université de BORDEAUX

Thèse présentée pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE BORDEAUX

École Doctorale Sociétés, Politique, Santé Publique
Spécialité Santé Publique, option Biostatistique

Par Corentin Segalas

Inférence dans les modèles à changement de pente aléatoire : application au déclin cognitif pré-démence

Inference on random changepoint models: application to pre-dementia cognitive decline

Sous la direction d'Hélène Jacqmin-Gadda

Soutenue le 3 Décembre 2019

Membres du jury

Leclercq Samson Adeline	Pr, Université Grenoble Alpes, Grenoble	Rapporteure
Muggeo Vito M. R.	PA, Université de Palerme, Italie	Rapporteur
Guedj Jérémie	CR, INSERM, Paris	Examinateur
Proust-Lima Cécile	CR, INSERM, Bordeaux	Examinatrice
Thiebaut Rodolphe	PU-PH, Université de Bordeaux	Président
Jacqmin-Gadda Hélène	DR, INSERM, Bordeaux	Directrice de thèse

université de BORDEAUX

Thèse présentée pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE BORDEAUX

École Doctorale Sociétés, Politique, Santé Publique
Spécialité Santé Publique, option Biostatistique

Par Corentin Segalas

Inférence dans les modèles à changement de pente aléatoire : application au déclin cognitif pré-démence

Inference on random changepoint models: application to pre-dementia cognitive decline

Sous la direction d'Hélène Jacqmin-Gadda

Soutenue le 3 Décembre 2019

Membres du jury

Leclercq Samson Adeline	Pr, Université Grenoble Alpes, Grenoble	Rapporteure
Muggeo Vito M. R.	PA, Université de Palerme, Italie	Rapporteur
Guedj Jérémie	CR, INSERM, Paris	Examinateur
Proust-Lima Cécile	CR, INSERM, Bordeaux	Examinatrice
Thiebaut Rodolphe	PU-PH, Université de Bordeaux	Président
Jacqmin-Gadda Hélène	DR, INSERM, Bordeaux	Directrice de thèse

Inférence dans les modèles à changement de pente aléatoire : application au déclin cognitif pré-démence

Résumé : Le but de ce travail a été de proposer des méthodes d'inférence pour décrire l'histoire naturelle de la phase pré-diagnostic de la démence. Durant celle-ci, qui dure une quinzaine d'années, les trajectoires de déclin cognitif sont non linéaires et hétérogènes entre les sujets. Pour ces raisons, nous avons choisi un modèle à changement de pente aléatoire pour les décrire. Une première partie de ce travail a consisté à proposer une procédure de test pour l'existence d'un changement de pente aléatoire. En effet, dans certaines sous-populations, le déclin cognitif semble lisse et la question de l'existence même d'un changement de pente se pose. Cette question présente un défi méthodologique en raison de la non-identifiabilité de certains paramètres sous l'hypothèse nulle rendant les tests standards inutiles. Nous avons proposé un supremum score test pour répondre à cette question. Une seconde partie du travail concernait l'ordre temporel du temps de changement entre plusieurs marqueurs. La démence est une maladie multidimensionnelle et plusieurs dimensions de la cognition sont affectées. Des schémas hypothétiques existent pour décrire l'histoire naturelle de la démence mais n'ont pas été éprouvés sur données réelles. Comparer le temps de changement de différents marqueurs mesurant différentes fonctions cognitives permet d'éclairer ces hypothèses. Dans cet esprit, nous proposons un modèle bivarié à changement de pente aléatoire permettant de comparer les temps de changement de deux marqueurs, potentiellement non gaussiens. Les méthodes proposées ont été évaluées sur simulations et appliquées sur des données issues de deux cohortes françaises. Enfin, nous discutons les limites de ces deux modèles qui se concentrent sur une accélération tardive du déclin cognitif précédant le diagnostic de démence et nous proposons un modèle alternatif qui estime plutôt une date de décrochage entre cas et non-cas.

Mots clés: Démence, modèles mixtes, données longitudinales multivariées, paramètres de nuisance non identifiables, changement de pente aléatoire, test du score.

Inference for random changepoint models: application to pre-dementia cognitive decline

Abstract: The aim of this work was to propose inferential methods to describe natural history of the pre-diagnosis phase of dementia. During this phase, which can last around fifteen years, the cognitive decline trajectories are nonlinear and heterogeneous between subjects. Because heterogeneity and nonlinearity, we chose a random changepoint mixed model to describe these trajectories. A first part of this work was to propose a testing procedure to assess the existence of a random changepoint. Indeed, in some subpopulations, the cognitive decline seems smooth and the question of the existence of a changepoint itself araises. This question is methodologically challenging because of identifiability issues on some parameters under the null hypothesis that makes standard tests useless. We proposed a supremum score test to answer this question. A second part of this work was the comparison of the temporal order of different markers changepoint. Dementia is a multidimensional disease where different dimensions of the cognition are affected. Hypothetic cascade models exist for describing this natural history but have not been evaluated on real data. Comparing change over time of different markers measuring different cognitive functions gives precious insight on this hypothesis. In this spirit, we propose a bivariate random changepoint model allowing proper comparison of the time of change of two cognitive markers, potentially non Gaussian. The proposed methodologies were evaluated on simulation studies and applied on real data from two French cohorts. Finally, we discussed the limitations of the two models we used that focused on the late acceleration of the cognitive decline before dementia diagnosis and we proposed an alternative model that estimates the time of differentiation between cases and non-cases.

Key words: Dementia, mixed models, multivariate longitudinal data, non identifiable nuisance parameters, random changepoint, score test.

INSERM U1219 Bordeaux Population Health Research Center
Equipe Biostatistique - 146 rue Léo Saignat - 33076 Bordeaux Cedex

À Juliette Boudon, née Sintès

Contents

Remerciements

Je tiens avant tout à remercier ma directrice, Hélène. Il y a plus de trois ans, tu m'as fait confiance en m'offrant l'opportunité d'entreprendre cette thèse. Ces trois années à tes côtés, dans un climat d'immuable bienveillance, ont été un apprentissage constant. Si j'ai beaucoup oscillé sur mes choix de carrière, j'ai au moins une certitude, celle d'espérer un jour pouvoir donner autant que ce tu m'as donné.

Je remercie les membres du jury. M. Thiebaud, Rodolphe, merci d'avoir accepté de présider ce jury et pour l'inépuisable énergie que vous déployez au service du centre. I also thank Mr. Muggeo and Mrs. Leclercq-Samson for evaluating my work. Je remercie mes examinateurs, M. Guedj et Mme Proust-Lima. Cécile, tu es de ces chercheurs à la vocation innée, ton jaillissement scientifique est une inspiration.

Maintenant, laissez-moi vous conter l'aventure d'un voyageur imprudent, qui, comme Ulysse, a fait un beau voyage. Ce voyage, mon voyage, n'a heureusement pas duré une décennie mais trois belles années, au temps désarticulé. Tout commença en octobre 2016 alors qu'un grand bateau se dessinait, là-bas. J'embarquai dans ce mystérieux navire à l'itinéraire incertain qui, selon la légende, devait à jamais m'ouvrir l'horizon. Dès mes premiers pas sur le pont, la pétillante Anaïs me guida et m'intégra à l'équipage. Après une longue escale dans la verte campagne d'une île mythologique, elle est revenue nous éclairer de son allégresse.

Cette aventure, je l'ai partagé avec une riche palette de personnages dont certains m'ont particulièrement cotoyé. Camille, avec qui, jour après jour, pause après pause, jusqu'au débarquement et en passant par l'inoubliable escale catalane, nos destins se sont croisés et soutenus. Sophie, corsaire bretonne, avec qui nous avons exploré zincs et pics au fil de l'eau. Et enfin Maude, bohème de la troupe, avec qui nous

avons parsemé l'écume de nos ondes radiophoniques.

Ma cabine, la numéro 40, a abrité nombre d'âmes. Les anciens, Rémi et Bachirou, sages parmi les sages, les nouveaux qui ne le sont plus vraiment, Mathilde et Denis ou Jocelyn, dernière recrue. Beaucoup de courts séjours, Julien et Céline avec qui j'ai partagé le goût de la plume ou Ina et Cristina venues de mers lointaines. Au delà des murs de la cabine, je pense à tous ceux qui ont partagé mon pont. L'efficace et souriante Sandrine, Viviane, experte *ès* pâtisserie et charmeuse d'Avakas, Casimir, comparse du gang des *cs*, Emilie camarade d'expéditions citadines et enfin Karen, qui était là quand est venue la tempête venue d'Orient.

Je n'oublie pas la cabine Michel Foucault qui abrite les éblouissantes Michettes, Gwendoline ou Gweuh pour les plus intimes, Leslie jamais sans Maddie, nos chères bretonnes Aline et Virginie, l'auréolée Perrine, la basque Bénédicte, Morgane des quatre coins de l'hexagone et l'ineffable Soufiane. Je n'oublie pas la cabine John Snow, qu'ont habité Chloé qui a depuis migré dans les nuages d'outre océan, les chantants Loïc et Perrine, le compatriote Henri, les inséparables autant que mystérieuses Myriam et Solenne, l'aérienne Laura, Bruno l'homme des machines et mes camarades Hadrien et Jean Noël embarqués en même temps que moi. Enfin, je n'oublie pas tous ceux rencontrés dans la guilde, au gré du vent.

A l'inverse de Major Tom, mes liens avec l'extérieur ne cessèrent jamais. Le plus ancien, Clément qui arpente maintenant le Piton. Les compagnons d'études, l'intendante Lola, le maître des cartes et jetons Pierre-Jakez et l'algébrique Emmanuelle. Les amis JE et tout particulièrement la paire William et Lorène et nos aventures sous pavillon vert. Puis les rencontres girondines, l'intrépide Noémie, Lola exilée en terres nordiques, Bérangère reine de la piste et dont on partage la passion du synthé avec Toscane et Hector et enfin le mythique équipage de *la Colo*.

Tout ceci n'aurait jamais pu commencer ni aboutir sans le soutien sans faille de ma tribu, nichée dans les vallées pyrénéennes. En particulier mon frère Robin qui a tracé le chemin et mes parents Hélène et Pascal, infatigables supports. Si je remercie tout le monde, c'est à eux que je dois le plus. Quel est ce parfum que le vent dépêche ? Plein de raisons d'être heureux, me voilà prêt à débarquer, les yeux fixés au large et les cheveux au vent, paré pour de nouvelles aventures...

Résumé substantiel

Motivations

La motivation principale de ce travail a été l'étude de l'histoire naturelle de la démence. La démence est un syndrome qui affecte les capacités cognitives et impacte la vie quotidienne des malades. Parmi les causes de démence, on retrouve principalement la maladie d'Alzheimer, 60 à 70% des cas selon l'OMS (WHO, 2017). La démence est le résultat d'un processus de dégradation progressif qui peut durer entre dix et quinze ans (Amieva *et al.*, 2008, 2014) et se distingue d'un vieillissement cognitif normal (Belleville *et al.*, 1996; Machulda *et al.*, 2013). L'OMS a classé la démence comme la cinquième cause de mortalité au monde (WHO, 2017). En 2018, selon l'association *Alzheimer's Disease International* (Patterson, 2018), 50 millions de personnes dans le monde étaient atteintes de démence et ce nombre devrait atteindre 152 millions d'ici 2050. Actuellement, un nouveau cas est identifié toutes les trois secondes.

Ces dernières décennies, la recherche médicale s'est intéressée au développement de traitements pour la maladie d'Alzheimer. Néanmoins, ceux-ci ont été conçus pour cibler les symptômes plus que les causes de la maladie qui restent mal identifiées. C'est pourquoi un champ majeur de la recherche actuelle se concentre à mieux comprendre l'histoire naturelle de la maladie d'Alzheimer. Cela permettrait de détecter la maladie à un stade suffisamment précoce pour permettre le développement de traitements préventifs ciblant une population identifiée comme à risque (Aisen *et al.*, 2011).

L'histoire naturelle de la démence a fait l'objet de nombreuses études (Hubbard

et al., 1990; Jost et Grossberg, 1995; Beker *et al.*, 1994). Il est désormais acquis que durant la phase de pré-démence, les trajectoires de déclins cognitifs mesurées par des marqueurs psychométriques sont non linéaires avec une accélération du déclin cognitif qui se manifeste par un changement de pente (Wilson *et al.*, 2012; Rajan *et al.*, 2017; Li *et al.*, 2017) et ces trajectoires sont très hétérogènes entre sujets (Amieva *et al.*, 2014). De plus, le processus de dégradation semble se dérouler en étapes successives (Godbolt *et al.*, 2004; Amieva *et al.*, 2008). À partir de ces observations, les chercheurs ont émis l'htpothèse de l'existence d'une cascade pathologique atteignant successivement différentes dimensions de la cognition dont les étapes ont été décrites dans des modèles théoriques (Jack *et al.*, 2010, 2013; Verlinden *et al.*, 2016). Néanmoins, ces derniers restent hypothétiques et des méthodes statistiques sont nécessaires pour les valider.

Dans ce travail, nous avons choisi les modèles non linéaires mixtes à changement de pente aléatoire pour données longitudinales afin de modéliser les trajectoires de déclin cognitif tout en prenant en compte l'hétérogénéité entre individus. Le premier objectif de la thèse était de proposer un test pour l'existence de changement de pente aléatoire afin de déterminer si le déclin cognitif présente effectivement une accélération individuelle. Un second travail a été de proposer un modèle bivarié permettant la comparaison de la date de changement de pente de plusieurs marqueurs mesurant différentes fonctions cognitives.

Test pour l'existence du changement de pente aléatoire

On note Y_{ij} la valeur du marqueur Y pour le sujet i au temps t_{ij} avec $1 \leq i \leq N$ et $1 \leq j \leq n_i$. Pour modéliser la trajectoire d'un marqueur cognitif nous avons choisi un modèle inspiré par Bacon et Watts (1971) avec la fonction de transition proposée par Griffiths et Miller (1973)

$$Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \beta_2 \sqrt{(t_{ij} - \tau_i)^2 + \gamma} + \varepsilon_{ij}$$

avec

$$\tau_i = \mu_\tau + \sigma_\tau \tilde{\tau}_i \text{ et } \tilde{\tau}_i \sim \mathcal{N}(0, 1),$$

$$\beta_{ki} = \beta_k^\top X_{ki} + \alpha_{ki} \text{ pour } k = 0, 1,$$

$$\alpha_i = (\alpha_{0i}, \alpha_{1i})^\top \sim \mathcal{N}(0, \Sigma) \text{ avec } \Sigma = \begin{pmatrix} \sigma_0 & \sigma_{01} \\ \sigma_{01} & \sigma_1 \end{pmatrix} \text{ et } \varepsilon_{ij} \sim \mathcal{N}(0, \sigma).$$

Nous proposons une procédure de test en deux temps. Premièrement nous testons s'il existe un changement de pente aléatoire puis, si l'existence de ce changement de pente aléatoire est confirmée, nous testons si la différence de pente entre les deux phases est aléatoire.

L'absence de changement de pente aléatoire est définie par l'hypothèse nulle $H_0 : \beta_2 = 0$. Pour tester cette hypothèse nulle contre l'alternative $H_1 : \beta_2 \neq 0$, nous avons choisi le test du score. Cependant, sous H_0, les paramètres liés au changement de pente aléatoire (μ_τ, σ_τ) sont non identifiables. On ne peut donc calculer leurs estimations du maximum de vraisemblance sous H_0 et il est donc impossible d'utiliser la statistique du test du score classique S_N qui en dépend. Ce problème d'identifiabilité présente un réel défi méthodologique. Nous avons proposé de considérer comme statistique de test le supremum de S_N en les paramètres de nuisances non identifiables (μ_τ, σ_τ). Si cela permet de lever le problème d'identifiabilité, il reste néanmoins à déterminer la distribution asymptotique sous H_0 pour pouvoir calculer une p-valeur mais cette distribution n'a pas de forme analytique connue. Hansen (1996) a proposé une procédure de perturbation des contributions individuelles au score par des variables gaussiennes permettant d'échantillonner selon cette distribution. Ces échantillonnages permettent de calculer une p-valeur empirique et de conclure sur le test de l'existence du changement de pente aléatoire.

Cette procédure de test a été implémentée en `R` dans la fonction `testRCPMM` du package `rcpm` (*cf.* Annexe B) et validée par une étude de simulations. Elle a été appliquée à la cohorte française Paquid (Letenneur *et al.*, 1994) et a permis de mettre en évidence l'existence de changement de pente chez les déments de haut et bas niveau d'étude pour le test d'Isaacs (Isaacs et Kennie, 1973) qui évalue la fluence verbale (*cf.* Chapitre 3).

Si le test révèle l'existence d'un changement de pente aléatoire, il est maintenant possible de tester si la différence de pentes entre les deux phases varie d'un individu à l'autre. Autrement dit, il s'agit ici de tester la présence d'un effet aléatoire sur β_2.

Notons $\beta_{2i} = \beta_2 + \alpha_{2i}$ avec $\alpha_i = (\alpha_{0i}, \alpha_{1i}, \alpha_{2i}) \sim \mathcal{N}(0, \Sigma)$, le test peut être formulé par les hypothèses

$$H_0 : \Sigma = \begin{pmatrix} \sigma_0 & \sigma_{01} & 0 \\ \sigma_{01} & \sigma_1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ vs. } H_1 : \Sigma = \begin{pmatrix} \sigma_0 & \sigma_{01} & \sigma_{02} \\ \sigma_{01} & \sigma_1 & \sigma_{12} \\ \sigma_{02} & \sigma_{12} & \sigma_2 \end{pmatrix}.$$

Bien que ce type de test soit courant dans les modèles mixtes, il ne s'agit pas d'un test standard car, sous l'hypothèse nulle, certains paramètres atteignent les bornes de leur espace de définition. Stram et Lee (1994) ont proposé une approche rigoureuse pour réaliser ce test et ont montré que la distribution nulle asymptotique suivait un mélange de distribution du χ^2. Le modèle à changement de pente aléatoire complet

$$Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \beta_{2i} \sqrt{(t_{ij} - \tau_i)^2 + \gamma} + \varepsilon_{ij}$$

peut alors être estimé en maximisant la log-vraisemblance

$$\ell_N(Y; \theta) = \sum_{i=1}^{N} \log \frac{1}{\pi} \int \prod_{j=1}^{n_i} f(Y_{ij} | \tilde{\tau}_i) \exp(-\tilde{\tau}_i^2) \mathrm{d}\tilde{\tau}_i.$$

où θ contient tous les paramètres du modèle. Cette maximisation peut se faire en utilisant l'algorithme de Marquardt-Levenberg (Levenberg, 1944; Marquardt, 1963) par exemple. L'intégrale sur les effets aléatoires est approchée par la méthode de quadrature de Gauss-Hermite. Cette procédure d'estimation a été implémentée en `R` dans la fonction `rcpme` du package `rcpm` (*cf.* Annexe B).

Modèle bivarié curvilinéaire à changement de pente aléatoire et comparaison temporelle

Afin de comparer les changements de pente de deux marqueurs, il est nécessaire de les estimer simultanément car l'estimation de la covariance des temps moyens de changement de pente μ_τ est nécessaire pour réaliser ce test statistique. Pour

cela, nous proposons un modèle bivarié à changement de pente aléatoire basé sur le modèle proposé par Bacon et Watts (1971) avec la fonction de transition proposée par Griffiths et Miller (1973). Ce modèle bivarié s'écrit en notation matricielle

$$Y_i = \Gamma_i \beta_i + \varepsilon_i$$

où $\beta_i = (\beta_{0i}^1, \beta_{1i}^1, \beta_{2i}^1, \beta_{0i}^2, \beta_{1i}^2, \beta_{2i}^2)^\top \sim \mathcal{N}(\beta, B)$ avec $\beta = (\beta_0^1, \beta_1^1, \beta_2^1, \beta_0^2, \beta_1^2, \beta_2^2)^\top$, $\tilde{\tau}_i = (\tilde{\tau}_i^1, \tilde{\tau}_i^2)^\top \sim \mathcal{N}(0, D)$ et $\varepsilon_i \sim \mathcal{N}_{n_i}(0, \Sigma_i)$ avec

$$\Gamma_i = \left[\begin{array}{c|c} T_i^1 & 0_{n_i^1 \times 3} \\ \hline 0_{n_i^2 \times 3} & T_i^2 \end{array}\right], B = \left[\begin{array}{c|c} B^1 & B^{12} \\ \hline B^{21} & B^2 \end{array}\right], D = \begin{bmatrix} 1 & d_{12} \\ d_{12} & 1 \end{bmatrix},$$

$$\Sigma_i = \left[\begin{array}{c|c} \sigma_{\varepsilon^1} I_{n_i^1} & 0_{n_i^1 \times n_i^2} \\ \hline 0_{n_i^2 \times n_i^1} & \sigma_{\varepsilon^2} I_{n_i^2} \end{array}\right], T_i^l = \begin{bmatrix} 1 & t_{i1} - \tau_i^l & \sqrt{(t_{i1} - \tau_i^l)^2 + \gamma} \\ \vdots & \vdots & \vdots \\ 1 & t_{in_i^l} - \tau_i^l & \sqrt{(t_{in_i^l} - \tau_i^l)^2 + \gamma} \end{bmatrix}.$$

où $\mathbb{I}_N$ est la matrice identité de taille N. Ce modèle s'estime également en maximisant la log-vraisemblance

$$\ell_N(\theta) = \sum_{i=1}^{N} \log \int f(Y_i|\tilde{\tau}_i) f(\tilde{\tau}_i) \mathrm{d}\tilde{\tau}_i$$

où le vecteur θ contient tous les paramètres du modèle. La distribution conditionnelle de $Y_i|\tilde{\tau}_i$ est une gaussienne multivariée définie par

$$Y_i|\tilde{\tau}_i \sim \mathcal{N}(\Gamma_i \beta,\ \Gamma_i B \Gamma_i^\top + \Sigma_i).$$

Cette maximisation peut être réalisée par l'algorithme de Marquardt-Levenberg (Levenberg, 1944; Marquardt, 1963). L'intégrale sur les deux changements de pente aléatoires est à nouveau approchée par une méthode de quadrature de Gauss-Hermite. Pour réduire le nombre de noeuds tout en gardant une bonne précision, on utilise la quadrature pseudo-adaptative (Rizopoulos, 2012). Plus précisément,

grâce à l'estimation des deux modèles univariés, nous pouvons estimer les effets aléatoires individuels pour chaque marqueur. Cette information peut alors être utilisée pour recentrer la grille de la quadrature de Gauss-Hermite avant de lancer l'optimisation. L'hypothèse de normalité du marqueur longitudinal, souvent peu réaliste avec les marqueurs psychométriques, est ici assouplie. Une transformation du marqueur brut, basée sur des I-splines, est estimée en même temps que tous les autres paramètres du modèle. Cette transformation estimée permet à la procédure d'estimation de s'appliquer à des marqueurs non gaussiens sans hypothèse *a priori* sur la transformation.

Une fois l'estimation du modèle réalisée, la comparaison des temps moyens de changements de pente entre les deux marqueurs revient à tester $H_0 : \mu_\tau^1 - \mu_\tau^2 = 0$ par un test de Wald classique dont la statistique de test suit sous H_0 une distribution du χ^2.

Cette procédure d'estimation d'un modèle bivarié curvilinéaire à changement de pente aléatoire a été implémentée en `R` dans la fonction `bircpme` du package `rcpm` (*cf.* Annexe B). La procédure a été validée sur des simulations et appliquée à la cohorte 3C (3C Study Group, 2003). Nous avons comparé chez les cas les dates de changements de pente moyens de deux des scores de Grober et Buschke (Grober et Buschke, 1987), le rappel immédiat et le rappel libre qui évaluent respectivement la capacité d'encodage d'une information et la capacité de la mémoriser. L'estimation du modèle a montré que la capacité de mémorisation d'une information déclinait avant la capacité d'encodage.

Discussion et perspectives

En comparant nos résultats d'estimation du changement de pente à d'autres résultats de la littérature (Amieva *et al.*, 2014), nous constatons que le changement de pente estimé par les modèles que nous proposons représente l'accélération tardive du déclin cognitif précédent le diagnostic de démence et non la premiere accélération du déclin qui correspond au moment où le déclin cognitif normal se distingue d'un déclin pathologique. Une perspective intéressante serait, dans une étude cas-témoin nichée

dans une cohorte, d'estimer cette date de décrochage. À cette fin, nous proposons un modèle à classes latentes avec une classe linéaire et une classe à changement de pente, où, à partir de τ_i, la différence entre la trajectoire dans la seconde phase et la trajectoire linéaire est modelisée par une fonction flexible basée sur des I-splines. Ce modèle est actuellement en cours de développement.

Un autre point de discussion est l'interprétation du changement de pente au regard du temps de base choisi. En effet, dans les travaux présentés dans cette thèse, nous avons considéré comme temps de base le délai à la démence en ne travaillant que sur les cas diagnostiqués pendant le suivi. Dans un schéma cas-témoin niché dans une cohorte, il est toujours envisageable d'utiliser le délai à la démence comme temps de base en choisissant, pour les témoins qui n'ont par définition pas de délai à la démence, le délai à la démence du cas apparié. Néanmoins, dans le cadre de modèles plus complexes, comme les modèles conjoints qui permettent de modéliser simultanément le temps jusqu'à la démence et l'évolution longitudinale d'un marqueur cognitif, le temps de base est l'âge et l'interprétation du changement de pente est moins aisée.

Ces travaux méthodologiques permettent de décrire l'histoire naturelle de la maladie d'Alzheimer et de répondre à des questions cliniques d'intérêt tout en proposant une solution aux défis méthodologiques que soulèvent ces questions. La méthodologie développée dans cette thèse peut s'appliquer à toute autre pathologie dont la progression clinique peut être mesurée par un marqueur et à la condition qu'un changement de pente dans la trajectoire longitudinale de ce marqueur ait un réel sens clinique.

Chapter 1

Introduction

Before introducing the statistical methods used and developed during this work, we outline in this introductory chapter the main clinical application that guided the statistical developments presented in this document.

1.1 Dementia, a global public health issue

The main motivation for this work was the study of dementia. Dementia is a syndrome that affects cognitive abilities and daily life. The main cause of dementia is known to be the Alzheimer's Disease (AD) that represents around 60% to 70% of the cases according to the WHO (2017), much more frequent than vascular dementia and dementia with Lewy bodies. However, distinguishing between the different types of dementia is difficult and often characteristics of AD and vascular dementia can coexist. Dementia is the result of a long and progressive degradation process that can last for around ten to fifteen years (Amieva *et al.*, 2008, 2014) and differs from a normal ageing process (Belleville *et al.*, 1996; Machulda *et al.*, 2013). According to the DSM-IV (2000), dementia is defined as the manifestation of symptoms that lead to the loss of cognitive functions, such as memory, at such a scale that activities of daily life are impacted. This loss of autonomy often causes institutionalization of the patient and leads to death. The WHO (2017) has ranked AD and other dementias as the fifth leading cause of death worldwide in 2016 and even the

third one in high-income countries.

In 2018, according to Alzheimer's Disease International (Patterson, 2018), 50 million people worldwide had dementia and this number is expected to rise to 152 million by 2050. Indeed, because of a global increasing life expectancy, especially in middle and low income countries, the prevalence of dementia will mechanically increase. Currently, one new case is identified every three seconds. It represents a huge and heavy challenge for modern society which needs to address it by proposing quality care for patients and appropriate support for their family. According to figures from the Alzheimer's Disease International (Patterson, 2018), the global annual burden of dementia amounts to a trillion United States dollars a year.

Medical research has been focused on developing new drugs to treat AD. However, since 1988, only four of the one hundred attempts have lead to an approved drug (Patterson, 2018). Moreover, these drugs were only designed to treat symptoms, not the causes of the disease themselves. New drugs, called *disease modifying drugs* have been recently developed in order to control the evolution of the disease. Unfortunately, most clinical trials have failed and no such drug yet has been approved (Salomone *et al.*, 2012). This is why a major field of AD research now focuses on understanding the natural history of the disease. This would help detecting the disease at an earlier stage and then developing new drugs that could target an early pre-dementia population (Aisen *et al.*, 2011), that is before dementia onset.

1.2 The natural history of Alzheimer's Disease

The natural history of AD has been studied for some years now. Many publications focused on confirmed cases only through anatomic brain studies coupled with retrospective assessment of cognitive deterioration (Hubbard *et al.*, 1990), by retrospectively reviewing medical records including psychometric tests (Jost and Grossberg, 1995) or by selecting cases on longitudinal studies and reviewing their evolution (Becker *et al.*, 1994). Since then, the pre-dementia evolution of subjects who were disease-free at the inclusion but could develop dementia during the follow-up have been explored through longitudinal cohort data. With such data, researchers

could explore pre-dementia evolution over time through the repeated measurement of markers.

It has emerged that during this pre-dementia phase, the cognitive decline trajectory of markers are nonlinear, generally with an acceleration of the cognitive decline several years before the diagnosis which manifests through a changepoint that might depends upon individual characteristics (Wilson *et al.*, 2012; Rajan *et al.*, 2017; Li *et al.*, 2017). Also, decline trajectories present a wide heterogeneity between individuals not only on the changepoint date, but also on slopes parameters, which needs to be taken into account, (Amieva *et al.*, 2014).

Moreover, it has been found that the predementia phase was quite long and that a temporal order existed in the degradation process (Godbolt *et al.*, 2004; Amieva *et al.*, 2008). By comparing the temporal decline of some abilities and anatomic functions, researchers were able to build hypothetical theoretical schemes of the degradation process which led them to consider the development of the pathology as a continuum (Jack *et al.*, 2010, 2013; Verlinden *et al.*, 2016). In Figure 1.1, the hypothetical cascade model proposed by Jack *et al.* (2013) states that accumulation of amyloid β and tau proteins is followed by brain lesions which are then manifested through cognitive decline. The continuum hypothesis and cascade model are widely accepted (Dubois *et al.*, 2016) and this framework helps researchers to better understand preclinical AD and to plan future research orientation. However, they remain hypothetical proposals and statistical methods are needed to validate these hypotheses.

The evolution of cognitive abilities during the preclinical phase of AD has been divided in three states: normal cognition, mild cognitive impairment (MCI) and then dementia. The transitional MCI state is the focus of many research efforts (Flicker *et al.*, 1991). Indeed, the construct of the MCI state is useful to identify patients at risk of developing dementia before substantial damage has already happened. However, the issue that it is not a valid pathological condition can be raised (Petersen *et al.*, 2001; Petersen, 2004) and there is a consensus that MCI must be well identified and its definition standardized (Winblad *et al.*, 2004).

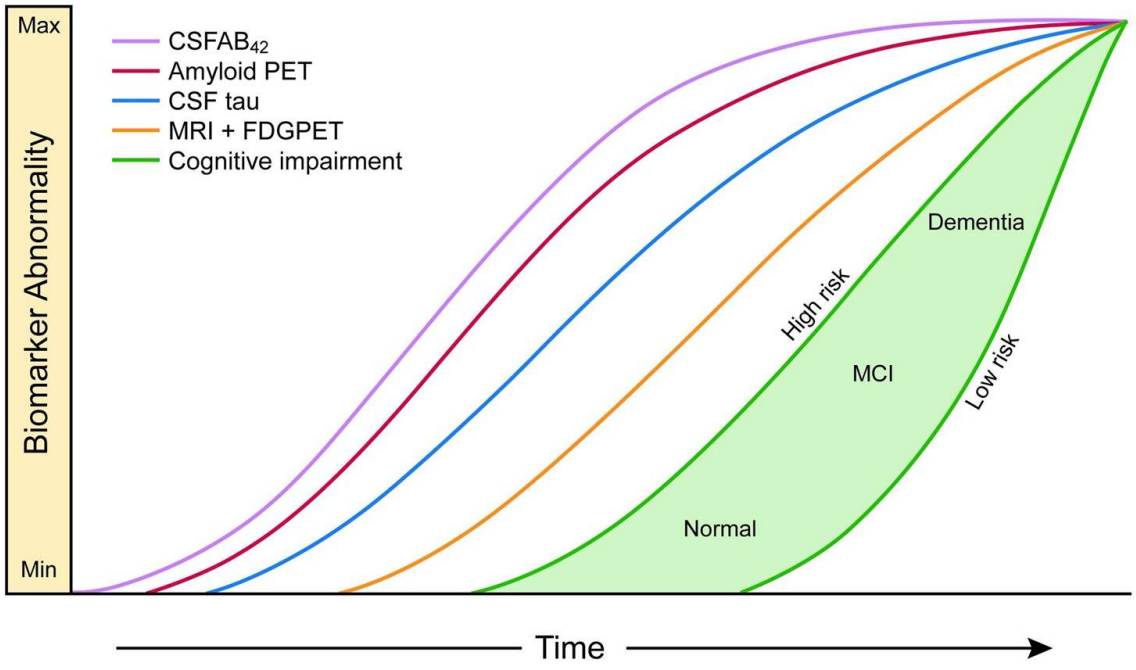

Figure 1.1: AD pathological cascade model from (Jack *et al.*, 2013)

1.3 Methodological challenges and objectives

The general motivation behind this work was to provide methodological tools to study the natural history of AD and in particular to identify and characterize a possible acceleration of the cognitive decline before dementia diagnosis. Previous studies using cohort data of subjects initially non-demented have shown that this decline is nonlinear and heterogeneous. To account for the correlation between repeated measures of cognitive markers and for the high between-subject variability in all the phases of the decline, we focused on mixed models with random changepoint. Indeed, compared to a polynomial mixed model, a subject-specific changepoint model has the advantage of directly estimating the time of acceleration of cognitive decline which can be interpreted as a delay to dementia diagnosis over the appropriate time scale.

However, the existence of a changepoint is not obvious for every subpopulations. For instance, the pre-dementia decline of subjects with low educational level is much smoother than the decline of highly educated subjects. This leads to an

interrogation about the existence of the random changepoint itself. A methodological issue is raised here as standard tests cannot be used for testing the existence of a fixed changepoint because of unidentifiability of some parameters under the null hypothesis. Moreover, in the literature, to our knowledge there is no proposal of a test for the existence of a random changepoint in longitudinal data. If a random changepoint exists, the time of change could be different according to some subjects characteristics or even for different cognitive markers. To tackle this interrogation, some methodological developments are required. Comparing the times of change of neuropsychological tests measuring different cognitive functions would give insight on the temporal order of decline of these abilities. This comparison can only be made by estimating a bivariate random changepoint models for the evolution of both markers.

The global methodological objective of this thesis was to propose inferential methods for mixed models with random changepoint. More specifically, the first objective was to propose a testing procedure to assess if there is an individual random changepoint in longitudinal cognitive decline trajectories. A second objective was to propose a methodology to compare the times of change for different markers based on a bivariate random changepoint model.

