Universidade Federal de Pernambuco
Centro de Ciências Exatas e da Natureza
Programa de Pós-Graduação em Estatística

ROMMY CAMASCA OLIVARI

LIKELIHOOD BASED INFERENCE FOR AUTOREGRESSIVE CENSORED MIXED-EFFECTS MODELS, WITH APPLICATIONS TO HIV VIRAL LOADS DATASET

Recife

2019

ROMMY CAMASCA OLIVARI

LIKELIHOOD BASED INFERENCE FOR AUTOREGRESSIVE CENSORED MIXED-EFFECTS MODELS, WITH APPLICATIONS TO HIV VIRAL LOADS DATASET

Dissertação apresentada ao Programa de Pós-Graduação em Estatística do Centro de Ciências Exatas e da Natureza da Universidade Federal de Pernambuco, como requisito parcial à obtenção do título de mestre em Estatística.

Área de Concentração: Estatística Aplicada

Orientador: Prof. Dr. Aldo William Medina Garay
Coorientador: Prof. Dr. Víctor Hugo Lachos Dávila

Recife

2019

Catalogação na fonte
Bibliotecária Elaine Freitas CRB 4-1790

O48l Olivari, Rommy Camasca
Likelihood based inference for autoregressive censored mixed-effects models, with applications to hiv viral loads dataset / Rommy Camasca Olivari. – 2019.
84 f.: fig., tab.

Orientador: Aldo William Medina Garay
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CCEN. Estatística. Recife, 2019.
Inclui referências e apêndices.

1. Probabilidade e estatística. 2. Modelos AR(p) autorregressivo. 3. Dados censurados. I. Garay, Aldo William Medina (orientador) II. Título.

519.5 CDD (22. ed.) UFPE-MEI 2019-34

ROMMY CAMASCA OLIVARI

LIKELIHOOD BASED INFERENCE FOR AUTOREGRESSIVE CENSORED MIXED-EFFECTS MODELS, WITH APPLICATIONS TO HIV VIRAL LOADS DATASET

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Estatística.

Aprovada em: 27 de fevereiro de 2019.

BANCA EXAMINADORA

Prof.(º) Aldo William Medina Garay
UFPE

Prof.(º) Roberto Ferreira Manghi
UFPE

Prof.(º) Celso Rômulo Barbosa Cabral
UFAM

À minha família, por me apoiar em todo momento e por ser um porto seguro ao qual eu posso retornar sempre. Mãe, Pai, Favio, Yuri, Negrita, Andy e Link seu amor e cuidado sempre me ajudam a superar tudo.

ACKNOWLEDGMENTS

No primeiro lugar agradeço aos meu pais Sofia e Jaime, meu imão Favio e meus gatos Negrita, Andy e Link pelo apoio e carinho incondicinoal para seguir meus sonhos. Obrigada por tudo.

Aos meus orientadores Aldo William Medina Garay e Víctor Hugo Lachos Dávila pela paciância e por acreditar no meu potencial. Muito obrigada pela dedicação e por todos os seus ensinamentos.

A professora Francyelle de Lima Medina pela grande ajuda com o idioma, suas opiniões e conselhos.

A todos os professores do CCEN, e em especial aos Professores Alex Dias Ramos, Aldo William Medina Garay, Audrey Helen Mariz de Aquino Cysneiros e Gauss Moutinho Cordeiro, vocês contribuíram muito para minha formação.

A Yuri Alves de Araujo por sua infinita paciência, carinho, amor e compreensão. Com teu apoio e humor consegui tornar as dificuldades em histórias para rir e recordar.

A minha tia Carolina, tio albeto, minha avó Betsabe e minha avó Emma os quais sempre acreditaram em mim.

A meus amigos do Perú Shirley Quispe Sigueñas, Kelly Figueroa Cabero, Lucerito Aguilar Contreras, Vilma Romero Romero, Erick Chacon Montalvan e Yeny Pillco Valencia pelo carinho e apoio, eu sempre posso contar como vocês.

A meus amigos do Brasil Roberta Tabaczenski, Anbeth Radünz, Vinícius Scher, César Diogo, Érica Vieira, Adenice Ferreira e Lays Alves por me mostrar suas costumbres e facilitar minha adaptação neste hermoso pais.

A senhora Elianice Araujo de Sousa e Ygor Alves de Araujo por me acolherem na sua familia e me dar seu apoio.

Aos servidores da Universidade Federal de Pernambuco (UFPE), principalmente a Valéria Bittencourt e Katia, pela paciência e cuidado ao longo desses dois anos de mestrado.

A CAPES pelo apoio financeiro, o qual me permitiu estudar neste pais.

Por fim, quero agradecer à todas as pessoas que contribuíram diretamente e indiretamente para meu sucesso e crescimento pessoal e profissional.

ABSTRACT

In AIDS clinical trials, the HIV-1 RNA measurements are often subject to some upper and lower detection limits, depending on the quantification assays. Linear and nonlinear mixedeffects models, with modifications to accommodate censored observations, are routinely used to analyze this type of data (VAIDA; LIU, 2009). This work presents a likelihood based approach for fitting Linear and nonlinear mixedeffects models, with modifications to accommodate censored observations and considering an structure autoregressive of order p (AR(p)) dependence on the error term. An EM-type algorithm is developed for computing the maximum likelihood estimates, obtaining as a byproduct the standard errors of the fixed effects and the likelihood value. Moreover, the constraints on the parameter space arising, from the stationarity conditions for the autoregressive parameters, in the EM algorithm are handled by a reparametrization scheme, as discussed by Lin e Lee (2007). Finally, the proposed algorithm is implemented in the R package ARpMMEC, which is available. It presents an application to real data and developed three simulation studies that show the relevance and applicability of the proposed model.

Keywords: Autoregressive AR(p) models. Censored data. EM Algorithm. Linear/nonlinear mixed models.

RESUMO

Ensaios clínicos de HIV, as medições de VIH-1 RNA são frequentemente restritas por alguns limites de detecção superiores e inferiores, dependendo do ensaio de quantificação. Modelos com efeitos mistos, lineares e não lineares, adaptados para observações censuradas, são rotineiramente utilizados para analisar este tipo de dados (VAIDA; LIU, 2009). Este trabalho apresenta, sobre uma abordagem baseada na maximização da função de verossimilhança, o ajuste dos modelos com efeitos mistos, lineares e não lineares, adaptados para observações censuradas e considerando uma estrutura de dependência autorregressiva de ordem p (AR(p)) para o erro do modelo. É desenvolvido um algoritmo tipo EM para calcular as estimativas por máxima verossimilhança, obtendo como resultado os erros padrões dos efeitos fixos e o valor da função de verossimilhança. Além disso, para lidiar com as restrições sobre o espaço paramétrico, que surgem na estimação dos parâmetros autorregresivos do modelo AR(p) no algoritmo EM, é utilizado um esquema de reparametrização, como discutido por Lin e Lee (2007). Finalmente, o algoritmo proposto é implementado no pacote *ARpLMEC*, o qual está disponível. É apresentada uma aplicação a dados reais e desenvolvido três estudos de simulação que evidenciam a relevância e aplicabilidade do modelo proposto.

Palavras-chave: Modelos AR(p) autorregressivos. Dados censurados. Algoritmo EM. Modelos mistos lineares/não lineares.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 INTRODUCTION

Neste capítulo será apresentada a motivação que levou ao desenvolvimento desta pesquisa, uma descrição e análise prévia do conjunto de dados reais utilizado, assim como os objetivos estabelecidos para este trabalho. Além disso, apresentaremos algumas definições prévias, que são necessárias para um melhor entendimento das metodologias aplicadas. Finalizaremos com uma descrição da organização desta dissertação.

1.1 MOTIVATION

Durante as últimas décadas, os métodos estatísticos para dados longitudinais contínuos, com medidas repetidas, receberam atenção considerável, como pode ser visto por meio de um detalhado levantamento bibliográfico (LAIRD; WARE, 1982; SCHAFER; YUCEL, 2002; WANG; FAN, 2010; LIN; WANG, 2013; SCHUMACHER; LACHOS; DEY, 2017). Os dados longitudinais são uma coleção de observações repetidas, dos mesmos indivíduos, retiradas de uma população, durante um período de tempo. Para a análise deste tipo de conjunto de dados, os modelos de regressão com efeitos mistos, lineares e não lineares, são as ferramentas comumente utilizadas (SCHAFER; YUCEL, 2002; LIN; LEE, 2007). Como foi discutido em Wang (2013), com estes modelos é possível considerar diferentes efeitos, intra-indivíduos e entre-indivíduos, por meio da inclusão de estruturas de dependência, tanto nos efeitos aleatórios quanto no erro do modelo. Laird e Ware (1982) desenvolveram o modelo de regressão linear misto, que incorpora efeitos aleatórios e permite um desenho não balanceado (LME). Este tipo de desenho é caracterizado por considerar que cada indivíduo possui diferente número de observações, ou seja, os indivíduos em estudo não precisam ter um mesmo número de medições ao longo do tempo. No entanto, em estudos longitudinais, as medições da variável resposta podem ser submetidas a algum limites de detecção superiores ou inferiores, abaixo ou acima do qual as medições não são quantificáveis, obtendo como resultado uma variável resposta censurada. Vaida e Liu (2009) estudaram o modelo proposto em Laird e Ware (1982), com extensão para respostas censuradas, com o intuito de desenvolver um algoritmo de estimação dos parametros mais eficiente, para efeitos mistos lineares e não lineares (denotado por LMEC / NLMEC). Além disso, os autores propuseram as extensões dos modelos LMEC / NLMEC para diferentes estruturas da variância do modelo, como erros heteroscedásticos, erros autocorrelacionados

e modelos multiníveis. Recentemente, Matos, Castro e Lachos (2016) apresentaram uma estrutura de correlação exponencial de decaimento (DEC), com o intuito de ajustar o modelo LMEC/NLMEC para variáveis respostas registradas em intervalos irregulares no tempo. Esta estrutura DEC, dependendo do valor de seus parâmetros, é equivalente a um processo autorregresivo de ordem 1 (AR(1)). Por outro lado, Schumacher, Lachos e Dey (2017) propuseram um modelo regressão linear com erros autorregressivos de ordem "p" (AR(p)-LR), para este mesmo tipo de variável resposta.

Esta estrutura de correlacão autorregressiva de ordem "p" fornece uma família mais ampla de estruturas de correlação que a estrutura DEC. No entanto, apesar de existir algumas propostas na literatura para lidar com o problema de respostas censuradas e observações correlacionadas, ainda não há estudos que considerem a inclusão de um estrutura autorregressiva para dados censurados, com efeitos mistos e respostas multivariadas registradas em intervalos irregulares no tempo. Nesta dissertação, utilizando uma abordagem baseada na maximização da função de verossimilhança, pretendemos desenvolver um método de estimação computacionalmente eficiente, por meio de um algoritmo tipo EM (DEMPSTER; LAIRD; RUBIN, 1977), em modelos de regressão censurados, com efeitos mistos, lineares e não lineares, considerando respostas multivariadas registradas em intervalos irregulares no tempo, com erros autorregressivos de ordem "p", denotado por AR(p)-LMEC/NLMEC respectivamente. Além disso, elaboraremos um pacote que estará disponível para utilização no software R (R Core Team, 2018).

Com a finalidade de motivar a aplicabilidade do modelo proposto na seguinte seção descreveremos uma análise preliminar dos dados, utilizados nesta pesquisa, que tem como principais características a censura e a coleta de observações em intervalos irregulares no tempo.

1.2 CASE STUDY: AIDS CLINICAL TRIAL

O protocolo "*ACTG-315*", é um conjunto de dados, estudado previamente por Wu (2002) e Matos, Castro e Lachos (2016), no qual são analisados 46 pacientes infectados pelo HIV-1, tratados com um potente coquetel de medicamentos anti-retrovirais, a base do inibidor da protease "ritonavir" e dois medicamentos inibidores da transcrição reversa (zidovudina e lamivudina). Antes de iniciar a terapia anti-retroviral, todos os pacientes interromperam seu próprio regime antirretroviral por cinco semanas, como um

período de "wash-out" (período necessário para que a concentração de um medicamento seja desprezível, depois da suspensão de seu uso). O objetivo deste esquema antirretroviral é mostrar que a imunidade pode ser parcialmente restaurada em pessoas com o virus do HIV, em um estágio moderadamente avançado. A carga viral dos pacientes foi quantificada nos dias 0, 2, 7, 10, 14, 21, 28, 56, 84, 168 e 196, após o início do tratamento. O conjunto de dados contém 361 observações.

