995) Change points in the series of T4 counts prior to AIDS. *Biometrics*, **51**, 236–48.

Lange N, Carlin BP and Gelfand AE (1992) Hierarchical bayes models for the progression of HIV infection using longitudinal CD4 T-cell numbers (with discussion). *Journal of American Statistical Association*, **87**, 615–32.

Morrell CH, Pearson JD, Carter HB and Brant LJ (1995) Estimating unknown transition times using a piecewise nonlinear mixed-effects model in men with prostate cancer. *Journal of the American Statistical Association*, **90**, 45–53.

Muggeo VMR (2003) Estimating regression models with unknown break-points. *Statistics in Medicine*, **22**, 3055–71.

Muggeo VMR (2008) Segmented: An R package to fit regression models with broken-line relationships. *R News*, **8**(1), 20–25.

Muniz Terrera G, van de Hout A and Matthews FE (2011) Random change point models: Investigating cognitive decline in the presence of missing data. *Journal of Applied Statistics*, **38**, 705–16.

Naumova EN, Musta A and Laird NM (2001) Evaluating the impact of critical periods in longitudinal studies of growth using piecewise mixed effects models. *International Journal of Epidemiology*, **30**, 1332–41.

Pinheiro J and Bates D (2000) *Mixed-effects models in S and S-PLUS*. Springer.

Scott MA, Norman RG and Berger K (2004) Modelling growth and decline in lung function in Duchenne's muscolar dystrophy with an augmented linear mixed effects model. *Applied Statistics*, **53**, 507–21.

van der Leeden R, Meijer E and Busing FM (2007) Resampling multilevel models. In J. de Leeuw and E Meijer, eds, *Handbook of multilevel analysis*, chapter 11. Springer.

Wolfinger RD and Lin X (1997) Two Taylor-series approximation methods for non-linear mixed models. *Computational Statistics and Data Analysis*, **25**, 465–90.

Wood SN (2001) Minimizing model fitting objectives that contain spurious local minima by bootstrap restarting. *Biometrics*, **57**, 240–44.