1.4 Cohorts on cognitive aging and dementia

An important part of the literature has already focused on studying the natural history of AD. For this purpose, several cohorts have been implemented with repeated measurements of psychometric scores and a longitudinal assessment of AD. During this work, two French cohorts have been used and are introduced below.

1.4.1 The Paquid cohort

The Paquid cohort is an epidemiologic study (Letenneur *et al.*, 1994) on cognitive ageing that was launched in 1988. A total of 3777 subjects from two French departments, Gironde and Dordogne, aged at least 65 years and living at home at the beginning of the study were included. At 1, 3, 5, 8, 10, 13, 15, 17, 20, 22 and

25 years after baseline, the participants completed a battery of neuropsychological tests during an interview with a psychologist. These interviews were held at home or at an institution if the subject had been transferred. A two-stage procedure was used to diagnose dementia. Subjects meeting the DSM IIIR criteria for dementia A, B and C (impairment of memory and at least one other cognitive function and interference with daily living) or subjects whose Mini Mental State Examination score had decreased by at least 3 points since the last visit were seen by a senior neurologist who made the final diagnosis. The cognitive marker we used was the Isaacs Set test (Isaacs and Kennie, 1973) of verbal fluency. This test requires the subject to quote a maximum of 10 words in 60 seconds from four different semantic categories: colors, animals, fruits, cities. The score is the number of words given by the subject and then ranges from 0 to 40. Due to the strong ceiling effect of the 60-sec test, we used the shortened 15-sec version (Proust-Lima *et al.*, 2007).

1.4.2 The Three-City study cohort

The Three-City Study (3C Study) is an observational cohort study started in France in 1999 aiming at understanding the link between dementia and vascular diseases in an elder population (3C Study Group, 2003). A total of 9.294 subjects from the French cities of Bordeaux, Montpellier and Dijon being at least 65 years old were recruited and followed-up over time 2, 4, 7, 10, 12 and 14 years after baseline. At each visit, subjects completed a battery of cognitive tests that partly differ between centres. The diagnosis of dementia was assessed at each visit in a two-step procedure: screening based on neuropsychological performance and final diagnosis made by a neurologist and evaluated by an independent committee. We were particularly interested in the results to the Grober and Bushke tests (GB) that measures memory functions (Grober and Buschke, 1987) through several recalls of 16 words. However those tests were not performed at baseline and at the fourth year visit so that only a maximum number of 5 measures per subjects was available in our sample.

1.5 Outline of the manuscript

In the next chapter, a state of the art reviewing the main statistical framework is presented. The third chapter deals with the testing procedure for the existence of a random changepoint for longitudinal data. The fourth chapter describes the bivariate random changepoint model that allows to compare the temporal order of decline of two different markers. In the fifth and last chapter, we discuss the proposed work and present some perspectives.

Chapter 2

State of the art

In this chapter, we describe the statistical challenges raised by our objectives and present some of the statistical tools proposed in the literature of random changepoint models to tackle them and discuss their interest. First, we quickly introduce mixed model theory and some numerical tools in Section 2.1. In Section 2.2, we introduce and discuss changepoint models and their application to neuropsychology. Section 2.3 focuses on literature on hypothesis testing for the existence of a changepoint from classic regression models to segmented regression and longitudinal data.

2.1 Mixed models for longitudinal data

2.1.1 From linear model to linear mixed models

The classic linear model

$$Y_i = X_i^\top \beta + \varepsilon_i$$

where for subject i, $i = 1, \dots, N$, $Y_i \in \mathbb{R}$ is the dependant variable, $X_i \in \mathbb{R}^p$ the vector of explanatory variables, $\beta \in \mathbb{R}^p$ the vector of regression coefficients and $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ the residual errors. Due to the independance of the residual errors, this model assumes that the observations Y_i are independent conditionally on X_i. However, in the context of repeated or grouped data, this assumption is violated and within subject or within group correlation must be taken into account. This

is why Laird and Ware (1982) proposed linear mixed models. They introduced into the model a random variable called the random effect that captures both the within subject or within group correlation and individual heterogeneity. From now on, we focus on repeated measures data only. Let Y_{ij} be the measure of subject i, $i = 1, \dots, N$ at time t_{ij}, $j = 1, \dots, n_i$ the classic linear mixed model formulation is

$$Y_{ij} = Y_i(t_{ij}) = X_{ij}^\top \beta + Z_{ij}^\top b_i + \varepsilon_{ij}. \tag{2.1}$$

where X_{ij} is the p-vector of regressors, β the p-vector of fixed effects, Z_{ij} is the q-vector of regressors, subvector of X_{ij} such as $q \leq p$, $b_i \sim \mathcal{N}(0, B)$ the q-vector of subject-specific random effects and $\varepsilon_i = (\varepsilon_{ij})_{j=1,\dots,n_i} \sim \mathcal{N}(0, \Sigma_i)$ the residual error assumed independent to b_i. As stated, the random effect b_i accounts for individual deviation from the mean trajectory but also for the within subject correlation.

For example, if Z_{ij} reduces to 1, then (2.1) is called the random intercept model. The random intercept measures the constant individual deviation from the mean trajectory over time. For a subject i and for $j \neq k$, the within subject covariance $cov(Y_{ij}, Y_{ik}) = var(b_i)$ remains constant over time which is not very realistic. If $Z_{ij}^\top = (1, t_{ij})$, (2.1) is a model with random intercept and slope, sometimes called growth curve model. In the latter case, not only the intercept but also the slopes varies between subjects and for $j \neq k$ the within subject correlation $cov(Y_{ij}, Y_{ik}) = Z_{ij}^\top var(b_i) Z_{ik}$ becomes time dependent.

2.1.2 Log-likelihood and estimation

Model (2.1) can be more conveniently written using matrix notation

$$Y_i = X_i \beta + Z_i b_i + \varepsilon_i. \tag{2.2}$$

where Y_i is the n_i-vector of observations for subject i, X_i is a $n_i \times p$ matrix of regressors, β the p-vector of fixed effects, Z_i is a $n_i \times q$ matrix of regressors, submatrix of X_i such as $q \leq p$, $b_i \sim \mathcal{N}(0, B)$ the q-vector of subject-specific random effects and $\varepsilon_i \sim \mathcal{N}(0, \Sigma_i)$ the n_i-vector of residual errors assumed independent from b_i and with Σ_i a positive definite matrix of size n_i. From (2.2), we can define the marginal

model

$$Y_i \sim \mathcal{N}(X_i\beta, V_i = Z_i B Z_i^\top + \Sigma_i) \tag{2.3}$$

from which a log-likelihood can be derived. We note α all the variance and covariance parameters intervening in V_i from (2.3) and $\theta = (\alpha^\top, \beta^\top)^\top$. We have a closed-form expression for the log-likelihood

$$\ell_N(\theta) = -\frac{1}{2}\sum_{i=1}^{N}\left\{n_i \log(2\pi) + \log|V_i(\alpha)| + (Y_i - X_i\beta)^\top V_i^{-1}(\alpha)(Y_i - X_i\beta)\right\} \tag{2.4}$$

which can be directly maximized in θ by an iterative procedure to obtain the maximum likelihood estimate $\hat{\theta}$. Note that, in the case of a known α, by solving the score equation

$$\frac{\partial \ell_N(\theta)}{\partial \beta} = 0$$

we get a closed-form expression for the estimate of β depending upon α

$$\hat{\beta}(\alpha) = \left(\sum_{i=1}^{N} X_i^\top V_i^{-1}(\alpha) X_i\right)^{-1} \sum_{i=1}^{N} X_i^\top V_i^{-1}(\alpha) y_i$$

where y_i are the observed values of Y_i and α, if not known, is replaced by its maximum likelihood estimate $\hat{\alpha}$ obtained by maximizing $\ell_N(\alpha, \hat{\beta}(\alpha))$ in α.

When using classic maximum likelihood to estimate α, the variances parameters of the nonlinear mixed model, we do not take into account the estimation of the fixed effects β from the mixed model. This lead to biased estimates $\hat{\alpha}^{ML}$ of the variance parameters (Verbeke and Molenberghs, 2000). To correct such bias, the restricted maximum likelihood estimation (REML) of the variance parameters $\hat{\alpha}^{REML}$ can be preferred. Its principle is to estimate α by maximising the likelihood of error contrasts which does not depend upon β. The obtained estimate $\hat{\alpha}^{REML}$ also does not depend upon the choice of the contrast. It has the advantage of correcting the bias on variance estimation but cannot be used to compare models without the same structure for the fixed effects.

For prediction purpose or fit analysis for example, it is also interesting to compute an estimate of individual random effects. The posterior distribution $f(b_i|y_i)$ follows a multivariate normal density and the individual random effect b_i is usually estimated by the mean of this posterior distribution which has a closed-form expression

$$\hat{b}_i(\theta) = \mathbb{E}(b_i|Y_i = y_i) = BZ_i^\top V_i^{-1}(\alpha)(y_i - X_i\beta).$$

This estimator of individual random effects is the best linear unbiased predictor (BLUP).

2.1.3 Nonlinear mixed model

In the model (2.1), the fixed and random effects are introduced through a linear predictor. A generalization of this model is the nonlinear mixed model

$$Y_i = g(X_i, \beta, b_i) + \varepsilon_i \tag{2.5}$$

where X_i is a $n_i \times p$ matrix of regressors, β the p-vector of fixed effects, $b_i \sim \mathcal{N}(0, B)$ the q-vector of subject-specific random effects, g an *a priori* specified parametric function and $\varepsilon_i \sim \mathcal{N}(0, \Sigma_i)$ the n_i-vector of residual errors assumed independent from b_i and with Σ_i a n_i positive definite matrix. Because of this nonlinearity, in most cases, the marginal log-likelihood

$$\ell_N(\theta) = \sum_{i=1}^{N} \log \int f(Y_i|b_i) f(b_i) \mathrm{d}b_i \tag{2.6}$$

has no analytic expression due to the integral over the random effects. This integral is approximated by numerical integration such as the Laplace approximation, the Gaussian quadrature or Markov chain Monte Carlo (MCMC) methods. The log-likelihood optimization has then no closed-form expression and has to be maximized with an iterative procedure. An estimation of individual random effects can also be computed from the nonlinear mixed model (2.5). However, no analytic solution exists for the mean which is approximated by the mode. It is obtained by maximizing

$f(b_i|Y_i = y_i)$ with an optimization algorithm using the relation from Bayes rule

$$f(b_i|Y_i = y_i) \propto f(Y_i = y_i|b_i)f(b_i). \tag{2.7}$$

2.1.4 Gauss-Hermite quadrature rule

Several methods exists for approximating the integral in (2.6) as detailed by Pinheiro and Bates (1995). We chose Gaussian quadrature rule for its mix of efficiency and accuracy when appropriately used. In particular, we did not choose methods based on MCMC because, as one of our objective is to build an hypothesis testing procedure, we wanted to avoid getting fluctuation of the test results coming from the numerical integration.

Gaussian quadrature rules are a family of numerical analysis techniques that are useful to approximate integrals. The general idea is to approximate the integral by an appropriate weighted sum

$$\int_a^b f(x)\varpi(x)\mathrm{d}x \simeq \sum_{l=1}^{Q} w_l f(x_l).$$

Many different types of Gaussian quadrature exists depending on the bounds values $a, b \in \mathbb{R} \cup \{\pm\infty\}$ and on the weighting function $\varpi : [a, b] \mapsto \mathbb{R}^+$. For a chosen type of Gaussian quadrature and for a fixed $Q \in \mathbb{N}^*$, the quadrature weights w_l and nodes x_l are known; they might be already tabulated or can be computed. In this work, we mainly used the Gauss-Hermite quadrature rule suited for integrals of the following form

$$\int_{-\infty}^{\infty} f(x)\exp(-x^2)\mathrm{d}x.$$

Weights and nodes for the Gauss-Hermite quadrature have been computed for different values of Q (Abramowitz and Stegun, 1970). Gaussian quadrature can be computationally heavy in practice when the integral is multidimensional. Indeed, for an integral of dimension $d \in \mathbb{N}^*$, the quadrature grid size is Q^d and the nodes of the Gaussian quadrature are d-vector. To tackle this issue, the adaptive Gauss-Hermite quadrature, which consists in centering and rescaling the quadrature points

at each iteration, may be used to increase the precision of the computation and reduce the number of quadrature points.

In the framework of the nonlinear mixed model (2.5), using the current estimate of θ at iteration k of the optimisation algorithm, $\hat{\theta}^{(k)}$, we can compute the BLUP of the random effects for each subject i, $\hat{b}_i^{(k)}$ and an estimate of its variance matrix $\hat{var}(\hat{b}_i^{(k)})$ by maximising (2.7). These estimates are used to center and rescale the nodes of the classic Gauss-Hermite grid $(b_l)_{1\leq l\leq Q}$ used for the approximation of the integral over the random effects b_i (2.6). At each iteration k, we get for each subject i an updated Gauss-Hermite grid $(b_{il}^{(k)})_{1\leq l\leq Q}$ where

$$b_{il}^{(k)} = \hat{b}_i^{(k)} + \sqrt{2}\hat{var}(\hat{b}_i^{(k)})^{-1/2} b_l.$$

And finally, by substituting the classic grid by this new grid, the adaptive Gauss-Hermite quadrature gives

$$\int f(Y_i|b_i)f(b_i)\mathrm{d}b_i \simeq 2^{d/2}|\hat{var}(\hat{b}_i^{(k)})|^{-1/2} \sum_{l=1}^{Q} w_l \exp(b_l^\top b_l) f(Y_i|b_{il}^{(k)}) f(b_{il}^{(k)}).$$

Therefore, we can reach the same level of precision than the classic approach using fewer quadrature nodes. By reducing Q, we also reduce the computational cost of the quadrature (Lesaffre and Spiessens, 2001).

However, the estimation of all subject-specific random-effects at each iteration k of the optimization algorithm is time-consuming. Some authors have then suggested different schemes. For example, Rizopoulos (2012) proposed the pseudo-adaptive Gauss-Hermite quadrature rule to estimate joint models for longitudinal markers and time-to-events. The general idea is to center and rescale the quadrature nodes and weights only once at the initial step of the optimisation algorithm using BLUP of the subject-specific random effects estimated from a simpler model, the linear mixed model in the framework of joint models. Then, nodes are not updated at each iteration of the optimization algorithm. However, to raise the accuracy of this approximation without too much heavier computational load, Ferrer *et al.* (2016)

have proposed a two-step pseudo-adaptive Gauss–Hermite quadrature rule. The idea is to estimate the BLUP and adapt the quadrature nodes and weights accordingly twice instead of just once.

2.1.5 Optimisation algorithms

In this section we present some of the existing iterative methods used to maximise the log-likelihood defined above to estimate mixed models. The main problem with the log-likelihood of the nonlinear mixed models is that it entails integrals over the random effects that have no closed-form expression and need to be numerically approximated, sometimes at a certain computational cost.

Expectation-Maximisation algorithm

One classical approach to maximise (2.6) is the expectation maximisation (EM) algorithm proposed by Dempster *et al.* (1977). We consider θ the set of all model parameters. The rationale for the EM algorithm is that if we had observed the complete data, i.e. in our framework the observations Y and the unknown random effects $b = (b_i)_{i=1,\dots,N}$, we could easily compute the log-likelihood of the complete data $\ell_N(\theta; Y, b)$. From this, an estimate of θ can be easily obtained. In practice, because we do not observe the random effects b, the log-likelihood $\ell_N(\theta; Y, b)$ cannot be computed but can be approximated by its expectancy conditionally on the observations Y. The EM algorithm consists into repeating the following steps:

- Expectation: From the previous estimate $\hat{\theta}^{k-1}$, we compute $\mathbb{E}_{\hat{\theta}^{k-1}}[\ell_N(\theta; Y, b)|Y]$
- Maximisation: $\hat{\theta}^k$ is obtained by maximizing $\mathbb{E}_{\hat{\theta}^{k-1}}[\ell_N(\theta; Y, b)|Y]$ over θ

These two steps are repeated until convergence. The convergence is generally assessed by evaluating the difference $\hat{\theta}^k - \hat{\theta}^{k-1}$. This algorithm is interesting as it avoids the heavy numerical integration from classic maximisation approaches. However, the computation of the expectancy can sometimes be difficult and might need further integral approximation techniques. Two majors drawbacks of the EM algorithm are its slow convergence and the fact that no estimation of the asymptotic

variance matrix of the estimates is directly provided by the algorithm. For these reasons, we chose not to use the EM algorithm in the work presented here.

Newton-like algorithms

The classic optimisation techniques for finding an optimum are the family of Newton-like algorithms. They range from the basic gradient method which can be quite slow to more efficient algorithms using not only the gradient but also the Hessian of the objective function like the Newton-Raphson method. An advantage of such method is that we can easily get estimates of the variance of the estimates from the Hessian of the last step of the algorithm. However, it can happen that sometimes, the Hessian used in the optimisation algorithm is not positive definite. To solve this issue, we can chose the Levenberg-Marquardt algorithm (Levenberg, 1944; Marquardt, 1963) designed to solve nonlinear problems and known to be robust with a good convergence rate. It is an iterative procedure where at each step $k \in \mathbb{N}$,

$$\theta^{(k+1)} = \theta^{(k)} - \alpha H^{\star(k)-1} \frac{\partial \ell_N(\theta)}{\partial \theta}\Big|_{\theta=\theta^{(k)}}$$

where the positive-definiteness is ensured by the inflation of the current Hessian matrix defined by $H^{\star(k)} = (H_{ij}^{\star(k)})$ where $H_{ij}^{\star(k)} = H_{ij}^{(k)}$ if $i \neq j$ and

$$H_{ii}^{\star(k)} = H_{ii}^{(k)} + \lambda \left[(1-\eta)|H_{ii}^{(k)}| + \eta \mathrm{tr}(H^{(k)}) \right].$$

The initial values for λ and η are 0.01, they are increased to ensure the positive definiteness of the Hessian matrix if necessary. First, the Hessian matrix is inflated by its diagonal and if this is not sufficient the Hessian matrix is inflated by its trace by an increase of both parameters. The parameter α is modified, if necessary, to ensure that each step improves the log-likelihood. An `R` package `MarqLevAlg` that proposes an implementation of this algorithm exists (Commenges *et al.*, 2016).

2.2 Changepoint models

In this section, we first review the broad literature of changepoint models and secondly, we limit ourselves to the main formulations tailored to answer our objectives. Last, we discuss their application in neuropsychology with a focus on cognitive decline studies.

2.2.1 The changepoint problem in the literature

Historically, the changepoint problem as defined by Hinkley (1970) consists in making inference about the point in a sequence of random variables at which the probability distribution changes. That is, if we observe independent outcomes $Y_1, \ldots, Y_N$, the goal is to find a unique fixed value τ such as $Y_1, \ldots, Y_\tau$ and $Y_{\tau+1}, \ldots, Y_N$ have two different distributions. These distributions could be assumed to have a parametric form $f(Y, \theta_1)$ and $g(Y, \theta_2)$ respectively, with f, g, θ_1 and θ_2 either known or unknown. Different approaches have been proposed to tackle such a problem, either with maximum likelihood approach (Hinkley, 1970) or with Bayesian approach (Smith, 1975). However, as put forward by Carlin *et al.* (1992), because of computational burden, Bayesian approach were barely used for some decades.

The changepoint detection issue has been investigated in more complex frameworks such as situations with multiple changepoints or non independent data. For example, Hawkins (2001) has proposed an algorithm for estimating multiple changepoints for exponential family distributed data and has described applications to stock-market data among others. Multiple changepoints models have also naturally arisen in time series literature, it has been and is still extensively studied. Several recent reviews exists on changepoint detection on time series (Aue and Horváth, 2013; Aminikhanghahi and Cook, 2017) with a lot of interest for climate data (Reeves *et al.*, 2007). Another important part of the literature about changepoint models is dealing with dynamic detection of a changepoint for sequentially obtained data (Lai, 1995). Here, we will not discuss such literature as these models are not suited to answer our objective to explore a unique changepoint in longitudinal cohort data.

In the context of our motivating application, the natural history of dementia, we

are interested in a specific class of changepoint models. First, this work relies on the analysis of cohort data that involve repeated measures of markers on the same subjects. We thus focused on mixed models to account for the intra-subject correlation. Secondly, because of the heterogeneity of the cognitive decline trajectories already mentioned (Amieva *et al.*, 2014), the time of change should be different between subjects. This is why a random effect on the changepoint is needed into the model. Also, as mentioned, cognitive decline trajectories are continuous and there is no gap at the time of change between the two distributions of the outcome. We also focus on models with only one changepoint to make the interpretation of the changepoint, as the time of acceleration of cognitive decline, more straightforward.

This is why, from now on, we only consider unique random changepoint mixed models. They are suited for our objectives as they take into account the within subject correlation and the between subject heterogeneity. In the literature, this model is sometimes called breakpoint mixed model, piecewise mixed model or segmented mixed model. From now on, we will only use the designation random changepoint mixed model.

2.2.2 Random changepoint mixed models

In the literature, several formulations have been proposed for the random changepoint mixed model. We describe four main formulations: the broken-stick model, the Bacon-Watts model, the bent-cable model and the polynomial model. In the following, we note $Y_{ij} = Y_i(t_{ij})$ the measure of the outcome Y for subject i at time t_{ij} with $i = 1, \ldots, N$ and $j = 1, \ldots, n_i$. The subject specific changepoint is noted τ_i. All the models introduced in this section are represented in Figure 2.1.

The broken-stick model

The broken-stick mixed model, or even more explicitly called the linear-linear mixed model consists of two straight lines intersecting at the changepoint (Hinkley, 1969).

$$Y_{ij} = \beta_{0i} + \beta_{1i}(t_{ij} - \tau_i) + \beta_{2i}(t_{ij} - \tau_i)\text{sgn}(t_{ij} - \tau_i) + \varepsilon_{ij} \tag{2.8}$$

with $\text{sgn}(x) = x/|x|$ if $x \neq 0$, $\text{sgn}(x) = 0$ otherwise, $\beta_{ki} = \beta_k + \alpha_{ki}$ for $k = 0, 1, 2$, $\tau_i = \mu_\tau + \sigma_\tau \tilde{\tau}_i$ where the vector of all the random effects $(\alpha_{0i}, \alpha_{1i}, \alpha_{2i}, \tilde{\tau}_i)^\top$ is multivariate normally distributed with null mean vector and variance matrix B a positive-definite matrix. With this formulation, β_{0i} is the value of the outcome of subject i at the changepoint, β_{1i} is the mean slope and β_{2i} half the difference of slope before and after the changepoint. In other words, the slope before the changepoint is $\beta_{1i} - \beta_{2i}$ and the slope after is $\beta_{1i} + \beta_{2i}$.

This model assumes a sharp change of slope at the changepoint which raises two important issues. First, it is not clinically realistic as cognitive decline trajectories are generally smooth, even at the changepoint. Second, if we choose a frequentist approach using optimization algorithm to maximize the log-likelihood, non-differentiability of the model at the changepoint τ_i can cause numerical troubles.

The Bacon-Watts model

To deal with these issues, a direct extension of the broken-stick model (2.8) has been proposed by Bacon and Watts (1971). They replaced the function sgn in (2.8) by a transition function trn which smooths the intersection between the two lines at the changepoint.

$$Y_{ij} = \beta_{0i} + \beta_{1i}(t_{ij} - \tau_i) + \beta_{2i}(t_{ij} - \tau_i)\text{trn}(t_{ij} - \tau_i) + \varepsilon_{ij} \tag{2.9}$$

The trn function must verify the following assumptions:

1. $\lim_{s \to \infty} \text{trn}(|s|/\gamma) = 1$
2. $\text{trn}(0) = 0$
3. $\lim_{\gamma \to 0} \text{trn}(s/\gamma) = \text{sgn}(s)$
4. $\lim_{s \to \infty} s\text{trn}(s/\gamma) = s.$

All these conditions allow the function trn to behave in a similar manner to the function sgn it approximates. The value of γ defines the smoothness of the transition. The closer to 0 γ is, the sharper the transition will be. The bigger γ is (according

to the range of the time variable), the smoother the transition will be. Bacon and Watts (1971) proposed several examples of such functions: $\text{trn}(s) = \tanh(s/\gamma)$, $\text{trn}(s) = 1 - \exp(|s|/\gamma)$, $\text{trn}(s) = (s/\gamma)^2/\{1 + (s/\gamma)^2\}$. Griffiths and Miller (1973) dropped out condition 2 in order to avoid a bulge in the trajectory that entails an increase just before the random changepoint which is not very realistic for dementia applications. They proposed instead to use $\text{trn}(s) = \sqrt{s^2 + \gamma}/x$ as a transition function

$$Y_{ij} = \beta_{0i} + \beta_{1i}(t_{ij} - \tau_i) + \beta_{2i}\sqrt{(t_{ij} - \tau_i)^2 + \gamma} + \varepsilon_{ij} \tag{2.10}$$

This bulge and the comparison between both formulations are illustrated in Figure 2.2. An interesting review of transition functions between linear regimes can be read in Seber and Wild (2005). This model has the advantages to solve the major issue of the broken-stick formulation without losing its nice interpretability. Indeed, for a small γ, β_{01}, β_{1i} and β_{2i} have the same interpretation than in the broken-stick model.

The bent-cable model

An alternative approach to solve the non-differentiability issue raised by the broken stick-model has been proposed by Tishler and Zang (1981) under the name of bent-cable model. Their idea is to smooth the trajectory on a neighbourhood around the changepoint using a quadratic transition between the two linear phases. With our notations, the bent-cable model is written

$$Y_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}q(t_{ij}; \tau_i, \gamma) + \varepsilon_{ij} \tag{2.11}$$

where

$$q(t; \tau, \gamma) = \frac{(t - \tau + \gamma)^2}{4\gamma}\mathbf{1}_{\{|t-\tau|\leq\gamma\}} + (t - \tau)\mathbf{1}_{\{t>\tau+\gamma\}}.$$

With this formulation, β_{0i} is the intercept and β_{1i} the slope of the first linear part before the transition phase, β_{2i} is the slope after the transition, with the transition spanning on the interval $[\tau_i - \gamma; \tau_i + \gamma]$ centered at τ_i and of length 2γ. If $\gamma = 0$, this model becomes a broken-stick model.

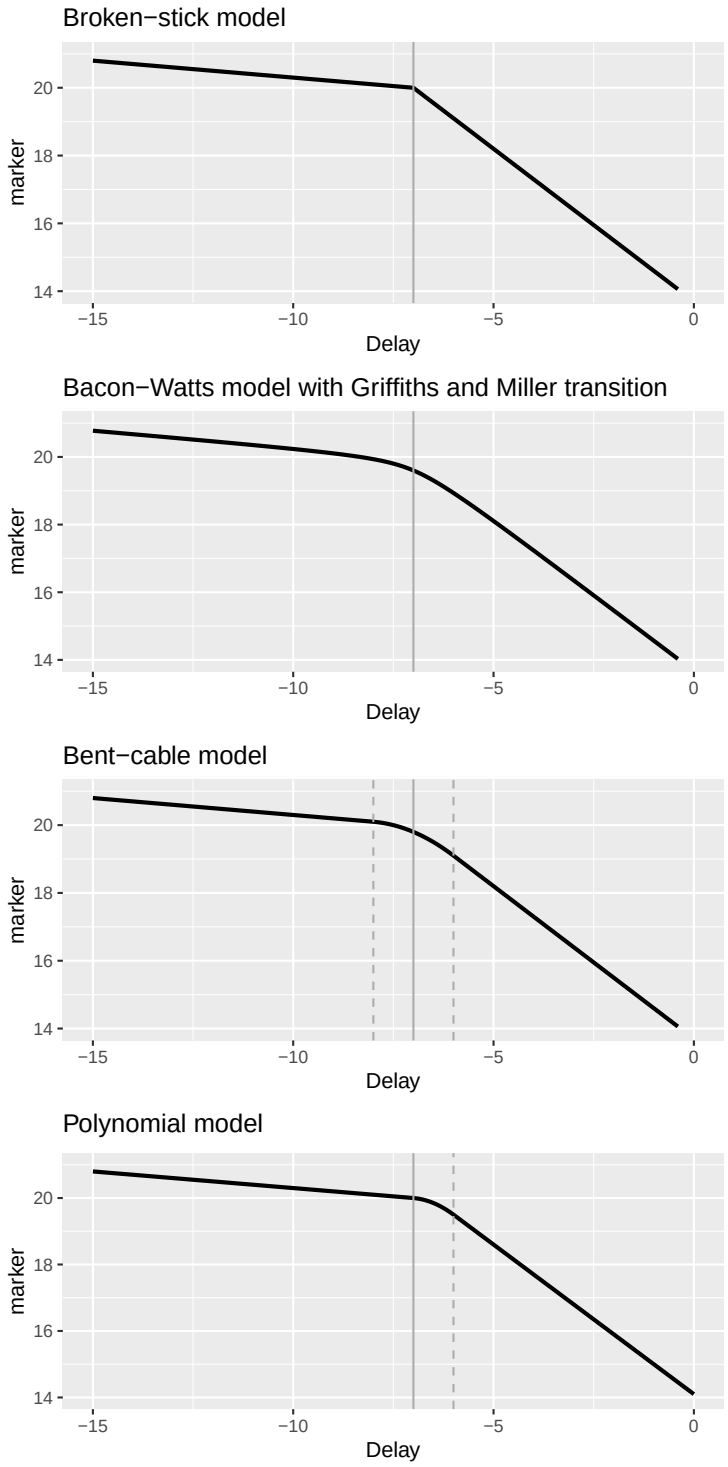

Figure 2.1: Example of trajectories according to the broken-stick model, the Bacon-Watts model for Griffiths and Miller (1973) transition, the bent-cable model and the polynomial model all with $\gamma = 1$. The grey solid line is the changepoint value, here at -7, and the grey dashed lines are the limit of the transition area. Adapted from van den Hout *et al.* (2013).

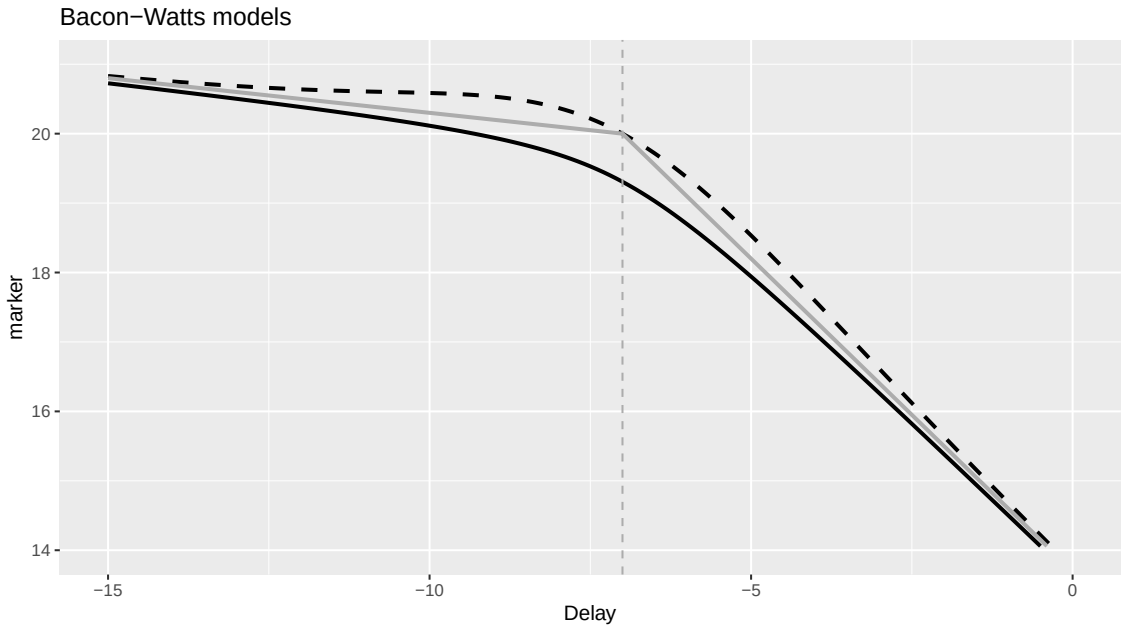

Figure 2.2: Bacon-Watts model with hyperbolic tangent transition function (dashed black) and Griffiths and Miller (1973) transition function (solid black) compared to the linear-linear trajectory (solid grey). The dashed grey line indicates the change-point and here $\gamma = 3$ to illustrate the behaviour of the transition functions.

In the case of a white noise and when there is no random effects, neither on the β_{ki}, $k = 0, 1, 2$ nor on the changepoint τ_i, Chiu *et al.* (2006) showed that the least squares estimators of the regression parameters were consistent and asymptotically normally distributed. In more recent work, Chiu and Lockhart (2010) extended their results for the case of auto-regressive noise. Note that sometimes in the literature, the Bacon-Watts model (2.9) is also called the bent-cable model.

The polynomial model

More recently, van den Hout *et al.* (2011) proposed an alternative to the previous models that they called the polynomial model. In their view, the Bacon-Watts model (2.9) with hyperbolic tangent transition does not always allow easy interpretation of β_{1i} and β_{2i}. In particular, for big γ values, β_{1i} and β_{2i} may not be the true slope values. However, when γ is small, the Bacon-Watts model parameters can be easily interpreted. Similarly to the bent-cable of Tishler and Zang (1981), van den Hout *et al.* (2011) modeled the transition between the two straight lines with a polynomial function. They chose a cubic polynomial function and their model is written

$$\mathbb{E}(Y_{ij}) = \begin{cases} \beta_{0i} + \beta_{1i} t_{ij} & t_{ij} < \tau_i \\ g(t_{ij} | \beta_{0i}, \beta_{1i}, \beta_{2i}, \gamma) & \tau_i \leq t_{ij} < \tau_i + \gamma \\ \lambda_i + \beta_{2i} t_{ij} & \tau_i + \gamma \leq t_{ij} \end{cases} \tag{2.12}$$

where the smoothness of the transition is ensured by the following constraints on the cubic polynomial function g

1. $g(\tau_i) = \beta_{0i} + \beta_{1i}\tau_i$
2. $g(\tau_i + \gamma) = \lambda_i + \beta_{2i}(\tau_i + \gamma)$
3. $\frac{\partial}{\partial t} g(\tau_i) = \beta_{1i}$
4. $\frac{\partial}{\partial t} g(\tau_i + \gamma) = \beta_{2i}$.