Um marcador imunológico conhecido como contagem de células CD4+ também foi obtido simultaneamente com a carga viral. Observou-se que 72 dos 361 valores de CD4+ (20 % aproximadamente) estavam faltando devido a uma incompatibilidade dos CD4 e os horários de medição de carga viral. Por outro lado, 40 das 361 medidas da carga viral (11 % aproximadamente) apresentaram um valor inferior ao valor limite detectável de 100 cópias/mL. Para poder comparar os valores das cargas virais entre eles, é aplicada uma transformação logaritmica em base 10 (NETO *et al.*, 2003). Na Figura 1 (a) mostra-se os perfis individuais para as cargas virais de HIV, medidas em diferentes intervalos no tempo. Foram destacados 3 indivíduos aleatoriamente: indivíduos 16 (linha verde) e 22 (linha azul) os quais não tem observações censuradas e o indivíduo 1 (linha vermelha), o qual tem a última observação censurada.

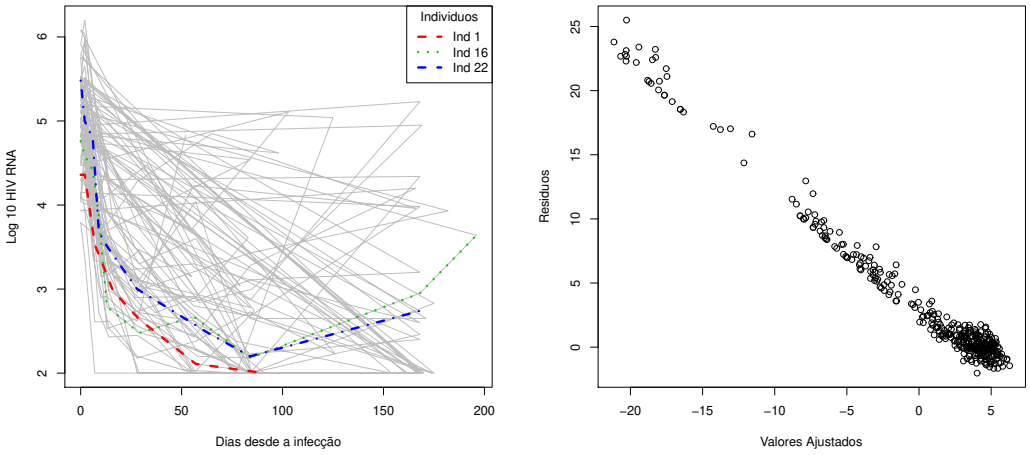

Figure 1 – Left: Profiles of individuals infected with HIV, at different time intervals, Right: Residues from the uncorrelated model vs estimated values of the logarithm of individuals viral loads

Wu (2002) analisou este conjunto de dados considerando o seguinte modelo

não linear:

$$\begin{aligned} y_{ij} &= \log_{10}(\exp(\lambda_1)+\exp(\lambda_2))+\epsilon_{ij}, \\ \lambda_1 &= \beta_1+b_{1i}-(\beta_2+b_{2i}t_{ij}), \\ \lambda_2 &= \beta_3+b_{3i}-(\beta_4+\beta_5 CD4_{ij}+b_{4i})t_{ij}, \end{aligned} \tag{1.1}$$

em que y_{ij} representa o logaritmo da carga viral observada da j-ésima repetição do indivíduo "i", no tempo "t_{ij}" ($i = 1,2,\ldots,n$, $j = 1,2,\ldots,n_i$), $CD4_{ij}$ indica o valor de CD4 observado no tempo t_{ij}, $\mathbf{b}_i = (b_{1i},\ldots,b_{4i})^\top$ é o vetor de efeitos aleatórios para o indivíduo "i" e $\epsilon_i = (\epsilon_{i1},\ldots,\epsilon_{in_i})^\top$ representa o vetor de erros aleatório. Ajustamos o modelo (1.1) deconsiderando a correlação intra-individuo, ou seja $\epsilon_i \sim N_{n_i}(\mathbf{0},\sigma^2\mathbf{I}_{\mathbf{n_i}})$, em que $\mathbf{I}_{\mathbf{n_i}}$ é a matriz identidade de dimensão $n_i \times n_i$. Na Figura 1 (b) é mostrada a relação entre os resíduos ordinarios e os valores estimados do modelo não correlacionado. Podemos observar que os resíduos têm uma tendencia decrescente, indicando que a aplicação de um modelo não correlacionado não seria adequada, pois os residuos não aprensetam um comportamento aleatório.

Este conjunto de dados também foi analisado em Matos, Castro e Lachos (2016), que ajusta modelos com 4 estruturas de correlação 1diferentes: (1) Estrutura não correlacionada (UNC), (2) Estrutura DEC, (3) Estrutura autorregressiva de ordem 1 e (4) Estrutura simétrica composta (SC). Na Tabela 1 são apresentados os resultados dos indicadores obtidos por Matos, Castro e Lachos (2016), AIC (AKAIKE, 1974), BIC (SCHWARZ *et al.*, 1978), e o valor do logaritmo da função de verossimilhança, na qual pode-se observar que as estruturas DEC e AR(1) apresentam os menores valores dos indicadores.

Table 1 – Indicators for the NLMEC model considering different correlation structures

	UNC	DEC	AR(1)	SC
loglik	-281.31	**-255.83**	**-264.99**	-279.33
AIC	594.61	**547.66**	**563.97**	592.33
BIC	656.83	**617.66**	**630.08**	658.77

Estes resultados indicam que aparentemente considerar uma estrutura de correlação pode ser mais adequado para modelar este tipo de conjunto de dados.

1.3 OBJECTIVES OF THE DISSERTATION

Este trabalho terá como objetivos:

- Apresentar um ajuste e análise inferencial para os modelos AR(p)-LMEC/NLMEC.
- Desenvolver estudos de simulação e aplicação em um conjunto de dados reais, para avaliar o desempenho e adequação do modelo AR(p)-MMEC.
- Disponibilizar os métodos propostos por meio de um pacote no CRAN do software R .

1.4 PRELIMINARY CONCEPTS

Com o intuito de facilitar a compreensão da formulação do modelo, serão apresentados alguns conceitos preliminares.

1.4.1 Regression models with mixed effects

Os modelos de regressão com efeitos mistos contém dois tipos de efeitos, fixos e aleatórios, além do erro experimental. Esses tipos de modelos são comumente utilizados nas aréas de física, biologia e ciências sociais, sendo mais adequados em estudos em que os dados apresentam medições repetidas sobre as mesmas unidades individuais (estudos longitudinais), já que estes modelos de regressão consideram dados latentes (SCHAFER; YUCEL, 2002), além de tambem considerarem uma estrutura de correlação intra-indivíduos (em termos da estrutura da dependência de erros) e a correlação entre-indivíduos (em termos do efeito aleatório).

Os modelos de regressão com efeitos mistos para dados longitudinais são definidos da seguinte forma:

$$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i + \epsilon_i, \tag{1.2}$$

em que $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$ é um vetor $n_i \times 1$ de respostas observadas contínuas, para o indivíduo i, avaliado nos pontos de tempo específicos $\mathbf{t}_i = (t_{i1}, \ldots, t_{n_i})^\top$; $\mathbf{X}_i$ é a matriz de desenho $n_i \times l$ correspondente ao vetor $l \times 1$ de efeitos fixos β; $\mathbf{Z}_i$ é a matriz de desenho $n_i \times q$ correspondente ao vetor $q \times 1$ de efeitos aleatórios $\mathbf{b}_i$ e ϵ_i é o vetor $n_i \times 1$ de erros aleatórios. $\mathbf{b}_i \overset{iid}{\sim} \mathrm{N}_q(\mathbf{0_q}, \mathbf{D})$ é independente de $\epsilon_i \overset{ind.}{\sim} \mathrm{N}_{n_i}(\mathbf{0_{n_i}}, \Omega_i)$, $\mathbf{0}_q$ representa um vetor

$q \times 1$ de zeros e $\mathbf{0}_{n_i}$ representa um vetor $n_i \times 1$ de zeros. A matriz de dispersão $\mathbf{D} = \mathbf{D}(\alpha)$ depende dos parâmetros desconhecidos α e Ω_i é a matriz de covariância do erro.

1.4.2 Censored Data

Pode-se definir como censura, a observação parcial de uma variável, ou seja, que para algumas unidades da amostra, os dados da variável resposta não estão completamente disponíveis, no entanto são conhecidos os valores das covariáveis para estas mesmas unidades. Esta limitação para a observação da variável pode ocorrer por diferentes razões, como por exemplo, a limitação de medição de uma ferramenta, erros de calibração, entre outros (EFRON, 1967). Existem três tipos de censura, definidos a seguir:
Seja y_t e v_t o valor real e valor observado, respectivamente, da uma variável de interesse no tempo t, e c_t um indicador de censura, denotado por:

$$c_t = \begin{cases} 1, & se\ for\ censurada \\ 0, & caso\ contrario. \end{cases}$$

Os tipo de censura são:

- Censura à esquerda: Ocorre quando o valor real da variável é menor do que o valor observado, o qual é denotado por:

$$\begin{cases} y_t \leq v_t & se \quad c_t = 1 \\ y_t = v_t & se \quad c_t = 0. \end{cases}$$

- Censura à direita: ocorre quando o valor real da variável resposta é maior do que o valor observado, o qual é denotado por:

$$\begin{cases} y_t \geq v_t & se \quad c_t = 1 \\ y_t = v_t & se \quad c_t = 0. \end{cases}$$

- Censura intervalar: ocorre quando só é possível observar um intervalo finito do tipo $[v_{t_l}, v_{t_u}]$, do valor real da variável, o qual é denotado por:

$$\begin{cases} v_{t_l} \leq y_t \leq v_{t_u} & se \quad c_t = 1 \\ y_t \leq v_{t_l} \ \text{ ou } \ y_t \geq v_{t_u} & se \quad c_t = 0. \end{cases}$$

1.4.3 EM/ECM Algorithm

Proposto por Dempster, Laird e Rubin (1977), o algoritmo Expectation-Maximization (EM) foi desenvolvido como uma ferramenta para obter as estimativas de máxima verossimilhança dos parâmetros θ, por meio de um algoritmo iterativo. Este algoritmo foi proposto como uma alternativa em caso nos quais o processo de maximização se torna complexo pelo comportamento da função de log-verossimilhança. De forma geral o algoritmo EM consiste em incluir uma ou mais variáveis latentes, obtendo assim uma função de log-verosimilhança aumentada para os parâmetros θ. Esta função é conhecida como função de log-verosimilhança completa. A seguir descreveremos os passos realizados em cada iteração do algoritmo EM.

Seja $y_{comp} = (y^{lat}; y^{obs})$ o vetor de dados completo, em que y^{lat} representa os dados faltantes, y^{obs} representa os dados observados, $l_c(\theta|y_{comp})$ representa a função de log-verossimilhança completa e θ os parâmetros da função de log-verossimilhança. Para a k-ésima iteração, serão realizados dois passos.

Passo-E (Expectation/Esperança): Seja $\widehat{\theta}^{(k)}$ a estimativa dos parâmetros θ na k-ésima iteração, podemos calcular a esperança da função de log-verosimilhança completa dado os valores observados, denotada por $\widehat{Q_k}(\theta)$, para a iteração k como,

$$\widehat{Q_k}(\theta) = E\left[l_c(\theta|y_{comp})|y^{obs}, \widehat{\theta}^{(k)}\right].$$

Passo-M (Maximization/Maximização): Maximizar $\widehat{Q_k}(\theta)$ com respeito a θ, obtendo $\widehat{\theta}^{(k+1)}$

Sob condições de regularidade apropiadas, a sequência de $\widehat{\theta}^{(k)}$, converge para o estimador de máxima verossimilhança (DEMPSTER; LAIRD; RUBIN, 1977). No entanto quando a maximização no passo M segue sendo complexa, como foi proposto por Meng e Rubin (1993), pode-se substituir o passo M por uma série de passos de maximizações condicionais computacionalmente mais simples. Cada maximização condicional é construída para ser um problema simples de otimização com restrições para os parâmetros θ que estão sendo estimados. Esta classe de algoritmo EM generalizado, é conhecido como algoritmo ECM (TESCHE *et al.*, 1997), *Espectation Condicional Maximization*. Este algoritmo conserva todas as propriedades de convergência do algoritmo EM (TESCHE *et al.*, 1997).