These conditions imply continuity and smoothness between the linear parts and the cubic transition. Note that here, contrarily to the previous models, the change-

point τ_i is defined as the beginning of the transition phase of length γ. In order to be closer to the broken-stick model, one might add the constraint that the two linear parts should intersect at the middle of the transition phase by imposing that $\lambda_i = \beta_{0i} + (\tau_i + \gamma/2)(\beta_{1i} - \beta_{2i})$. The parameters of the model from both the linear parts have the advantage of having a direct interpretation, contrarily to the Bacon-Watts model with large γ. Over the bent-cable model, this model has the advantage of estimating a cubic transition rather than a quadratic one which allows smoother transition regimes. In our application however, such smoothness is not necessary and a quadratic transition is sufficient to model cognitive decline trajectories. One drawback of this model is that the third degree polynomial function g must be estimated by solving a system of differential equation.

2.2.3 Changepoint models in neuropsychology

We now review applications of models with fixed or random changepoint in neuropsychology mainly with a focus on dementia studies.

Fixed changepoint and profile likelihood approach

In the first attempts to model cognitive trajectory of demented subjects, a profile likelihood approach with fixed changepoint was preferred to estimate the time of change as it avoids estimation of nonlinear mixed models. However, some authors were aware that the assumption of a fixed changepoint was not very realistic regarding the known heterogeneity in cognitive decline trajectories.

Hall *et al.* (2000) proposed a changepoint model to study the natural history of dementia based on a sample of 365 subjects including 72 cases from the Bronx Aging Study cohort (Katzman *et al.*, 1989). They modelled retrospectively the individual evolution of the Buschke Selective Reminding Test (Buschke, 1973), a score that measures episodic memory, as a function of age and time to dementia. The trajectory over age was at first assumed quadratic-quadratic, with, for the cases only, a shift at a fixed changepoint relative to time to diagnosis. They finally dropped all the quadratic terms as they were non significant and studied the reduced model, a

broken-stick model (2.8) with a fixed changepoint for the cases and a linear mixed model for the control. The authors chose a frequentist approach with gaussian assumption to estimate the model by profiling on a grid of changepoint values. They discarded the random changepoint model approach for several reasons. First their main goal was to determine when cases and non-cases trajectories differ and, according to them, heterogeneity on this date would not be helpful. Second, estimating such a model would need a Bayesian approach but they highlight the lack of prior information on most of the parameters. Last, they put forward that their data are not sufficient to model correctly the changepoint distribution. In their application, the changepoint was found to happen on average 5.1 years before the diagnosis of dementia. A few years later, the same authors Hall *et al.* (2003), aware of their strong assumption of a fixed changepoint explored it further. For the same cohort data, they modeled again the evolution of the Buschke Selective Reminding Test as a function of age with a changepoint relative to time before diagnosis. A fixed changepoint model estimated with a profile likelihood approach and a random changepoint model estimated with a Bayesian approach were proposed. A Bayesian model selection procedure based on the pseudo-Bayes factor was proposed to compare the reduced model of common changepoint to the full model where at least one individual changepoint differs from others. Surprisingly they found that heterogeneity in the changepoint did not improve the fit of the model. This conclusion was unexpected because of the known heterogeneity in cognitive decline. The fact that they modeled the evolution of the marker as a function of time to diagnosis might partly explain this result. The authors also put forward that their surprising conclusion matched previous results on the same cohort data.

Carlson *et al.* (2008) studied the evolution of the ventricular volume over age for MCI and non-MCI subjects from the Oregon Brain Aging Study. A fixed changepoint relative to time to MCI diagnosis was assumed for all cases. They found that the annual rate of expansion of ventricular volume globally decreased with age and that for MCI subjects it accelerated 2.3 years before diagnosis. Thorvaldsson *et al.* (2008) evaluated the evolution of cognitive abilities before death. They used age as a timescale with a changepoint relative to time to death. They found an acceleration

of this decline 6.6 years prior to death for verbal ability, 7.8 years for spatial ability, and 14.8 years for perceptual speed. On this study, the changepoint appeared earlier than what some previous work have suggested (Wilson *et al.*, 2003; Sliwinski *et al.*, 2006). This difference might be due to different lengths of follow-up and differences in health characteristics of the sample. More recently, Bartolucci *et al.* (2009) used a changepoint model to evaluate how the Mini Mentale State Examination (MMSE), a cognitive score introduced by Folstein *et al.* (1975), evolved over time in a cohort of diagnosed AD subjects. Using Bayesian techniques with non-informative priors and a uniform prior for the changepoint, they computed posterior estimates of the changepoint and of the two slopes. In their data, the changepoint was estimated at 2.4 years over the 4 year follow-up of the cohort, after adjustment on age, sex and education.

Howieson *et al.* (2008) evaluated the trajectory of verbal memory, animal fluency, and visuospatial constructions abilities over time before MCI onset. To do so, he estimated a linear mixed model for non-cases and a broken-stick model (2.8) for MCI subjects using a profile likelihood approach. The changepoint for verbal memory was estimated to be around 3 years before MCI onset and about 4 years for animal fluency, and visuospatial constructions abilities. After these changepoints, the decline was significantly increased for all cognitive abilities. Johnson *et al.* (2009) evaluated in separated analyses the trajectory of global, verbal, visuospatial and working memory as a function of time before diagnosis. A linear mixed model, a broken-stick model (2.8) and a linear-quadratic model were estimated on the data. The changepoint was estimated using a profile likelihood approach. They found the linear model to be the best for non-cases and the broken-stick model for the cases with the quadratic term being non significant as in Hall *et al.* (2000). They could estimate for each marker a changepoint which happened 3 years before dementia for visuospatial memory, 2 years for global memory and 1 year for verbal and working memory.

The results of Howieson *et al.* (2008), Thorvaldsson *et al.* (2008) and Johnson *et al.* (2009) are interesting as they compared trajectories of different markers which give insight on the multidimensional aspects of the natural history of the disease.

However, because they modeled the markers independently, comparing the times of change of different markers is impossible because to do so their covariance is needed. All of the above models assumed a common changepoint for all individuals, mainly because of computational limitations or too short follow-up. Despite the results of Hall *et al.* (2003), that the authors themselves have discussed, it is mainly acknowledged that the cognitive evolution is very heterogeneous between subjects. To handle this heterogeneity while profiling, all of the above authors, except Bartolucci *et al.* (2009) have considered the changepoint as a time to a specific event: MCI diagnosis, AD diagnosis or death. This way, they diminished the heterogeneity around the changepoint which allowed them to assume a fixed changepoint. However, it should be wiser to properly take into account the known subject-specific variability of the changepoint by using a full random changepoint model, even when the timescale is time to diagnosis.

Random changepoint model: frequentist and Bayesian approach

Recent advances in computational efficiency has made possible the estimation of mixed model with a subject-specific random changepoint leading to more realistic models. Two main approaches have been used: either a frequentist approach where the log-likelihood of the nonlinear mixed model is maximized, entailing the numerical approximation of an integral over all the random effects, or a Bayesian approach using MCMC techniques for posterior elicitation and the necessary specification of prior distribution.

Dominicus *et al.* (2008) compared the performance of the full broken-stick model (2.8) including all four random effects model to a linear mixed model and a quadratic mixed model. They used data from a Swedish Adoption Twin Study of Aging cohort (Pedersen *et al.*, 1992) focusing on the evolution over age of the symbol digit test which assesses the ability of a subject to quickly compare numbers and symbols. Estimation was done in the Bayesian framework using conjugate priors and MCMC simulations through Gibbs sampling to approximate the posterior distribution of the parameters. The three models were compared with the deviance information criterion (Spiegelhalter *et al.*, 2002) and the best model was found to be the random

changepoint model, whatever the hyperparameters values among three different scenarios. For these scenarios, the mean random changepoint was estimated between 71.4 and 74.3 years old.

van den Hout *et al.* (2011) compared a frequentist and a Bayesian approach for the estimation of the broken-stick (2.8), the Bacon-Watts (2.9) with hyperbolic tangent transition function and the polynomial model (2.12) they proposed. They applied these models to highlight a terminal decline of the MMSE score over time to death in the UK cohort CC75C (Brayne *et al.*, 1992). Standard software routine from the `R` package `lme4` was used for the frequentist estimation of the Bacon-Watts and polynomial model. As their routine needed the derivative of the model, they could not estimate the broken-stick model with the frenquentist approach since it has no derivative. They used a profile likelihood approach to find the optimal value of the smoothing parameters γ in (2.9) and (2.12). Bayesian inference was performed using `WinBUGS` software that allows easy implementation of MCMC methods. The Bayesian approach gave better variance estimation and was less sensitive to starting values. However the Bayesian approach is slow and model comparison is not straightforward. Using the AIC obtained from the frequentist approach, they compared the fit of the Bacon-Watts and polynomial model. Both gave similar results for the estimation of the changepoint, around 6 years before death, but the Bacon-Watts model was preferred. Moreover, the Bacon-Watts model remains easier to implement and faster to run while keeping nice interpretability. van den Hout *et al.* (2013) extended their previous work by dropping the Gaussian assumption on the outcome and on the distribution of the random effects. They proposed a semi-parametric nonlinear random changepoint model to study the evolution of MMSE over time to death in the Origins of Variance in the Old–old (OCTO-Twin) study (McClearn *et al.*, 1997). Their model is plugged into a latent class model with two classes: one with a change in the trajectory and one without a change. Models were estimated using maximum likelihood approach with a Nelder-Mead algorithm (Nelder and Mead, 1965) for the maximisation and were compared using BIC. The bent cable model was preferred and they found a drop of the MMSE score 5.8 years before death. Authors discarded the Bacon-Watts model because using the hyperbolic transition function, it

presents an increase just before the changepoint that the authors judged unsuitable to model cognitive decline and that makes parameters interpretation less obvious. Note however that using a different transition function in the Bacon-Watts model, like the one proposed by Griffiths and Miller (1973) in (2.10), rules out this issue.

An alternative estimation procedure was proposed by Muggeo *et al.* (2014) and applied to compare the performance of three treatments over time on the longitudinal evolution of the Beck Depression Inventory, a marker that measures depressive symptoms. He used the broken-stick model (2.8) with all four random effects assumed correlated. A reparametrization of the changepoint distribution is used so that it can be contained in a chosen bounded interval. This non linear mixed model is estimated by a maximum likelihood approach using a linearisation by a first order Taylor expansion. This linearisation makes possible the use of standard estimation routines for linear mixed model but it may lead to biased estimates or less efficient estimator than methods based on numerical integration (Molenberghs and Verbeke, 2005).

Because they were able to estimate full random changepoint models, the authors mentioned above were able to confirm the superiority of random changepoint model over non random changepoint models contrarily to the result obtained by Hall *et al.* (2003). They could also compare the frequentist and the Bayesian approach. The main issues with the Bayesian approach are its slowness, the non trivial choice of hyperparameters and priors which can influence the results, the absence of proper statistical tests and the randomness that comes from MCMC integration techniques. The main issue with the frequentist approach lies on the integral over the random effects. However, this can be ruled out using adaptive Gauss quadrature (see Section 2.1.4). In the frequentist approach, another issue is the transition window parameters γ. It is generally *a priori* fixed or estimated by profile likelihood. However, clinical knowledge about the marker can be sufficient to chose a realistic value for γ. Above authors also compared the various formulations of the random changepoint models described in Section 2.2.2. The broken-stick model is discarded for its lack of realism and its non differentiability. The polynomial model is generally considered harder to implement and to estimate. The bent-cable and the Bacon-Watts

model performs well especially with a small γ and the Griffiths and Miller (1973) formulation (2.10) which ease the interpretation.

Beyond the simple random changepoint mixed model

In this section, we explore extensions of the random changepoint model. First, as already mentioned, Howieson *et al.* (2008), Thorvaldsson *et al.* (2008) and Johnson *et al.* (2009) all described the evolution over time of several markers assumed independent and with a fixed changepoint. To explore the multidimensional aspect of the cognitive decline a multivariate random changepoint model that allows proper comparison between different markers would be more appropriate. Secondly, as cognitive change over time is linked to dementia and death, these two events may induce informative dropout in the cohorts. To avoid biases in the estimation, joint modeling of the cognitive trajectories and the time to dementia and/or death is required.

To our knowledge, only two bivariate random changepoint mixed models have been proposed in the literature and both assumed very restrictive correlation structure between the markers. Hall *et al.* (2001) proposed a bivariate model to compare the changepoint in both the decline of accelerated memory over age, assessed by the Buschke Selective Reminding Test (Buschke, 1973), and the decline of perfomance on speeded tasks over age assessed by the WAIS performance IQ test (Wechsler, 1955). They compared a profile likelihood approach with common changepoint to a full Bayesian approach with subject-specific changepoint where the changepoints are relative to time to diagnosis. Only the intercepts of the two markers were correlated. They found that performance on memory declines before performance on speeded tasks (7.5 years before diagnosis versus 2.1 years) and, quite surprisingly, that neither the randomness of the changepoint neither the correlation between markers improved the fit of the model. The result about the changepoint was similar to Hall *et al.* (2003) and appeared also to be inherent to this cohort. The absence of correlation between the two markers did not surprise the authors as these markers are known to show very little within-subject correlation. More recently, Yang and Gao (2013) also proposed a bivariate random changepoint model and compared the

performance of the broken-stick 2.8, the Bacon-Watts 2.9 with hyperbolic tangent function and the polynomial mixed model formulations. They found that the polynomial model performed better but their conclusion has to be tempered because they simulated data from this model only in order to compare all the formulations. In their implementation, only the changepoints were correlated. It was estimated with Bayesian techniques using MCMC methods and prior sensitivity analysis. Their methodology was applied on Indianapolis–Ibadan Dementia Study (Hendrie *et al.*, 1995) to compare evolution of Body Mass Index (BMI) and cognitive functions over age and it was found that BMI declines 16 years before cognitive functions.

Jacqmin-Gadda *et al.* (2006) proposed a joint model which combines a linear-polynomial model with a smooth transition between the two phases to model the longitudinal evolution of cognition and a log-normal model including the changepoint as a covariate for the time to dementia. Maximisation of the joint log-likelihood is performed by a Levenberg-Marquardt algorithm (Levenberg, 1944; Marquardt, 1963). They applied the model to the Paquid cohort to compare the cognitive trajectories of the Benton Visual Retention Test over age according to educational level. Their results confirmed the hypothesis of cognitive reserve stated by Stern *et al.* (1994) among high education subjects. Indeed, before the changepoint their cognitive decline is slight, but afterwards, when defence mechanisms fails, the decline is more dramatic compared to low education subjects. However, in this model all subjects were assumed to be at risk of dementia and to have a changepoint in their cognitive decline trajectory. Another limit was that potential informative censoring due to death was not taken into account but this can be fixed by extending the model to a multi-state model.

Yu and Ghosh (2010) proposed a joint model where a mixture survival model takes into account two competing risks: dementia versus dementia-free death using a logistic model for the class membership. The longitudinal evolution of the Cognitive Abilities Screening Instrument score over age from the Honolulu Asia Ageing Study is modelled using a piecewise polynomial model with a random changepoint. The risk of each event follows a Weibull model with the risk of dementia depending upon the changepoint value. Their model also takes into account the uncertainty on the

time of dementia onset due to the interval censoring but not the missing information regarding the health status between the last visit without dementia and the death. The model was estimated using a Bayesian approach with weakly informative priors.

Recently, to address most of the previously mentioned issues, Dantan *et al.* (2011) proposed a joint multi-state model. The longitudinal part was modeled using the Bacon-Watts formulation of the random changepoint model (2.10) with Griffiths and Miller (1973) transition. For the survival part, a multi-state model was proposed with four states: healthy, pre-diagnosis, dementia and death. The changepoint defines the entrance to the pre-diagnosis state and the risk of dementia before this state is null and increases with the time spent in this phase. The risk of death depends on the individual health status only through the current expected marker value. The authors proposed to consider the pre-diagnosis state as the MCI state. A frequentist approach was proposed to estimate the model and an application on the Paquid cohort to evaluate evolution of the Benton Visual Retention Test over age was presented.

These extensions are useful to answer important limits of previous models. The multivariate modelisation allows a proper estimation of correlated markers. However, in the above mentioned article, the correlation structures remain too simple and may not capture the whole association: only the intercepts are correlated in Hall *et al.* (2001) and only the changepoints in Yang and Gao (2013). In dementia studies, where cognitive dimensions are often strongly correlated, a more complex between markers correlation structure should be considered. The extension to joint models is important to take into account informative dropout due do death or dementia but raises two main issues. First, all of the proposed models do not allow testing for the existence of a random changepoint because the absence of changepoint entails independence between the risk of dementia and the cognitive evolution which is not realistic at all. Secondly, because joint models are applied to the whole population, the used timescale is age and the changepoint interpretation becomes more subtile as detailed in the concluding remarks.

Concluding remarks

We have explored the use of changepoint models in the literature of dementia. In practise, both the Bayesian and frequentist approach have advantages and drawbacks. However, we chose a frequentist approach which allows proper statistical testing and used appropriate Gaussian quadrature to handle the heavy numerical integration.We also chose to use Bacon-Watts type formulation because this model is easy to implement and if the appropriate formulation (2.10) is used, its parameters are easily interpretable.

One important aspect that emerges from this state of the art is the importance of chosing an appropriate timescale. When the changepoint is fixed, the chosen timescale is the time to dementia because the delay between acceleration of cognitive decline and dementia may be expected to be more homogeneous than the age at acceleration of the decline which can be at least as variable as the age at dementia. When the changepoint is random, authors generally considered age or time spent in the cohort as timescale because the inclusion of a random effect on the changepoint allows for the expected great heterogeneity between the age of change. In joint models, the timescale cannot be a delay to diagnosis because all subjects are included and such a delay cannot be computed for non-cases. The interpretation of the time of change directly depends upon this timescale. When the timescale is the delay to diagnosis, a time of change represents the length of the phase of accelerated cognitive decline until diagnosis. If the timescale is age, the changepoint represents the age at which the cognitive decline begins to accelerate which is highly variable between subjects because strongly correlated with the age at dementia. In our case, the delay to diagnosis sounds more appropriate to describe the natural history of dementia. As mentioned however, such a timescale imposes working with cases only because non-cases have no delay to dementia. This leads to a selection bias due to the interval censored nature of dementia diagnosis because only subjects who remain in the study until diagnosis are considered while subjects who dropped out before diagnosis or where diagnosed after the end of the study are excluded.

2.3 Testing the existence of a changepoint

In the previous section, we described the main changepoint models proposed in the literature. However, in dementia, it can happen that for certain subpopulations, the trajectory appears smoother which arises the question of the existence of the random changepoint. The formulations of such a test are multiples and various null hypothesis are plausible. However, under these null hypothesis, some parameters of the models vanishes making them unidentifiable under the null hypothesis and standard testing methods non applicable. In this section, we describe the existing approaches proposed in the literature, first to tackle the more general problem of detecting a change in a parameter of a regression model, i.e. a structural change, and secondly for detecting a structural change in segmented regression model specifically. Finally, we are looking into the quite limited literature discussing approaches suited for longitudinal data.

2.3.1 Tests for structural change in regression models

Here, the changepoint is defined as a unique fixed value τ such as, if we observe independent outcomes $Y_1, \ldots, Y_N$, the subvectors $Y_1, \ldots, Y_\tau$ and $Y_{\tau+1}, Y_N$ have two different distributions F and G. These distributions might have a non parametric form or a parametric form $f(Y, \theta_1)$ and $g(Y, \theta_2)$ respectively, with f, g, θ_1 and θ_2 either known or unknown. No assumption of continuity is made at the changepoint τ, contrarily to segmented regression model. Several methods have been proposed to test for a shift in a regression model for independent data.

The non parametric case

Some of them are suited to the non parametric case and the proposed tests are based on rank statistics. Among these, Pettitt (1979) proposed, for the case of continuous F and G, a procedure to test the null hypothesis of no change versus the alternative of exactly one change. Its test statistics is based on the Mann-Whitney statistics. He derived exacts significance probabilities for the case of F and G being Bernoulli distributions and computed approximations for the case of continuous distributions.

Lombard (1987) extended this work by considering both the abrupt case where no assumption was made on the values of $F(\tau)$ and $G(\tau)$ and a continuous case with the constraint $F(\tau) = G(\tau)$. For the abrupt case, he considered the *one changepoint alternative* as well as the *multiple changepoints alternative*. His test was built on quadratic form rank statistics from which he derived the asymptotic null distribution and tabulated significance probabilities. More recently, an interesting particular case has been treated by Aly *et al.* (2003) who considered, as the alternative hypothesis, the existence of ordered multiple changepoints. That means, that on each of the $k+1$ segments formed by k changepoints, they considered a partially ordered relation $F_i \prec F_{i+1}, i = 1, \ldots, k$ as the alternative. The asymptotic null distribution was given and tables of critical values were computed through a Monte Carlo approach.

The parametric case

For the parametric case, methods based on the likelihood ratio tests have been extensively used in the literature and different tests have been proposed depending on the nature of the distribution: gaussian data or a logistic model framework.

For normally distributed data, Quandt (1960) studied the behaviour of the likelihood ratio statistic for testing that no switch occurred versus the alternative that one switch occurred in a linear regression model after which intercept, slope and residual variance might change. He showed that the commonly used chi squared approximation of the asymptotic distribution of the likelihood ratio under the null failed completely and computed its empirical distribution from simulations so that he could build an empirical test. Kim and Siegmund (1989) proposed a likelihood ratio test procedure for normally homoscedastic distributed data in a more constrained framework. On a linear regression model with one dependent variable, they tested the null hypothesis of no change versus a change in the intercept or versus a change either in the intercept or in the slope. Approximations were given for the significance levels of their likelihood ratio test and accuracy was assessed in simulation studies. Horvath (1993) proposed a likelihood ratio test to study the more general case of a sample of independent random Gaussian variables. He proposed a procedure to test that all variables are sampled from the same Gaussian distribution, i.e.

with same mean and variance, versus the alternative that, from some changepoint τ, the sampling distribution changes either in mean or in variance. The asymptotic distribution was computed using asymptotic results from stochastic process theory.

In order to detect a change in the parameters of a logistic model, a likelihood ratio test statistic that makes no assumption on the covariates behaviour was proposed by Gurevich and Vexler (2005) inspired by a sequential procedure introduced by Robbins and Siegmund (1972). Fong *et al.* (2015) proposed to test for a threshold effect in a logistic model, i.e. to test if a covariate has no effect before an unknown changepoint and has a constant effect afterwards. They used a supremum score test approach because of identifiability issues and proposed approximations of the p-values. This supremum score test will be discussed more thoroughly in Chapter 3. An additional case was considered by including an interaction term between some covariates and the changepoint variable. This test is built on the maximum of likelihood ratio statistic and a sampling procedure of its asymptotic distribution under the null is provided. Their methods have been implemented in a `R` package (Fong *et al.*, 2017).

2.3.2 Tests for structural change in segmented regression models for independent data

Here, we specifically focus on segmented regression model, that is, a linear-linear model where at the changepoint there is continuity between the two straight lines. The smooth model of Lombard (1987) introduced previously could fit in this section. We only discuss the case of independent data for the moment.

Farley and Hinich (1970) proposed a method to detect a change in the slope coefficents of a linear regression model with the condition that this change has to be small relative to the residual variance. The changepoint was assumed uniformly distributed on all the range of possible values. They used a likelihood ratio test statistic whose critical values were approximated by a first order approximation around the null. For the same statistical problem, Feder (1975) gave more theoretical insights for the asymptotic behaviour of the test statistic under the null. They

showed that, under the null, this test statistic behaves actually like the maximum of correlated $\chi^2(1)$ and $\chi^2(2)$ random variables.

Muggeo (2016b) proposed a framework for testing the existence of a fixed change for longitudinal Gaussian homoscedastic data. A score test approach that avoids the possibly heavy estimation of the alternative model is chosen. However, because some parameters vanish under the null, they become unidentifiable and the classic score test approach cannot be used as the MLE of these unidentifiable nuisance parameters cannot be computed under the null. In a similar spirit than Andrews and Ploberger (1994), he replaced the quantity that vanishes under the null by its average over a pre-specified grid of values. The asymptotic null distribution was given and a simulation assessed this test performance. We refer to Section 3.1.4 for more details about the problem of nonidentifiable nuisance parameters under the null.

2.3.3 Changepoint detection for segmented longitudinal data

All of the previously described literature only focused on independent data. When working on longitudinal data, within subject correlation needs to be taken into account. However, if the literature about tests for independent data is quite developed as we have seen, there is very few literature on dependent data.

Juang and Wolfe (1990) proposed test statistics based on the Mann-Whitney statistics generalizing a previous work of Pettitt (1979) to test for the existence of at most one changepoint for repeated data where each individual has the same number of measures taken at the same time occurrences. They studied the asymptotic behaviour of their statistics and proposed approximations of critical values for small samples. However, BuHamra (1997), who worked in the same framework, evaluated the performance of these statistics and found that some performed poorly because they were not integrating between subjects information. She then proposed new test statistics and by deriving its asymptotic distribution and by computing small samples critical values using MCMC techniques she found that her new test statistics performed better.

In a purely frequentist approach, Piepho and Ogutu (2003) proposed a test for

the existence of a fixed changepoint in a segmented regression model for repeated measures. They used a likelihood ratio approach where the bounds of the critical values of the test were computed using the approximation proposed by Davies (1977) and Davies (1987). Ramanayake and Gupta (2010) proposed a likelihood ratio type statistic T_1 to detect a changepoint in the parameter of exponentially distributed data assuming uniform prior for the changepoint and compared it to T_2 the supremum of the classic likelihood ratio test. They computed the distribution under the alternative to perform a power comparison of both tests and found that T_1 outperformed T_2 when the true changepoint lies in the middle of the sequence of observations.

As we have seen, the literature for testing a breakpoint in a regression model is well developed (Bhattacharya, 1994) but not the literature for identifying a breakpoint in a segmented model and especially in the case of repeated data. Several challenges exists for developing such test. A first challenge is the identifiability issue that arises when testing for the existence of a changepoint in segmented regression models. As we have seen, some authors proposed way to tackle this issue and we described this matter more precisely in Chapter 3. A second challenge is that the distribution under the null can be intractable and needs to be approximated. Finally, the third challenge lies in the longitudinal nature of the data. To our knowledge, there is in fact no developed methodology to assess the existence of a random changepoint for repeated measures. This is why in Chapter 3, we propose a testing procedure for the existence of a random changepoint in a mixed model.

Chapter 3

Testing the existence of a random changepoint

In this chapter, we describe the methodology we propose to answer our first objective: testing the existence of a random changepoint for longitudinal data. In the next section, we describe the mixed model with random changepoint and the test procedure we propose for testing the existence of a random changepoint. Then we discuss tests for the variability of this changepoint. The procedure is evaluated in a simulation study in Section 3.2 and applied to real data on dementia among the elderly in Section 3.3. We finally discuss the method in Section 3.4. This work has been detailed in an article published in *Statistics in Medicine* (Segalas *et al.*, 2019).

3.1 Methodology

3.1.1 The mixed model with random changepoint

Let us denote $Y_{ij} = Y_i(t_{ij})$, the marker measure for subject i at time t_{ij} with $1 \leq i \leq N$ and $1 \leq j \leq n_i$ where n_i is the number of measures for subject i. Among all the possible formulations of the random changepoint mixed model presented in Section 2.2.2, we chose the Bacon-Watts formulation (2.10) with Griffiths and Miller (1973) transition. Already used by Dantan *et al.* (2011), it has the advantage of

being easily implemented and computationally attractive as stated by van den Hout *et al.* (2011) while keeping easy interpretability properties when γ is small.

The test we aim to develop can be formulated using a simple null hypothesis "*there is no random changepoint, i.e., the trajectory is linear*" and a simple alternative hypothesis "*the trajectory follows a random changepoint mixed model*". We then need a formulation of the model that allows to translate mathematically these null and alternative hypotheses in the most convenient way. Unfortunately, this cannot be done directly with formulation (2.10) because of the way τ_i intervenes as a product of both β_{1i} and β_{2i}. Also, when testing for the existence of the random changepoint, we first assume there is no random effect on β_2, that is the difference between the two slopes is assumed to be fixed if the random changepoint exists. The reasons behind this assumption are twofold. First are the computational reasons as removing one random effect is useful for numerical integration. Secondly, it appears logic to test an eventual interindividual variation of the difference of slopes around the changepoint only if this changepoint exists. This is why, we slightly modify (2.10) into

$$Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \beta_2 \sqrt{(t_{ij} - \tau_i)^2 + \gamma} + \varepsilon_{ij} \tag{3.1}$$

with

$$\tau_i = \mu_\tau + \sigma_\tau \tilde{\tau}_i \text{ and } \tilde{\tau}_i \sim \mathcal{N}(0,1),$$

$$\beta_{ki} = \beta_k^\top X_{ki} + \alpha_{ki} \text{ for } k = 0, 1,$$

$$\alpha_i = (\alpha_{0i}, \alpha_{1i})^\top \sim \mathcal{N}(0, \Sigma) \text{ with } \Sigma = \begin{pmatrix} \sigma_0 & \sigma_{01} \\ \sigma_{01} & \sigma_1 \end{pmatrix} \text{ and } \varepsilon_{ij} \sim \mathcal{N}(0, \sigma).$$

Thanks to this new formulation, τ_i only intervenes as a product of β_2 and the null and alternative hypotheses can be easily written into the following mathematical equations:

$$H_0 : \beta_2 = 0 \quad \text{vs.} \quad H_1 : \beta_2 \neq 0. \tag{3.2}$$

Therefore, our testing procedure is built as a two-step procedure. First, we test if a random changepoint exists (3.2) and then, if it exits, we can test if, around

this random changepoint, variation of the difference of slopes is subject-specific or covariate dependent using classical tests detailed in Section 3.1.6. However, because of this formulation, parameter β_{0i} loses its interpretation as the value of the marker at the changepoint. But as our main goal here is to test for the existence of the changepoint, this is an acceptable compromise. In (3.1), τ_i is the individual random changepoint, β_{1i} is the mean slope over the two phases and β_2 is half the difference of slopes between the two phases. The slope before the changepoint τ_i is $\beta_{1i} - \beta_2$ and the slope after τ_i is $\beta_{1i} + \beta_2$. In addition to the random changepoint, the intra-subject correlation is accounted for by including subject-specific random intercept and random slope, α_{0i} and α_{1i}. The model may include covariates in the vectors X_{0i} and X_{1i}, respectively associated with the mean level and with the mean slope.

In some applications, the changepoint could be constrained to be positive, for example when age is the time basis. In such a case, we could use a log-normal changepoint instead of a gaussian one. The changepoint support might also be fixed to a closed interval by chosing an appropriate probability law such as a truncated normal distribution as in van den Hout *et al.* (2013). These non gaussian distributions might be handled by suitable reparametrization so that the random effects could still be considered as normal variables (log-transformation for instance or changing the bounds and rescaling the interval).

The random coefficient α_{ki} are assumed to be independent from the random changepoint $\tilde{\tau}_i$ because it was previously observed that without this assumption, random changepoint mixed models were hardly identifiable (Jacqmin-Gadda *et al.*, 2006). Note, however that *a priori* independence does not imply the independence of the *posterior* random effects predictions. The individual deviation from the mean slope is assumed to be unchanged after the random changepoint because the difference of slope β_2 is not random. The inclusion of a random effect on β_2 as well as inclusion of covariates for the difference of slopes and the time of change are discussed in Section 3.1.6. However, we stress the fact that the time of the changepoint itself is random.

As advised by Dantan *et al.* (2011), we chose a small γ ($\gamma = 0.1$) to stay close to the linear-linear model while ensuring the derivability condition. The size of γ has

to be considered in relation to the time scale of the measured markers in our real data. This parameter could also be estimated by profile likelihood but we emphasize that large values of γ would make the interpretation of the changepoint and slopes more questionable as underlined by van den Hout *et al.* (2011).

3.1.2 Estimation procedure

The log-likelihood of model (3.1) takes the following form

$$\begin{aligned}\ell_N(Y;\beta_2,\theta) &= \sum_{i=1}^{N} \log f_i(Y_i;\beta_2,\theta) \\ &= \sum_{i=1}^{N} \log \iint \prod_{j=1}^{n_i} f(Y_{ij}|\alpha_i,\tilde{\tau}_i) f(\alpha_i) f(\tilde{\tau}_i) \mathrm{d}\alpha_i \mathrm{d}\tilde{\tau}_i. \end{aligned} \tag{3.3}$$

with $\theta = (\beta_0^\top, \beta_1^\top, \mu_\tau, \sigma, \sigma_0, \sigma_1, \sigma_{01}, \sigma_\tau)^\top$ the vector of all model parameters except β_2, $f_i(Y_i;\beta_2,\theta)$ the contribution to the likelihood of subject i, $f(Y_{ij}|\alpha_i,\tilde{\tau}_i)$ the univariate gaussian density for Y_{ij} given the random effects and $f(\alpha_i)$ and $f(\tilde{\tau}_i)$ the densities of the random effects. The log-likelihood (3.3) involves a three-dimensional integral on the random effects but conditionally on $\tilde{\tau}_i$, the log-likelihood is linear according to α_i and therefore the integral in α_i is analytic and we have

$$\ell_N(Y;\beta_2,\theta) = \sum_{i=1}^{N} \log \int \prod_{j=1}^{n_i} f(Y_{ij}|\tilde{\tau}_i) f(\tilde{\tau}_i) \mathrm{d}\tilde{\tau}_i. \tag{3.4}$$

Because $\tilde{\tau}_i \sim \mathcal{N}(0,1)$, if we substitute $\tilde{\tau}_i$ by $\breve{\tau}_i = \tilde{\tau}_i/\sqrt{2}$ in the integral of the log-likelihood (3.4), we have

$$\ell_N(Y;\beta_2,\theta) = \sum_{i=1}^{N} \log \frac{1}{\pi} \int \prod_{j=1}^{n_i} f(Y_{ij}|\breve{\tau}_i) \exp(-\breve{\tau}_i^2) \mathrm{d}\tilde{\tau}_i. \tag{3.5}$$

which makes the use of Gaussian quadrature straightforward. To estimate model (3.1) we chose a frequentist approach based on the maximization of the log-likelihood (3.5) using the Marquardt-Levenberg algorithm (Levenberg, 1944; Marquardt, 1963). The estimation algorithm based on this marginalized version of the log-likelihood

has been implemented in `R` (R Core Team, 2018) and `Rcpp` (Eddelbuettel and Balamuta, 2017) using the classic Gauss-Hermite quadrature with 20 nodes for numerical integration. The algorithm is freely available on the `R` package `rcpm` available on `GitHub` at https://github.com/crsgls/rcpm. Details are given in Appendix B.

3.1.3 Score test statistic

Our objective is to test for the existence of a random changepoint in the mixed model (3.1). As we said, thank to the new formulation we have proposed, the null hypothesis of no random changepoint may be defined by

$$H_0 : \beta_2 = 0 \quad \text{vs.} \quad H_1 : \beta_2 \neq 0.$$

Under the null hypothesis $H_0 : \beta_2 = 0$, the random changepoint model (3.1) reduces to a simple linear mixed model. Alternatively, the null hypothesis of absence of random changepoint could have been specified by $H_0 : \sigma_\tau = 0$ and $\mu_\tau \pm \infty$ (or any values outside the range of time) but this formulation of the null hypothesis is too complex to directly derive a test. Our objective, here, is not to test if the changepoint is random vs. the changepoint is fixed but to test for the existence of a changepoint that can be subject specific.

Among all the possible approach to perform this test, we chose a score test approach. It has the interesting advantage of avoiding the estimation of the alternative model, a random changepoint model in our case, which can be computationally expensive.

Taking the derivative of the log-likelihood (3.3) at β_2, we can compute the score

$$\begin{aligned} U_N(\beta_2;\theta) &= \frac{\partial \ell_N(Y;\beta_2,\theta)}{\partial \beta_2} \\ &= \sum_{i=1}^{N} \frac{\partial}{\partial \beta_2} \log f_i(Y_i;\beta_2,\theta) \\ &= \sum_{i=1}^{N} \frac{1}{f_i(Y_i;\beta_2,\theta)} \underbrace{\frac{\partial}{\partial \beta_2} f_i(Y_i;\beta_2,\theta)}_{(\star)} . \end{aligned} \tag{3.6}$$

Denoting

$$\tilde{Y}_{ij} = \mathbb{E}(Y_{ij}|\alpha_i,\tilde{\tau}_i) = \beta_{0i} + \beta_{1i} t_{ij} + \beta_2 \sqrt{(t_{ij}-\tau_i)^2+\gamma},$$

we have

$$(\star) = \iint f(\alpha_i) f(\tilde{\tau}_i) (\sqrt{2\pi}\sigma)^{-n_i} \underbrace{\frac{\partial}{\partial \beta_2} \prod_{j=1}^{n_i} \exp\left\{ -\frac{1}{2\sigma^2} (Y_{ij} - \tilde{Y}_{ij})^2 \right\}}_{(\star\star)} \mathrm{d}\alpha_i \mathrm{d}\tilde{\tau}_i. \tag{3.7}$$

Using the equality $(\prod_j u_j(\theta))' = \sum_j u_j'(\theta) \prod_{k\neq j} u_k(\theta)$, the derivative of the product is

$$\begin{aligned} (\star\star) = &\sum_{j=1}^{n_i} \frac{1}{\sigma^2} \exp\left\{ -\frac{1}{2\sigma^2} (Y_{ij} - \tilde{Y}_{ij})^2 \right\} (Y_{ij} - \tilde{Y}_{ij}) \sqrt{(t_{ij}-\tau_i)^2+\gamma} \\ &\times \prod_{k\neq j} \exp\left\{ -\frac{1}{2\sigma^2} (Y_{ik} - \tilde{Y}_{ik})^2 \right\}, \end{aligned} \tag{3.8}$$

Finally, by combining (3.6), (3.7) and (3.8) and by taking $\beta_2 = 0$, the null score has

the following form:

$$
\begin{aligned}
U_N(0;\theta) &= \sum_{i=1}^{N} u_i(0;\theta) \\
&= \sum_{i=1}^{N} \left[\int f(\tilde{\tau}_i) \int f(\alpha_i) \prod_{j=1}^{n_i} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2\sigma^2}(Y_{ij} - \beta_{0i} - \beta_{1i} t_{ij})^2 \right\} \mathrm{d}\alpha_i \mathrm{d}\tilde{\tau}_i \right]^{-1} \\
&\times \iint f(\alpha_i) f(\tilde{\tau}_i) (\sqrt{2\pi}\sigma)^{-n_i} \sum_{j=1}^{n_i} \left[\frac{1}{\sigma^2} \exp\left\{ -\frac{1}{2\sigma^2}(Y_{ij} - \beta_{0i} - \beta_{1i} t_{ij})^2 \right\} \right. \\
&\times (Y_{ij} - \beta_{0i} - \beta_{1i} t_{ij}) \sqrt{(t_{ij} - \tau_i)^2 + \gamma} \\
&\times \left. \prod_{k \neq j} \exp\left\{ -\frac{1}{2\sigma^2}(Y_{ik} - \beta_{0i} - \beta_{1i} t_{ik})^2 \right\} \right] \mathrm{d}\alpha_i \mathrm{d}\tilde{\tau}_i. \qquad (3.9)
\end{aligned}
$$

The observed score test statistic is

$$
S_N(0;\hat{\theta}_0) = \frac{U_N(0;\hat{\theta}_0)^2}{I_N(0;\hat{\theta}_0)} \qquad (3.10)
$$

where $\hat{\theta}_0$ is the maximum likelihood estimate (MLE) of all nuisance parameters θ under the null and $I_N(0;\hat{\theta}_0)$ is the variance of the score function given by Cox and Hinkley (1979)

$$
I_N(\beta_2, \theta) = I_{\beta_2\beta_2} - I_{\beta_2\theta} I_{\theta\theta}^{-1} I_{\beta_2\theta}^{\top},
$$

with

$$
\begin{aligned}
I_{\beta_2\beta_2} &= \sum_{i=1}^{N} E\left[\frac{\partial \log f_i}{\partial \beta_2} \right]^2, \\
I_{\theta\theta} &= \sum_{i=1}^{N} E\left[\left(\frac{\partial \log f_i}{\partial \theta} \right) \left(\frac{\partial \log f_i}{\partial \theta} \right)^{\top} \right], \\
I_{\beta_2\theta} &= \sum_{i=1}^{N} E\left[\left(\frac{\partial \log f_i}{\partial \beta_2} \right) \left(\frac{\partial \log f_i}{\partial \theta} \right)^{\top} \right].
\end{aligned}
$$

However, analytic computation of $I_N(0;\hat{\theta}_0)$ is intractable and numerical computation would be not precise enough without heavy computation due to numerical

integrals. Thus, as suggested by Freedman (2012), the variance is approximated by its empirical version

$$\hat{I}_N(0;\hat{\theta}_0) = \sum_{i=1}^{N} u_i(0;\hat{\theta}_0)^2$$

where u_i is the individual contribution to the score. When testing a random effect, this approximation may lead to a slightly conservative test (Commenges *et al.*, 1994) but simulation results in Section 3.2 show that the type I errors are acceptable.

3.1.4 The identifiability issue

Following the classic score-test approach, to compute the observed test statistic (3.10), we first need $\hat{\theta}_0$, the MLE of θ under the null. However, in our case, the whole set of parameters θ cannot be estimated under the null. The mean and variance of the changepoint, μ_τ and σ_τ, are indeed unidentifiable if $\beta_2 = 0$ as τ_i vanishes from model (3.1). We note η_0 the vector of identifiable nuisance parameters under the null, so that $\theta = (\eta_0, \mu_\tau, \sigma_\tau)$, and $\hat{\eta}_0$ its MLE under the null which therefore does not depend upon (μ_τ, σ_τ). The presence of unidentifiable parameters under the null makes that the classic approach does not hold anymore.

In the literature, three main ideas have been proposed to tackle the unidentifiability issue in score tests, generally in the framework of mixture models. First, the unidentifiable nuisance parameters may be replaced by a specific value, for example the MLE under the alternative hypothesis as suggested by Conniffe (2001). One disadvantage of such a strategy is that it requires estimating the alternative model which we wanted to avoid. Moreover, no theoretical result is proposed about the distribution of $S_N(0, \hat{\mu}_\tau, \hat{\sigma}_\tau, \hat{\eta}_0)$ where $\hat{\mu}_\tau, \hat{\sigma}_\tau$ are the MLE of μ_τ and σ_τ under the alternative. Even numerically, the distribution under the null remains difficult to compute and needs intense simulations.

A second method was proposed by Muggeo (2016b) to test for the existence of a fixed changepoint for longitudinal data. This test also only requires the estimation of the null model and the issue of unidentifiable nuisance parameters under the null is ruled out by replacing the quantity that vanishes under the null by its average over a pre-specified grid of values. The null distribution of the statistic is given and

its behaviour on finite sample assessed through simulation studies. Their approach to deal with unidentifiability is similar on philosophy to previous theoretical work by Andrews and Ploberger (1994) who also proposed to use a weighting function over a grid of nuisance parameter values from which they build test statistics. However they do not discussed how critical values could be computed in practice.

The last method is to consider the supremum of the test statistic over the unidentifiable nuisance parameters as the new test statistic. It has first been proposed by Davies (1977, 1987). They also proposed bounds for the probability of the critical region which can help building a test procedure. Unfortunately, as Hansen (1996) highlighted, they make a strong theoretical assumption about the derivative of the asymptotic of the test statistic that is often violated for large sample sizes and results in greater error. A resampling perturbation procedure based on the multiplier bootstrap (Van Der Vaart and Wellner, 1996) has been proposed by (Hansen, 1996). Since then, this procedure has been applied, for example, by Hsu *et al.* (2016) for building a supremum score test for testing homogeneity in mixture models.

3.1.5 The supremum score-test statistic

Following the idea of Hansen (1996), we consider the supremum of the score-test statistic over the unidentifiable nuisance parameters (μ_τ, σ_τ) as our test statistic, i.e.

$$T_N = \sup_{(\mu_\tau, \sigma_\tau)} S_N(0; \mu_\tau, \sigma_\tau, \eta_0). \tag{3.11}$$

To compute the observed value of T_N, denoted $T_N^{(obs)}$, the identifiable parameters η_0 are estimated from the null model, which is a standard linear mixed model. Then, replacing η_0 by its MLE $\hat{\eta}_0$, $S_N(0; \mu_\tau, \sigma_\tau, \hat{\eta}_0)$ is maximized over (μ_τ, σ_τ) by a Newton-like algorithm. We then compare this observed value $T_N^{(obs)}$ to the theoretical distribution of T_N under the null in order to compute the p-value. As this distribution is unknown and has no analytical expression, we use the resampling perturbation procedure proposed by Hansen (1996). To do so, we repeat the following steps for $b = 1, \ldots, B$:

1. we simulate N standard normal variables $(\xi_1^{(b)}, \ldots, \xi_N^{(b)})$;

2. we compute the sample statistic $T_N^{(b)}$

$$T_N^{(b)} = \sup_{(\mu_\tau, \sigma_\tau)} \frac{\left(\sum_{i=1}^N u_i(0; \mu_\tau, \sigma_\tau, \hat{\eta}_0)\xi_i^{(b)}\right)^2}{\sum_{i=1}^N u_i(0; \mu_\tau, \sigma_\tau, \hat{\eta}_0)^2}$$

Then, we obtain a sample of B realizations from the distribution of T_N under the null and the empirical p-value can be computed

$$\hat{p} = \sum_{b=1}^{B} \frac{1}{B} \mathbf{1}_{\{T_N^{(b)} > T_N^{(obs)}\}}.$$

The computation of the score statistic involves a maximization procedure performed with the BFGS quasi-Newton algorithm already implemented in the `R` (R Core Team, 2018) function `optim`. The three dimension integrals over $(\alpha_{0i}, \alpha_{1i}, \tilde{\tau}_i)$ are computed with a pseudo-adaptive gaussian quadrature approach. The estimates of the individual random effects from the null linear mixed model were used to rescale and center the quadrature nodes for $(\alpha_{0i}, \alpha_{1i})$. For $\tilde{\tau}_i$, standard Gauss-Hermite nodes were used. In the simulation study of Section 3.2 and in the application of Section 3.3, integrals are computed with five nodes for each dimension. The computation of the empirical p-value entails B optimizations of the perturbed sample statistics $T_N^{(b)}$ with $b = 1, \ldots, B$. This algorithm was implemented using `R` (R Core Team, 2018) in the function `testRCPMM` of the package `rcpm` that is available on `GitHub` and described in Appendix B.

3.1.6 Standard tests for the heterogeneity of β_{2i} and τ_i

Heterogeneity in the difference of slopes β_{2i}

In model (3.1), the time of change τ_i, the intercept β_{0i} and the mean slope β_{1i} are subject-specific but the slope difference between the two phases, β_2 is common for all subjects. If the previous test had concluded to the existence of a random changepoint, it might be worthwile testing whether there is variability in the parameter β_2, due either to covariates or to a random effect. Indeed, it would be plausible that the

heterogeneity between subjects increases in this second phase, i.e. the pathological phase, as opposed to the healthy phase.

If we want to test for a covariate effect, replacing β_2 by $\beta_{2i} = \beta_{20} + \beta_{21} X_i$ in model (3.1), the standard Wald test may be used to test $H_0 : \beta_{21} = 0$ vs. $H_1 : \beta_{21} \neq 0$.

To introduce a random difference in slopes, we replaced β_2 by $\beta_{2i} = \beta_2 + \alpha_{2i}$ with $\alpha_i = (\alpha_{0i}, \alpha_{1i}, \alpha_{2i}) \sim \mathcal{N}(0, \Sigma)$ in model (3.1). This model may be estimated by maximizing the likelihood (3.4), marginalized over α_i, since the integral over the three-dimensional α_i still has an analytical solution. To test the null hypothesis that β_2 is fixed versus random, the null and alternative hypotheses are defined by:

$$H_0 : \Sigma = \begin{pmatrix} \sigma_0 & \sigma_{01} & 0 \\ \sigma_{01} & \sigma_1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ vs. } H_1 : \Sigma = \begin{pmatrix} \sigma_0 & \sigma_{01} & \sigma_{02} \\ \sigma_{01} & \sigma_1 & \sigma_{12} \\ \sigma_{02} & \sigma_{12} & \sigma_2 \end{pmatrix}.$$

. A standard test statistic cannot be applied because, under the null, the parameter σ_2 lies on the boundaries of the parameter space $\mathbb{R}^+$. A corrected test for variance components was proposed by Stram and Lee (1994) who showed that, under the null, the likelihood ratio statistic followed a mixture of chi-squared distribution. For this test, with three non-independent random effects, it then follows a $0.5\chi^2_3 + 0.5\chi^2_2$ distribution. In practice, the p-value for this test is computed as the mean of the two p-values obtained by assuming that the likelihood ratio statistic has either a χ^2_2 or χ^2_3 distribution under the null.

Heterogeneity in the changepoint dates τ_i

It is also interesting to test the dependence of the mean and variance of the changepoint τ_i on some covariates. This can be achieved by estimating model (3.1) while including the covariate of interest X_i in the model for τ_i, i.e.

$$\tau_i = \mu_{\tau 0} + \mu_{\tau 1} X_i + (\sigma_{\tau 0} + \sigma_{\tau 1} X_i)\tilde{\tau}_i \text{ and } \tilde{\tau}_i \sim \mathcal{N}(0, 1).$$

Then again, standard Wald tests can be used to test the association of X_i with the mean time of change ($H_0 : \mu_{\tau 1} = 0$) and with the variance of the time of change

$(H_0 : \sigma_{\tau 1} = 0)$.

3.2 Simulations

3.2.1 Scenarios

We performed simulations studies to validate our testing procedure. The procedure was applied to data simulated under the null and the alternative hypotheses to compute the size and the power of the test, respectively. Under the null hypothesis, data were simulated according to the linear mixed model (M_0) $Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \varepsilon_{ij}$, while under the alternative, they were simulated according to model (3.1) with $\beta_{ki} = \beta_k + \alpha_{ki}$ for $k = 0, 1$. We generated samples of size $N = 50$ or $N = 100$, with 8 measures per individual at times $0, 3, 6, 9, 12, 15, 18, 21$.

For all generated models, the parameter values were $\beta_0 = 20$, $\beta_1 = -0.3$, $\sigma = 1$, $\sigma_0 = 1$, $\sigma_1 = 0.2$, $\rho_{01} = 0.5$, $\gamma = 0.1$. Several models were considered under the alternative hypothesis by varying β_2 (difference between the mean slopes of the two phases), μ_τ (mean time of change) and σ_τ (inter-individual variance in time of change). We used three different values for β_2 : -0.05, -0.075 and -0.1 which gave three models denoted M_1, M_2 and M_3 with respective slopes $(-0.25, -0.35)$, $(-0.225, -0.375)$ and $(-0.2, -0.4)$ (see Figure 3.1). For each of these three models, we used three different sets of values for (μ_τ, σ_τ) : $(10; 2)$, $(10; 4)$ and $(15; 2)$ leading to a total of nine alternative models. For each of these models, two scenarios were considered: one without any dropout and one with a dropout probability of 0.1 at each visit so that around half the initial sample remained in the cohort at the last visit. The empirical distribution of the null test statistic was simulated with $K = 500$ perturbed samples. Empirical powers and sizes of the test were computed using 1000 replicates.

3.2.2 Results

Table 3.1 presents the sizes and powers of the test for the various scenarios. First, the type I error was correct even if it might be slightly undersized for small samples with

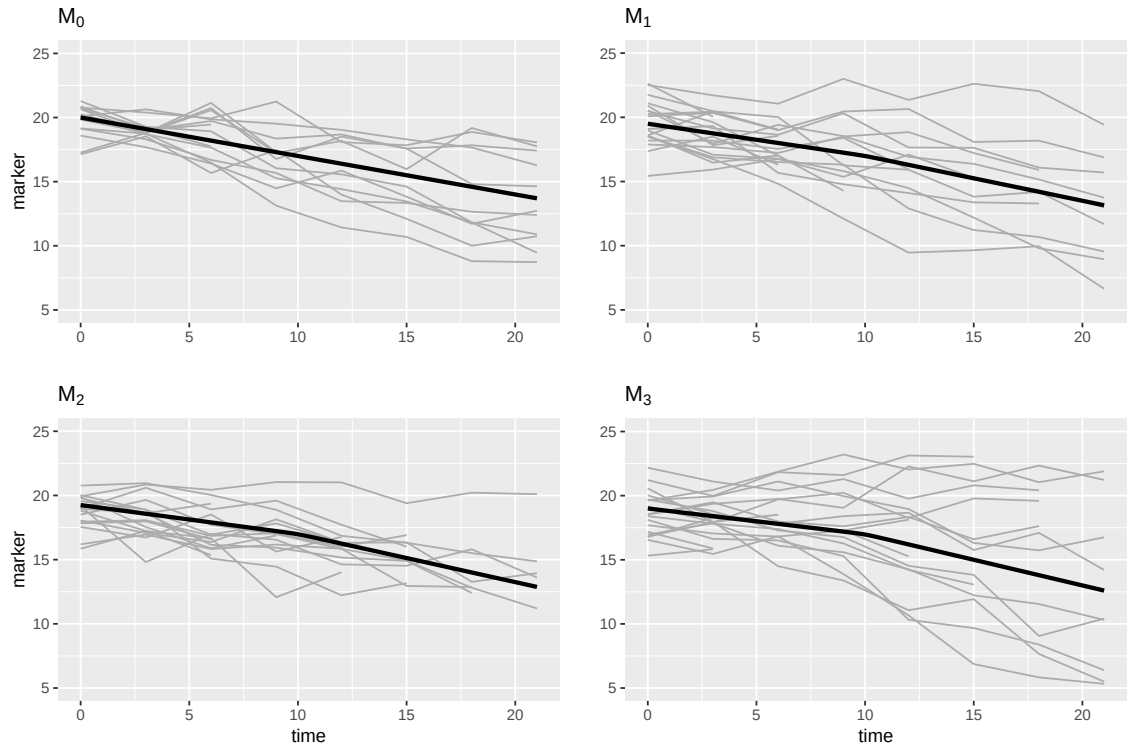

Figure 3.1: The null and the three alternative scenarios used for the simulations. The mean trajectory with $\mu_\tau = 10$ is plotted in solid black and individual trajectories with a dropout probability of 0.1 at each visit for 30 randomly generated subjects are plotted in grey.

N		50		100	
dropout		0	0.1	0	0.1
$(\mu_\tau, \sigma_\tau) = (10, 2)$	M_0	0.041	0.030	0.038	0.040
	M_1	0.630	0.304	0.966	0.680
	M_2	0.967	0.678	1	0.973
	M_3	1	0.945	1	1
$(\mu_\tau, \sigma_\tau) = (10, 4)$	M_1	0.470	0.185	0.864	0.501
	M_2	0.873	0.527	0.998	0.902
	M_3	0.980	0.791	1	0.993
$(\mu_\tau, \sigma_\tau) = (15, 2)$	M_1	0.303	0.071	0.626	0.207
	M_2	0.615	0.215	0.967	0.545
	M_3	0.917	0.438	0.999	0.869

Table 3.1: Size and power of the test computed on 1000 replicates of each scenarios with $K = 500$ perturbations.

dropout. As expected, the power increased with the absolute value of β_2 (from M_1 to M_3), i.e. with the intensity of the change in slopes, and with sample size ($N = 50$ vs. $N = 100$). Whatever the model and sample size, the power was lower for samples with dropout, owing to the loss of information and the shorter mean follow-up time. Indeed, dropout induces a strong imbalance between the number of measurements before and after the changepoint. Similarly, when the changepoint was later, i.e. $\mu_\tau = 15$, the power of the test decreased. This behaviour is expected because this shift in the changepoint time involves a loss of information on what happens next it compared to the case where $\mu_\tau = 10$. When the variance of the changepoint time σ_τ increased, the power also decreased, probably because more subjects had an unequal number of measures before and after their individual changepoint.

3.2.3 Power of the test when the difference in slopes is random

In our two-step procedure, the difference between the two slopes in the two phases, β_2, is assumed to be a fixed parameter. The test for its variability is only performed after the test for the existence of a random changepoint. It could thus be interesting to evaluate the behaviour of the test of $\beta_2 = 0$ when this assumption is violated,

N		50		100	
dropout		0	0.1	0	0.1
$(\mu_\tau, \sigma_\tau) = (10, 2)$	M_1	0.361	0.153	0.708	0.390
	M_2	0.732	0.407	0.986	0.863
	M_3	0.955	0.754	1	0.986
$(\mu_\tau, \sigma_\tau) = (10, 4)$	M_1	0.268	0.121	0.579	0.300
	M_2	0.623	0.328	0.952	0.729
	M_3	0.894	0.590	0.999	0.944
$(\mu_\tau, \sigma_\tau) = (15, 2)$	M_1	0.187	0.061	0.421	0.147
	M_2	0.457	0.140	0.859	0.426
	M_3	0.752	0.307	0.991	0.752

Table 3.2: Power of the test computed on 1000 replicates of each scenarios with $K = 500$ perturbations with data simulated with a random β_{2i}, $\sigma_2 = 0.1$.

i.e., when the parameter β_2 is in fact random. To do so, the score test was applied to data simulated with a subject-specific random effect $\beta_{2i} = \beta_2 + \alpha_{2i}$. All previous values of parameters were kept and we chose $\sigma_2 = 0.1$, $\rho_{02} = corr(\alpha_{0i}, \alpha_{2i}) = 0.5$ and $\rho_{12} = corr(\alpha_{1i}, \alpha_{2i}) = -0.2$. Results for all previous simulated scenarios are shown in Table 3.2. As expected, the power globally decreased when β_2 was random compared to when it was fixed, but the obtained powers remained very satisfactory. These results suggest that our procedure give good results even when the assumption of a fixed β_2 is violated.

3.2.4 Power analyses for changepoint models

The simulation procedure, also implemented in the package, may be used to conduct a power analysis for the detection of a difference in slopes. It is of clinical interest to know what sample size is needed to detect a random changepoint for a given difference in slopes and a given power. The idea is to compute the empirical power for different sample sizes N and to select the minimal value of N for which the empirical power is at the desired level. As an illustration, we computed the empirical power for a difference of slopes of 0.15, corresponding to scenario M_2, with $(\mu_\tau, \sigma_\tau) = (15, 2)$. We used 1000 replicates for each of $N = 30, 40, 50, 60, 70, 100, 125, 150$. We plotted the results obtained for a fixed β_2 as well as for a random β_{2i} in Figure 3.2. The

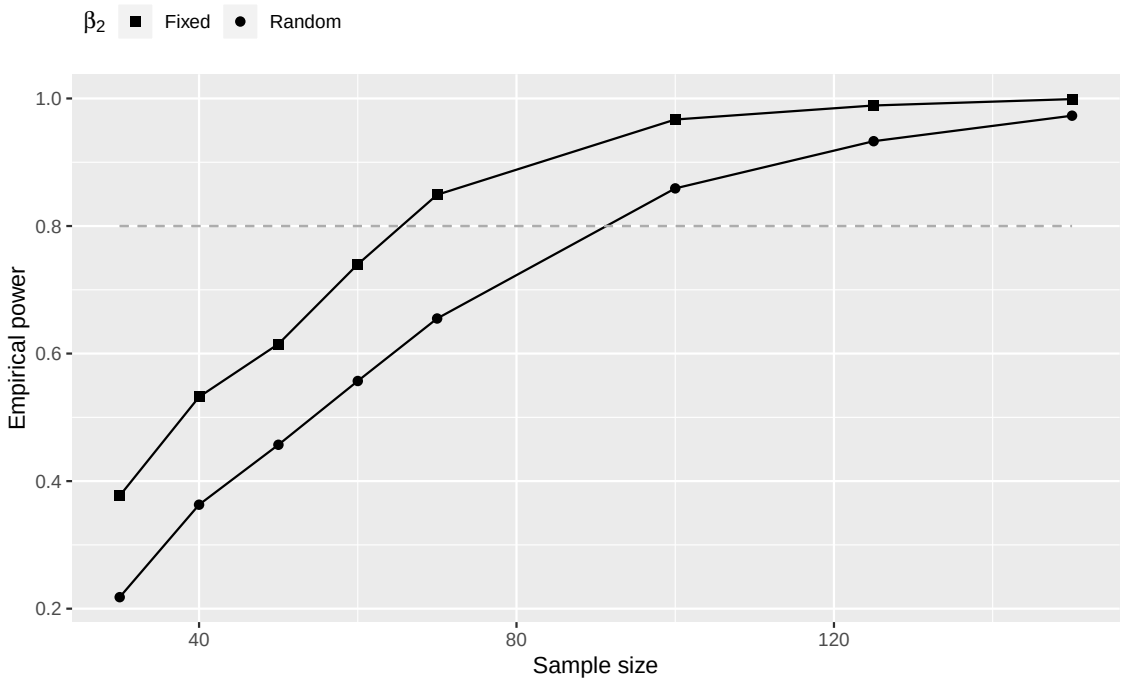

Figure 3.2: Empirical power obtained for a mean difference of slopes of 0.15 assuming either a fixed difference of slopes (β_2) or a random difference of slopes (β_{2i}) for different sample size with 1000 replicates for each. The horizontal dashed grey line represents a power of 0.8.

minimum sample size for a power of 0.8 would then be $N = 70$ if β_2 is fixed and $N = 100$ if it is random.

3.3 Application

We applied the proposed inference procedure on real data about dementia in the elderly. We tested the existence of an acceleration of cognitive decline before the diagnosis of dementia whatever the educational level and, if changepoints were identified, compared them according to educational level.

3.3.1 Sample

For a general presentation of the Paquid cohort, see Section 1.4.1. In our analysis, only subjects free of dementia at baseline and diagnosed as demented during the 25 years of follow-up were selected because our objective was here to study the trajectory of cognitive decline before dementia. Subjects without any measures were also excluded. The final sample included 880 incident cases of dementia, of whom 522 had a primary school diploma or higher educational level. The mean number of measures for each subject was 5.8 (from 1 to 12) for high educational subjects and 4.6 (from 1 to 12) for low educational ones. Delay to dementia in years was used as the time basis and the date of dementia was estimated by the mean, denoted D_i, between the date of diagnosis and the date of the last visit without dementia. The time scale used for the analysis is the difference between the measurement time for the cognitive score and D_i. Thus, time 0 is the imputed time of dementia. It is negative in the pre-dementia phase and positive thereafter.

3.3.2 Test for the random changepoint

We performed the analyses separately for subjects with high educational level (HEL) and subjects with low educational level (LEL). We tested $H_0 : \beta_2 = 0$ vs. $H_1 : \beta_2 \neq 0$ without any adjustment in the model defined by (3.1). To compute the empirical test statistic distribution under the null, we applied the perturbation procedure with $K = 500$. For the HEL sample, the observed test statistic was 143.7 and the estimated p-value $p < 0.001$. For the LEL sample, the observed test statistic was 56.9 with $p < 0.001$. We then rejected the null hypothesis of no random changepoint for both educational levels. There was statistical evidence for a subject-specific breakpoint in the trajectory of the Isaacs Set test before dementia for both educational levels.

As mentioned above, in the presence of a random changepoint, we can test whether there is inter-individual variability in the parameter β_{2i} by including the random effect α_{2i}. To compute the likelihood ratio statistic (LRS) for $H_0 : \sigma_2 = 0$, the alternative random changepoint model with four random effects was estimated. For the HEL, the LRS was -139.1 ($p < 0.001$) while it was -20.4 ($p < 0.001$)

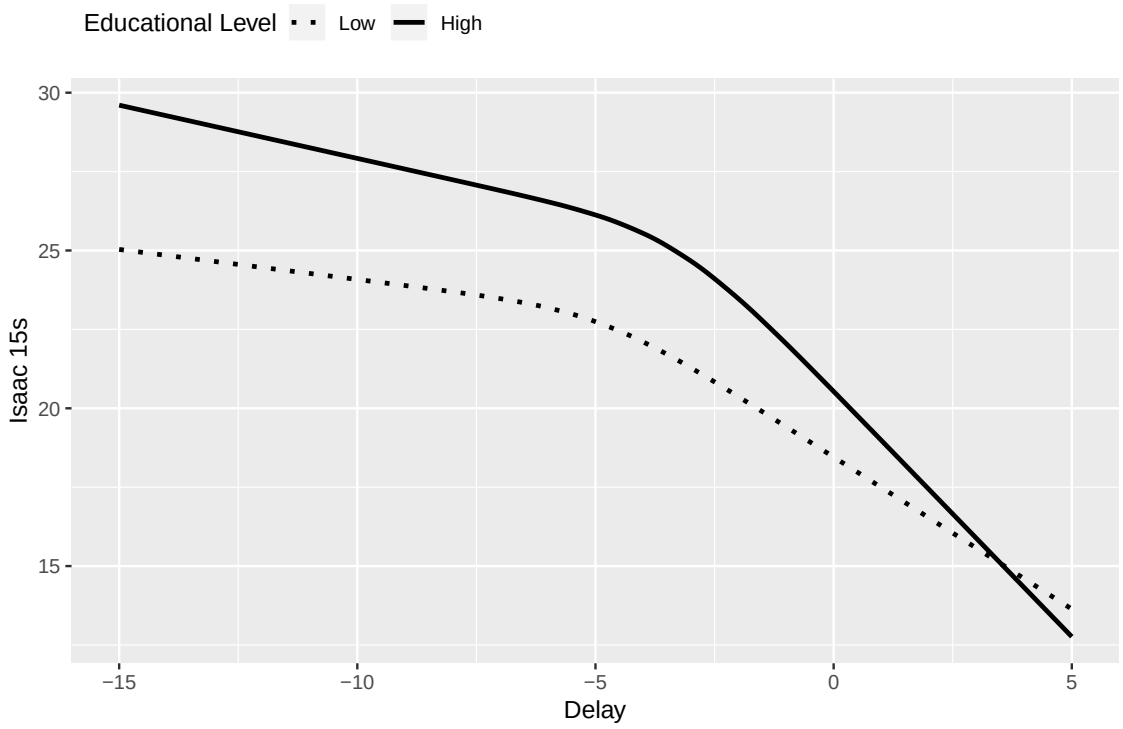

Figure 3.3: Marginal estimated trajectories of the random changepoint mixed models for HEL and LEL subjects.

for the LEL. There was statistical evidence for an inter-individual variation in the difference of slopes. From now on, we included this random effect in the model.

3.3.3 Estimation of the mixed model with random changepoint

The mixed model with random changepoint

$$Y_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \beta_{2i} \sqrt{(t_{ij} - \tau_i)^2 + \gamma} + \varepsilon_{ij}. \tag{3.12}$$

was then estimated on each subsample with $\beta_{ki} = \beta_{k0} + \alpha_{ki}$ for $k = 0, 1, 2$ with $\alpha_i = (\alpha_{0i}, \alpha_{1i}, \alpha_{2i})^\top \sim \mathcal{N}(0, \Sigma)$, $\gamma = 0.1$ and $\tau_i = \mu_\tau + \sigma_\tau \tilde{\tau}_i$. The model was estimated by maximum likelihood using a Levenberg-Marquardt algorithm (Levenberg,

	Paquid sample of demented subjects			
	High education (N = 522)		Low education (N = 358)	
	Esti.	95%CI	Esti.	95%CI
β_0	22.56	[22.08,23.03]	20.32	[19.69,20.96]
β_1	-0.94	[-1.03,-0.86]	-0.58	[-0.66,-0.50]
β_2	-0.61	[-0.69,-0.52]	-0.39	[-0.46,-0.32]
μ_τ	-3.31	[-4.06,-2.55]	-4.82	[-6.10,-3.55]
σ	3.26	[3.15,3.37]	3.26	[3.11,3.41]
σ_τ	1.85	[1.25,2.45]	1.92	[0.88,2.95]

Table 3.3: Parameter estimates of the mixed model with random changepoint on Paquid data stratified on educational level. Numerical integration uses classic gaussian quadrature with 20 nodes.

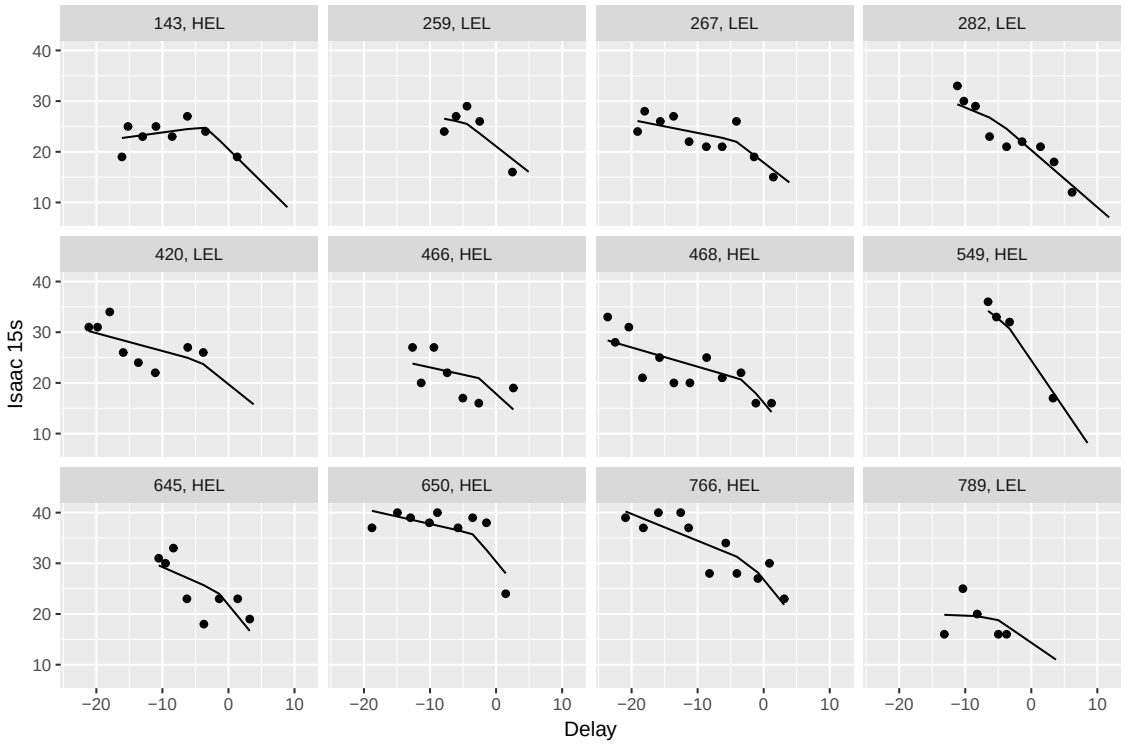

Figure 3.4: Estimated individual trajectories (solid line) for randomly selected subjects of either educational level (HEL or LEL) from the Paquid cohort compared to their observed values (dots).