1.5 ORGANIZATION OF THE DISSERTATION

Esta dissertação está organizada em quatro capítulos. No Capítulo 2, apresentaremos os modelos AR(p)-LMEC/NLMEC, as estimações dos parâmetros por máxima verossimilhança (MV), via algoritmo ECM. Discutiremos como obter os erros padrão e a estratégia previsão. Para concluir esse capítulo, avaliaremos o desempenho dos métodos propostos por meio de estudos de simulação e análise de um conjunto de dados reais. No Capítulo 3, apresentaremos o pacote ARpLMEC, elaborado com a metodologia proposta no capítulo anterior, descrevendo as funções e proporcionando exemplos para uma melhor compreensão. Finalmente no Capítulo 4, serão apresentadas algumas considerações finais, a produção técnica como resultado desta dissertação e algumas ideias para futuras pesquisas.

2 AUTOREGRESSIVE CENSORED MIXED-EFFECTS MODELS

Neste capítulo formularemos o modelo AR(p)-LMEC, para o caso linear e não linear, desenvolvendo cálculos para obter os estimadores dos parâmetros por máxima verossimilhança via algoritmo ECM, obtendo os erros padrões e a predição de valores futuros. Finalizaremos avaliando o modelo por meio de estudos de simulação e aplicação em dados reais.

2.1 INTRODUCTION

The most popular analytic tools for longitudinal data analysis with continuous outcomes are the linear and nonlinear mixed-effects models (LME/NLME). However, in such longitudinal studies, such as those on acquired immunodeficiency syndrome (AIDS) and environmental pollution, some variables may have certain threshold values below or above which the measurements are not quantifiable. For instance, viral load measures the amount of actively replicating virus and, depending on the diagnostic assays used, its measurement over time may be subject to detection limits, below or above which they are not quantifiable. Linear and nonlinear mixed-effects models, with modifications to accommodate censored observations (LMEC/NLMEC), have been proposed to fit this kind of data. In this context, Vaida e Liu (2009) proposed an exact EM-type algorithm for LMEC/NLMEC models that uses closed-form expressions at the E-step as opposed to the Monte Carlo EM algorithm, proposed by Hughes (1999) and Vaida, Fitzgerald e DeGruttola (2007). Matos *et al.* (2013a) provided additional tools, including influence diagnostics, for LMEC/NLMEC models. In the context of heavy–tailed LMEC/NLMEC, Lachos, Bandyopadhyay e Dey (2011) advocated the use of the normal/independent (NI) class of distributions, proposed by Liu (1996), and adopted a Bayesian framework to carry out posterior inference. Recently, Matos *et al.* (2013b) and Matos *et al.* (2015) proposed a likelihood-based estimation and influence analysis for Student-t LMEC/NLMEC models, respectively. A common feature of these classes of LMEC/NLMEC models is to assume that the correlation structure is induced just by the random effects term. However, in longitudinal studies the repeated measures of each subject are collected over time and hence the random errors tend also to be serially correlated.

Recently, some alternatives for modeling the correlation observation responses

and correlations induced by longitudinal data have been proposed in the statistical literature. These proposals consider not only measurements of some variables may have be subjected to certain threshold values below or above which the measurements are not quantifiable. the correlation structure induced by the random effects term, but also by other types of correlation in the error term. Particularly, Wang (2013) introduced the Student-t LME model for outcome variables recorded on irregular occasions, considering a damping exponential correlation (DEC) structure, as proposed by Muñoz *et al.* (1992). This correlation structure takes into account the autocorrelation generated by the within-subject dependence among irregular occasions and has as a special case the AR(1) correlation structure. Matos, Castro e Lachos (2016) consider the LMEC/NLMEC model with DEC structure, where an EM-type algorithm is also developed for computing the maximum likelihood (ML) estimates. On the other hand, Wang e Fan (2011) consider the Student-t LME model, with autoregressive dependence structure, of order p (AR(p)), for the within-subject errors.

Even though, some proposal have been made to deal with the problem of serial correlation among the observations in LMEC/NLMEC models, to the best of our knowledge, there are no studies of the LMEC/NLMEC with AR(p) errors. Thus, in this paper we develop a full likelihood-based approach for LMEC/NLMEC modeling with AR(p) errors, hereafter AR(p)-LMEC/NLMEC, including the implementation of a computationally efficient estimation method, via the EM algorithm, with the likelihood function, predictions of unobservable values of the response and the asymptotic standard errors as byproducts. The results developed here are an extension to those presented by Vaida e Liu (2009) and Matos, Castro e Lachos (2016) for the analysis of mixed-effects models with censored responses and HIV data.

2.2 MODEL FORMULATION

The autoregressive censored linear mixed-effects model (AR(p)-LMEC) is defined by:

$$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i + \epsilon_i, \tag{2.1}$$

where $\mathbf{y}$, $\mathbf{X}_i$, β, $\mathbf{Z}_i$, $\mathbf{b}_i$ and ϵ_i are as defined in Section (1.4.1) and the correlation

structure of the error vector is assumed to be $\Omega_i = \sigma^2 M_{n_i}(\phi)$, where the $n_i \times n_i$ matrix $M_{n_i}(\phi)$ incorporates a time-dependence structure. In this paper, we consider $M_{n_i}(\phi)$ with a structured $AR(p)$ dependence matrix for the within-subject errors, denoted by AR(p)-LME model. Specifically,

$$M_{n_i}(\phi) = [\rho_{|r-s|}(\phi)],$$

where $r, s = 1, \ldots, n_i$ and $\rho'_k s$ is the autocorrelation function and can be written as a function of the parameters $\phi = (\phi_1, \ldots, \phi_p)^\top$ and satisfy the Yule–Walker equation Box *et al.* (2015), i.e.,

$$\rho_k(\phi) = \phi_1 \rho_{k-1} + \ldots + \phi_p \rho_{k-p},\ \rho_o = 1,\ (k = 1, \ldots, n_i - 1).$$

In addition, the roots of $1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p = 0$ must lie outside the unit circle to ensure the stationarity of the model. For the pure AR model, admissible values of ϕ are limited in a p-dimensional hypercube $\mathbb{C}_p$.

In order to simplify the estimation procedure and ensure the admissibility of ϕ, we follow Barndorff-Nielsen e Schou (1973), to reparameterize ϕ as

$$\begin{aligned} \phi_p^{(p)} &= \pi_p, \\ \phi_v^{(p)} &= \phi_v^{(p-1)} - \pi_p \phi_{p-v}^{(p-1)}, \end{aligned} \tag{2.2}$$

where $\phi_v^{(p)}$ is the v-th AR parameter in the AR(p) model given by Equation (2.1), and $\pi_v = \phi_v^{(v)}$ represents the partial autocorrelation function, of lag v, for $v = 1, \ldots, p-1$. With this notation, the matrix $M_{n_i}(\phi)$ can be written as:

$$M_{n_i}(\phi) = \begin{bmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{n_i-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n_i-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{n_i-1} & \gamma_{n_i-2} & \cdots & \gamma_0 \end{bmatrix}, \text{ for } i = 1, \ldots, n.$$

The recursion given in Equation (2.2), can be used to define a transformation

$$\mathcal{B} : \gamma = (\pi_1, \ldots, \pi_p)^\top \to \phi = (\phi_1, \ldots, \phi_p)^\top \tag{2.3}$$

which is one-to-one, continuous and differentiable inside the admissible region. This parameterization, has the advantage that, in the γ-space the admissible region is simply the p-dimensional cube with boundary surfaces corresponding to ± 1, while in

the ϕ-space it is very complicated. As an illustration, for $p=2$ the transformation is $\phi_1=\pi_1(1-\pi_2)$ and $\phi_2=\pi_2$. For $p=3$, it can be written as $\phi_1=\pi_1(1-\pi_2)-\pi_2\pi_3$, $\phi_2=\pi_2(1+\pi_1\pi_3)-\pi_1\pi_3$ and $\phi_3=\pi_3$ (SCHUMACHER; LACHOS; DEY, 2017).

As mentioned above, the proposed model also considers censored observations (for more details see defined in Section (1.4.2)), *i.e.*, we assume that the response Y_{ij} is not fully observed for all i,j. Let $(\mathbf{V}_i^\top,\mathbf{C}_i^\top)^\top$ be the observed data for the i-th subject, where $\mathbf{V}_i$ represents the vector of uncensored readings or censoring level and $\mathbf{C}_i$ is the vector of censoring indicators, such that:

$$\begin{aligned} y_{ij} &\leq v_{ij} \quad \text{if } c_{ij}=1, \\ y_{ij} &= v_{ij} \quad \text{if } c_{ij}=0. \end{aligned} \tag{2.4}$$

In fact, the right-censored problem can be represented by a left-censored problem by simultaneously transforming the response y_{ij} and censoring level v_{ij} to $-y_{ij}$ and $-v_{ij}$. The model defined in Equations (2.1)-(2.4), is henceforth called **AR(p)-LMEC** model.

2.2.1 The log-likelihood function

Following Vaida e Liu (2009), classic inference on the parameter vector $\theta=(\beta^\top,\sigma^2,\alpha^\top,\phi^\top)^\top$ is based on the marginal distribution of $\mathbf{y}_i$. For complete data, the marginal distribution of the vector $\mathbf{y}_i$, follows a multivariate normal distribution $\mathrm{N}_{n_i}(\mathbf{X}_i\beta,\Sigma_i)$, where $\Sigma_i=\Omega_i+\mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top$, for $i=1,\ldots,n$. The strategy to compute the likelihood function associated with the AR(p)-LMEC model, defined by Equations (2.1)-(2.4), is to treat separately the observed and censored components of $\mathbf{y}_i$.

Let $\mathbf{y}_i^o$ be the n_i^o-vector of observed outcomes and $\mathbf{y}_i^c$ be the n_i^c-vector of censored observations for subject i $(n_i=n_i^o+n_i^c)$, such that $C_{ij}=0$ for all elements in $\mathbf{y}_i^o$, and $C_{ij}=1$ for all elements in $\mathbf{y}_i^c$. After reordering, $\mathbf{y}_i$, $\mathbf{V}_i$, $\mathbf{X}_i$ and Σ_i can be partitioned as follows:

$$\mathbf{y}_i=vec(\mathbf{y}_i^o,\mathbf{y}_i^c),\ \mathbf{V}_i=vec(\mathbf{V}_i^o,\mathbf{V}_i^c),\ \mathbf{X}_i^\top=(\mathbf{X}_i^o,\mathbf{X}_i^c)\ \text{ and}$$

$$\Sigma_i=\begin{pmatrix} \Sigma_i^{oo} & \Sigma_i^{oc} \\ \Sigma_i^{co} & \Sigma_i^{cc} \end{pmatrix}.$$

In this setup, the operator $vec(\cdot)$ denotes the function with stack vectors or matrices of the same number of columns. Consequently, from the marginal-conditional decomposition of the multivariate normal distribution, $\mathbf{y}_i^o \sim \mathrm{N}_{n_i^o}(\mathbf{X}_i^o\beta, \Sigma_i^{oo})$ and $\mathbf{y}_i^c \mid \mathbf{y}_i^o \sim \mathrm{N}_{n_i^c}(\mu_i, \mathbf{S}_i)$, where $\mu_i = \mathbf{X}_i^c\beta + \Sigma_i^{co}(\Sigma_i^{oo})^{-1}(\mathbf{y}_i^o - \mathbf{X}_i^o\beta)$ and $\mathbf{S}_i = \Sigma_i^{cc} - \Sigma_i^{co}(\Sigma_i^{oo})^{-1}\Sigma_i^{oc}$. Now, let $\Phi_{n_i}(\mathbf{u};\mathbf{a},\mathbf{A})$ and $\phi_{n_i}(\mathbf{u};\mathbf{a},\mathbf{A})$ be the *cdf* (left tail) and *pdf*, respectively, of the multivariate normal distribution $\mathrm{N}_{n_i}(\mathbf{a},\mathbf{A})$, computed at vector $\mathbf{u}$. From Vaida e Liu (2009) and Matos *et al.* (2013a), the likelihood function for subject i (using conditional probability arguments) is given by

$$\begin{aligned} L_i(\theta) &= f(\mathbf{y}_i^c \le \mathbf{V}_i^c \mid \mathbf{y}_i^o = \mathbf{V}_i^o, \theta) f(\mathbf{y}_i^o = \mathbf{V}_i^o \mid \theta) \\ &= f(\mathbf{y}_i^c \le \mathbf{V}_i^c \mid \mathbf{y}_i^o, \theta) f(\mathbf{y}_i^o \mid \theta) \\ &= \Phi_{n_i^c}(\mathbf{V}_i^c; \mu_i, \mathbf{S}_i)\phi_{n_i^o}(\mathbf{y}_i^o; \mathbf{X}_i^o\beta, \Sigma_i^{oo}), \end{aligned} \tag{2.5}$$

which would be easily evaluated computationally.

The log-likelihood function for the observed data, given by $\ell(\theta|\mathbf{y}) = \sum_{i=1}^n \log L_i(\theta)$, is used to compute different model selection criteria, such as AIC, AICc (HURVICH; TSAI, 1989) and BIC defineds by:

$$\mathrm{AIC} = 2m - 2\ell_{max}\,,\ \mathrm{AICc} = 2m - 2\ell_{max}\ + 2\frac{m(m+1)}{N-m-1} \text{ and}$$

$$\mathrm{BIC} = m\,log\,N - 2\ell_{max}\ ,$$

where m is the number of the model parameters, $N = \sum_{i=1}^n n_i$ and ℓ_{max} is the maximized log-likelihood value and $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$.

2.3 ML ESTIMATION VIA THE ECM ALGORITHM

In this section, we describe in detail, how the parameters of the proposed model specified in Equations (2.1)-(2.4) can be fitted by using the ECM algorithm (MENG; RUBIN, 1993) (for more details see defined in Section (1.4.3))).

Let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top)^\top$, $\mathbf{V} = vec(\mathbf{V}_1, \ldots, \mathbf{V}_n)$ and $\mathbf{C} = vec(\mathbf{C}_1, \ldots, \mathbf{C}_n)$. Considering $\mathbf{b}$ as the hypothetical missing data, the complete data are denoted by

$$\mathbf{y}_c = (\mathbf{C}^\top, \mathbf{V}^\top, \mathbf{y}^\top, \mathbf{b}^\top)^\top.$$

Hence, the ECM algorithm is applied to the complete data log-likelihood function

$$\begin{aligned}\ell_{c_i}(\theta \mid \mathbf{y}_c) &= -\frac{1}{2}\Big[n_i \log\sigma^2 + \log(|M_{n_i}(\phi)|) \\ &+ \frac{1}{\sigma^2}(\mathbf{y}_i - \mathbf{X}_i\beta - \mathbf{Z}_i\mathbf{b}_i)^\top M_{n_i}^{-1}(\phi)(\mathbf{y}_i - \mathbf{X}_i\beta - \mathbf{Z}_i\mathbf{b}_i) \\ &+ \log|\mathbf{D}| + \mathbf{b}_i^\top \mathbf{D}^{-1}\mathbf{b}_i\Big] + K, \end{aligned} \tag{2.6}$$

with K being a constant that does not depend on the parameter vector θ. Given the current estimate $\theta = \widehat{\theta}^{(k)}$, the E-step calculates the conditional expectation of the complete data log-likelihood function, given by

$$\begin{aligned} Q\left(\theta \mid \widehat{\theta}^{(k)}\right) &= E\left(\ell_c(\theta \mid \mathbf{y}_c) \mid \mathbf{V}, \mathbf{C}, \widehat{\theta}^{(k)}\right) \\ &= \sum_{i=1}^{n} Q_{1i}\left(\beta, \sigma^2, \phi \mid \widehat{\theta}^{(k)}\right) + \sum_{i=1}^{n} Q_{2i}\left(\alpha \mid \widehat{\theta}^{(k)}\right), \end{aligned}$$

where

$$\begin{aligned} Q_{1i}\left(\beta, \sigma^2, \phi \mid \widehat{\theta}^{(k)}\right) &= -\frac{n_i}{2}\log\sigma^{2(k)} - \frac{1}{2}\log(|M_{n_i}^{(k)}(\phi)|). \\ &- \frac{1}{2\sigma^{2(k)}}\left[\widehat{a}_i^{(k)}(\phi^{(k)}) - 2\beta^{(k)\top}\mathbf{X}_i^\top M_{n_i}^{-1(k)}(\phi)\left(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\,\widehat{\mathbf{b}}_i^{(k)}\right)\right] \\ &+ \frac{1}{2\sigma^{2(k)}}\left[\beta^{(k)\top}\mathbf{X}_i^\top M_{n_i}^{-1(k)}(\phi)\mathbf{X}_i\beta^{(k)}\right], \end{aligned} \tag{2.7}$$

$$Q_{2i}\left(\alpha \mid \widehat{\theta}^{(k)}\right) = -\frac{1}{2}\log|\mathbf{D}^{(k)}| - \frac{1}{2}tr\left(\widehat{\mathbf{b}_i\mathbf{b}_i^\top}^{(k)}\mathbf{D}^{-1(k)}\right), \tag{2.8}$$

with $\widehat{a}_i^{(k)}(\phi) = tr\left(\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} M_{n_i}^{-1}(\phi) - 2\widehat{\mathbf{y}_i\mathbf{b}_i^\top}^{(k)}\mathbf{Z}_i^\top M_{n_i}^{-1}(\phi) + \widehat{\mathbf{b}_i\mathbf{b}_i^\top}^{(k)}\mathbf{Z}_i^\top M_{n_i}^{-1}(\phi)\mathbf{Z}_i\right)$, and

$$\begin{aligned} \widehat{\mathbf{b}_i}^{(k)} &= E\left(\mathbf{b}_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\theta}^{(k)}\right) = \widehat{\varphi}_i^{(k)}(\widehat{\mathbf{y}_i}^{(k)} - \mathbf{X}_i\widehat{\beta}^{(k)}), \\ \widehat{\mathbf{b}_i\mathbf{b}_i^\top}^{(k)} &= E\left(\mathbf{b}_i\mathbf{b}_i^\top \mid \mathbf{V}_i, \mathbf{C}_i,\ \widehat{\theta}^{(k)}\right) \\ &= \widehat{\Lambda}_i^{(k)} + \widehat{\varphi}_i^{(k)}(\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} - \widehat{\mathbf{y}_i}^{(k)}\widehat{\beta}^{(k)\top}\mathbf{X}_i^\top - \mathbf{X}_i\widehat{\beta}^{(k)}\widehat{\mathbf{y}_i}^{(k)\top} + \mathbf{X}_i\widehat{\beta}^{(k)}\widehat{\beta}^{(k)\top}\mathbf{X}_i^\top)\widehat{\varphi}_i^{(k)\top}, \\ \widehat{\mathbf{y}_i\mathbf{b}_i^\top}^{(k)} &= E\left(\mathbf{y}_i\mathbf{b}_i^\top \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\theta}^{(k)}\right) = (\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} - \widehat{\mathbf{y}_i}^{(k)}\widehat{\beta}^{(k)\top}\mathbf{X}_i^\top)\widehat{\varphi}_i^{(k)\top}, \end{aligned}$$

with $\widehat{\Lambda}_i^{(k)} = (\widehat{\mathbf{D}}^{-1(k)} + \mathbf{Z}_i^\top \widehat{M}_{n_i}^{-1(k)}(\phi)\mathbf{Z}_i/\widehat{\sigma^2}^{(k)})^{-1}$, $\widehat{\varphi}_i^{(k)} = \widehat{\Lambda}_i^{(k)}\mathbf{Z}_i^\top \widehat{M}_{n_i}^{-1(k)}(\phi)/\widehat{\sigma^2}^{(k)}$ and $tr(\mathbf{A})$ being the trace function of the matrix $\mathbf{A}$

It is easy to see, from Equations (2.7) and (2.8), that the E-step reduces only to the computation of

$$\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} = E\left(\mathbf{y}_i\mathbf{y}_i^\top \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\theta}^{(k)}\right) \text{ and } \widehat{\mathbf{y}_i}^{(k)} = E\left(\mathbf{y}_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\theta}^{(k)}\right).$$

These conditional expectations rely on the first and second moments of a multivariate truncated normal distribution and can be determined in closed-form. For more details on the computation of these moments, see Vaida e Liu (2009) and Matos *et al.* (2013a).

The conditional maximization step (CM) conditionally maximizes $Q(\theta|\widehat{\theta}^{(k)})$ function, with respect to θ, obtaining a new estimate $\widehat{\theta}^{(k+1)}$, as follow

$$\begin{aligned}
\widehat{\beta}^{(k+1)} &= \left(\sum_{i=1}^{n}\mathbf{X}_i^\top \widehat{M}_{n_i}^{-1(k)}(\phi)\mathbf{X}_i\right)^{-1}\sum_{i=1}^{n}\mathbf{X}_i^\top \widehat{M}_{n_i}^{-1(k)}(\phi)\left(\widehat{\mathbf{y}_i}^{(k)} - \mathbf{Z}_i\widehat{\mathbf{b}_i}^{(k)}\right),\\
\widehat{\sigma^2}^{(k+1)} &= \frac{1}{N}\sum_{i=1}^{n}\left[\widehat{a}_i^{(k)} - 2\widehat{\beta}^{(k+1)\top}\mathbf{X}_i^\top \widehat{M}_{n_i}^{-1(k)}(\phi)\left(\widehat{\mathbf{y}_i}^{(k)} - \mathbf{Z}_i\widehat{\mathbf{b}_i}^{(k)}\right)\right.\\
&+ \left.\widehat{\beta}^{(k+1)\top}\mathbf{X}_i^\top \widehat{M}_{n_i}^{-1(k)}(\phi)\mathbf{X}_i\widehat{\beta}^{(k+1)}\right],\\
\widehat{\mathbf{D}}^{(k+1)} &= \frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{b}_i\mathbf{b}_i^\top}^{(k)},\\
\widehat{\gamma}^{(k+1)} &= \underset{\gamma\in(-1,1)^p}{\operatorname{argmax}}\left(-\frac{1}{2}\log(|M_{n_i}(\phi)|) - \frac{1}{2\widehat{\sigma^2}^{(k+1)}}\left[\widehat{a}_i^{(k)}(\phi) + \widehat{\beta}^{(k+1)\top}\mathbf{X}_i^\top M_{n_i}^{-1}(\phi)\mathbf{X}_i\widehat{\beta}^{(k+1)}\right.\right.\\
&- \left.\left.2\widehat{\beta}^{(k+1)\top}\mathbf{X}_i^\top M_{n_i}^{-1}(\phi)\left(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{\mathbf{b}}_i^{(k)}\right)\right]\right),\\
\hat{\phi}^{(k+1)} &= \mathcal{B}(\hat{\gamma}^{(k+1)}),
\end{aligned}$$

where $\mathcal{B}$ is the transformation defined in the equation (2.3). This process is iterated until some distance between two successive parameter estimations, such as $||\theta^{(k+1)} - \theta^{(k)}||$, becomes small enough, where $||\theta||$ denotes the Euclidean norm of the vector θ. The initial values can be calculated by taking the censored values to be observed ones and proceeding as an usual LME model.