1944; Marquardt, 1963) and gaussian quadrature with 20 nodes to compute the marginalized log-likelihood (3.4). Estimated parameters are shown in Table 3.3. Confidence intervals (CI) were computed by standard maximum likelihood theory that can be applied here given that we know from the test results that β_2 is non null and that the changepoint parameters are identifiable. The variances are estimated from the inverse of the observed Hessian matrix and the CI are computed using asymptotic normal distribution. The marginal estimated trajectories are plotted in Figure 3.3 for each educational level and were computed using classic Gauss-Hermite quadrature to integrate

$$\mathbb{E}(Y(t), \hat{\theta}, \hat{\beta}_2) = \int \mathbb{E}(Y(t)|\tau_i, \hat{\theta}, \hat{\beta}_2) f(\tau_i|\hat{\theta}, \hat{\beta}_2) \mathrm{d}\tau_i.$$

Table 3.3 highlights differences in the IST trajectories according to educational level. Indeed, the mean slope over the two phases appeared steeper ($\beta_1 = -0.94$ vs. -0.58). Regarding the changepoint, it appeared closer to the time of dementia onset ($\mu_\tau = -3.31$ vs. -4.82) for HEL and the difference between the two slopes was also larger ($\beta_2 = -0.61$ vs -0.39). All these differences are noticeable on the plotted estimated trajectories of the two groups in Figure 3.3. On the other hand, the variance in the changepoint time and the residual variances were similar between the two groups. From the previously estimated parameters, we can compute the first and second slopes and their 95% CI for LEL subjects: -0.19 $[-0.27;-0.1]$ and -0.96 $[-1.09;-0.84]$ as well as for HEL subjects: -0.34 $[-0.39;-0.28]$ and -1.55 $[-1.71;-1.39]$. These results enlightened the difference in trajectories.

The estimated variance matrices for the random effects among LEL subjects Σ_0 and among HEL subjects Σ_1 also suggested some differences between educational levels:

$$\hat{\Sigma}_0 = \begin{pmatrix} 14.96 & -0.01 & -0.01 \\ -0.01 & 0.08 & 0.02 \\ -0.01 & 0.02 & 0.01 \end{pmatrix} \quad \hat{\Sigma}_1 = \begin{pmatrix} 18.93 & -0.08 & -0.03 \\ -0.08 & 0.26 & -0.07 \\ -0.03 & -0.07 & 0.03 \end{pmatrix}$$

The inter-subject variability was greater among HEL which is expected because

	Paquid sample of demented subjects				
	High education (N=522)		Low education (N=358)		Test for equality
	Esti.	95%CI	Esti.	95%CI	p-value
β_0	22.553	[22.08,23.03]	20.323	[19.69,20.96]	< 0.001
β_1	-0.945	[-1.03,-0.86]	-0.577	[-0.66,-0.50]	< 0.001
β_2	-0.608	[-0.69,-0.53]	-0.388	[-0.46,-0.32]	< 0.001
μ_τ	-3.295	[-4.06,-2.53]	-4.823	[-6.09,-3.56]	0.042
σ	3.259	[3.17,3.35]	3.259	[3.17,3.35]	
σ_τ	1.852	[1.24,2.46]	1.919	[0.89,2.95]	0.913

Table 3.4: Parameter estimates of the mixed model with random changepoint on Paquid data adjusted on educational level. Numerical integration uses classic gaussian quadrature with 20 nodes.

the shift used to distinguish the two subsamples, the primary school certificate, leads to more heterogeneity among HEL.

To test the apparent difference between the two groups, we estimated the complete model (3.12) on the whole sample with both HEL and LEL subjects and including educational level as a covariate in the modelling of each regression parameter β_k and of the mean μ_k and standard error σ_k of the changepoint. In addition, the covariance matrix Σ for the random effects was specific for each educational level. According to results from the stratified analysis in Table 3.3, the residual variance was assumed identical for the two educational levels. Results of the adjusted model, see Table 3.4, show that β_0, β_1 and β_2 were significantly different between educational levels ($p < 0.001$ for the three tests). The mean time of change occured significantly later for HEL ($p = 0.04$) while the variance in the changepoint time σ_τ was not different between the groups ($p = 0.9$). The delayed and steeper acceleration of cognitive decline among HEL supports the hypothesis of their greater cognitive reserve (Stern *et al.*, 1994; Stern, 2009). Figure 3.4 displays the subject-specific estimated trajectories for 12 randomly selected subjects with their observed values, showing a good fit of individual trajectories.

3.4 Discussion

We have proposed a procedure to test the existence of a random changepoint for longitudinal data accounting for intra-subject correlation. Our model allows the trajectory parameters to depend upon covariates and subject-specific random effects to account for inter-individual variations. A normal distribution was chosen for the random effects, but another type of distribution could be considered such as a log-normal or a truncated normal distribution. The approach used to perform the test is a supremum score test coupled with a perturbation method to sample the empirical distribution of the test statistic under the null. The performance of our procedure was studied through simulation. Results of the application to the Paquid cohort demonstrated the existence of a change of slopes in the cognitive decline before dementia for both educational levels. We detected a later changepoint for HEL subjects which is consistent with the cognitive reserve hypothesis (Stern, 2009). Both the testing procedure and the estimation algorithm have been made available on a `R` package `rcpm` available on the `GitHub` platform and are described in Appendix B. Note that another package exists to estimate linear mixed models with random changepoint (Muggeo, 2016a).

Another possible choice for the testing procedure is the approach proposed by Conniffe (2001) which replaces the unidentifiable parameters under the null by their estimates under the alternative. As pointed out, it needs the alternative model estimation and the null statistic distribution still remains analytically intractable. Moreover, unlike the supremum score test, no theoretically founded simulation method has been proposed to approximate the null statistic distribution. Wald and likelihood ratio tests were also ruled out for several reasons. First, both require the alternative model estimation which can be quite time-consuming. Second, there are no asymptotic results for the null asymptotic distribution with unidentifiable parameters. Moreover, any bootstrap-based simulation method to sample this distribution would need intense numerical computations. These reasons and the abundant literature on the supremum score test motivated our choice.

We propose a step-by-step procedure to investigate the structure of the random

changepoint model. First, we tested whether a random changepoint exists assuming that the difference of slopes between the two phases β_2 was identical for all subjects, contrary to the intercept and the mean slope which are subject-specific. This is legitimate because the randomness of β_{2i} does not make sense without the existence of the changepoint. We carefully evaluated the impact of this assumption. Our complementary simulation study showed that the test performs well in terms of power even when β_{2i} is actually random. Also, we emphasize that this test is designed to detect a non-zero mean difference of slopes i.e. a change in the mean trajectory. If our procedure detects that this changepoint exists, the between-subject variability in the slope difference can be tested. It may depend upon some covariates or upon a subject-specific random effect. Finally, dependence of the mean and variance of the changepoint time upon covariates may be tested by standard tests such as the Wald test, as we did for the educational level.

Testing the existence of a random changepoint in a mixed model is a topic of wide interest in biomedical studies, especially with regards to the natural history of chronic diseases or marker changes under treatment. This question also arises in the more complex context of joint models for time-to-events and longitudinal markers. For example, testing for a random changepoint in the cognitive decline trajectory by modelling jointly the age at dementia and death would avoid a selection bias due to dropout and death. However, in such a joint analysis with age as time scale, the difference in detected changepoints would not reflect a difference in delay to dementia but rather a difference in terms of age at dementia. Nevertheless, the score-test procedure we propose is valid for joint models but requires some changes in the score statistic formula (3.9) to include terms regarding survival submodels.

Chapter 4

Bivariate random changepoint model and temporal order

On the previous chapter, we proposed a test procedure to assess the existence of a random changepoint for longitudinal data. If the test was positive, we proposed a way to estimate the random changepoint model and to compare its amplitude according to a covariate. However, we might want to compare the time of change between different markers in order to compare their temporal order of decline. This would help to better understand the natural history of the disease. Such a comparison implies the two markers to be modelled jointly in a bivariate random changepoint model. In the next section we described the proposed methodology to proceed to this time of change comparison. We evaluate our procedure through simulation studies in Section 4.2 and apply it to real data in Section 4.3. We discuss this methodology in Section 4.4.

4.1 Methodology

4.1.1 Model formulation

Let us denote $Y_i = (Y_i^1, Y_i^2)^\top$ the n_i-vector of all measures for subject i where Y_i^1 and Y_i^2 are the n_i^1-vector and n_i^2-vector of measures of subject i from marker 1 and 2

respectively with $n_i = n_i^1 + n_i^2$. We assume that each of these markers is described by a random changepoint mixed model. Again, we chose the model (2.10) introduced by Bacon and Watts (1971) with the transition function proposed by Griffiths and Miller (1973). Here, we do not use the reformulation of the model that was needed for testing the existence of the changepoint in the previous chapter. The model is written

$$Y_{ij}^l = Y^l(t_{ij}^l) = \beta_{0i}^l + \beta_{1i}^l(t_{ij}^l - \tau_i^l) + \beta_{2i}^l\sqrt{(t_{ij}^l - \tau_i^l)^2 + \gamma} + \varepsilon_{ij}^l \tag{4.1}$$

for marker $l = 1, 2$, subject $i = 1, \ldots, N$ and measure $j = 1, \ldots, n_i^l$. With these notations, Y_{ij}^l is the value of marker l for subject i at time t_{ij}^l. We assume that $\varepsilon_{ij}^l \sim \mathcal{N}(0, \sigma_{\varepsilon^l})$ and that the $(\varepsilon_{ij}^l)_{ijl}$ are independent according to both i, j and l. In this work, the smoothing parameter γ is assumed identical for the two markers because we are comparing markers with similar timescales and we chose $\gamma = 0.1$.

Each regression coefficient is written as the sum of a fixed effect and a random effect $\beta_{ki}^l = \beta_k^l + b_{ki}^l$, with $b_i^l = (b_{0i}^l, b_{1i}^l, b_{2i}^l)^\top \sim \mathcal{N}(0, B^l)$ where B^l is a positive definite matrix of dimension 3, and the random changepoint $\tau_i^l = \mu_\tau^l + \sigma_{\tau^l}\tilde{\tau}_i^l$, with $\tilde{\tau}_i^l \sim \mathcal{N}(0, 1)$. We assume that $\tilde{\tau}_i{}^l$ and b_i^l are independant. The bivariate model is defined by considering the covariances between the marker-specific random effects: $B^{12} = Cov(b_i^1, b_i^2)$ and $d_{12} = Cov(\tilde{\tau}_i{}^1, \tilde{\tau}_i{}^2)$.

In matrix notation, the bivariate model (4.1) is defined by:

$$Y_i = \Gamma_i\beta_i + \varepsilon_i \tag{4.2}$$

where $\beta_i = (\beta_{0i}^1, \beta_{1i}^1, \beta_{2i}^1, \beta_{0i}^2, \beta_{1i}^2, \beta_{2i}^2)^\top \sim \mathcal{N}(\beta, B)$ with $\beta = (\beta_0^1, \beta_1^1, \beta_2^1, \beta_0^2, \beta_1^2, \beta_2^2)^\top$ and $\tilde{\tau}_i = (\tilde{\tau}_i^1, \tilde{\tau}_i^2)^\top \sim \mathcal{N}(0, D)$ and $\varepsilon_i \sim \mathcal{N}_{n_i}(0, \Sigma_i)$ with

$$\Gamma_i = \left[\begin{array}{c|c} T_i^1 & 0_{n_i^1 \times 3} \\ \hline 0_{n_i^2 \times 3} & T_i^2 \end{array}\right], B = \left[\begin{array}{c|c} B^1 & B^{12} \\ \hline B^{21} & B^2 \end{array}\right], D = \begin{bmatrix} 1 & d_{12} \\ d_{12} & 1 \end{bmatrix},$$

$$\Sigma_i = \left[\begin{array}{c|c} \sigma_{\varepsilon^1}\mathbb{I}_{n_i^1} & 0_{n_i^1\times n_i^2} \\ \hline 0_{n_i^2\times n_i^1} & \sigma_{\varepsilon^2}\mathbb{I}_{n_i^2} \end{array}\right], T_i^l = \left[\begin{array}{ccc} 1 & t_{i1}-\tau_i^l & \sqrt{(t_{i1}-\tau_i^l)^2+\gamma} \\ \vdots & \vdots & \vdots \\ 1 & t_{in_i^l}-\tau_i^l & \sqrt{(t_{in_i^l}-\tau_i^l)^2+\gamma} \end{array}\right].$$

where $\mathbb{I}_N$ is the identity matrix of size N.

4.1.2 Estimation procedure

We choose to estimate model (4.2) by directly maximizing the log-likelihood and use the maximum likelihood theory to derive tests for comparing trajectories of the markers. The log-likelihood of model (4.2) is:

$$\begin{aligned} \ell_N(\theta) &= \sum_{i=1}^{N} \log f(Y_i,\theta) \\ &= \sum_{i=1}^{N} \log \iint f(Y_i|\tilde{\tau}_i,b_i) f(\tilde{\tau}_i,b_i)\mathrm{d}\tilde{\tau}_i\mathrm{d}b_i \\ &= \sum_{i=1}^{N} \log \int f(Y_i|\tilde{\tau}_i) f(\tilde{\tau}_i)\mathrm{d}\tilde{\tau}_i \end{aligned} \tag{4.3}$$

where the vector θ includes all model parameters (fixed effects and variance parameters). Given $\tilde{\tau}_i$, the model is linear and the integral over the random coefficients b_i has a closed form. The conditional distribution $Y_i|\tilde{\tau}_i$ is multivariate Gaussian with mean and variance defined by:

$$Y_i|\tilde{\tau}_i \sim \mathcal{N}(\Gamma_i\beta,\ \Gamma_i B\Gamma_i^\top + \Sigma_i)$$

The integral of size 2 over $\tilde{\tau}_i$ in the last term of (4.3) does not have an analytical solution and needs to be approximated numerically. We use pseudo-adaptive Gauss-Hermite quadrature to estimate joint models. We first estimate the two univariate random changepoint models separately and compute predictions of the individual random changepoints $\hat{\tilde{\tau}}_i^l$ by empirical Bayes estimates. The *posterior* expectation $E(\tilde{\tau}_i^l|Y_i^l)$ is approximated by the mode of the distribution of $f(Y_i^l|\tilde{\tau}_i^l)f(\tilde{\tau}_i^l)$.

The modes are computed with the Newton-like Levenberg-Marquardt optimization algorithm (Levenberg, 1944; Marquardt, 1963) and their variance matrix are estimated from the observed Hessian matrix. Then, we center and rescale the nodes of the Gauss-Hermite quadrature using $\hat{\tilde{\tau}}_i^l$ for $l = 1, 2$ and their variance, allowing to reduce the number of quadrature nodes to 10. Finally, the Newton-like Levenberg-Marquardt optimization algorithm is used again to maximize the log-likelihood of the bivariate model (4.2).

In some scenario with great variability, it might be necessary to increase the number of nodes or to use a two-step pseudo-adaptive Gauss-Hermite quadrature as proposed by Ferrer *et al.* (2016). In the case of the bivariate random changepoint model, it means that after a first run of the optimisation algorithm, we estimate the BLUP $\hat{\tilde{\tau}}_i^1$ and $\hat{\tilde{\tau}}_i^2$ for both markers from the estimated bivariate random changepoint model by taking the mode of $f(Y_i^1, Y_i^2|\tilde{\tau}_i^1, \tilde{\tau}_i^2)f(\tilde{\tau}_i^1, \tilde{\tau}_i^2)$. Then, we update nodes and weights of the Gaussian quadrature and run a second optimisation algorithm.

The optimization is performed on a non-constrained space using a parametrization that ensures positive-definiteness of the variance matrices Σ_i, B and D. Instead of estimating these matrices directly, we estimated U the Cholesky matrix of B such as $U^\top U = B$ where U is an upper triangular matrix, V the Cholesky matrix of D, σ_{ε^1} and σ_{ε^2} the residual standard deviations from the matrix Σ_i. Variances of all the estimated parameters are computed directly from the inverse of the Hessian matrix and the delta-method is used to compute the variance of the estimates of the untransformed parameters in Sigma, B and D (Oehlert, 1992). Let us note $\tilde{\theta}$ the Cholesky and standard deviation parameters and h a function such that $h(\tilde{\theta})$ encompass all the variance parameters of the model. We can deduce the variance of $h(\tilde{\theta})$ from the asymptotic variance of $\tilde{\theta}$,

$$var(h(\tilde{\theta})) \approx \nabla h(\tilde{\theta})^\top var(\tilde{\theta}) \nabla h(\tilde{\theta})$$

where ∇h denotes the gradient of the function h.

4.1.3 Curvilinearity

Model (4.2) assumes that the two markers are Gaussian whereas many psychometric tests have asymmetric distribution that often highlights ceiling or floor effect. To take into account such curvilinearity, a smooth transformation of each marker can be used, instead of their crude values. Following Proust-Lima *et al.* (2013), we extend the model to include monotonic marker specific transformations of the crude markers defined on a basis of I-splines (Ramsay, 1988). They are defined as an integral of M-splines so that the transformation is bijective. Explicit formulation of the splines are given on Appendix C. The curvilinear model is defined by the transformation

$$\tilde{Y}_{ij}^l = g^l(Y_{ij}^l, \eta^l) = \eta_0^l + \sum_{k=1}^{q} \eta_k^{l2} I_k^l(Y_{ij}^l) \tag{4.4}$$

where the transformed variable $\tilde{Y}_{ij}^l$ follows model (4.1). The parameters $(\eta_k^l)_k$ have to be estimated and I_k^l is the marker specific I-spline basis depending on pre-specified degree and number and location of knots. In this work, I-splines of degree 2 with 2 internal knots located at the terciles were used. To ensure identifiability of the model, we added the constraints $\beta_0^l = 0$ and $\sigma_{\varepsilon^l} = 1$: the intercept values are captured by η_0^l while the residual variances are captured by the $(\eta_k^l)_k$.

The log-likelihood of the curvilinear model defined by (4.2) and (4.4) has the following form:

$$\ell_N(\theta) = \sum_{i=1}^{N} \log \int f(\tilde{Y}_i|\tilde{\tau}_i) f(\tilde{\tau}_i) \mathrm{d}\tilde{\tau}_i + N \log |J_g^1||J_g^2|$$

where θ now also includes I-spline parameters $(\eta_k^l)_k$, $\tilde{Y}_i = (\tilde{Y}_i^1, \tilde{Y}_i^2)$ and $|J_g^l|$ is the determinant of the Jacobian matrix of g^l with the constraints $\beta_0^l = 0$ and $\sigma_{\varepsilon^l} = 1$ for $l = 1, 2$.

4.1.4 Comparing the mean times of change

In order to identify the temporal order of cognitive impairments in dementia, the comparison of the mean times of change of the two markers was a major objective of the bivariate modelling. The null hypothesis to be tested is $H_0 : \mu_\tau^1 - \mu_\tau^2 = 0$. Thanks to the maximum likelihood theory, a simple Wald test can be used:

$$\frac{\hat{\mu}_\tau^1 - \hat{\mu}_\tau^2}{\text{var}(\hat{\mu}_\tau^1 - \hat{\mu}_\tau^2)}$$

which under the null follows a $\chi^2(1)$ distribution. The variance is estimated by

$$\hat{\text{var}}(\hat{\mu}_\tau^1 - \hat{\mu}_\tau^2) = \hat{\text{var}}(\hat{\mu}_\tau^1) + \hat{\text{var}}(\hat{\mu}_\tau^2) - 2 \times \hat{\text{cov}}(\hat{\mu}_\tau^1, \hat{\mu}_\tau^2)$$

computed from the observed Hessian matrix and the delta-method for the transformed variance parameters.

4.2 Simulation

4.2.1 Scenarios

Simulations were performed to assess our estimation procedure and validate the test for comparing the mean time of change. We simulated the longitudinal trajectories of two correlated markers, according to model (4.2), measured at seven equally spaced times from $t = -25$ to $t = 0$. These measurement times were assumed identical for all subjects. To evaluate how the sample size impacts the behaviour of the test procedure and the estimation quality, we simulated data for $N = 100$ and $N = 500$ subjects.

For all scenarios and for the first marker the intercept was $\beta_0^1 = 10$ and slopes parameters were $\beta_1^1 = -0.3$, $\beta_2^1 = -0.3$ while for the second marker they were $\beta_0^2 = 20$, $\beta_1^2 = -0.9$ and $\beta_2^2 = -0.6$. The residual variance was 1 for both markers and we chose $\sigma_{\tau^1}^2 = 2$, $\sigma_{\tau^2}^2 = 3$ and $\rho_{\tau^{12}} = 0.5$. Variance parameters for β_i are given below:

$$B^1 = B^2 = \begin{bmatrix} 1 & 0.5\sqrt{1 \times 0.1} & 0.5\sqrt{1 \times 0.1} \\ 0.5\sqrt{1 \times 0.1} & 0.1 & 0.5\sqrt{0.1 \times 0.1} \\ 0.5\sqrt{1 \times 0.1} & 0.5\sqrt{0.1 \times 0.1} & 0.1 \end{bmatrix},$$

$$B^{12} = B^{21\top} = \begin{bmatrix} 0.5 & 0.2\sqrt{1 \times 0.1} & 0.2\sqrt{1 \times 0.1} \\ 0.2\sqrt{1 \times 0.1} & 0.5\sqrt{0.1 \times 0.1} & 0.2\sqrt{0.1 \times 0.1} \\ 0.2\sqrt{1 \times 0.1} & 0.2\sqrt{0.1 \times 0.1} & 0.5\sqrt{0.1 \times 0.1} \end{bmatrix}.$$

We considered two different scenarios for the mean time of change: the first corresponds to the null hypothesis $\mu_\tau^1 = \mu_\tau^2$ and the second corresponds to the alternative hypothesis $\mu_\tau^1 \neq \mu_\tau^2$. For these two scenarios, we first generated markers using a gaussian model and then using a curvilinear model so that gaussian assumption is violated. For the latter one, markers were $\sqrt{10\tilde{y}}$ where $\tilde{y}$ was generated from the gaussian model. For the gaussian scenario, we generated two variants: a first with centered changepoints ($\mu_\tau^1 = \mu_\tau^2 = -10$ for the null and $\mu_\tau^1 = -10$, $\mu_\tau^2 = -8$ for the alternative) and a second with late changepoints ($\mu_\tau^1 = \mu_\tau^2 = -5$ for the null and $\mu_\tau^1 = -5$, $\mu_\tau^2 = -3$ for the alternative). For the curvilinear scenario, only the centered case was considered. Finally, we have six main different scenarios: gaussian null, gaussian alternative, late gaussian null, late gaussian alternative, curvilinear null and curvilinear alternative.

From 500 replicates for all scenarios, estimation quality was assessed through the bias, the comparison of asymptotic and empirical standard error and the coverage rate of the 95% confidence interval. Then from the alternative and null scenarios, we computed empirical powers and sizes of the test for comparing the time of change. The optimisation was performed using a pseudo-adaptive Gaussian quadrature with 10 nodes.

4.2.2 Results

All results from simulations are reported in Table 4.1 for the gaussian model with centered changepoints, in Table 4.2 for the gaussian model with late changepoints

and in Table 4.3 for the curvilinear model (β_0^l and σ_{ε^l} do not appear in Table 4.3 because their value are constrained to be zero and one respectively).

As we can see from these tables, the overall estimation quality is good with satisfying coverage rates and no bias for all scenarios and all parameters. As expected, the results are better when the sample size increases and with $N = 500$ coverage rates are very good. The sizes of the comparison test for the time of change were close to the nominal value 0.05 and the empirical powers of the test were excellent. Table 4.2 shows that even when changepoints are shifted towards the right, i.e. when the information before and after the changepoint is imbalanced, the estimation procedure performs well. However, due to this imbalance, the size of the test slightly increases around the value of 0.07. Table 4.3 shows that both our estimation procedure and the test for comparison of the time of change are valid for non-gaussian markers thanks to the I-spline link transformation. Figure 4.1 displays the estimated transformation and the true transformation of the marker for all curvilinear scenarios. The estimated link functions fit very well the true ones.

		$N = 100$					$N = 500$				
	θ	$\hat{\theta}$	bias%	SdEmp	SdAs	CR	$\hat{\theta}$	bias%	SdEmp	SdAs	CR
gaussian null scenario: $\mu_\tau^1 = \mu_\tau^2$											
β_0^1	10.000	10.008	0.081	0.139	0.143	94.8	9.998	0.018	0.064	0.064	94.6
β_1^1	-0.300	-0.300	0.088	0.032	0.033	95.0	-0.302	0.563	0.014	0.015	95.0
β_2^1	-0.300	-0.299	0.361	0.034	0.033	94.8	-0.300	0.003	0.014	0.015	96.6
μ_τ^1	-10.000	-10.006	0.056	0.289	0.288	95.6	-9.994	0.064	0.134	0.130	94.2
$\sigma^2_{\tau^1}$	2.000	1.963	1.843	0.811	0.790	90.2	2.039	1.954	0.395	0.362	93.0
$\sigma^2_{\varepsilon^1}$	1.000	1.000	0.001	0.079	0.074	93.6	0.998	0.177	0.034	0.033	94.4
β_0^2	20.000	20.005	0.025	0.181	0.183	94.8	20.000	0.002	0.083	0.082	94.6
β_1^2	-0.900	-0.899	0.103	0.034	0.033	95.0	-0.901	0.138	0.015	0.015	96.4
β_2^2	-0.600	-0.598	0.412	0.034	0.033	93.8	-0.600	0.018	0.015	0.015	94.0
μ_τ^2	-10.000	-10.008	0.078	0.255	0.245	94.0	-9.998	0.024	0.113	0.110	94.2
$\sigma^2_{\tau^2}$	3.000	2.963	1.231	0.659	0.640	92.6	3.010	0.327	0.292	0.289	94.4
$\sigma^2_{\varepsilon^2}$	1.000	0.997	0.275	0.075	0.076	95.6	1.000	0.028	0.033	0.034	96.6
$\sigma_{\tau^{12}}$	1.225	1.222	0.239	0.532	0.530	95.4	1.237	1.015	0.246	0.241	94.8
empirical size of the test $H_0 : \mu_\tau^1 = \mu_\tau^2$											
				0.052					*0.050*		
gaussian alternative scenario: $\mu_\tau^1 \neq \mu_\tau^2$											
β_0^1	10.000	10.003	0.033	0.151	0.142	94.0	10.001	0.007	0.064	0.064	95.0
β_1^1	-0.300	-0.300	0.089	0.032	0.033	95.4	-0.302	0.559	0.015	0.015	95.2
β_2^1	-0.300	-0.302	0.766	0.032	0.033	94.8	-0.300	0.052	0.015	0.015	93.0
μ_τ^1	-10.000	-9.999	0.007	0.313	0.287	93.2	-10.000	0.005	0.135	0.129	94.4
$\sigma^2_{\tau^1}$	2.000	1.992	0.416	0.884	0.782	88.0	1.967	1.647	0.375	0.357	92.6
$\sigma^2_{\varepsilon^1}$	1.000	0.994	0.568	0.079	0.074	92.4	1.002	0.168	0.034	0.033	95.4
β_0^2	20.000	19.997	0.014	0.181	0.180	95.8	20.001	0.006	0.088	0.081	92.6
β_1^2	-0.900	-0.902	0.225	0.037	0.035	93.8	-0.901	0.086	0.017	0.016	92.4
β_2^2	-0.600	-0.604	0.627	0.037	0.035	93.2	-0.600	0.077	0.016	0.016	94.2
μ_τ^2	-8.000	-7.985	0.190	0.277	0.254	92.4	-8.003	0.037	0.118	0.115	92.8
$\sigma^2_{\tau^2}$	3.000	2.945	1.847	0.656	0.641	92.4	3.002	0.070	0.306	0.289	94.6
$\sigma^2_{\varepsilon^2}$	1.000	0.990	1.013	0.078	0.075	92.6	1.000	0.000	0.035	0.034	94.6
$\sigma_{\tau^{12}}$	1.225	1.252	2.219	0.574	0.543	93.2	1.216	0.737	0.226	0.245	95.8
empirical power of the test $H_0 : \mu_\tau^1 = \mu_\tau^2$											
				0.998					*1.000*		

SdEmp: Empirical standard deviation; SdAs: Mean asymptotic standard deviation; CR: coverage rate of the 95% confidence interval.

Table 4.1: Results of the simulation study for the gaussian model with centered changepoints.

		$N = 100$					$N = 500$				
	θ	$\hat{\theta}$	bias%	SdEmp	SdAs	CR	$\hat{\theta}$	bias%	SdEmp	SdAs	CR
late gaussian null scenario: $\mu_\tau^1 = \mu_\tau^2$											
β_0^1	10.000	10.012	0.117	0.142	0.139	95.2	10.001	0.013	0.062	0.062	94.4
β_1^1	-0.300	-0.300	0.058	0.039	0.039	95.2	-0.299	0.243	0.018	0.018	94.6
β_2^1	-0.300	-0.298	0.747	0.038	0.038	93.6	-0.300	0.023	0.018	0.018	94.2
μ_τ^1	-5.000	-5.021	0.420	0.416	0.372	89.6	-5.005	0.110	0.179	0.173	94.4
$\sigma^2_{\tau^1}$	2.000	1.925	3.766	0.988	0.828	84.8	1.998	0.097	0.416	0.400	94.2
$\sigma^2_{\varepsilon^1}$	1.000	0.997	0.262	0.073	0.072	95.6	1.000	0.031	0.032	0.033	94.6
β_0^2	20.000	19.992	0.038	0.191	0.186	94.4	20.005	0.026	0.083	0.083	94.8
β_1^2	-0.900	-0.906	0.632	0.049	0.045	93.4	-0.899	0.083	0.021	0.020	92.6
β_2^2	-0.600	-0.605	0.849	0.046	0.044	92.8	-0.600	0.077	0.021	0.020	94.2
μ_τ^2	-5.000	-4.963	0.743	0.332	0.310	93.2	-4.997	0.051	0.144	0.139	93.6
$\sigma^2_{\tau^2}$	3.000	3.047	1.572	0.856	0.786	90.0	3.016	0.534	0.369	0.356	94.0
$\sigma^2_{\varepsilon^2}$	1.000	0.994	0.563	0.072	0.073	94.8	1.001	0.140	0.035	0.033	94.4
$\sigma_{\tau^{12}}$	1.225	1.241	1.309	0.690	0.612	89.0	1.223	0.149	0.288	0.280	94.6
empirical size of the test $H_0 : \mu_\tau^1 = \mu_\tau^2$											
				0.076					*0.072*		
late gaussian alternative scenario: $\mu_\tau^1 \neq \mu_\tau^2$											
β_0^1	10.000	9.999	0.013	0.140	0.139	95.4	10.000	0.001	0.065	0.062	94.6
β_1^1	-0.300	-0.302	0.749	0.038	0.040	95.8	-0.300	0.050	0.018	0.018	95.6
β_2^1	-0.300	-0.304	1.257	0.040	0.040	96.2	-0.300	0.143	0.018	0.018	94.2
μ_τ^1	-5.000	-5.001	0.015	0.401	0.377	91.6	-4.998	0.036	0.181	0.175	93.6
$\sigma^2_{\tau^1}$	2.000	1.978	1.120	0.950	0.866	88.4	2.007	0.353	0.420	0.406	94.4
$\sigma^2_{\varepsilon^1}$	1.000	0.998	0.242	0.071	0.073	94.0	1.000	0.026	0.033	0.033	95.6
β_0^2	20.000	19.975	0.123	0.230	0.205	90.4	19.984	0.081	0.101	0.091	92.4
β_1^2	-0.900	-0.919	2.141	0.091	0.076	89.8	-0.912	1.342	0.039	0.034	89.6
β_2^2	-0.600	-0.620	3.406	0.090	0.075	89.4	-0.613	2.117	0.038	0.034	90.8
μ_τ^2	-3.000	-2.933	2.223	0.498	0.406	88.6	-2.954	1.538	0.207	0.183	91.8
$\sigma^2_{\tau^2}$	3.000	2.939	2.027	0.835	0.738	90.6	3.034	1.137	0.343	0.338	93.4
$\sigma^2_{\varepsilon^2}$	1.000	0.994	0.551	0.075	0.070	91.6	1.001	0.117	0.032	0.032	94.0
$\sigma_{\tau^{12}}$	1.225	1.201	1.936	0.700	0.609	88.8	1.234	0.764	0.283	0.281	94.8
empirical power of the test $H_0 : \mu_\tau^1 = \mu_\tau^2$											
				0.942					*1.000*		

SdEmp: Empirical standard deviation; SdAs: Mean asymptotic standard deviation; CR: coverage rate of the 95% confidence interval.

Table 4.2: Results of the simulation study for the gaussian model with late changepoints.