2.3.1 Estimation of random effects

To estimate the random effects, we consider the conditional mean of $\mathbf{b}_i$ given the observed data $\mathbf{V}_i$ and $\mathbf{C}_i$, that is, $E(\mathbf{b}_i \mid \mathbf{V}_i, \mathbf{C}_i)$. Thus, for a given value of $\theta = (\beta^\top, \sigma^2, \alpha^\top, \phi^\top)^\top$, the conditional mean of $\mathbf{b}_i$ given $\mathbf{V}_i$ and $\mathbf{C}_i$ is

$$\widehat{\mathbf{b}}_i(\theta) = E(\mathbf{b}_i \mid \mathbf{V}_i, \mathbf{C}_i) = \varphi_i(\widehat{\mathbf{y}}_i - \mathbf{X}_i\beta), \tag{2.9}$$

where $\varphi_i = \Lambda_i \mathbf{Z}_i^\top M_{n_i}^{-1}(\phi)/\sigma^2$ and $\Lambda_i = (\mathbf{D}^{-1} + \mathbf{Z}_i^\top M_{n_i}^{-1}(\phi)\mathbf{Z}_i/\sigma^2)^{-1}$. Note that $\widehat{\mathbf{y}}_i = E\left(\mathbf{y}_i | Q_{1i}, Q_{2i}, \mathbf{C}_i\right)$, with $Q_{1i} = Q_{1i}(\beta, \sigma^2, \phi \mid \widehat{\theta}^{(k)})$ and $Q_{2i} = Q_{2i}(\alpha \mid \widehat{\theta}^{(k)})$ as diffined in 2.7 and 2.8, is given by the first moment of a multivariate truncated normal distribution. In practice, the estimator of $\mathbf{b}_i$, $\widehat{\mathbf{b}}_i$, can be obtained by substituting the ML estimates $\widehat{\theta}$ into Equation (2.9), leading to $\widehat{\mathbf{b}}_i = \widehat{\mathbf{b}}_i(\widehat{\theta})$. On the other hand, the conditional covariance matrix of $\mathbf{b}_i$, given $\mathbf{V}_i$ and $\mathbf{C}_i$, is

$$Var\left(\mathbf{b}_i \mid Q_{1i}, Q_{2i}, \mathbf{C}_i\right) = E\left(\mathbf{b}_i \mathbf{b}_i^\top \mid Q_{1i}, Q_{2i}, \mathbf{C}_i\right) - \widehat{\mathbf{b}}_i(\theta)\widehat{\mathbf{b}}_i(\theta)^\top = \Lambda_i + \varphi_i Var\left(\mathbf{y}_i \mid \mathbf{V}_i, \mathbf{C}_i\right)\varphi_i^\top.$$

Note that $Var(\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i)$ can be easily obtained as a byproduct of the proposed ECM algorithm developed previously.

2.4 STANDARD ERROR

Following Lin (2010), we compute the asymptotic covariance of the ML estimates, through the empirical information matrix ($\mathbf{IM}_e$), which is computed as Meilijson (1989)

$$\mathbf{IM}_e(\theta \mid \mathbf{y}) = \sum_{i=1}^{n} \mathbf{s}(\mathbf{y}_i \mid \theta)\mathbf{s}^\top(\mathbf{y}_i \mid \theta) - \frac{1}{n}\mathbf{S}(\mathbf{y}_i \mid \theta)\mathbf{S}^\top(\mathbf{y}_i \mid \theta), \tag{2.10}$$

where $\mathbf{S}(\mathbf{y}_i \mid \theta) = \sum_{i=1}^{n} \mathbf{s}(\mathbf{y}_i \mid \theta)$ and $\mathbf{s}(\mathbf{y}_i \mid \theta)$ is the empirical score function for subject 'i'. According to Louis (1982), it is possible to relate the score function of the incomplete data log-likelihood, with the conditional expectation of the complete data log-likelihood function. Therefore, the individual score can be determined as

$$\mathbf{s}(\mathbf{y}_i \mid \theta) = \frac{\partial \log f(\mathbf{y}_i \mid \theta)}{\partial \theta} = E\left(\frac{\partial \ell_{c_i}(\theta \mid \mathbf{y}_c)}{\partial \theta} \mid \mathbf{V}_i, \mathbf{C}_i, \theta\right), \tag{2.11}$$

where $\ell_{c_i}(\theta \mid \mathbf{y}_c)$ is the complete data log-likelihood, formed from the observation 'i'. Using the ML estimates $\widehat{\theta}$, that is, $\mathbf{S}(\mathbf{y}_i \mid \widehat{\theta}) = 0$, it follows that Equation (2.10), can be approximated by

$$\mathbf{IM}_e(\widehat{\theta} \mid \mathbf{y}) = \sum_{i=1}^{n} \widehat{\mathbf{s}}_i \widehat{\mathbf{s}}_i^\top, \tag{2.12}$$

where $\widehat{\mathbf{s}}_i = \mathbf{s}(\mathbf{y}_i \mid \widehat{\theta}) = \left(\widehat{\mathbf{s}}_{i,\beta}^\top, \widehat{\mathbf{s}}_{i,\sigma^2}, \widehat{\mathbf{s}}_{i,\alpha}^\top, \widehat{\mathbf{s}}_{i,\phi}^\top\right)^\top$ has elements given by

$$\widehat{\mathbf{s}}_{i,\beta} = E\left(\frac{\partial \ell_{c_i}(\theta \mid \mathbf{y}_c)}{\partial \beta} \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\theta}\right) = \frac{1}{\widehat{\sigma^2}}\left[\mathbf{X}_i^\top \widehat{M}_{n_i}(\phi)^{-1}\left(\widehat{\mathbf{y}_i} - \mathbf{Z}_i\widehat{\mathbf{b}_i}\right) - \mathbf{X}_i^\top \widehat{M}_{n_i}(\phi)^{-1}\mathbf{X}_i\widehat{\beta}\right],$$

$$\begin{aligned}\widehat{\mathbf{s}}_{i,\sigma^2} &= E\left(\frac{\partial \ell_{c_i}(\theta \mid \mathbf{y}_c)}{\partial \sigma^2} \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\theta}\right) \\ &= -\frac{n_i}{2\widehat{\sigma^2}} + \frac{1}{2\widehat{\sigma^4}}\left[\widehat{h}_i - 2\widehat{\beta}^\top \mathbf{X}_i^\top \widehat{M}_{n_i}(\phi)^{-1}\left(\widehat{\mathbf{y}}_i - \mathbf{Z}_i\widehat{\mathbf{b}}_i\right) + \widehat{\beta}^\top \mathbf{X}_i^\top \widehat{M}_{n_i}(\phi)^{-1}\mathbf{X}_i\widehat{\beta}\right],\end{aligned}$$

$$\widehat{\mathbf{s}}_{i,\alpha} = E\left(\frac{\partial \ell_{c_i}(\theta \mid \mathbf{y}_c)}{\partial \alpha} \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\theta}\right) = -\frac{1}{2} tr\left(\widehat{D}^{-1}\frac{\partial D}{\partial \alpha}\widehat{D}^{-1}(\widehat{D} - \widehat{\mathbf{b}}_i\widehat{\mathbf{b}}_i^\top)\right),$$

with $\widehat{h}_i = tr\left(\widehat{\mathbf{y}_i\mathbf{y}_i^\top}\widehat{M}_{n_i}(\phi)^{-1} - 2\widehat{\mathbf{y}_i\mathbf{b}_i^\top}\mathbf{Z}_i^\top \widehat{M}_{n_i}(\phi)^{-1} + \widehat{\mathbf{b}_i\mathbf{b}_i^\top}\mathbf{Z}_i^\top \widehat{M}_{n_i}(\phi)^{-1}\mathbf{Z}_i\right)$.

Considering the reparametrization (2.2), the parts of the log-likelihood function that depends on ϕ can be written as $|\mathbf{M}_{n_i}(\phi)| = |\mathbf{M}_{p_i}(\phi)| = \prod_{j=1}^{p}(1-\gamma_j^2)^{-j}$ and $(\mathbf{y}_i - \mathbf{X}_i\beta - \mathbf{Z}_i\mathbf{b}_i)^\top M_{n_i}^{-1}(\phi)(\mathbf{y}_i - \mathbf{X}_i\beta - \mathbf{Z}_i\mathbf{b}_i) = \lambda^\top D(\mathbf{y}, \beta)\lambda$, with $\lambda^\top = (-1, \phi^\top) = (-1, \mathcal{B}(\gamma)^\top)$ and $D(\mathbf{y}, \beta)$ being the $(p+1)\times(p+1)$ matrix with the (r,s)-entry defined by

$$D_{r,s} = D_{s,r} = \mathbf{d}_{(r,s)} \cdot \mathbf{d}_{(s,r)}, \tag{2.13}$$

where the vector $\mathbf{d}_{(r,s)} = (e_r, \ldots, e_{n+1-s})$, such that $\mathbf{e} = (\mathbf{y}_i - \mathbf{X}_i^\top\beta - \mathbf{Z}_i\widehat{\mathbf{b}}_i)$ and $A \cdot B$ is the internal product of vector A and B.

Now to calculate the derive $\frac{\partial \ell_{c_i}(\theta|\mathbf{y}_c)}{\partial \phi}$, we partitioned $D = D(\mathbf{y}_c, \beta)$ as

$$D = \begin{bmatrix} D_{11} & D_{\phi 1}^\top \\ D_{\phi 1} & D_{\phi\phi} \end{bmatrix},$$

such that D_{11} is 1×1, $D_{\phi 1}$ is $p\times 1$ and $D_{\phi\phi}$ is $p\times p$. Then, the sum of squares in Equation (2.13) can be written as

$$\lambda^\top D\lambda = \begin{bmatrix} -1 & \phi^\top \end{bmatrix}\begin{bmatrix} D_{11} & D_{\phi 1}^\top \\ D_{\phi 1} & D_{\phi\phi} \end{bmatrix}\begin{bmatrix} -1 \\ \phi \end{bmatrix} = D_{11} - 2\phi^\top D_{\phi 1} + 2\phi^\top D_{\phi\phi}\phi.$$

Therefore, we have

$$\frac{\partial}{\partial \phi}\lambda^\top D\lambda = -2D_{\phi 1} + 2D_{\phi\phi}\phi.$$

Finally, the vector $\widehat{\mathbf{s}}_{i,\phi}$ will be defined as

$$\widehat{\mathbf{s}}_{i,\phi} = E\left(\frac{\partial \ell_{c_i}(\theta \mid \mathbf{y}_c)}{\partial \phi} \mid \mathbf{V}_i, \mathbf{C}_i, \theta\right) = \frac{1}{\widehat{\sigma^2}}(-D_{\phi 1} + D_{\phi\phi}\phi) + \frac{1}{2}\widehat{M}_{p_i}(\phi)^{-1}\frac{\partial M_{p_i}}{\partial \phi}.$$

2.5 PREDICTION OF FUTURE OBSERVATIONS

The problem related to the prediction of future values has a great impact on many practical applications. Rao (1987) pointed out that the predictive accuracy of future observations can be taken as an alternative measure of "goodness-of-fit". In order to propose a strategy to generate predicted values from the AR(p)-LMEC model, we use the approach proposed by Wang (2013). Thus, let $\mathbf{y}_{i,obs}$ be an observed response vector of dimension $n_{i,obs} \times 1$ for a new subject i over the first portion of time and $\mathbf{y}_{i,pred}$ be the corresponding $n_{i,pred} \times 1$ response vector over the future portion of time. Let $\tilde{\mathbf{X}}_i = (\mathbf{X}_{i,obs}, \mathbf{X}_{i,pred})$ be the $(n_{i,obs} + n_{i,pred}) \times p$ design matrix corresponding to $\tilde{\mathbf{y}}_i = (\mathbf{y}_{i,obs}^\top, \mathbf{y}_{i,pred}^\top)$.

To deal with the censored values existing in $\mathbf{y}_{i,obs}$, we use the imputation procedure, by replacing the censored values by $\widehat{\mathbf{y}_i} = E\{\mathbf{y}_i | \mathbf{V}_i, \mathbf{C}_i, \widehat{\theta}\}$ obtained from the EM algorithm. Therefore, when the censored values are imputed, a complete dataset, denoted by $\mathbf{y}_{i,obs^*}$, is obtained. The reason to use the imputation procedure is that it avoids computing truncated conditional expectations of the multivariate normal distribution originated by the censoring scheme. Hence, we have that

$$\tilde{\mathbf{y}}_i^* = \left(\mathbf{y}_{i,obs^*}^\top, \mathbf{y}_{i,pred}^\top\right)^\top \sim N_{n_{i,obs}+n_{i,pred}}\left(\mathbf{X}_i\beta, \Sigma_i\right),$$

where the matrix Σ_i, can be represented by $\Sigma_i = \begin{pmatrix} \Sigma_i^{obs^*,obs^*} & \Sigma_i^{obs^*,pred} \\ \Sigma_i^{pred,obs^*} & \Sigma_i^{pred,pred} \end{pmatrix}$. As mentioned in Wang (2013), the best linear predictor of $\mathbf{y}_{i,pred}$ with respect to the minimum mean squared error (MSE) criterion is the conditional expectation of $\mathbf{y}_{i,pred}$ given $\mathbf{y}_{i,obs^*}$, which is given by

$$\widehat{\mathbf{y}}_{i,pred}(\theta) = \mathbf{X}_{i,pred}\beta + \Sigma_i^{pred,obs^*}\Sigma_i^{obs^*,obs^*-1}\left(\mathbf{y}_{i,obs^*} - \mathbf{X}_{i,obs^*}\beta\right). \tag{2.14}$$

Therefore, $\mathbf{y}_{i,pred}$ can be estimated directly, by substituting $\widehat{\theta}$ into Equation (2.14), leading to $\widehat{\widehat{\mathbf{y}}_{i,pred}} = \widehat{\mathbf{y}}_{i,pred}(\widehat{\theta})$.

2.6 THE NONLINEAR CASE

As mentioned in the Introduction, some approximations based on the EM algorithm have been proposed in the statistical literature to deal with NLME models. In this context, we use an approximation of the nonlinear functions mentioned by Vaida e Liu (2009). This approximation (2.16) was considered by Matos *et al.* (2013a) in the context of

censored nonlinear mixed-effects models. In that paper, simulation studies revealed that the approximation can efficiently estimate the model parameters. Recently, Wang (2013) used this approximation to implement an ECM algorithm to carry out ML estimation in Student-t nonlinear mixed-effects models for multi-outcome longitudinal data with missing values. Consequently, we conclude that this approximation is robust, stable, and does not produce any severe consequences in inference when applied to other types of (censored) nonlinear models.

The NLME (without censoring) of Pinheiro e Bates (2000) is defined as

$$\mathbf{y}_i \;=\; \eta(\psi_i, \mathbf{X}_i) + \epsilon_i,\ \psi_i = \mathbf{A}_i\beta + B_i\mathbf{b}_i,\ i = 1,\ldots,n, \tag{2.15}$$

where $\mathbf{b}_i \overset{iid}{\sim} \mathrm{N}_q(0,\mathbf{D})$ and $\epsilon_i \overset{ind}{\sim} \mathrm{N}_{n_i}(0,\sigma^2\mathrm{E}_i)$ are independent; $\mathbf{y}_i$ is an $(n_i \times 1)$ vector of observed responses for subject i; η is a nonlinear function of the individual random parameter ψ_i; $\mathbf{A}_i$ and $\mathbf{B}_i$ are known design matrices of dimensions $r \times p$ and $r \times q$, respectively, possibly depending on some covariate values; and β is the $(p \times 1)$ vector of fixed effects and $\mathbf{b}_i$ is the $(q \times 1)$ vector of random effects.

As mentioned by Vaida e Liu (2009), the linearization (L) procedure to obtain the approximate MLE of $\theta = (\beta^\top, \sigma^2, \alpha^\top, \phi^\top)^\top$ involves taking the first-order Taylor expansion of η_i around the current parameter estimate $\widetilde{\beta}$ and the random effect estimates $\widetilde{\mathbf{b}}_i$ (empirical predictors). This procedure is equivalent to iteratively solving the following LME model (L-step)

$$\widetilde{\mathbf{Y}}_i = \widetilde{\mathbf{W}}_i\beta + \widetilde{\mathbf{H}}_i\mathbf{b}_i + \epsilon_i,\ i = 1,\ldots,n, \tag{2.16}$$

where $\mathbf{b}_i \overset{iid}{\sim} \mathrm{N}_q(0,\mathbf{D})$ and $\epsilon_i \overset{ind}{\sim} \mathrm{N}_{n_i}(\mathbf{0},\sigma^2\mathrm{E}_i)$; and $\widetilde{\mathbf{Y}}_i = \mathbf{Y}_i - \eta(\psi(\widetilde{\beta},\widetilde{\mathbf{b}}_i),\mathbf{X}_i)$, with

$$\widetilde{\mathbf{H}}_i = \frac{\partial\eta(\mathbf{A}_i\widetilde{\beta} + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_i)}{\partial\mathbf{b}_i^\top}\Big|_{\mathbf{b}_i=\widetilde{\mathbf{b}}_i} \quad \text{and} \quad \widetilde{\mathbf{W}}_i = \frac{\partial\eta(\mathbf{A}_i\beta + \mathbf{B}_i\widetilde{\mathbf{b}}_i, \mathbf{X}_i)}{\partial\beta_i^\top}\Big|_{\beta_i=\widetilde{\beta}_i}.$$

Thus, in the censored case, the model in (2.16) is an LME with censored data that can be fitted using the strategy explained in Section 2.3. The model matrices in (2.16) depend on the current parameter value, and need to be recalculated at each iteration. The algorithm iterates between the L-, E- and CM-steps until convergence.

2.7 SIMULATION STUDIES

In order to examine the performance of our proposed model, we develop three simulation studies, considering the linear and nonlinear cases: (i) the first simulation

study shows the asymptotic behavior of the parameter estimates, when the sample size is increasing; (ii) the goal of the second simulation is to study the consistency of the ML estimates and (iii) the third simulation shows the performance of the prediction of future values.

2.7.1 The Linear Case

For the linear case, we perform a simulation studies based on the AR(p)-LMEC model defined in (2.1)-(2.4), considering a left-censored setting with same number of measurements for each individual i, i.e., $n_i = 6$. As in Schumacher, Lachos e Dey (2017), we assume an autoregressive dependence for the error term of order $p = 2$ and covariate vector for each individual given by $\mathbf{X}_i = (\mathbf{1}_6,\ \mathbf{x}_{1i},\ \mathbf{x}_{2i})$, where each vector $\mathbf{x}_{ti}$ was generated independently, from a uniform distribution $U(-1,1)$ and $\mathbf{1}_{n_i}$ represents a $1 \times n_i$ vector of ones. We consider $\mathbf{Z}_i = (\mathbf{1}_6,\ z_{1i})$, with $z_{1i} = (1,\ 2,\ 3,\ 4,\ 5,\ 6)^\top$, as the covariate vector associated with the random effect. It is important to stress that, all these values were fixed throughout the replications.

For all simulation schemes, the values of the parameters were set at: $\beta = (\beta_1, \beta_2, \beta_3)^\top = (1, 2, 1)^\top$, $\sigma^2 = 0.48$ and $\phi = (\phi_1, \phi_2)^\top = (0.48, -0.2)^\top$. The random effect $\mathbf{b}_i = (b_{1i}, b_{2i})^\top$ was generated, previously, from a bivariate normal distribution $\mathrm{N}_2(\mathbf{0}_2, \mathbf{D})$, with:

$$\mathbf{D} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} = \begin{bmatrix} 0.049 & 0.001 \\ 0.001 & 0.002 \end{bmatrix},$$

as presented by Matos, Castro e Lachos (2016). We fix a left censoring level at $l\%$ (that is, $l\%$ of the observations in each data set were left censored, respectively).

- **Simulation Study 1:**

  We generated $R = 100$ datasetfor each one of the combinations, of different samples size $n = \{50,\ 100,\ 200,\ 400,\ 600\}$ and censoring levels $l\% = \{0\%,\ 5\%,\ 20\%,\ 40\%\}$. Considering our proposed EM algorithm, we compute the Bias and the mean squared error (MSE) for the parameter estimates $\widehat{\theta}_i$, over the $R = 100$ samples and twenty different situations. The measures are defined by:

$$Bias\left(\widehat{\theta}_i\right) = \frac{1}{R}\sum_{j=1}^{R}\left(\widehat{\theta}_i^{(j)} - \theta_i\right) \text{ and } MSE\left(\widehat{\theta}_i\right) = \frac{1}{R}\sum_{j=1}^{R}\left(\widehat{\theta}_i^{(j)} - \theta_i\right)^2,$$

where $\widehat{\theta}_i^{(j)}$ denotes the ML estimate of parameter θ_i, for the j-th replication, with $j = 1, \ldots, R$.

The results for the fixed parameters are shown in Figures 2 and Figure 4. As a general rule, we can observe that the Bias and MSE tend to zero when the sample size increases, indicating that the estimates based on our proposed EM-type algorithm have good asymptotic properties. The results for the random effect are shown in Figures 3 and Figure 5. Here is observed that the random effect present Bias and MSE tend to decrease when the sample size increases.

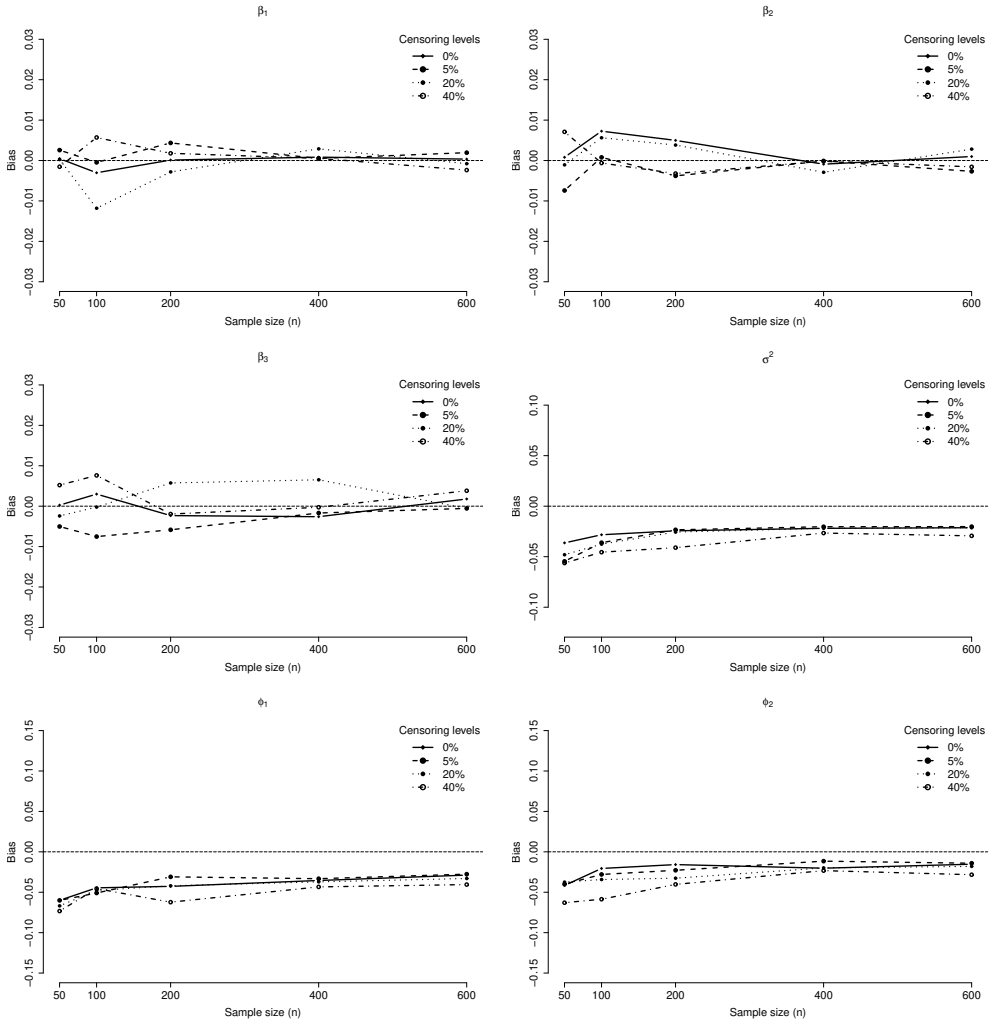

Figure 2 – Average Bias of parameter estimates of the AR(p)-LMEC model, considering different sample sizes "n" and censoring levels "l".