			$N = 100$					$N = 500$		
θ	$\hat{\theta}$	bias%	SdEmp	SdAs	CR	$\hat{\theta}$	bias%	SdEmp	SdAs	CR
curvilinear null scenario: $\mu_\tau^1 = \mu_\tau^2$										
β_0^1	-	-	-	-	-	-	-	-	-	-
β_1^1	-0.300	-0.298	0.579	0.034	0.036	96.4	-0.297	0.846	0.016	0.016	95.2
β_2^1	-0.300	-0.298	0.747	0.038	0.038	93.6	-0.298	0.690	0.017	0.017	93.8
μ_τ^1	-10.000	-10.025	0.255	0.285	0.296	94.6	-10.024	0.242	0.137	0.133	92.6
$\sigma^2_{\tau^1}$	2.000	1.919	4.063	0.850	0.786	88.6	1.974	1.289	0.356	0.361	94.0
$\sigma^2_{\varepsilon^1}$	-	-	-	-	-	-	-	-	-	-	-
β_0^2	-	-	-	-	-	-	-	-	-	-	-
β_1^2	-0.900	-0.897	0.298	0.057	0.055	95.0	-0.896	0.430	0.023	0.025	96.0
β_2^2	-0.600	-0.598	0.390	0.055	0.054	94.2	-0.596	0.601	0.022	0.024	95.8
μ_τ^2	-10.000	-10.070	0.699	0.257	0.263	94.8	-10.030	0.300	0.119	0.119	93.8
$\sigma^2_{\tau^2}$	3.000	2.989	0.363	0.711	0.654	93.0	2.972	0.930	0.273	0.292	95.6
$\sigma^2_{\varepsilon^2}$	-	-	-	-	-	-	-	-	-	-	-
$\sigma_{\tau^{12}}$	1.225	1.204	1.665	0.539	0.528	92.8	1.201	1.959	0.222	0.239	97.0
empirical size of the test $H_0 : \mu_\tau^1 = \mu_\tau^2$										
			0.052					*0.064*		
curvilinear alternative scenario: $\mu_\tau^1 \neq \mu_\tau^2$										
β_0^1	-	-	-	-	-	-	-	-	-	-	-
β_1^1	-0.300	-0.298	0.800	0.037	0.036	94.0	-0.298	0.806	0.016	0.016	93.8
β_2^1	-0.300	-0.295	1.591	0.039	0.038	93.8	-0.297	1.066	0.017	0.017	94.4
μ_τ^1	-10.000	-10.049	0.493	0.299	0.295	94.2	-10.005	0.052	0.137	0.134	94.2
$\sigma^2_{\tau^1}$	2.000	1.905	4.736	0.865	0.782	87.0	1.972	1.388	0.372	0.363	92.6
$\sigma^2_{\varepsilon^1}$	-	-	-	-	-	-	-	-	-	-	-
β_0^2	-	-	-	-	-	-	-	-	-	-	-
β_1^2	-0.900	-0.890	1.121	0.058	0.059	94.0	-0.893	0.762	0.026	0.026	94.8
β_2^2	-0.600	-0.587	2.225	0.055	0.057	94.0	-0.593	1.197	0.025	0.025	94.4
μ_τ^2	-8.000	-8.061	0.757	0.290	0.286	95.2	-8.045	0.565	0.133	0.126	91.4
$\sigma^2_{\tau^2}$	3.000	2.989	0.358	0.710	0.660	91.2	2.999	0.033	0.286	0.293	93.4
$\sigma^2_{\varepsilon^2}$	-	-	-	-	-	-	-	-	-	-	-
$\sigma_{\tau^{12}}$	1.225	1.202	1.875	0.589	0.545	93.4	1.220	0.363	0.250	0.247	95.4
empirical power of the test $H_0 : \mu_\tau^1 = \mu_\tau^2$										
			0.998					*1.000*		

SdEmp: Empirical standard deviation; SdAs: Mean asymptotic standard deviation; CR: coverage rate of the 95% confidence interval.

Table 4.3: Results of the simulation study for the curvilinear model.

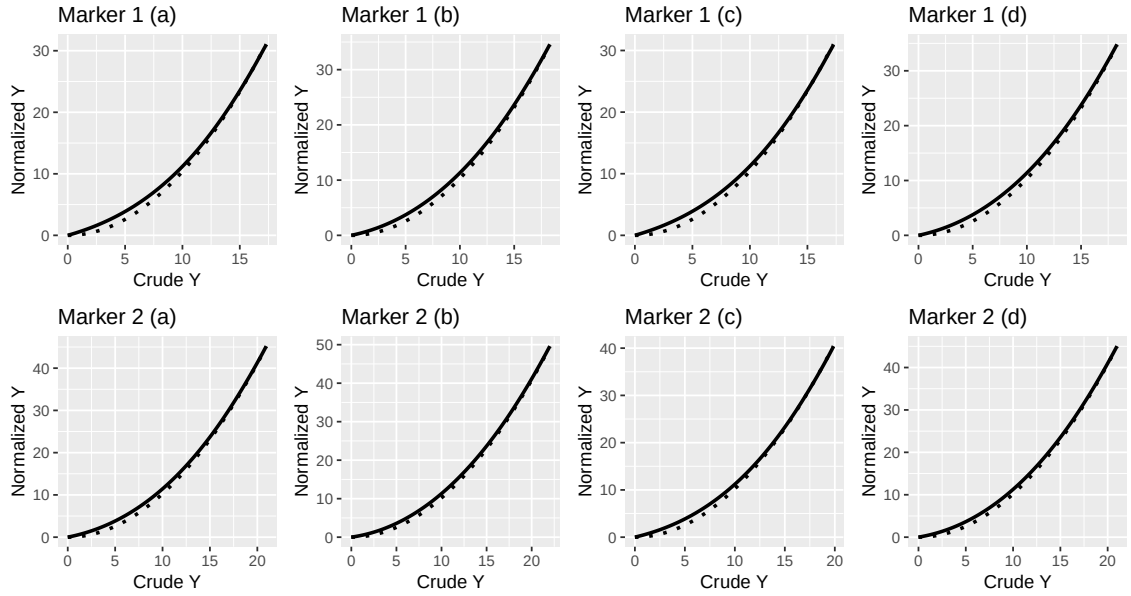

Figure 4.1: Estimated link function (solid) vs. true link function (dotted) for both markers for the four following scenarios: (a) $N = 100$, $\mu_\tau^1 = \mu_\tau^2 = -10$; (b) $N = 500$, $\mu_\tau^1 = \mu_\tau^2 = -10$; (c) $N = 100$, $\mu_\tau^1 = -10$, $\mu_\tau^2 = -8$; (d) $N = 500$, $\mu_\tau^1 = -10$, $\mu_\tau^2 = -8$.

4.2.3 Comparison to Yang and Gao (2013)

In their work, Yang and Gao (2013) compared different formulations of the bivariate mixed model with random changepoint: the broken-stick (2.8), the Bacon-Watts (2.9) with hyperbolic tangent transition and the polynomial model (2.12). In their simulation studies, they discarded the Bacon-Watts model that gave poor results and preferred the smooth polynomial model. However, in their simulation, data were simulated only from a polynomial model. Two scenarios were considered, one with small variances and one with greater variances for the changepoint and residual errors with the following values:

$$\text{Small variances: } \sigma^2_{\tau 1} = 16;\ \sigma^2_{\tau 2} = 4;\ \sigma^2_{\varepsilon 1} = 5;\ \sigma^2_{\varepsilon 2} = 1$$
$$\text{Great variances: } \sigma^2_{\tau 1} = 64;\ \sigma^2_{\tau 2} = 16;\ \sigma^2_{\varepsilon 1} = 20;\ \sigma^2_{\varepsilon 2} = 5$$

In order to check if the misspecification of the Bacon-Watts model had an impact on the results they obtained, we proceed to some further simulations studies. We simulated scenarios from the Bacon-Watts model and chose parameters as close as possible as those used by Yang and Gao (2013) that were chosen from the fit of their model on the Indianapolis–Ibadan Dementia Study (Hendrie *et al.*, 1995).

We simulated data from the Bacon-Watts model (2.10) for $N = 238$ subjects followed over time during twenty years with seven equally spaced visits, assumed equals for all subjects and without any missing data. Both the small and great variance scenarios of Yang and Gao (2013) were considered. Results from 500 replicates of our model estimation are shown in Table 4.4 and Table 4.5 for the small and great variance scenarios respectively. The optimisation was performed using a pseudo-adaptive Gaussian quadrature with 10 nodes.

For the small variance scenario, results are correct with no bias and good coverage rates except for the mean time of change for the first marker which shows low coverage rate due to a large underestimation of its standard error and a slight bias for the variance of this random changepoint. For the great variance scenario, results are globally less satisfying for the first marker changepoint. Concerning the empirical power of the test, for both scenarios we get very satisfying power equal to one for the small variance scenario and with only a slight decrease for the great variance scenario. Compared to what was obtained by Yang and Gao (2013), our estimation results are much better. It suggests that, as expected, the way they simulated data has a clear impact on their results and their comparison. Also, we can note that their changepoint variance parameters are quite huge for both the small and great variances scenarios (see our estimated parameters on real data in Section 4.3 for comparison). As they do not propose a statistical procedure to compare the mean

			$N = 238$			
	θ	$\hat{\theta}$	bias%	SdEmp	SdAs	CR
β_0^1	70.000	70.001	0.001	0.355	0.341	92.8
β_1^1	-1.600	-1.605	0.303	0.047	0.047	94.0
β_2^1	-1.400	-1.397	0.185	0.099	0.099	94.4
μ_τ^1	15.000	14.933	0.450	0.423	0.267	80.0
σ_τ^1	16.000	15.253	4.667	1.669	1.711	90.2
σ_ε^1	5.000	4.973	0.542	0.238	0.239	94.0
β_0^2	28.000	28.015	0.054	0.270	0.276	95.2
β_1^2	-0.300	-0.298	0.673	0.031	0.030	94.4
β_2^2	-0.100	-0.102	1.866	0.029	0.030	95.0
μ_τ^2	10.000	9.987	0.131	0.260	0.256	94.4
σ_τ^2	4.000	4.139	3.466	1.005	0.946	94.0
σ_ε^2	1.000	1.001	0.133	0.049	0.050	95.6
$\sigma_{\tau^{12}}$	3.200	3.233	1.037	1.014	1.021	95.2
empirical power of the test $H_0 : \mu_\tau^1 = \mu_\tau^2$						
			1			

SdEmp: Empirical standard deviation; SdAs: Mean asymptotic standard deviation; CR: coverage rate of the 95% confidence interval.

Table 4.4: Results for the small variances scenario of the simulation study inspired from Yang and Gao (2013).

changepoint, we cannot compare the power of the test.

4.3 Application

4.3.1 The Three-City Study

As detailed in Section 1.4.2, the Three-City Study (3C Study) is an observational cohort of elderly involving repeated measures of cognitive tests over time and assessment of dementia diagnosis at each visit started in France in 1999. In this analysis, we focused on the Grober and Buschke test (GB), a French adaptation of the Free and Cued Selective Reminding Test that measures memory function (Grober and Buschke, 1987) through several recalls of 16 words. This test was proposed only in the Bordeaux center at visits at 2, 7, 10, 12 and 14 years so that only a maximum number of 5 measures per subject was available. The present analysis was

			$N = 238$			
	θ	$\hat{\theta}$	bias%	SdEmp	SdAs	CR
β_0^1	70.000	69.963	0.053	0.630	0.549	91.0
β_1^1	-1.600	-1.605	0.340	0.086	0.071	90.0
β_2^1	-1.400	-1.414	0.979	0.123	0.106	89.8
μ_τ^1	15.000	14.726	1.826	1.047	0.448	62.8
σ_τ^1	64.000	54.408	14.987	6.868	6.062	61.2
σ_ε^1	20.000	19.821	0.895	0.927	0.903	92.6
β_0^2	28.000	27.955	0.162	0.413	0.380	92.8
β_1^2	-0.300	-0.301	0.191	0.032	0.033	95.6
β_2^2	-0.100	-0.101	1.121	0.036	0.036	94.4
μ_τ^2	10.000	10.103	1.027	0.881	0.753	88.8
σ_τ^2	16.000	15.281	4.492	4.190	3.854	91.2
σ_ε^2	5.000	4.982	0.352	0.233	0.238	94.8
$\sigma_{\tau^{12}}$	3.200	2.490	22.177	4.107	3.787	88.6
empirical power of the test $H_0 : \mu_\tau^1 = \mu_\tau^2$						
			0.968			

SdEmp: Empirical standard deviation; SdAs: Mean asymptotic standard deviation; CR: coverage rate of the 95% confidence interval.

Table 4.5: Results for the great variances scenario of the simulation study inspired from Yang and Gao (2013).

performed on the sample of 401 incident cases of dementia diagnosed over the 14 years of follow-up in the center of Bordeaux. Our objective was to compare the immediate recall and the free recall which respectively measures the ability to encode an information and to memorize it.

The timescale used in this application is the time to the diagnosis, the diagnosis being considered as the time 0. We kept the measures up to five years after the diagnosis to get enough information on individual trajectories but we discarded measures beyond 5 years because the rate of missing measures were very high and the missingness mechanism was probably informative. We used the curvilinear model defined by (4.2) and (4.4) because the gaussian assumption was not valid with these two markers as it can be seen on the histogram of the two markers displayed in Figure 4.2. The estimation procedure described in Section 4.1.2 was run with 10 quadrature nodes using pseudo-adaptive Gaussian quadrature.

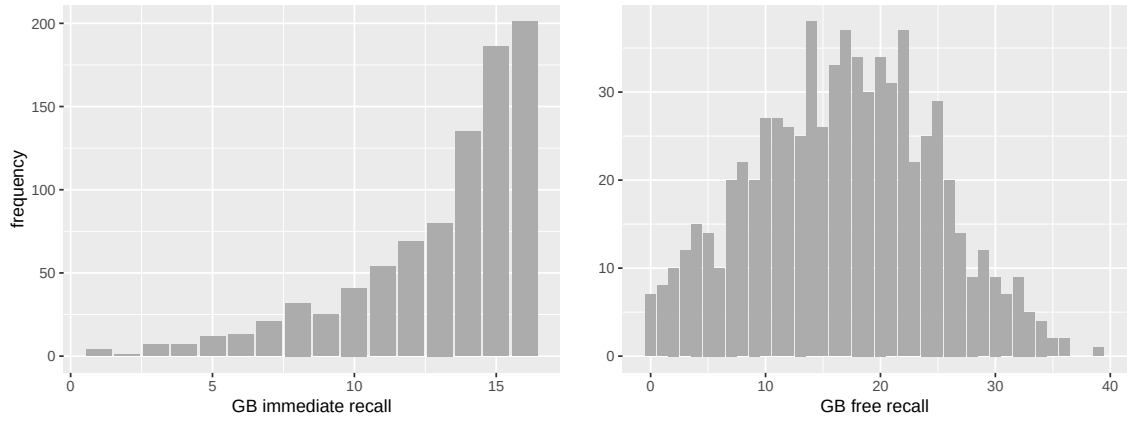

Figure 4.2: Histogram of GB immediate and free recall.

4.3.2 Results

We estimated a bivariate random changepoint model for these two markers. Table 4.6 shows the estimates for the trajectory parameters of both markers on the transformed scale. Below is $\hat{B}$, the estimated variance covariance matrix of the six random effects b_i on β_{0i}^l, β_{1i}^l and β_{2i}^l, three for each transformed marker. It was computed from the estimated Cholesky parameters of this matrix.

$$\hat{B} = \begin{pmatrix} 0.452 & -0.011 & -0.047 & 0.804 & 0.000 & -0.061 \\ -0.011 & 0.004 & 0.003 & -0.018 & 0.006 & 0.006 \\ -0.047 & 0.003 & 0.008 & -0.080 & 0.001 & 0.013 \\ 0.804 & -0.018 & -0.080 & 1.780 & -0.025 & -0.150 \\ 0.000 & 0.006 & 0.001 & -0.025 & 0.011 & 0.007 \\ -0.061 & 0.006 & 0.013 & -0.150 & 0.007 & 0.030 \end{pmatrix}$$

For the changepoint, variances were estimated at 6.791 and 0.726 respectively for the GB immediate and free recall with an estimated correlation of 0.922. For each marker, we plotted in Figure 4.3 all the individual trajectories and the marginal trajectory estimated from the bivariate random changepoint model obtained by

$$E(\tilde{Y}^l(t), \hat{\theta}^l) = \int E(\tilde{Y}^l(t)|\tau_i^l, \hat{\theta}^l) f(\tau_i^l|\hat{\theta}^l) \mathrm{d}\tau_i^l$$

whose integral was computed with a classic Gauss-Hermite quadrature with 20 nodes. In Table 4.6, the mean time of change was estimated at -3.177 $(-3.856; -2.498)$ for the immediate recall and at -5.82 $(-6.954; -4.685)$ years before diagnosis for the free recall. These mean times of change were significantly different according to the Wald test ($p = 0.047$). This means that the ability to memorize declines before the ability to encode an information. Figure 4.3 highlights the differences in the estimated trajectories of decline of the two markers with an earlier changepoint for the GB free recall.

	GB immediate recall			GB free recall			Wald test	
	$\hat{\beta}$	se($\hat{\beta}$)	95%CI	$\hat{\beta}$	se($\hat{\beta}$)	95%CI	Stat.	pvalue
β_1	-0.286	0.023	[-0.331;-0.242]	-0.262	0.037	[-0.334;-0.189]	0.589	0.443
β_2	-0.230	0.022	[-0.272;-0.187]	-0.229	0.029	[-0.285;-0.173]	0.024	0.877
μ_τ	-3.177	0.347	[-3.856;-2.498]	-5.820	0.579	[-6.954;-4.685]	3.937	0.047

se: standard error; 95%CI: 95% confidence interval.

Table 4.6: Results of the estimation of the bivariate random changepoint model on the immediate and free GB recall of the 3C Bordeaux demented subjects.

The fit of the bivariate random changepoint mixed model was evaluated in Figure 4.4. In the upper panel, we compared the individual observed values to the individual predicted values in the transformed scale for all subjects while the lower panel displays the observed and estimated trajectories for 8 randomly selected subjects. Individual predictions were computed as $E(Y_{ij}^l | b_i, \tilde{\tau}_i, \hat{\theta}^l)$ from (4.1) replacing b_i and $\tilde{\tau}_i$ by their empirical Bayes estimates. Figure 4.4 shows that the individual fit is good with individual predicted trajectories matching the observed marker values.

4.4 Discussion

We proposed in this article a procedure to estimate a bivariate random changepoint mixed model for two correlated longitudinal markers and to test for the difference in the times of change. This methodology has been implemented in a `R` (R Core Team, 2018) package `rcpm` that is freely available on the `GitHub` platform and whose main functions are described in Appendix B. We assessed the performance of the

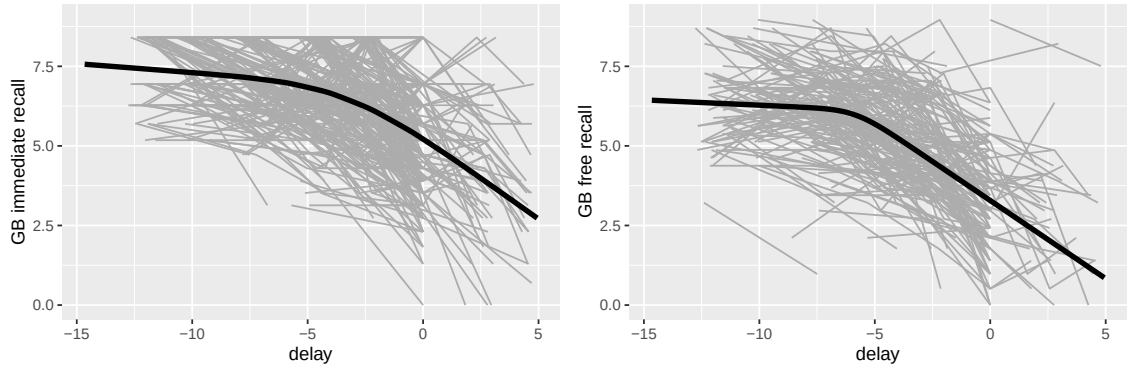

Figure 4.3: All individual GB immediate recall and GB free recall trajectories on the transformed scale compared to the estimated marginal trajectory $E(\tilde{Y}^l(t))$

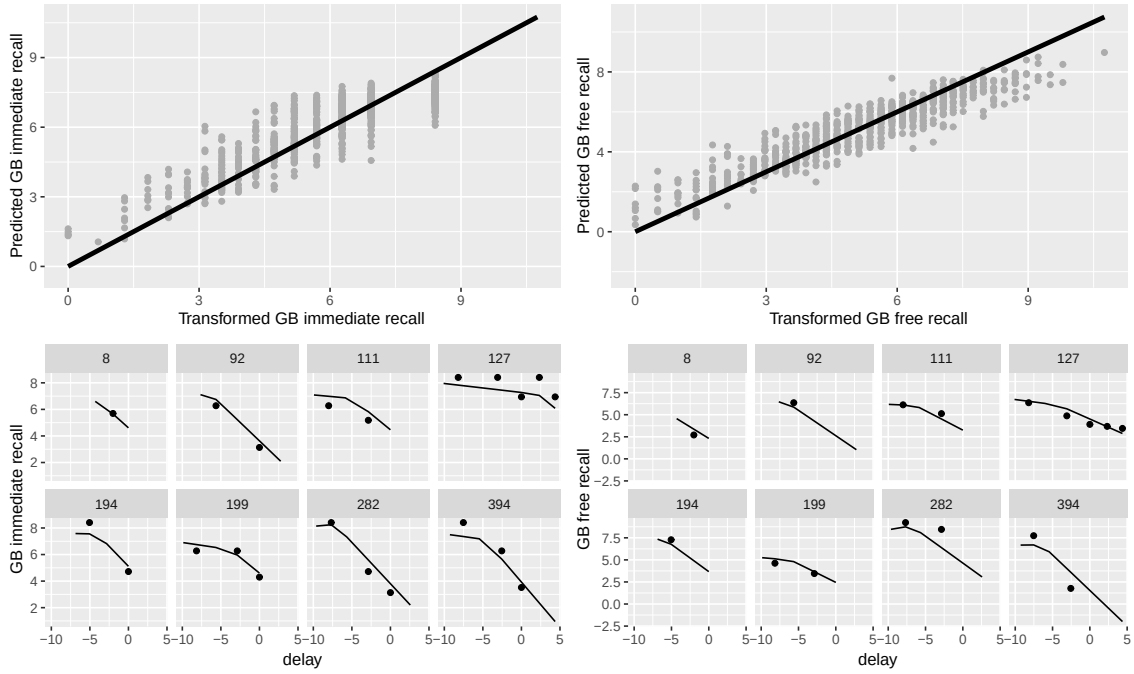

Figure 4.4: Fit of the bivariate random changepoint model for both markers (left: GB immediate recall; right: GB free recall). On the upper panel are plotted the true transformed observations vs. the predicted observations and on the lower panels are plotted individual observations (dots) vs. their predicted trajectories (solid line) for 8 randomly selected subjects

inference procedure in several simulated scenarios and we obtained good results for the estimation as well as for the test. The procedure was applied to the French 3C cohort of elderly and we showed that the GB free recall (memorizing ability) dropped on average more than 2 years before the GB immediate recall (encoding ability) (p=0.047).

This work completes previous work about bivariate random changepoint (Yang and Gao, 2013) by proposing a complete frequentist estimation procedure. Especially, we proposed a simple test to assess the temporal order of decline. Moreover, the model we presented takes into account potential deviations from normality which are very common for psychometric scores (floor or ceiling effect). Additionally, the methods described in this article have been implemented and are freely available for the `R` community. Yang and Gao (2013) compared several formulations for the changepoint model and discarded the Bacon-Watts model. However, their simulation studies were not very realistic regarding our cohort data as they had great variances for the random changepoint and a correlation structure where only the time of change were correlated between the two markers but not the slopes. Moreover, their conclusion on the Bacon-Watts model were based on data simulated from a different model and the simulations we performed to check the consequence of this misspecification led to better results.

In our application, only the evolution of subjects diagnosed with dementia has been studied. This may lead to a selection bias because subjects who dropped out before dementia diagnosis were excluded. Such a bias could be avoided by taking into account dementia and death. A joint modelling approach with an illness-death model for the survival part could be used. However, the delay to diagnosis could not be used as the timescale any more and the changepoint would measure an age of change rather than the delay between acceleration of the decline and diagnosis.

The methodology presented here has been developed for two markers. Its principle can be directly extended to a multivariate random changepoint models. However, this would drastically increase the number of random effects and therefore the computational burden. We rather advise to perform a pairwise comparison of the time of change of the different markers using our bivariate random changepoint model

and therefore deducing the temporal order of all the studied markers.

Chapter 5

Discussion and perspectives

In this section, we highlight the strengths and limitations of the models proposed in this thesis and we outline and discuss alternative modelling approaches that could solve some of these limitations.

5.1 Strengths of the proposed approaches

We proposed in this thesis methods for testing the existence of a random changepoint for longitudinal data and proposed way of estimating and comparing such changepoints for different markers measuring different cognitive functions.

The development of the testing procedure raised important statistical challenges because of the nonidentifiability of some parameters under the null. We proposed an approach based on the supremum score test using a perturbation algorithm to sample the empirical distribution of the test statistic under the null. To our knowledge, they were no statistical test available for testing the existence of a random changepoint in longitudinal data. The curvilinear bivariate random changepoint models we developed makes possible the comparison of the time of change in order to answer questions of particular clinical interest. The developed methodology was applied on real data from two french cohorts Paquid and 3C (see Sectionss 3.3 and 4.3). We assessed the existence of a random changepoint for several cognitive markers according to individual characteristics such as educational level. The bivariate model

allowed us to show that among future demented subjects, the ability to memorize an information declines before the ability to encode it.

5.2 Late acceleration versus time of differentiation

However, if we compare our estimation results to curves of cognitive decline previously estimated on the Paquid cohort using a semi-parametric approach (Amieva *et al.*, 2014), we observe that we are identifying later changepoints. Actually, the changepoint that we identified represents the late acceleration of cognitive decline just before dementia diagnosis rather than a time of differentiation between cases and non-cases which happens earlier than the late acceleration. The reason is that all our analyses presented on Sections 3.3 and 4.3 were done retrospectively on diagnoses cases only making impossible the identification of the time of differentiation between case and non-cases. In order to identify this time of differentiation, it is necessary to model cases and non-cases together.

To tackle this problem, we could consider a nested case-control study where incident cases of dementia diagnosed during the follow-up would be matched to controls according to age, sex and educational level with the condition that controls are observed and free of dementia at the visit of diagnosis of the matching case. The delay for a control would be the delay to diagnosis of the matching case. In the following, δ_i is the case indicator that equals 1 for cases and 0 for controls.

Cognitive decline of cases and controls could be modelled together using the delay to the diagnosis of the case as the timescale and assuming a two-class model with a linear trend for a class (typically the controls) and the same linear trend up to a certain date where the decline accelerates for a second class (typically the cases). From this date and up to the diagnosis, there is a quite long phase during which cognitive decline trajectories are nonlinear with a late acceleration just before diagnosis. Thus, for this phase, a linear trend cannot be assumed but rather a nonlinear decreasing trajectory that may be modelled using splines functions. The next section described this model with two alternative formulations: a model where

the class are defined by the status (case and control) and a latent-class model.

5.3 A proposal of a latent class random changepoint model

The two-class model described above is written

$$Y_i(t_{ij}) = \beta_{0i} + \beta_{1i} t_{ij} + c_i \beta_{2i} f(t_{ij} - \tau_i, \eta) + \varepsilon_{ij} \tag{5.1}$$

where c_i is an indicator that equals 1 for subjects with a random changepoint trajectory and 0 otherwise, f is a function which depends upon some parameters η that represent the difference from the linear trajectory after the time of differentiation τ_i. We assume here that $\beta_{ki} = \beta_k + b_{ki}$ where $b_i = (b_{0i}, b_{1i}, b_{2i})^\top \sim \mathcal{N}(0, B)$ with B a positive matrix and that $\tau_i = \mu_\tau + \sigma_\tau \tilde{\tau}_i$ where $\tilde{\tau}_i \sim \mathcal{N}(0,1)$ is independent from b_i. The residual errors ε_i are assumed to follow a centered Gaussian distribution with diagonal variance matrix $\sigma_\varepsilon \mathbb{I}_{n_i}$ and are assumed independent from all the random effects. Here, t_{ij} denotes the delay as defined is our nested case-control study design. For cases, it is directly the delay to dementia; for controls, it is the delay to dementia of the matched case.

For the function f we chose to use a function based on I-spline of order 3 which have the advantages of being monotonous, smooth and such as that at zero, they are null, differentiable and of null slopes. These properties ensure a smooth transition between the two phases and make sure that the second phase correspond to an acceleration of the decline. Only one internal knot is chosen so that

$$f(t_{ij} - \tau_i, \eta) = \sum_{k=1}^{3} \eta_k^2 I_k(t_{ij} - \tau_i)$$

where $(I_k)_{k=1,\dots,3}$ denotes the I-spline basis. For the model to be identifiable, we assume the constraint $\beta_2 = -1$ and let the splines parameters η_k to be unconstrained.

In this model, β_0 is the mean value of the marker for subjects in the linear class at the time of the case diagnosis, β_1 is the mean slope of the cognitive decline

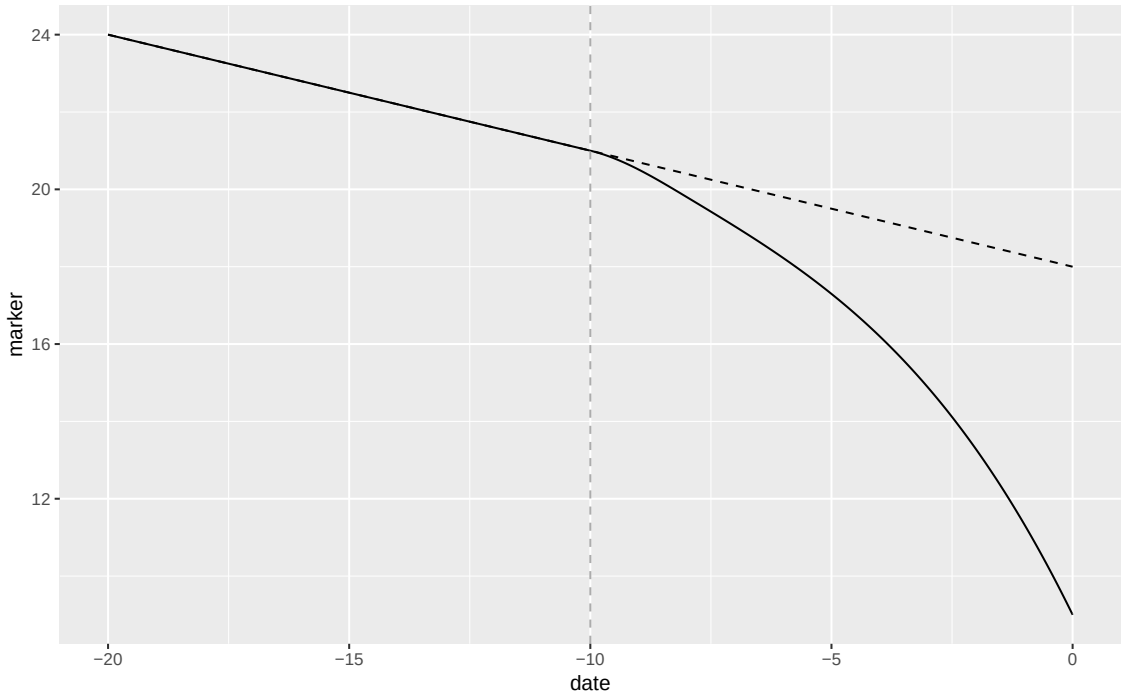

Figure 5.1: Example of two mean trajectories from model (5.1): a linear trajectory without any changepoint (dashed black) and a changepoint trajectory (solid) where, from $\tau_i = -10$, a differentiation from the linear trajectory modelled by an I-spline function arises. The I-spline function is exactly the difference between the dashed black line and the solid line.

during the normal cognitive ageing phase. For subjects whose trajectory presents a random changepoint, f models the difference between this normal cognitive ageing and a pathologic cognitive decline while β_2 measures its mean intensity. Examples of trajectories generated by model (5.1) are plotted in Figure 5.1. From here, two approaches could be considered.

5.3.1 *A priori* defined class

First, we could assume both classes to be *a priori* known meaning that c_i is known for all subjects by fixing $c_i = \delta_i$. The N_0 controls would have a linear trajectory and the N_1 cases a changepoint trajectory with $N_0 + N_1 = N$. The log-likelihood is then directly the sum of the log-likelihood of the linear mixed model for the controls and

of the log-likelihood of the random changepoint mixed model for the cases

$$\ell_N(Y;\theta) = \ell^0_{N_0}(Y;\theta) + \ell^1_{N_1}(Y;\theta). \tag{5.2}$$

We have

$$\ell^0_{N_0}(Y;\theta) = \sum_{i=1}^{N_0} \log f(Y_i|c_i = 0;\theta)$$

where $f(Y_i|c_i = 0;\theta)$ is a multivariate Gaussian density with mean 0 and variance $Z_{0i}B_0Z_{0i}^\top + \sigma_\varepsilon^2\mathbb{I}_{n_i}$ where Z_{0i} is a $n_i \times 2$ matrix with rows $(1, t_{ij})_{j=1,\dots,n_i}$ and B_0 a 2×2 definite positive matrix, variance of the random effects $(b_{0i}, b_{1i})^\top$. And we have

$$\ell^1_{N_1}(Y;\theta) = \sum_{i=1}^{N_1} \log f(Y_i|c_i = 1;\theta) = \sum_{i=1}^{N_1} \log \int f(Y_i|c_i = 1, \tilde{\tau}_i;\theta) f(\tilde{\tau}_i)\mathrm{d}\tilde{\tau}_i$$

the log-likelihood of the nonlinear mixed model defined by (5.1) when $c_i = 1$.

5.3.2 A latent class approach

Alternatively, we could assume a latent class model where controls have a non-null probability of having a changepoint whereas all the cases belong to the class with changepoint. Indeed, in the nested-case-cohort framework, some subjects are considered as control because they are free of dementia at this date even though they might develop dementia at a later visit. Thus, we could model the probability for a control of having a changepoint trajectory by a logistic model

$$\pi_i = \mathbb{P}(c_i = 1|X_i, \delta_i) = \left(\frac{\exp(\eta^\top X_i)}{1 + \exp(\eta^\top X_i)}\right)^{1-\delta_i}$$

where X_i are some covariates. In this case, the log-likelihood of the model is written

$$\ell_N(Y;\theta) = \sum_{i=1}^{N_0} \log[(1-\pi_i) f(Y_i|c_i = 0, X_i) + \pi_i f(Y_i|c_i = 1, X_i)] + \ell^1_{N_1}(Y;\theta) \tag{5.3}$$

where $\ell^1_{N_1}(Y;\theta)$, $f(Y_i|c_i = 0, X_i)$ and $f(Y_i|c_i = 1, X_i)$ are all defined as above.