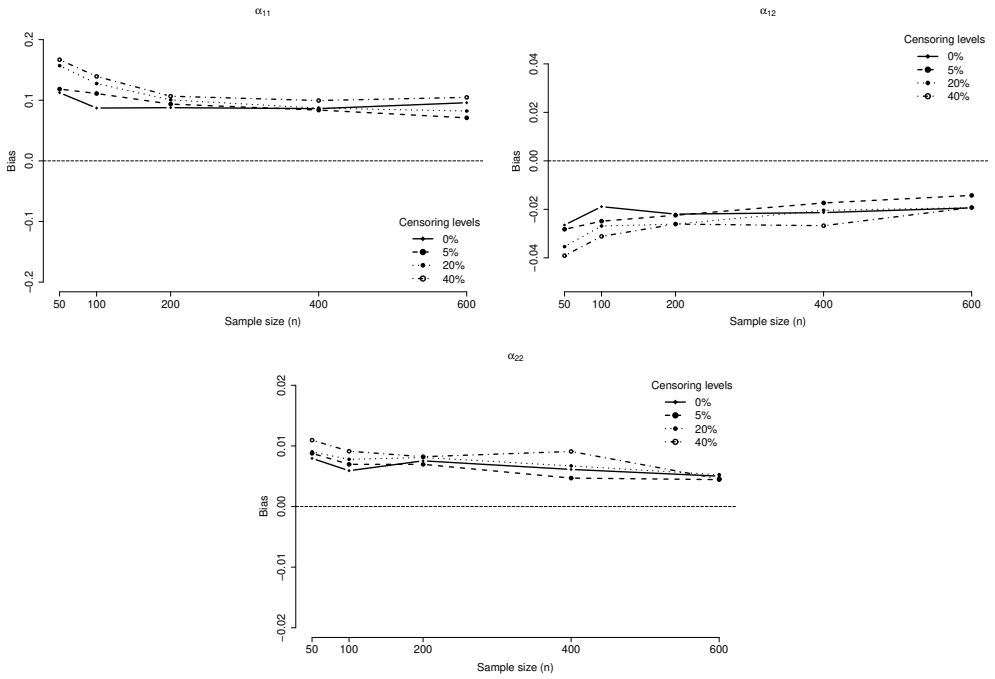

Figure 3 – Average Bias of random effect estimates of the AR(p)-LMEC model, considering different sample sizes "n" and censoring levels "l".

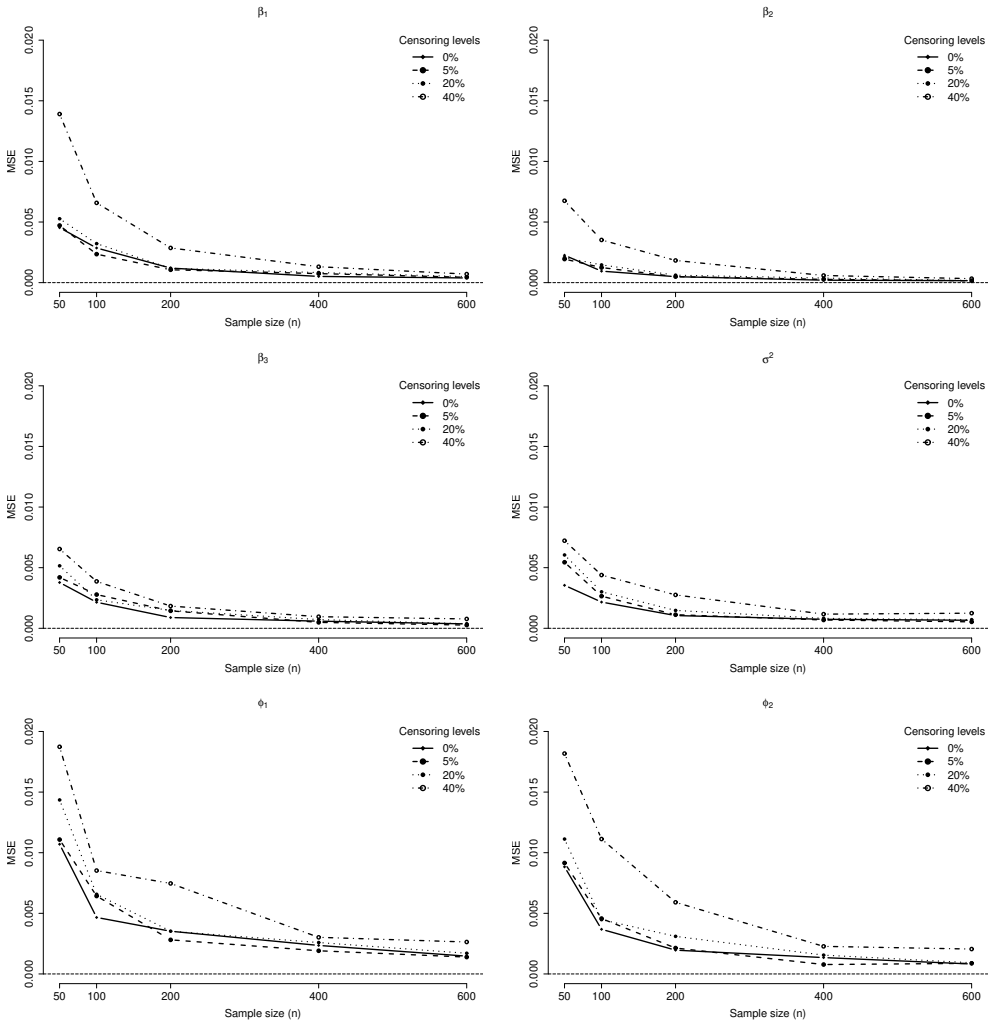

Figure 4 – Average MSE of parameter estimates of the AR(p)-LMEC model, considering different sample sizes "n" and censoring levels "l".

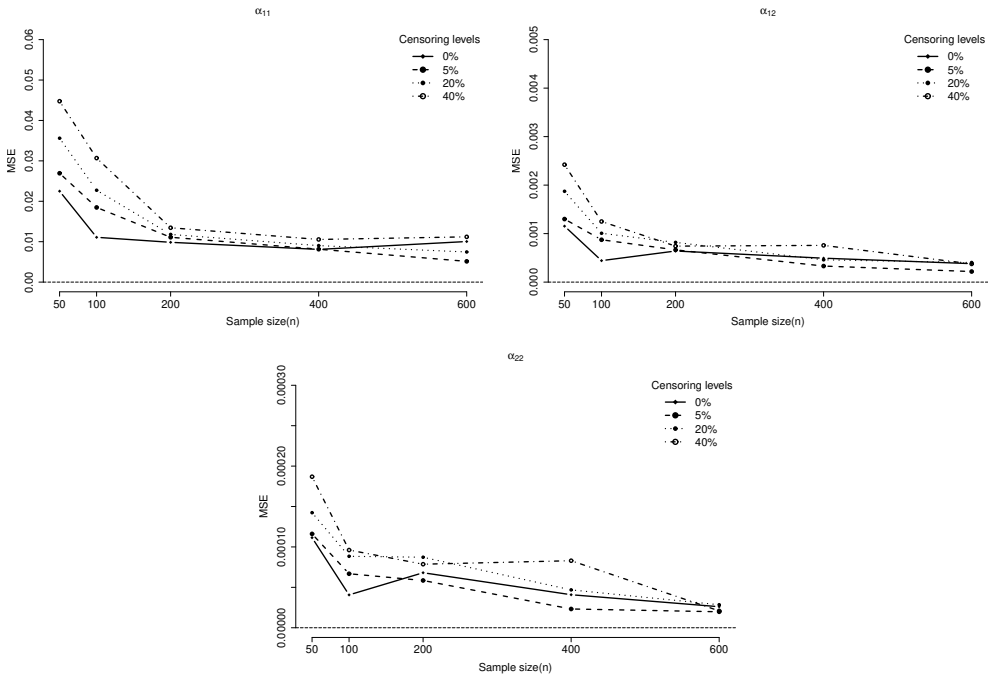

Figure 5 – Average MSE of random effect estimates of the AR(p)-LMEC model, considering different sample sizes "n" and censoring levels "l".

- **Simulation Study 2:**

  In this second simulation study, we fixed the sample size in $n = 100$ and generated $R = 100$ replications, in different censoring levels $l\% = \{0\%,\ 5\%,\ 10\%,\ 20\%,\ 40\%\}$. Thus, as proposed by Matos, Castro e Lachos (2016), Garay *et al.* (2017b) and Schumacher, Lachos e Dey (2017), in order to show that the method, suggested in Section (2.4), to approximate the SE of the ML estimates has good asymptotic properties, we analyzed the standard errors of the parameter estimates ($MC\ SE$), the average values of the standard errors computed using the empirical information matrix ($MC\ IM\ SE$) and the percentage of times, over the R samples, that the 95% asymptotic confidence intervals ($\theta \pm 1.95 * SE$) contain the true parameter values ($COV\ MC$). These measures were defined as:

$$MC\ IM\ SE(\theta_i) = \frac{1}{R}\sum_{j=1}^{R} \widehat{SE\left(\widehat{\theta}_i\right)}^{(j)} \text{ and}$$

$$MC\ SE(\theta_i) = \frac{1}{R-1}\left(\sum_{j=1}^{R}\left(\widehat{\theta}_i^{(j)}\right)^2 - \left(\frac{1}{R}\sum_{j=1}^{R}\widehat{\theta}_i^{(j)}\right)^2\right),$$

where $\widehat{\theta}_i^{(j)}$ denotes the ML estimates of parameter θ_i and $\widehat{SE\left(\widehat{\theta}_i\right)}^{(j)}$ represents the

SE estimation of $\widehat{\theta}_i^{(j)}$, obtained by the method suggested in Section 2.4, for the j-th replication with $j = 1, \ldots, R$.

From Table 2, we note a reasonable ($COV\ MC$) for the both β and ϕ_2; although, the values for σ^2 and ϕ_1 tend to be lower the nominal level 95% . For the case of the estimations of the SE, the measures of $MC\ SE$ and $MC\ IM\ SE$ are small and close. Taking into account the moderate sample size ($n = 100$), we consider these results quite satisfactory. Similarly results were obtained by Schumacher, Lachos e Dey (2017).

Table 2 – Standard errors of parameter estimates ($MC\ SE$), average values of the standard errors ($MC\ IM\ SE$) and $COV\ MC$

Censoring levels "l%"		Parameters					
		β_1	β_2	β_3	σ^2	ϕ_1	ϕ_2
0%	MC IM SE	0.063	0.030	0.047	0.030	0.050	0.055
	MC SE	0.048	0.034	0.054	0.039	0.066	0.061
	COV MC	94%	97%	95%	87%	87%	93%
5%	MC IM SE	0.065	0.032	0.047	0.031	0.051	0.056
	MC SE	0.053	0.032	0.045	0.038	0.066	0.059
	COV MC	97%	93%	94%	88%	93%	90%
10%	MC IM SE	0.065	0.032	0.045	0.031	0.054	0.056
	MC SE	0.052	0.027	0.051	0.038	0.056	0.073
	COV MC	95%	96%	96%	86%	93%	92%
20%	MC IM SE	0.069	0.035	0.050	0.031	0.057	0.061
	MC SE	0.052	0.036	0.054	0.036	0.063	0.055
	COV MC	95%	95%	94%	87%	89%	90%
40%	MC IM SE	0.085	0.050	0.057	0.027	0.076	0.073
	MC SE	0.068	0.048	0.060	0.054	0.087	0.088
	COV MC	95%	94%	96%	89%	88%	92%

- **Simulation Study 3:**

  Here, we fix the sample size "n" and compared the prediction value, under two different orders "p" of correlation structure: $AR(1)$ and $AR(2)$, and under an uncorrelated structure. Thus, we generated $R = 100$ samples of sample size $n = 100$, under the AR(2)-LMEC model, considering two censoring levels $l\% = \{5\%,\ 20\%\}$. We turn our attention to the one and two step ahead forecast of future observations using the approach proposed in Section 2.5.

  In order to compare performance of the predictions one an two step ahead forecast

considering the UNC-LMEC, AR(1)-LMEC and AR(2)-LMEC models under the censoring level $l\%$, as proposed by Schumacher, Lachos e Dey (2017) and Garay *et al.* (2017a), we utilized two empirical discrepancy measures, called the mean absolute prediction error $MAPE$ and mean square prediction error $MSPE$ defined by:

$$\text{MAPE} = \frac{1}{R \times h} \sum_i \sum_j |y_{ij} - \widehat{y}_{ij}| \quad \text{and} \quad \text{MSPE} = \frac{1}{R \times h} \sum_i \sum_j (y_{ij} - \widehat{y}_{ij})^2,$$

where "h" represents the step number ahead, y_{ij} and $\widehat{y}_{ij}$ represent the original and predicted value of the j-th observation for the i-th subject, respectively, where $i = 1, \ldots, n$ and $j = 5,\ 6$.