5.3.3 Estimation and discussion

The log-likelihoods (5.2) and (5.3) can be maximised using the Levenberg-Marquardt optimization algorithm (Levenberg, 1944; Marquardt, 1963) and Gaussian quadrature to approximate the integrals over the random effect $\tilde{\tau}_i$. This new I-spline random changepoint model is currently under development. For the *a priori* defined class model, some preliminary results are detailed in Appendix D.

In the above formulation of the model we assume that during the normal cognitive ageing phase, the mean trajectory is the same for subjects from both classes as illustrated in the trajectories plotted in Figure 5.1. In practice, the mean cognitive level at midlife of future demented subjects tends to be lower than normal subjects because their characteristics differ. This difference may be accounted for either by a careful adjustment on known risk factors for dementia (education, occupation, *e.g.*) or by adding a class specific intercept $\beta_3 c_i$.

The interest of the nested case-control study design is that it reduces bias due to differential dropout rates between cases and controls as they are selected at the same visits. However, because at each visit, we select only cases and controls seen and observed up to this visit, there is an inevitable selection bias. Also, because of the long pre-diagnosis phase of dementia, many subjects are classified as control whereas they might develop dementia at a later stage leading to a loss of power. One might increase the number of control per cases to improve the power. However, to avoid the selection bias while increasing power, a more appropriate approach might be to jointly model the longitudinal trajectory of the cognitive marker and the time-to-dementia for controls.

5.4 A new joint random changepoint model

Jacqmin-Gadda *et al.* (2006) and Dantan *et al.* (2011) proposed joint models to study the cognitive decline trajectories of cases and non-cases simultaneously. In both these models, it is assumed that all the subjects will present a changepoint in their cognitive decline trajectories while it could be expected that some subjects among non-cases would present a linear trend only. Moreover, the models proposed by

Jacqmin-Gadda *et al.* (2006) and Dantan *et al.* (2011) does not allow testing for the existence of the random changepoint because its absence entails the independence of the event and the longitudinal marker which is not realistic.

We could define a shared random effects joint model linking the longitudinal model (5.1) with non-null probability for controls to be in the changepoint class and a class specific survival model for time to dementia. We note $Y_i(t_{ij}) = \tilde{Y}_i(t_{ij}) + \varepsilon_{ij}$ and define the survival model

$$\lambda_g(t_{ij}) = \lambda_{0g}(t_{ij}) \exp(\nu_g^\top Z_{ig} + \gamma_g \tilde{Y}_i(t_{ij}))$$

where $g = 0, 1$ denotes the class dependence, λ_g are the hazard functions, λ_{0g} denotes the baseline hazards, Z_{ig} the sets of covariates and ν_g and γ_g some regression parameters. We insist on distinguishing this joint latent class model to a cure model. A cure model would assume a null risk of developing dementia for a portion of the subjects, typically those with a linear cognitive trend over time in our context. If this null-risk portion existed, a plateau should be observed on Kaplan-Meier survival curves of dementia. As this plateau is not observed, the null risk portion is not realistic and a cure model is not appropriate here. In the joint latent class model we propose, subjects from both class are considered at risk of developing dementia which is a more realistic assumption. However, these risks are class specific and we can expect the risk of dementia in the linear class to be way lower than in the random changepoint class.

This model would have some advantages compared to the previous models proposed by Jacqmin-Gadda *et al.* (2006) and Dantan *et al.* (2011). First, because of the two latent classes approach, it allows a portion of the subjects to have a linear trend and does not make the assumption that all subjects have a changepoint. This joint random changepoint model has also the advantage of allowing to test for the existence of a random changepoint using a similar approach than the one developed in Chapter 3.

As in Jacqmin-Gadda *et al.* (2006) and Dantan *et al.* (2011), the joint model approach makes necessary the use of age as timescale because the delay to dementia

cannot be computed for non-cases in such a cohort design. This leads to a change in the interpretation of the time of differentiation which becomes the age at which, on average, subjects of the changepoint class begin to differ from subjects having a linear cognitive decline trajectory. Testing for differences of the mean age at differentiation according to individual characteristic such as educational level would not be relevant here as the detected difference would mainly capture the heterogeneity of age at dementia according to educational level. A comparison of age at differentiation is however possible on a same sample therefore allowing to compare the mean age at differentiation between different markers as in Chapter 4. Here, however, we are mainly interested in the length of the phase of accelerated decline before dementia diagnosis compared to normal ageing. For instance, it is known that subjects with high education have a lower risk of dementia and thus a later age at acceleration of cognitive decline than subjects with low education but it is still unclear if the shape of the decline before diagnosis differs according to educational level. To investigate this question, we would like to know if the delay between the differentiation from normal ageing and the diagnosis differs according to the educational level. If this joint model does not allow comparing delay between the time of differentiation and the diagnosis of dementia, this delay could still be computed *a posteriori*.

This joint model could also be used to make prediction of the risk of dementia. Therefore, an interesting question would be to investigate how this proposed joint model would perform compared to joint models with, for example, a quadratic or a spline evolution for the longitudinal cognitive marker. In other words, does the random changepoint model for the cognitive marker improve the prediction of the event in a shared random effects model?

5.5 A revival process

Recently, Dempsey and McCullagh (2018) have proposed a general framework for modelling a survival process with a final absorbing state and a sequence of health measurements using reverse alignment. Rather than considering the longitudinal

health sequence over time $Y_i(t)$, they rather propose to study the revival process $Z_i(s) = Y(T_i - s)$ where T_i is the survival time of subject i. The main interest of their approach is the alignment of timescale and patient records to an event of interest. It is similar to the models we proposed where the timescale used is the delay do dementia diagnosis. Interpretation becomes easier and it helps finding structure in the longitudinal trajectories. The difference with our models is that the survival process is also modelled. As long as T_i is not observed however, the revival process is not observable component-wise. This makes the proposed methodology as it stands non applicable to incomplete records, which is the case in our application where time to dementia is right censored for a large part of the sample.

5.6 Towards a smoother model?

As we highlighted in this discussion, several changepoints could be considered in the cognitive decline trajectory before diagnosis. We identified at least two: the time of differentiation, from which on average cases decline trajectories begin to derive from a linear trajectory considered as normal ageing, and the late cognitive acceleration that happens a few years before the diagnosis.

Changepoint models seems a natural way to model this progressive aspect of cognitive decline trajectories. We could consider multiple changepoint models which would have the main advantage of allowing clinically relevant interpretation of the parameters. By directly modelling the changepoints, it would allow the comparison of the time of change according to some individual characteristics or between different cognitive markers. However estimation of mixed models with multiple random changepoint would be computationally challenging.

It could be certainly possible to take into account this progressive aspect of cognitive decline using a more flexible and simple model for the longitudinal marker. A spline model could perfectly suit this purpose for example. Of course, by doing so, we would loose the interpretation of the parameters, but for doing prediction of time to dementia or describing cognitive decline before dementia without investigating clinical hypotheses, a smooth formulation of the longitudinal marker could be

sufficient.

5.7 Other medical applications

Strictly speaking, the statistical methods described and developed here, in the particular context of dementia, could be applied to any other diseases. The only needed condition is that the evolution of the disease is measurable through a marker and, obviously, that a changepoint in this marker trajectory is of clinical interest. For example, in Human Immunodeficiency Viruses studies, the rate of CD4 T-lymphocytes and the viral load are very useful to assess the disease progression. In prostate cancer studies, a useful marker for cancer recurrence is the prostate specific antigen. In chronic kidney disease studies, the evolution of their glomerular filtration rate is very informative of the stage of kidney disease. For all these markers, a changepoint in the trajectory may highlight a clinical progression of the disease.

5.8 On statistical models

When presenting this work, one criticism we got pointed out the heavy hypotheses made by the random changepoint model on cognitive decline trajectories. Our most honest answer is to say that, three years ago, these hypotheses actually led us to chose this specific model. This is the ordinary way of scientific method. In statistics, it begins with a question, here relative to the shape of cognitive decline trajectories before dementia diagnosis. With this question in mind and knowledge from previous epidemiological findings, we formulate one or several hypotheses. A statistical model is then chosen or developed and applied to real data, here an observational study satisfying scientific good practices on collecting data. From the interpretation of the results, the initial hypotheses can be modified and new ones can emerge. It happened in this manuscript. We first developed a model to identify and compare a changepoint that happened to be the late cognitive decline. This led us to suggest an alternative modelling to identify the time of differentiation between normal and pathological cognitive decline while keeping the advantages of our first model. This

cycle of confrontation between hypotheses and real world has fed scientific literature for decades. In the particular case of statistics, this has led to the development of certainly as many statistical models as statisticians.

A famous aphorism attributed to the British statistician George E. P. Box about statistical modelling states that *"All models are wrong but some are useful."* Pretending that all models are wrong might seem provocative while it is just a tautology. A statistical model is a simplified transposition of a real world phenomenon onto the abstract world of mathematics. It cannot be true by itself. There is this joke about an engineer, a physicist and a mathematician who are on a train heading north, and have just crossed the border into Scotland. The engineer looks out of the window and said "Look! Scottish sheep are black!". The physicist says, "No, no. Some Scottish sheep are black." The mathematician looks irritated. "There is at least one field, containing at least one sheep, of which at least one side is black." This joke reflects how a model depends directly upon who formulates it and what knowledge he has. Following scientific method principles, the confrontation of a model to reality will led to improved knowledge and better models. Any model can be useful as long as scientific principles are respected. If not, this will lead to biased knowledge and biased models. Alas, these days, some people, even in the highest position, tend to deny scientific findings by putting them on an equal footing to ideological models that only answer a political agenda.

To conclude with a sense of irony, we might say that *all models are useful but some modellers are wrong.*

List of Abbreviations and Symbols

Abbreviations

3C Three-City french cohort

AD Alzheimer's Disease

BLUP Best Linear Unbiased Predictor

BMI Body Mass Index

CI Confindence Interval

CR Coverage Rate

DSM Diagnostic and Statistical Manual of Mental Disorders

EM Expectation Maximisation

GB Grober and Buschke tests

HEL High Educational Level subjects

LEL Low Educational Level subjects

LRS Log-likelihood Ratio Statistic

MCI Mild-Cognitve Impairement

MCMC Monte-Carlo Markov Chain

MLE Maximum Likelihood Estimation

MMSE Mini Mental State Examination

REML Restricted Maximum Likelihood

Symbols

$.^\top$ Transpose operator

$\ell_N(\theta)$ Log-likelihood depending upon unknown θ parameters from a N-sample

$\hat{\theta}$ Estimator of unknown θ parameter

$\mathbb{I}_N$ Identity matrix of size N

$\mathbb{R}^+$ Set of real positive numbers

$\mathbb{R}^p$ Set of real p-vectors

$\mathbf{1}_A(.)$ Indicator function of set A

$\text{sgn}(x)$ Sign function

$\text{trn}(x)$ Transition function

τ_i Random changepoint

$x \prec y$ Order relation between x and y

$x \propto y$ x is proportional to y

Appendix A

Scientific valorisation

A.1 Scientific publications

- **Segalas C**, Helmer C, Jacqmin-Gadda H. A curvilinear bivariate random changepoint model to assess temporal order of markers. Submitted to *Statistical Methods in Medical Research*, first revision in progress.
- **Segalas C**, Amieva H, Jacqmin-Gadda H. A hypothesis testing procedure for random changepoint mixed models. *Statistics in Medicine*. 2019;38:3791–3803. https://doi.org/10.1002/sim.8195

A.2 Oral communications

- **Segalas C**, Jacqmin-Gadda H, Inferential methods for random changepoint models, *40th ISCB Conference*, Leuven, Belgium, july 2019
- **Segalas C**, Jacqmin-Gadda H, Curvilinear bivariate mixed model with random changepoint to compare times of change between cognitive markers in Alzheimer's disease, *7th IBS Channel Network Conference*, Rothamstead Research, UK, july 2019
- **Segalas C**, Jacqmin-Gadda H, Testing the existence of a random changepoint in a mixed model, *29th IBS Conference*, Barcelona, Spain, july 2018

- **Segalas C**, Jacqmin-Gadda H, Testing the existence of a random changepoint in a mixed model, *50th annual conference of the French Statistical Society*, Paris, France, may 2018

- **Segalas C**, Jacqmin-Gadda H, Test d'un changemnt de pente aléatoire dans un modèle mixte, *Joint conference of the research group Statistics and Health and of the French Biometrics Society*, Bordeaux, France, october 2017

Appendix B

The rcpm package

This R package proposes different functions to make inference in a random changepoint models for longitudinal data as described all along this work. The development is still in progress and the package is not yet fully functional. Already implemented or being implemented are the following functions

- `testRCPMM`: a test for the existence of a random changepoint for longitudinal data
- `rcpme`: an estimation algorithm for random changepoint mixed models
- `bircpme`: an estimation algorithm for bivariate random changepoint mixed models taking into account an eventual correlation between two markers

B.1 The testRCPM function

This function realizes the supremum score test on `longdata` according to `formu`.

```
testRCPMM(longdata, formu, gamma, nbnodes, nbpert, covariate)
```

- `longdata`: A longitudinal dataset containing all variables used in the formula `formu`.

- `formu`: A formula object describing which variables are to be used. The formula has to be of the following form `markervar` $\sim$ `scorevar | groupvar` for the function to work.
- `covariate`: An eventual covariate dependence of all the parameters in the model. `NULL` by default. *Not implemented yet.*
- `gamma`: A smoothing parameter for the transition on the changepoint date. `0.1` by default.
- `nbnodes`: Number of pseudo-adaptive Gaussian quadrature nodes used to compute the numeric integrals. `5` by default.
- `nbpert`: Number of perturbations used to compute the empirical p-value. `500` by default.

The output contains a list with the computed empirical p-value and the observed test statistic.

B.2 The `rcpme` function

This function estimates univariate random changepoint models developed and detailed in Chapter 3, Chapter 4 and the fixed two-class random changepoint model from the discussion in Chapter 5. `rcpme(longdata, formu, covariate, REadjust, gamma, nbnodes, param, model, link, statut)`

- `longdata`: A longitudinal dataset containing all variables used in the formula `formu`.
- `formu`: A formula object describing which variables are to be used. The formula has to be of the following form `markervar` $\sim$ `scorevar | groupvar` for the function to work.
- `covariate`: An optional string indicating a binary covariate to add on the fixed effects, i.e. intercept, mean slope, difference of slopes and changepoint

date. The parameter `REadjust` indicates how this covariate influences the random effects variance structure. Default to `NULL`, i.e. no covariates.

- `REadjust:` An optional string value indicating how the random effects variance structure depends on `covariate`. Default to "no".
  - `no` means that the structure doesn't depend upon `covariate`
  - `prop` indicates that the random effects variance structure is proportional according to `covariate` value
  - `yes` indicates that there is two different random effects variance structures, i.e. one for each level of `covariate`
- `gamma:` A numeric parameter indicating how smooth the trajectory is on the changepoint date. It should be small according to the time variable scale. Default to `0.1`.
- `nbnodes:` A numeric parameter indicating how many nodes are to be used for the gaussian quadrature for numerical integration. Default to `10`.
- `param:` An optional vector parameter that contains initial parameter for the optimization of the log-likelihood. Default to `NULL`.
- `model:` An optional string indicating which formulation of the random changepoint exists. When used for estimation purpose, you should either `bw` or `isplines` which has better interpretability properties. Default to `bw`.
  - `test` is used by the `testRCPMM`
  - `bw` stands for the Bacon-Watts formulation
  - `isplines` stands for the I-spline formulation
- `link:` An optional string indicating which link function is to be used. This link function is used to deal with non-Gaussian data. Default to `linear`.
  - with `splines` the model estimates an appropriate I-spline link function g so that $g($`scorevar`$)$ is a Gaussian variable

  - with `linear` no link function will be estimated and data is assumed Gaussian

- `statut:` An optional string indicating a binary variable from which two class are considered: a linear class for subjects with `statut=0` and a random changepoint class for subjects with `statut=1`. Default to `NULL`.

The output contains several objects : `call` is the function call; `Loglik` is the value of the log-likelihood at the optimum; `formula` is the formula describing which variables are used in the model; `fixed` contains all fixed parameters estimates, standard errors, CIs, wald test statistic and corresponding p-value when possible; `sdres` the estimated residual error; `VarEA` a 4×4 matrix or a list of 4×4 matrices - if there is some covariate for example - containing the estimated random effects covariance matrix; `optpar` the optimal parameters maximizing the log-likelihood; `covariate` the covariate declared in the function call; `REadjust` the string indicating how random effects structure is handled as declared in the function call, `invhessian` the covariance matrix containing all the standard errors and correlations of the parameter estimates; `conv` an index of successful convergence, equals to 1 if success; `init` the initial values vector; `model` the model used during estimation; `gamma` the value of gamma used during estimation; `link` the link function used during estimation.

B.3 The `bircpme` function

This function estimates the bivariate random changepoint models developed and detailed in Chapter 4. `bircpme(longdata, formu, covariate, REadjust, gamma, nbnodes, adapt, param, nproc, model, link1, link2, twostep)`

- `longdata:` A longitudinal dataset containing all variables used in the formula `formu`.

- `formu:` A formula object describing which variables are to be used. The formula has to be of the following form `markervar` $\sim$ `scorevar | groupvar` for the function to work.

- `covariate`: An optional string indicating a binary covariate to add on the fixed effects, i.e. intercept, mean slope, difference of slopes and changepoint date. The parameter `REadjust` indicates how this covariate influences the random effects variance structure. Default to `NULL`, i.e. no covariates.

- `REadjust`: An optional string value indicating how the random effects variance structure depends on `covariate`. Default to "no".

  - `no` means that the structure doesn't depend upon `covariate`
  - `prop` indicates that the random effects variance structure is proportional according to `covariate` value
  - `yes` indicates that there is two different random effects variance structures, i.e. one for each level of `covariate`

- `gamma`: A numeric parameter indicating how smooth the trajectory is on the changepoint date. It should be small according to the time variable scale. Default to `0.1`.

- `nbnodes`: A numeric parameter indicating how many nodes are to be used for the gaussian quadrature for numerical integration. Default to `10`.

- `adapt`: A boolean indicating whether adaptive gaussian quadrature should be used for numerical integration. Default to `FALSE`.

- `param`: An optional vector parameter that contains initial parameter for the optimization of the log-likelihood. Default to `NULL`.

- `nproc`: An optional integer specifying the number of processors for parallelisation of the optimization algorithm. Default to `1`. An optional string indicating which formulation of the random changepoint exists. When used for estimation purpose, you should either `bw` or `isplines` which has better interpretability properties. Default to `bw`.

  - `test` is used by the `testRCPMM`
  - `bw` stands for the Bacon-Watts formulation

  - `isplines` stands for the I-spline formulation which is not yet implemented in the bivariate estimation

- `link1`: An optional string indicating which link function is to be used. This link function is used to deal with non-Gaussian data. Default to `linear`.
  - with `splines` the model estimates an appropriate I-spline link function g so that $g($`scorevar`$)$ is a Gaussian variable
  - with `linear` no link function will be estimated and data is assumed Gaussian

- `link2`: Same as `link1` but for the second marker. Default to `linear`.

- `twostep`: An optional boolean to specify if a two-step pseudo adaptive Gaussian quadrature should be used. Currently not working. Default to `FALSE`.

The output contains several objects : `loglik` is the value of the log-likelihood at the optimum; `fixed` contains all fixed parameters estimates, standard errors, CIs, wald test statistic and corresponding p-value when possible; `sdres` the estimated residual error; `VarEA` a matrix containing the estimated random effects covariance matrix of the eight random effects: four for each marker with a general correlation structure between them; `optpar` the optimal parameters maximizing the log-likelihood; `covariate` the covariate declared in the function call; `REadjust` the string indicating how random effects structure is handled as declared in the function call, `invhessian` the covariance matrix containing all the standard errors and correlations of the parameter estimates; `conv` an index of successful convergence, equals to 1 if success; `init` the initial values vector; `niter` the number of iterations before convergence; `model` the model used during estimation; `gamma` the value of gamma used during estimation; `link1` and `link2` the link functions used during estimation for respectively the first and the second marker, `paramSpl` the parameters of the I-spline transformations η_k.

Appendix C

Doing splines with `spleen`

During this work, we used at several occasions function defined by a basis of I-splines functions as described by Ramsay (1988). Here we define the I-splines and M-splines and discuss the package `spleen` we developed in `Rcpp` and compared it to the `R` package `splines2` (Wang and Yan, 2018).

C.1 From M-spline to I-spline

Spline functions are piecewise polynomial functions often used for interpolation of an unknown function. We first define the M-spline basis for a fixed order k, i.e. for a fixed degree $k-1$ and for n free parameters. We define the sequence of knots $t = t_1 \leq \cdots \leq t_{n+k}$ such as $t_1 = \cdots = t_k$ and $t_{n+1}, \ldots, t_{n+k}$ and with a strict inequality for the $n-k$ internal knots. For $x \in \mathbb{R}$, the basis of M-splines contains n members $M_i(x|k,t)$ defined for $k=1$ as

$$M_i(x|1,t) = \begin{cases} \frac{1}{t_{i+1}-t_i} & \text{if} \quad t_i \leq x < t_{i+1} \\ 0 & \text{else} \end{cases}$$

and for $k > 1$ as

$$M_i(x|k,t) = \frac{k\left[(x-t_i)M_i(x|k-1,t) + (t_{i+k}-x)M_{i+1}(x|k-1,t)\right]}{(k-1)(t_{i+k}-t_i)}.$$

Because of this definition, $M_i(x|k,t)$ is positive and null outside $[t_i, t_{i+k}]$. The integral of an M-spline function is therefore a monotonous increasing function and is called I-spline of order k where k is the order of the M-splines

$$I_i(x|k,t) = \int_{t_1}^{x} M_i(u|k,t)\mathrm{d}u.$$

For $x \in \mathbb{R}$ and j such as $t_j \leq x < t_{j+1}$, I-splines can be easily computed from M-splines

$$I_i(x|k,t) = \begin{cases} 0 & \text{if} \quad i > j \\ \sum_{m=i}^{j} (t_{m+k+1} - t_m) M_m(x|k+1,t)/(k+1) & \text{if} \quad j-k+1 \leq j \leq i \\ 1 & \text{if} \quad j-k+1 > i \end{cases}$$

Example of generated splines basis are plotted in Figure C.1.

C.2 Using the `spleen` package

As most of our likelihood computation routines were developed directly in `Rcpp` and not in `R`, we developed a `Rcpp` routine to generate spline basis as we were not aware of the existence of such a routine. This routine is called `spleen`, not because of a typo, but to the memory of the many hours of struggle with programming and as a tribute to the French poet Charles Baudelaire and its famous and so enthusiastic *Spleen*. This routine is available on my `GitHub` page at https://github.com/crsgls/spleen. The package makes available two functions `mspline(x, tmin, tmax, tint, k, intercept)` and `ispline(x, tmin, tmax, tint, k, intercept)` whose arguments are:

- `x`: A real vector containing the abscissa at which the spline basis should be computed.
- `tmin`: A real number indicating the lower boundary knot.
- `tmax`: A real number indicating the upper boundary knot.

- `tint`: A real vector indicating the internal knots.
- `k`: An integer indicating the order of the spline. An M-spline of order k is a function of degree $k-1$. An I-spline of order k is the integral of an M-spline of order k and is a function of degree k.
- `intercept`: A boolean indicating if an intercept should be included in the spline basis.

We note l_x the size of `x` and l_{int} the size of `tint`, the output is a matrix of size $l_x \times (l_{\text{int}} + k)$ if `intercept = TRUE` and $l_x \times (l_{\text{int}} + k - 1)$ if `intercept = FALSE` containing the spline basis.

C.3 Comparison to `splines2` package

The splines basis built by the `spleen` package are exactly the same as the ones built using the `splines2` `R` package that provides M-spline and I-spline implementation. We compared the efficiency of both packages by generating 1000 M-splines and I-splines basis using functions from the two packages and plotted the execution times in Figure C.2. We observe that our package outperforms `splines2` and especially for I-splines.

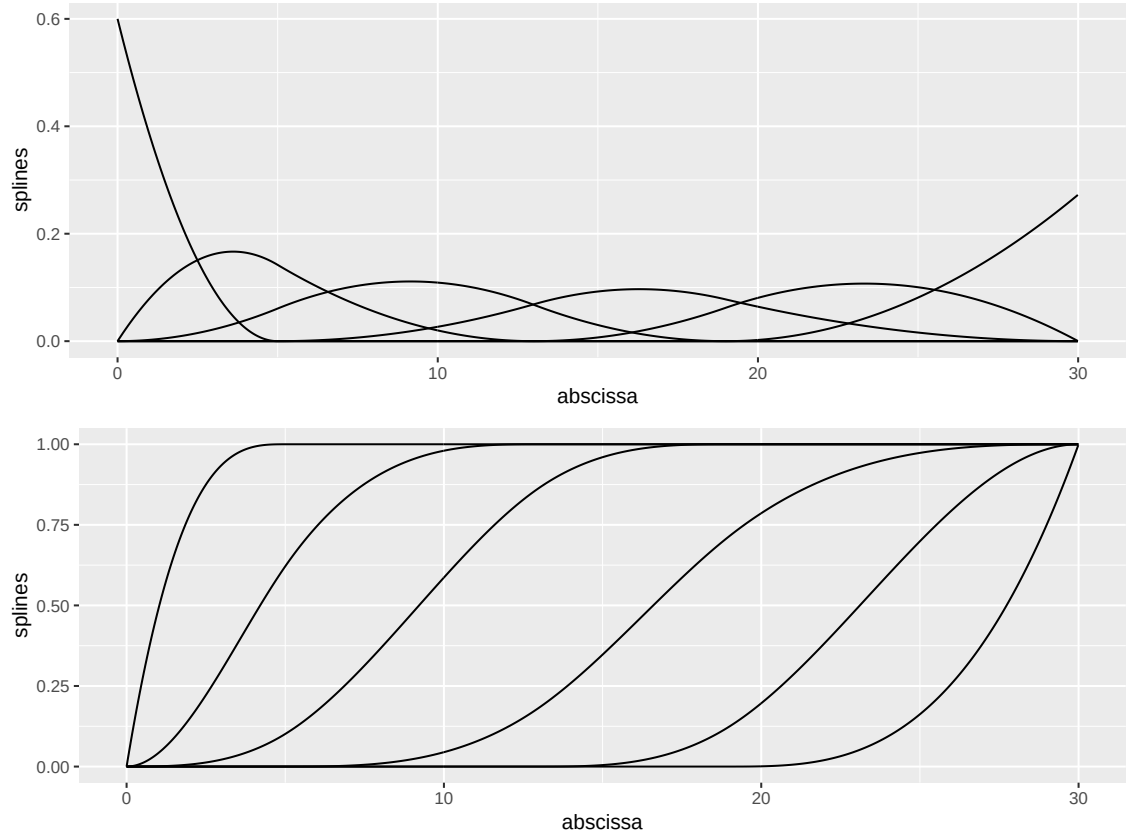

Figure C.1: Example of M-spline (upper panel) and I-spline (lower panel) basis with `tmin` $= 0$, `tmax` $= 30$, `k` $= 3$ and $5, 3, 19$ as internal knots with the intercept spline included.

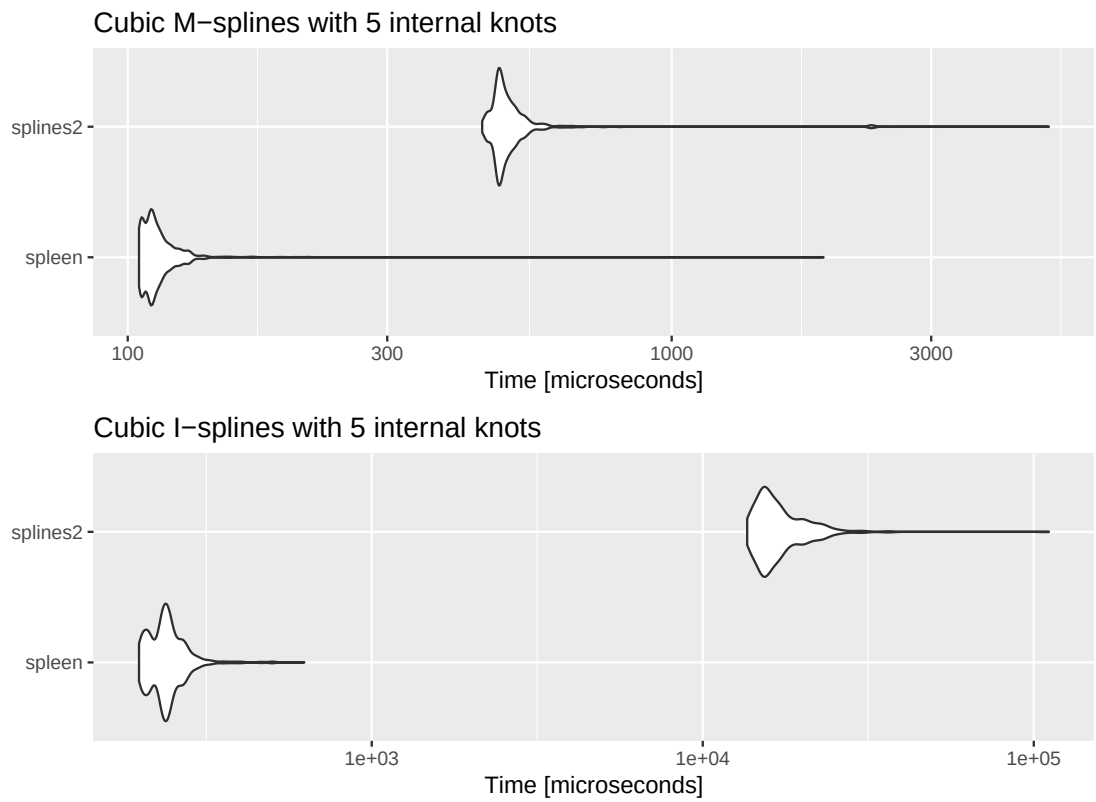

Figure C.2: Benchmarking of splines function from `splines2` and `spleen` packages over 1000 computations of cubic M-splines and cubic I-splines basis with 5 internal knots and a vector of abscissa of length 1000.

Appendix D

Preliminary results from I-spline model

In this Appendix, we detail some preliminary results obtained from the two-class random changepoint model detailed in Section 5.3.1 of the discussion where the two classes are fixed according to the subject status, case or control.

D.1 Simulation study

We ran 500 replicates of a simulated cohort of 1000 subjects all having 8 measurements of a simulated marker. Among these 1000 subjects, half were simulated as controls having a linear trajectory for the marker and the other half were simulated having a random changepoint model for the marker as defined by the model (5.2). Results of the simulation study are shown in Table D.1. The results are very satisfying with very little bias and good coverage rates.

D.2 Application

From the Paquid cohort, we built a nested case-control study from the 901 incident cases of dementia. For each of these cases, we matched one control with the same age (± 2 years), same educational level, same sex and with the condition that the control

	θ	$\hat{\theta}$	bias%	sdEmp	sdAs	CR
β_0	20.000	19.999	-0.000	0.038	0.039	0.96
β_1	-0.500	-0.500	-0.001	0.007	0.007	0.95
μ_τ	10.000	9.948	-0.005	0.464	0.421	0.90
σ^2_ε	1.000	0.998	-0.002	0.020	0.020	0.94
σ^2_0	1.000	1.000	0.000	0.065	0.068	0.95
σ^2_1	0.100	0.099	-0.005	0.008	0.008	0.95
σ^2_2	0.100	0.102	0.017	0.022	0.021	0.94
σ_{01}	0.040	0.040	-0.000	0.002	0.004	1.00
σ_{02}	-0.008	-0.008	0.039	0.004	0.008	1.00
σ_{12}	0.040	0.041	0.017	0.012	0.007	0.77
σ^2_τ	4.000	3.977	-0.006	0.794	0.753	0.94
η_1	2.000	1.992	-0.004	0.201	0.180	0.93
η_2	2.000	1.926	-0.037	0.389	0.367	0.93
η_3	2.000	1.943	-0.029	0.855	0.858	0.89

SdEmp: Empirical standard deviation; SdAs: Mean asymptotic standard deviation; CR: coverage rate of the 95% confidence interval.

Table D.1: Results of the preliminary simulation study for the two-class random changepoint model (5.2) over 500 replicates of simulated cohort with $N = 1000$.

has to be observed non demented at the visit of diagnosis of the case. On this nested case-control design, we estimated the two-class model where the class membership was fixed according to the status, controls being in the linear class and cases being in the random changepoint class. We estimated a mean time of differentiation -11.094 years before diagnosis with a 95%CI of $[-12.522; -9.667]$. Estimated mean trajectories from the Paquid cohort are plotted in Figure D.1. These findings are consistent to previous results on the time of differentiation (Amieva *et al.*, 2014). As we can see from Figure D.1, the late cognitive acceleration that we identified among cases appears approximately 3 years before diagnosis.

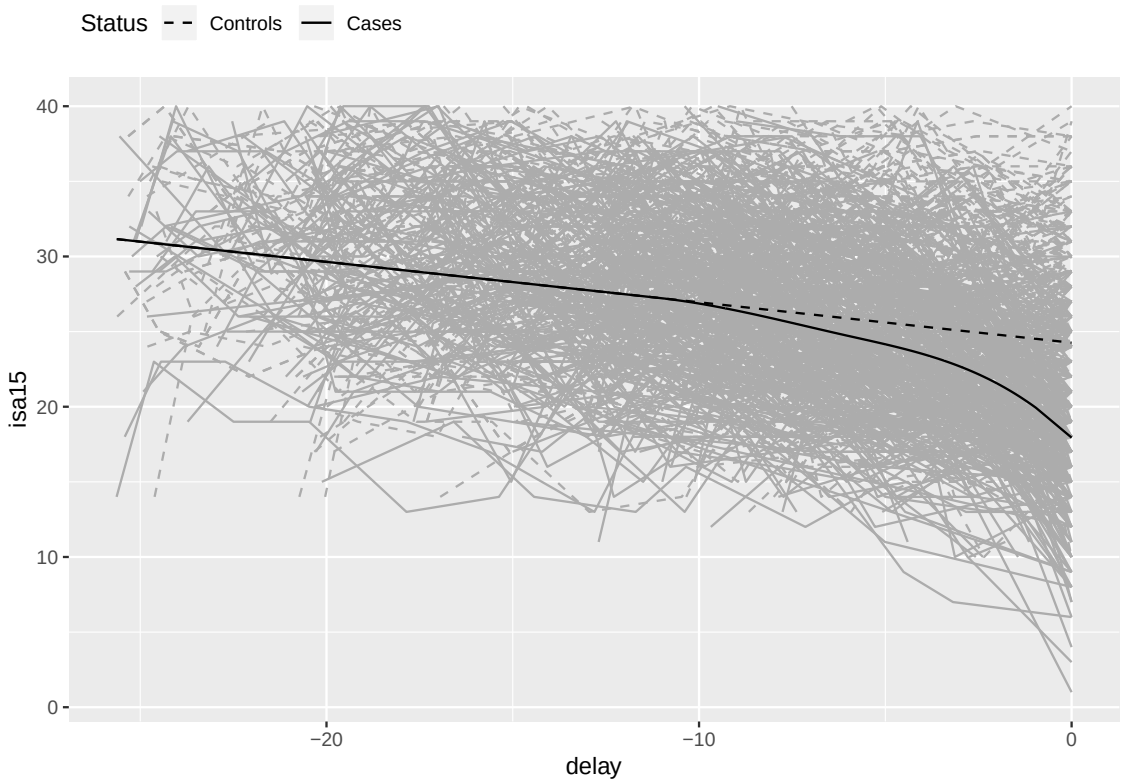

Figure D.1: Estimated mean trajectories of the two-class model assuming that the changepoint happens at the mean of all individual changepoints for controls (dashed black) and cases (solid black) for 500 randomly selected cases and 500 randomly selected controls whose longitudinal trajectories are plotted in grey.

Bibliography

3C Study Group. Vascular factors and risk of dementia: design of the Three-City Study and baseline characteristics of the study population. *Neuroepidemiology*, 22(6):316–325, December 2003. doi: 10.1159/000072920.

Abramowitz, M. and Stegun, I. A. *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*. Dover Publications, New York, 1970. ISBN 978-0-486-61272-0. OCLC: 18003605.

Aisen, P., Andrieu, S., Sampaio, C., Carrillo, M., Khachaturian, Z., Dubois, B., Feldman, H., Petersen, R., Siemers, E., Doody, R., Hendrix, S., Grundman, M., Schneider, L., Schindler, R., Salmon, E., Potter, W., Thomas, R., Salmon, D., Donohue, M., Bednar, M., Touchon, J., and Vellas, B. Report of the task force on designing clinical trials in early (predementia) AD. *Neurology*, 76(3):280–286, January 2011. doi: 10.1212/WNL.0b013e318207b1b9.

Aly, E.-E. A. A., Abd-Rabou, A.-E. S., and Al-Kandari, N. M. Tests for multiple change points under ordered alternatives. *Metrika*, 57(3):209–221, July 2003. doi: 10.1007/s001840200209.

Amieva, H., Le Goff, M., Millet, X., Orgogozo, J. M., Pérès, K., Barberger-Gateau, P., Jacqmin-Gadda, H., and Dartigues, J. F. Prodromal Alzheimer's disease: successive emergence of the clinical symptoms. *Annals of Neurology*, 64(5):492–498, November 2008. doi: 10.1002/ana.21509.

Amieva, H., Mokri, H., Le Goff, M., Meillon, C., Jacqmin-Gadda, H., Foubert-Samier, A., Orgogozo, J.-M., Stern, Y., and Dartigues, J.-F. Compensatory mechanisms in higher-educated subjects with Alzheimer's disease: a study of 20 years of cognitive decline. *Brain: A Journal of Neurology*, 137(Pt 4):1167–1175, April 2014. doi: 10.1093/brain/awu035.

Aminikhanghahi, S. and Cook, D. J. A survey of methods for time series change point detection. *Knowledge and Information Systems*, 51(2):339–367, May 2017. doi: 10.1007/s10115-016-0987-z.

Andrews, D. W. K. and Ploberger, W. Optimal Tests when a Nuisance Parameter is Present Only Under the Alternative. *Econometrica*, 62(6):1383–1414, 1994. doi: 10.2307/2951753.

Aue, A. and Horváth, L. Structural breaks in time series. *Journal of Time Series Analysis*, 34(1):1–16, 2013. doi: 10.1111/j.1467-9892.2012.00819.x.

Bacon, D. W. and Watts, D. G. Estimating the Transition between Two Intersecting Straight Lines. *Biometrika*, 58(3):525–534, 1971. doi: 10.2307/2334387.

Bartolucci, A., Bae, S., Singh, K., and Griffith, H. R. An examination of Bayesian statistical approaches to modeling change in cognitive decline in an Alzheimer's disease population. *Mathematics and computers in simulation*, 80(3):561–571, November 2009. doi: 10.1016/j.matcom.2009.09.002.

Becker, J. T., Boller, F., Lopez, O. L., Saxton, J., and McGonigle, K. L. The natural history of Alzheimer's disease. Description of study cohort and accuracy of diagnosis. *Archives of Neurology*, 51(6):585–594, June 1994.

Belleville, S., Peretz, I., and Malenfant, D. Examination of the working memory components in normal aging and in dementia of the Alzheimer type. *Neuropsychologia*, 34(3):195–207, March 1996. doi: 10.1016/0028-3932(95)00097-6.

Bhattacharya, P. K. *Some aspects of change-point analysis*. Institute of Mathematical Statistics, 1994. ISBN 978-0-940600-34-8. doi: 10.1214/lnms/1215463112.

Brayne, C., Huppert, F., Paykel, E., and Gill, C. The Cambridge Project for Later Life: design and preliminary results. *Neuroepidemiology*, 11 Suppl 1:71–75, 1992. doi: 10.1159/000110983.

BuHamra, S. S. Testing for a change in repeated measures data. *Communications in Statistics - Simulation and Computation*, 26(3):841–872, January 1997. doi: 10.1080/03610919708813415.

Buschke, H. Selective reminding for analysis of memory and learning. *Journal of Verbal Learning and Verbal Behavior*, 12(5):543–550, October 1973. doi: 10.1016/S0022-5371(73)80034-9.

Carlin, B. P., Gelfand, A. E., and Smith, A. F. M. Hierarchical Bayesian Analysis of Changepoint Problems. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 41(2):389–405, 1992. doi: 10.2307/2347570.

Carlson, N. E., Moore, M. M., Dame, A., Howieson, D., Silbert, L. C., Quinn, J. F., and Kaye, J. A. Trajectories of brain loss in aging and the development of cognitive impairment. *Neurology*, 70(11):828, March 2008. doi: 10.1212/01.wnl.0000280577.43413.d9.

Chiu, G. S. and Lockhart, R. A. Bent-cable regression with autoregressive noise. *Canadian Journal of Statistics*, 38(3):386–407, 2010. doi: 10.1002/cjs.10070.

Chiu, G. S., Lockhart, R. A., and Routledge, R. Bent-Cable Regression Theory and Applications. *Journal of the American Statistical Association*, 101(474):542–553, June 2006. doi: 10.1198/016214505000001177.

Commenges, D., Prague, M., and Diakite, A. *marqLevAlg: An algorithm for least-squares curve fitting*, 2016. R package version 2.0.

Commenges, D., Letenneur, L., Jacqmin, H., Moreau, T., and Dartigues, J.-F. Test of Homogeneity of Binary Data with Explanatory Variables. *Biometrics*, 50(3):613–620, 1994. doi: 10.2307/2532776.

Conniffe, D. Score tests when a nuisance parameter is unidentified under the null hypothesis. *Journal of Statistical Planning and Inference*, 97(1):67–83, August 2001. doi: 10.1016/S0378-3758(00)00346-3.

Cox, D. R. and Hinkley, D. V. *Theoretical Statistics*. Chapman and Hall/CRC, Boca Raton, 1 edition edition, September 1979. ISBN 978-0-412-16160-5.

Dantan, E., Joly, P., Dartigues, J.-F., and Jacqmin-Gadda, H. Joint model with latent state for longitudinal and multistate data. *Biostatistics (Oxford, England)*, 12(4):723–736, October 2011. doi: 10.1093/biostatistics/kxr003.

Davies, R. B. Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative. *Biometrika*, 64(2):247–254, 1977. doi: 10.2307/2335690.

Davies, R. B. Hypothesis Testing when a Nuisance Parameter is Present Only Under the Alternatives. *Biometrika*, 74(1):33–43, 1987. doi: 10.2307/2336019.

Dempsey, W. and McCullagh, P. Survival models and health sequences. *Lifetime Data Analysis*, 24(4):550–584, 2018. doi: 10.1007/s10985-018-9424-9.

Dempster, A., Laird, N., and B. Rubin, D. Maximum Likelihood From Incomplete Data Via The EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39:1–38, January 1977.

Dominicus, A., Ripatti, S., Pedersen, N. L., and Palmgren, J. A random change point model for assessing variability in repeated measures of cognitive function. *Statistics in Medicine*, 27(27):5786–5798, November 2008. doi: 10.1002/sim.3380.

DSM-IV, A. P. A. T. F. o. *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition: DSM-IV-TR®*. American Psychiatric Association, 2000. ISBN 978-0-89042-025-6. Google-Books-ID: 3SQrtpnHb9MC.

Dubois, B., Hampel, H., Feldman, H. H., Scheltens, P., Aisen, P., Andrieu, S., Bakardjian, H., Benali, H., Bertram, L., Blennow, K., Broich, K., Cavedo, E., Crutch, S., Dartigues, J.-F., Duyckaerts, C., Epelbaum, S., Frisoni, G. B., Gauthier, S., Genthon, R., Gouw, A. A., Habert, M.-O., Holtzman, D. M., Kivipelto, M., Lista, S., Molinuevo, J.-L., O'Bryant, S. E., Rabinovici, G. D., Rowe, C., Salloway, S., Schneider, L. S., Sperling, R., Teichmann, M., Carrillo, M. C., Cummings, J., and Jack, C. R. Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. *Alzheimer's & Dementia*, 12(3):292–323, March 2016. doi: 10.1016/j.jalz.2016.02.002.

Eddelbuettel, D. and Balamuta, J. J. Extending R with C: A Brief Introduction to Rcpp. *PeerJ Preprints*, 5: e3188v1, aug 2017. doi: 10.7287/peerj.preprints.3188v1.

Farley, J. U. and Hinich, M. J. A Test for a Shifting Slope Coefficient in a Linear Model. *Journal of the American Statistical Association*, 65(331):1320–1329, 1970. doi: 10.2307/2284298.

Feder, P. I. The Log Likelihood Ratio in Segmented Regression. *The Annals of Statistics*, 3(1):84–97, January 1975. doi: 10.1214/aos/1176343000.

Ferrer, L., Rondeau, V., Dignam, J., Pickles, T., Jacqmin-Gadda, H., and Proust-Lima, C. Joint modelling of longitudinal and multi-state processes: application to clinical progressions in prostate cancer. *Statistics in medicine*, 35(22):3933–3948, 2016.

Flicker, C., Ferris, S. H., and Reisberg, B. Mild cognitive impairment in the elderly: Predictors of dementia. *Neurology*, 41(7):1006–1006, July 1991. doi: 10.1212/WNL.41.7.1006.

Folstein, M. F., Folstein, S. E., and McHugh, P. R. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3):189–198, November 1975.

Fong, Y., Di, C., and Permar, S. Change point testing in logistic regression models with interaction term. *Statistics in Medicine*, 34(9):1483–1494, April 2015. doi: 10.1002/sim.6419.

Fong, Y., Huang, Y., Gilbert, P. B., and Permar, S. R. chngpt: threshold regression model estimation and inference. *BMC Bioinformatics*, 18(1):454, October 2017. doi: 10.1186/s12859-017-1863-x.

Freedman, D. A. How Can the Score Test Be Inconsistent? *The American Statistician*, January 2012. doi: 10.1198/000313007X243061.

Godbolt, A. K., Cipolotti, L., Watt, H., Fox, N. C., Janssen, J. C., and Rossor, M. N. The natural history of Alzheimer disease: a longitudinal presymptomatic and symptomatic study of a familial cohort. *Archives of Neurology*, 61(11):1743–1748, November 2004. doi: 10.1001/archneur.61.11.1743.

Griffiths, D. A. and Miller, A. J. Hyperbolic regression - a model based on two-phase piecewise linear regression with a smooth transition between regimes. *Communications in Statistics*, 2(6):561–569, January 1973. doi: 10.1080/03610927308827098.

Grober, E. and Buschke, H. Genuine Memory Deficits in Dementia. *Developmental Neuropsychology*, 3:13–36, January 1987. doi: 10.1080/87565648709540361.

Gurevich, G. and Vexler, A. Change point problems in the model of logistic regression. *Journal of Statistical Planning and Inference*, 131(2):313–331, May 2005. doi: 10.1016/j.jspi.2004.03.004.

Hall, C. B., Lipton, R. B., Sliwinski, M., and Stewart, W. F. A change point model for estimating the onset of cognitive decline in preclinical Alzheimer's disease. *Statistics in Medicine*, 19(11-12):1555–1566, June 2000.

Hall, C. B., Ying, J., Kuo, L., Sliwinski, M., Buschke, H., Katz, M., and Lipton, R. B. Estimation of bivariate measurements having different change points, with application to cognitive ageing. *Statistics in Medicine*, 20 (24):3695–3714, December 2001.

Hall, C. B., Ying, J., Kuo, L., and Lipton, R. B. Bayesian and profile likelihood change point methods for modeling cognitive function over time. *Computational Statistics & Data Analysis*, 42(1):91–109, February 2003. doi: 10.1016/S0167-9473(02)00148-2.

Hansen, B. E. Inference When a Nuisance Parameter Is Not Identified under the Null Hypothesis. *Econometrica*, 64(2):413–430, 1996.

Hawkins, D. M. Fitting multiple change-point models to data. *Computational Statistics & Data Analysis*, 37 (3):323–341, September 2001. doi: 10.1016/S0167-9473(00)00068-2.

Hendrie, H. C., Osuntokun, B. O., Hall, K. S., Ogunniyi, A. O., Hui, S. L., Unverzagt, F. W., Gureje, O., Rodenberg, C. A., Baiyewu, O., and Musick, B. S. Prevalence of Alzheimer's disease and dementia in two communities: Nigerian Africans and African Americans. *The American Journal of Psychiatry*, 152(10): 1485–1492, October 1995. doi: 10.1176/ajp.152.10.1485.

Hinkley, D. V. Inference about the Intersection in Two-Phase Regression. *Biometrika*, 56(3):495–504, 1969. doi: 10.2307/2334655.

Hinkley, D. V. Inference about the change-point in a sequence of random variables. *Biometrika*, 57(1):1–17, April 1970. doi: 10.1093/biomet/57.1.1.

Horvath, L. The Maximum Likelihood Method for Testing Changes in the Parameters of Normal Observations. *The Annals of Statistics*, 21(2):671–680, 1993.

Howieson, D. B., Carlson, N. E., Moore, M. M., Wasserman, D., Abendroth, C. D., Payne-Murphy, J., and Kaye, J. A. Trajectory of mild cognitive impairment onset. *Journal of the International Neuropsychological Society: JINS*, 14(2):192–198, March 2008. doi: 10.1017/S1355617708080375.

Hsu, W.-W., Todem, D., and Kim, K. A sup-score test for the cure fraction in mixture models for long-term survivors. *Biometrics*, 72(4):1348–1357, December 2016. doi: 10.1111/biom.12514.

Hubbard, B. M., Fentonm, G. W., and Anderson, J. M. A quantitative histological study of early clinical and preclinical Alzheimer's disease. *Neuropathology and Applied Neurobiology*, 16(2):111–121, 1990. doi: 10.1111/j.1365-2990.1990.tb00940.x.

Isaacs, B. and Kennie, A. T. The Set Test as an Aid to the Detection of Dementia in Old People. *The British Journal of Psychiatry*, 123(575):467–470, October 1973. doi: 10.1192/bjp.123.4.467.

Jack, C. R., Knopman, D. S., Jagust, W. J., Shaw, L. M., Aisen, P. S., Weiner, M. W., Petersen, R. C., and Trojanowski, J. Q. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet. Neurology*, 9(1):119–128, January 2010. doi: 10.1016/S1474-4422(09)70299-6.

Jack, C. R., Knopman, D. S., Jagust, W. J., Petersen, R. C., Weiner, M. W., Aisen, P. S., Shaw, L. M., Vemuri, P., Wiste, H. J., Weigand, S. D., Lesnick, T. G., Pankratz, V. S., Donohue, M. C., and Trojanowski, J. Q. Update on hypothetical model of Alzheimer's disease biomarkers. *Lancet neurology*, 12(2):207–216, February 2013. doi: 10.1016/S1474-4422(12)70291-0.

Jacqmin-Gadda, H., Commenges, D., and Dartigues, J.-F. Random change point model for joint modeling of cognitive decline and dementia. *Biometrics*, 62(1):254–260, March 2006. doi: 10.1111/j.1541-0420.2005.00443.x.

Johnson, D. K., Storandt, M., Morris, J. C., and Galvin, J. E. Longitudinal Study of the Transition From Healthy Aging to Alzheimer Disease. *Archives of Neurology*, 66(10):1254–1259, October 2009. doi: 10.1001/archneurol.2009.158.

Jost, B. C. and Grossberg, G. T. The Natural History of Alzheimer's Disease: A Brain Bank Study. *Journal of the American Geriatrics Society*, 43(11):1248–1255, November 1995. doi: 10.1111/j.1532-5415.1995.tb07401.x.

Juang, S.-H. and Wolfe, D. A. Nonparametric changepoint procedures for repeated measures data. *Communications in Statistics - Simulation and Computation*, 19(1):263–286, January 1990. doi: 10.1080/03610919008812856.

Katzman, R., Aronson, M., Fuld, P., Kawas, C., Brown, T., Morgenstern, H., Frishman, W., Gidez, L., Eder, H., and Ooi, W. L. Development of dementing illnesses in an 80-year-old volunteer cohort. *Annals of Neurology*, 25(4):317–324, 1989. doi: 10.1002/ana.410250402.

Kim, H.-J. and Siegmund, D. The Likelihood Ratio Test for a Change-Point in Simple Linear Regression. *Biometrika*, 76(3):409–423, 1989. doi: 10.2307/2336108.

Lai, T. L. Sequential Changepoint Detection in Quality Control and Dynamical Systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(4):613–644, November 1995. doi: 10.1111/j.2517-6161.1995.tb02052.x.

Laird, N. M. and Ware, J. H. Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974, December 1982.

Lesaffre, E. and Spiessens, B. On the Effect of the Number of Quadrature Points in a Logistic Random-Effects Model: An Example. *Journal of the Royal Statistical Society Series C*, 50:325–335, February 2001. doi: 10.1111/1467-9876.00237.

Letenneur, L., Commenges, D., Dartigues, J.-F., and Barberger-Gateau, P. Incidence of dementia and Alzheimer's disease in elderly community residents of south-western France. *International Journal of Epidemiology*, 23(6):1256–1261, December 1994.

Levenberg, K. A method for the solution of certain non-linear problems in least squares. *Quarterly of Applied Mathematics*, 2(2):164–168, 1944. doi: 10.1090/qam/10666.

Li, G., Larson, E. B., Shofer, J. B., Crane, P. K., Gibbons, L. E., McCormick, W., Bowen, J. D., and Thompson, M. L. Cognitive trajectory changes over 20 years prior to dementia diagnosis: a large cohort study. *Journal of the American Geriatrics Society*, 65(12):2627–2633, December 2017. doi: 10.1111/jgs.15077.

Lombard, F. Rank Tests for Changepoint Problems. *Biometrika*, 74(3):615–624, 1987. doi: 10.2307/2336701.

Machulda, M. M., Pankratz, V. S., Christianson, T. J., Ivnik, R. J., Mielke, M. M., Roberts, R. O., Knopman, D. S., Boeve, B. F., and Petersen, R. C. Practice Effects and Longitudinal Cognitive Change in Normal Aging vs. Incident Mild Cognitive Impairment and Dementia in The Mayo Clinic Study of Aging. *The Clinical Neuropsychologist*, 27(8):1247–1264, November 2013. doi: 10.1080/13854046.2013.836567.

Marquardt, D. W. An Algorithm for Least Square Estimation of Non-Linear Parameters. *SIAM Journal on Applied Mathematics*, 11:431–441, June 1963. doi: 10.1137/0111030.

McClearn, G. E., Johansson, B., Berg, S., Pedersen, N. L., Ahern, F., Petrill, S. A., and Plomin, R. Substantial genetic influence on cognitive abilities in twins 80 or more years old. *Science (New York, N.Y.)*, 276(5318): 1560–1563, June 1997. doi: 10.1126/science.276.5318.1560.

Molenberghs, G. and Verbeke, G. The Generalized Linear Mixed Model (GLMM). In *Models for Discrete Longitudinal Data*, Springer Series in Statistics, pages 265–280. Springer New York, New York, NY, 2005. ISBN 978-0-387-28980-9. doi: 10.1007/0-387-28980-1_14.

Muggeo, V. M. R. Segmented mixed models with random changepoints in r. Working Paper, 02 2016a.

Muggeo, V. M. R. Testing with a nuisance parameter present only under the alternative: a score-based approach with application to segmented modelling. *Journal of Statistical Computation and Simulation*, 86(15):3059–3067, October 2016b. doi: 10.1080/00949655.2016.1149855.

Muggeo, V. M., Atkins, D. C., Gallop, R. J., and Dimidjian, S. Segmented mixed models with random changepoints: a maximum likelihood approach with application to treatment for depression study. *Statistical Modelling*, 14(4):293–313, August 2014. doi: 10.1177/1471082X13504721.

Nelder, J. A. and Mead, R. A Simplex Method for Function Minimization. *The Computer Journal*, 7(4): 308–313, January 1965. doi: 10.1093/comjnl/7.4.308.

Oehlert, G. W. A Note on the Delta Method. *The American Statistician*, 46(1):27–29, 1992. doi: 10.2307/2684406.

Patterson, C. *World Alzheimer Report 2018 - The state of the art of dementia research: New frontiers.* Alzheimer's Disease International, 2018.

Pedersen, N., Plomin, R., Nesselroade, J., and McClearn, G. A Quantitative Genetic Analysis of Cognitive Abilities during the Second Half of the Life Span. *Psychological Science*, 3(6):346–353, November 1992. doi: 10.1111/j.1467-9280.1992.tb00045.x.

Petersen, R. C. Mild cognitive impairment as a diagnostic entity. *Journal of Internal Medicine*, 256(3):183–194, 2004. doi: 10.1111/j.1365-2796.2004.01388.x.

Petersen, R. C., Doody, R., Kurz, A., Mohs, R. C., Morris, J. C., Rabins, P. V., Ritchie, K., Rossor, M., Thal, L., and Winblad, B. Current Concepts in Mild Cognitive Impairment. *Archives of Neurology*, 58(12): 1985–1992, December 2001. doi: 10.1001/archneur.58.12.1985.

Pettitt, A. N. A Non-Parametric Approach to the Change-Point Problem. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(2):126–135, 1979. doi: 10.2307/2346729.

Piepho, H. P. and Ogutu, J. O. Inference for the Break Point in Segmented Regression with Application to Longitudinal Data. *Biometrical Journal*, 45(5):591–601, 2003. doi: 10.1002/bimj.200390035.

Pinheiro, J. C. and Bates, D. M. Approximations to the Log-Likelihood Function in the Nonlinear Mixed-Effects Model. *Journal of Computational and Graphical Statistics*, 4(1):12–35, 1995. doi: 10.2307/1390625.

Proust-Lima, C., Amieva, H., Dartigues, J.-F., and Jacqmin-Gadda, H. Sensitivity of Four Psychometric Tests to Measure Cognitive Changes in Brain Aging-Population–based Studies. *American Journal of Epidemiology*, 165(3):344–350, February 2007. doi: 10.1093/aje/kwk017.

Proust-Lima, C., Amieva, H., and Jacqmin-Gadda, H. Analysis of multivariate mixed longitudinal data: a flexible latent process approach. *The British Journal of Mathematical and Statistical Psychology*, 66(3): 470–487, November 2013. doi: 10.1111/bmsp.12000.

Quandt, R. E. Tests of the Hypothesis that a Linear Regression System Obeys Two Separate Regimes. *Journal of the American Statistical Association*, 55(290):324–330, 1960. doi: 10.2307/2281745.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018.

Rajan, K. B., Wilson, R. S., Barnes, L. L., Aggarwal, N. T., Weuve, J., and Evans, D. A. A Cognitive Turning Point in Development of Clinical Alzheimer's Disease Dementia and Mild Cognitive Impairment: A Biracial Population Study. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 72(3): 424–430, March 2017. doi: 10.1093/gerona/glw246.

Ramanayake, A. and Gupta, A. K. Testing for a change point in a sequence of exponential random variables with repeated values. *Journal of Statistical Computation and Simulation*, 80(2):191–199, February 2010. doi: 10.1080/00949650802586384.

Ramsay, J. O. Monotone Regression Splines in Action. *Statistical Science*, 3(4):425–441, November 1988. doi: 10.1214/ss/1177012761.

Reeves, J., Chen, J., Wang, X. L., Lund, R., and Lu, Q. Q. A Review and Comparison of Changepoint Detection Techniques for Climate Data. *Journal of Applied Meteorology and Climatology*, 46(6):900–915, June 2007. doi: 10.1175/JAM2493.1.

Rizopoulos, D. Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. *Computational Statistics & Data Analysis*, 56:491–501, March 2012. doi: 10.1016/j.csda.2011.09.007.

Robbins, H. and Siegmund, D. A class of stopping rules for testing parametric hypotheses. The Regents of the University of California, 1972.

Salomone, S., Caraci, F., Leggio, G. M., Fedotova, J., and Drago, F. New pharmacological strategies for treatment of Alzheimer's disease: focus on disease modifying drugs. *British Journal of Clinical Pharmacology*, 73(4):504–517, April 2012. doi: 10.1111/j.1365-2125.2011.04134.x.

Seber, G. A. F. and Wild, C. J. Multiphase and Spline Regressions. In *Nonlinear Regression*, pages 433–489. John Wiley & Sons, Ltd, 2005. ISBN 978-0-471-72531-2. doi: 10.1002/0471725315.ch9.

Segalas, C., Amieva, H., and Jacqmin-Gadda, H. A hypothesis testing procedure for random changepoint mixed models. *Statistics in Medicine*, 38(20):3791–3803, 2019. doi: 10.1002/sim.8195.

Sliwinski, M. J., Stawski, R. S., Hall, C. B., Katz, M., Verghese, J., and Lipton, R. Distinguishing preterminal and terminal cognitive decline. *European Psychologist*, 11(3):172–181, 2006. doi: 10.1027/1016-9040.11.3.172.

Smith, A. F. M. A Bayesian Approach to Inference about a Change-Point in a Sequence of Random Variables. *Biometrika*, 62(2):407–416, 1975. doi: 10.2307/2335381.

Spiegelhalter, D., Best, N., Carlin, B., and van der Linde, A. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, 64:1–34, January 2002. doi: 10.1111/1467-9868.02022.

Stern, Y., Gurland, B., Tatemichi, T. K., Tang, M. X., Wilder, D., and Mayeux, R. Influence of education and occupation on the incidence of Alzheimer's disease. *JAMA*, 271(13):1004–1010, April 1994.

Stern, Y. Cognitive reserve. *Neuropsychologia*, 47(10):2015–2028, August 2009. doi: 10.1016/j.neuropsychologia.2009.03.004.

Stram, D. O. and Lee, J. W. Variance Components Testing in the Longitudinal Mixed Effects Model. *Biometrics*, 50(4):1171–1177, 1994. doi: 10.2307/2533455.

Thorvaldsson, V., Hofer, S. M., Berg, S., Skoog, I., Sacuiu, S., and Johansson, B. Onset of terminal decline in cognitive abilities in individuals without dementia. *Neurology*, 71(12):882–887, September 2008. doi: 10.1212/01.wnl.0000312379.02302.ba.

Tishler, A. and Zang, I. A New Maximum Likelihood Algorithm for Piecewise Regression. *Journal of the American Statistical Association*, 76(376):980–987, 1981. doi: 10.2307/2287599.

van den Hout, A., Muniz-Terrera, G., and Matthews, F. E. Smooth random change point models. *Statistics in Medicine*, 30(6):599–610, 2011. doi: 10.1002/sim.4127.

van den Hout, A., Muniz-Terrera, G., and Matthews, F. E. Change point models for cognitive tests using semi-parametric maximum likelihood. *Computational Statistics & Data Analysis*, 57(1):684–698, January 2013. doi: 10.1016/j.csda.2012.07.024.

Van Der Vaart, A. W. and Wellner, J. A. *Weak Convergence and Empirical Processes*. Springer Series in Statistics. Springer New York, New York, NY, 1996. ISBN 978-1-4757-2547-6 978-1-4757-2545-2. DOI: 10.1007/978-1-4757-2545-2.

Verbeke, G. and Molenberghs, G. Estimation of the Marginal Model. In *Linear Mixed Models for Longitudinal Data*, Springer Series in Statistics, pages 41–54. Springer New York, New York, NY, 2000. ISBN 978-0-387-22775-7. doi: 10.1007/978-0-387-22775-7-5.

Verlinden, V. J. A., van der Geest, J. N., de Bruijn, R. F. A. G., Hofman, A., Koudstaal, P. J., and Ikram, M. A. Trajectories of decline in cognition and daily functioning in preclinical dementia. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, 12(2):144–153, February 2016. doi: 10.1016/j.jalz.2015.08.001.

Wang, W. and Yan, J. *splines2: Regression Spline Functions and Classes*, 2018. R package version 0.2.8.

Wechsler, D. *Manual for the Wechsler Adult Intelligence Scale.* Manual for the Wechsler Adult Intelligence Scale. Psychological Corp., Oxford, England, 1955.

WHO. *Global action plan on the public health response to dementia, 2017-2025.* World Health Organization, 2017.

Wilson, R. S., Beckett, L. A., Bienias, J. L., Evans, D. A., and Bennett, D. A. Terminal decline in cognitive function. *Neurology*, 60(11):1782–1787, June 2003. doi: 10.1212/01.wnl.0000068019.60901.c1.

Wilson, R. S., Segawa, E., Boyle, P. A., Anagnos, S. E., Hizel, L. P., and Bennett, D. A. The Natural History of Cognitive Decline in Alzheimer's Disease. *Psychology and aging*, 27(4):1008–1017, December 2012. doi: 10.1037/a0029857.

Winblad, B., Palmer, K., Kivipelto, M., Jelic, V., Fratiglioni, L., Wahlund, L.-O., Nordberg, A., Bäckman, L., Albert, M., Almkvist, O., Arai, H., Basun, H., Blennow, K., Leon, M. D., DeCarli, C., Erkinjuntti, T., Giacobini, E., Graff, C., Hardy, J., Jack, C., Jorm, A., Ritchie, K., Duijn, C. V., Visser, P., and Petersen, R. C. Mild cognitive impairment – beyond controversies, towards a consensus: report of the International Working Group on Mild Cognitive Impairment. *Journal of Internal Medicine*, 256(3):240–246, 2004. doi: 10.1111/j.1365-2796.2004.01380.x.

Yang, L. and Gao, S. Bivariate random change point models for longitudinal outcomes. *Statistics in medicine*, 32(6):1038–1053, March 2013. doi: 10.1002/sim.5557.

Yu, B. and Ghosh, P. Joint modeling for cognitive trajectory and risk of dementia in the presence of death. *Biometrics*, 66(1):294–300, March 2010. doi: 10.1111/j.1541-0420.2009.01261.x.

Inférence dans les modèles à changement de pente aléatoire : application au déclin cognitif pré-démence

Résumé : Le but de ce travail a été de proposer des méthodes d'inférence pour décrire l'histoire naturelle de la phase pré-diagnostic de la démence. Durant celle-ci, qui dure une quinzaine d'années, les trajectoires de déclin cognitif sont non linéaires et hétérogènes entre les sujets. Pour ces raisons, nous avons choisi un modèle à changement de pente aléatoire pour les décrire. Une première partie de ce travail a consisté à proposer une procédure de test pour l'existence d'un changement de pente aléatoire. En effet, dans certaines sous-populations, le déclin cognitif semble lisse et la question de l'existence même d'un changement de pente se pose. Cette question présente un défi méthodologique en raison de la non-identifiabilité de certains paramètres sous l'hypothèse nulle rendant les tests standards inutiles. Nous avons proposé un supremum score test pour répondre à cette question. Une seconde partie du travail concernait l'ordre temporel du temps de changement entre plusieurs marqueurs. La démence est une maladie multidimensionnelle et plusieurs dimensions de la cognition sont affectées. Des schémas hypothétiques existent pour décrire l'histoire naturelle de la démence mais n'ont pas été éprouvés sur données réelles. Comparer le temps de changement de différents marqueurs mesurant différentes fonctions cognitives permet d'éclairer ces hypothèses. Dans cet esprit, nous proposons un modèle bivarié à changement de pente aléatoire permettant de comparer les temps de changement de deux marqueurs, potentiellement non gaussiens. Les méthodes proposées ont été évaluées sur simulations et appliquées sur des données issues de deux cohortes françaises. Enfin, nous discutons les limites de ces deux modèles qui se concentrent sur une accélération tardive du déclin cognitif précédant le diagnostic de démence et nous proposons un modèle alternatif qui estime plutôt une date de décrochage entre cas et non-cas.

Mots clés : Démence, modèles mixtes, données longitudinales multivariées, paramètres de nuisance non identifiables, changement de pente aléatoire, test du score.

Inference for random changepoint models: application to pre-dementia cognitive decline

Abstract: The aim of this work was to propose inferential methods to describe natural history of the pre-diagnosis phase of dementia. During this phase, which can last around fifteen years, the cognitive decline trajectories are nonlinear and heterogeneous between subjects. Because heterogeneity and nonlinearity, we chose a random changepoint mixed model to describe these trajectories. A first part of this work was to propose a testing procedure to assess the existence of a random changepoint. Indeed, in some subpopulations, the cognitive decline seems smooth and the question of the existence of a changepoint itself araises. This question is methodologically challenging because of identifiability issues on some parameters under the null hypothesis that makes standard tests useless. We proposed a supremum score test to answer this question. A second part of this work was the comparison of the temporal order of different markers changepoint. Dementia is a multidimensional disease where different dimensions of the cognition are affected. Hypothetic cascade models exist for describing this natural history but have not been evaluated on real data. Comparing change over time of different markers measuring different cognitive functions gives precious insight on this hypothesis. In this spirit, we propose a bivariate random changepoint model allowing proper comparison of the time of change of two cognitive markers, potentially non Gaussian. The proposed methodologies were evaluated on simulation studies and applied on real data from two French cohorts. Finally, we discussed the limitations of the two models we used that focused on the late acceleration of the cognitive decline before dementia diagnosis and we proposed an alternative model that estimates the time of differentiation between cases and non-cases.

Keywords: Dementia, mixed models, multivariate longitudinal data, non identifiable nuisance parameters, random changepoint, score test.

INSERM U1219 Bordeaux Population Health Research Center

Equipe Biostatistique - 146 rue Léo Saignat - 33000 Bordeaux