Table 3 presents the comparison between the predicted values (one and two step ahead) with the real ones, considering the UNC-LMEC, AR(1)-LMEC and AR(2)-LMEC models, under the censoring levels $l\% = \{5\%,\ 20\%\}$. Thus, from this Table, as expected, we can conclude that the AR(2)-LMEC model generates better predictive results than the UNC-LMEC and AR(1)-LMEC model, under both censoring levels $l\%$ considered.

Table 3 – Mean absolute prediction error $MAPE$ and mean square prediction error $MSPE$ for different correlation structures

	$l\% = 5\%$					
	MAPE			MSPE		
	UNC	AR(1)	AR(2)	UNC	AR(1)	AR(2)
One step	0.70	0.61	**0.59**	0.80	0.57	**0.57**
Two step	0.66	0.60	**0.57**	0.69	0.55	**0.53**
	$l\% = 20\%$					
	MAPE			MSPE		
	UNC	AR(1)	AR(2)	UNC	AR(1)	AR(2)
One step	0.69	0.66	**0.65**	0.83	0.70	**0.54**
Two step	0.70	0.64	**0.63**	0.78	0.66	**0.63**

2.7.2 The nonlinear case

For the nonlinear case, we performed three simulations studies based in the AR(p)-NLME model, defined in Equations (2.15)-(2.4), considering $l\%$ level of right censoring and the same number of measurements $n_i = 10$, for each i-th subject.

As proposed by Vaida e Liu (2009) and Matos, Castro e Lachos (2016), we consider the

following nonlinear curve model:

$$y_{ij} = \delta_{1i} + \frac{\delta_2}{1 + \exp\left(\frac{x_{ij} - \delta_3}{\delta_{4i}}\right)} + \epsilon_{ij},$$

where $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{in_i})^\top$ and $\epsilon_i = (\epsilon_{i1}, \epsilon_{i2}, \ldots, \epsilon_{in_i})^\top \sim \mathrm{N}_{n_i}(\mathbf{0}_{n_i}, \sigma^2 M_{n_i}(\phi))$. $M_{n_i}(\phi)$ assumes a first-order auto-regressive structure (AR(1)), $\mathbf{x}_{ij}$ is generated from a discrete uniform distribution $U_d(0, 90)$ and were fixed for all replications, $\delta_{1i} = \exp(\beta_1 + b_{1i})$, $\delta_2 = \exp(\beta_2)$, $\delta_3 = \exp(\beta_3)$ and $\delta_{4i} = \exp(\beta_4 + b_{2i})$.

As presented by Matos, Castro e Lachos (2016), for all simulation schemes, the parameter values considered were: $\beta = (\beta_1, \beta_2, \beta_3, \beta_4)^\top = (1.6094,\ 0.6931,\ 3.8067,\ 2.3026)^\top$, $\sigma = 0.55$ and $\phi_1 = 0.58$. The random effect vector $\mathbf{b}_i = (b_{1i},\ b_{2i})^\top$ was generated, previously, from a bivariate normal distribution $\mathrm{N}_2(\mathbf{0}, \mathbf{D})$, with

$$\mathbf{D} = \begin{bmatrix} \alpha_{11} & \\ \alpha_{21} & \alpha_{22} \end{bmatrix} = \begin{bmatrix} 0.0025 & -0.0010 \\ -0.0010 & 0.010 \end{bmatrix}$$

as presented by Vaida e Liu (2009). We fixed the right censoring level at $l\%$.

- **Simulation Study 1:**

  We generated $R = 100$ datasets, for each one of the combinations of different samples size $n = \{50,\ 100,\ 200,\ 400,\ 600\}$ and censoring levels $l\% = \{5\%,\ 10\%,\ 20\%\}$. As in the linear case, we estimate the parameter models using our proposed EM algorithm and compute the Bias and the mean squared error (MSE), for $\widehat{\theta}$, over the R samples with $\theta = (\beta^\top, \sigma^2, \alpha^\top, \phi^\top)^\top$.

  Figures 6 and 8 show, that the Bias and MSE, for the fixed parameters,tends to zero when the sample size increases, indicating that the estimates based on the proposed EM-type algorithm provides good asymptotic properties. Figures 7 and 9 show, that the Bias and MSE, for the random effects. Here is observed that the random effects present Bias and MSE tend to decrease when the sample size increases.

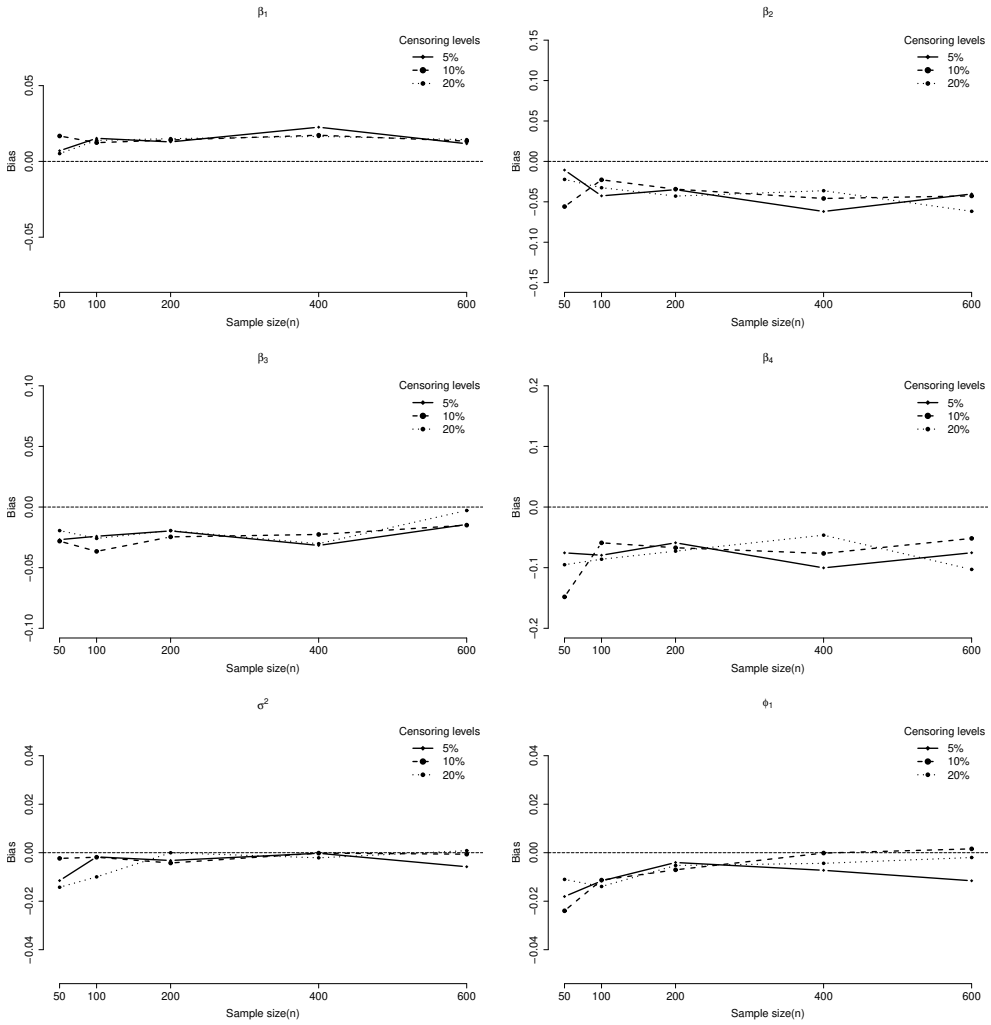

Figure 6 – Average Bias of parameter estimates of the AR(p)-NLMEC model, considering different sample sizes and censoring levels.

- **Simulation Study 2:**

  In this simulation study, we fix the sample size in $n = 100$ and generated $R = 100$ replications, in different censoring levels $l\% = \{5\%,\ 10\%,\ 20\%\}$. As a linear case, we analyzed the MC-SE, MC-IM-SE and COV-MC.

  From Table 4, we observe a reasonable MC coverage for the all parameters, including σ^2 and ϕ_1. For the case of the estimations of the SE, the measures of $MC\ SE$ and $MC\ IM\ SE$ are small and close. Thus, taking into account the moderate sample size ($n = 100$), we consider these results satisfactory.

- **Simulation Study 3:**

  In this case, we fix the sample size in "n" and compare the prediction values under two different orders "p", of correlation structures, AR(1) and AR(2), and under an

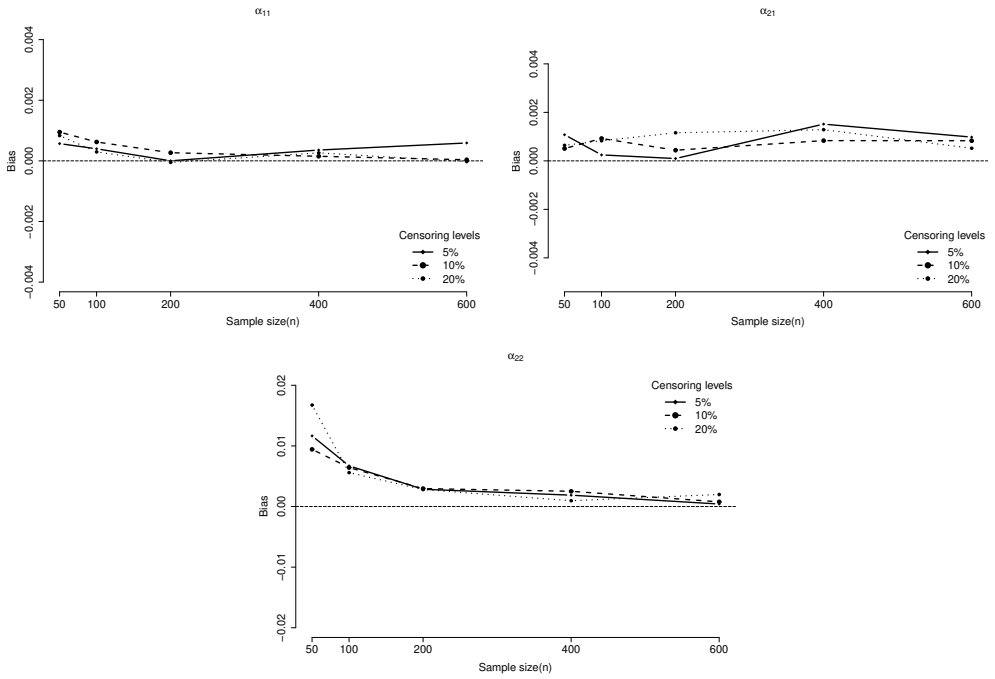

Figure 7 – Average Bias of random effect estimates of the AR(p)-NLMEC model, considering different sample sizes and censoring levels.

Table 4 – Standard errors of parameter estimates (MC-SE), average values of the standard errors (MC-IM-SE) and COV-MC

Censoring levels "l%"		Parameters					
		β_1	β_2	β_3	β_4	σ^2	ϕ_1
5%	MC-IM-SE	0.031	0.142	0.059	0.245	0.03	0.003
	MC-SE	0.03	0.128	0.057	0.231	0.028	0.034
	COV-MC	96%	93%	95%	93%	95%	95%
10%	MC-IM-SE	0.031	0.144	0.061	0.248	0.03	0.003
	MC-SE	0.026	0.125	0.057	0.233	0.027	0.032
	COV-MC	91%	95%	90%	92%	92%	95%
20%	MC-IM-SE	0.03	0.137	0.061	0.233	0.029	0.003
	MC-SE	0.029	0.138	0.062	0.241	0.028	0.04
	COV-MC	96%	95%	94%	92%	94%	94%

uncorrelated structure. Thus, initially, we generated $R = 100$ dataset, of size $n = 100$, following the AR(2)-NLMEC model, with $\phi = (\phi_1, \phi_2)^\top = (0.58, -0.1)^\top$, with two censoring levels $l\% = \{5\%,\ 20\%\}$. As in linear case, we compared the performance of the predicted values (one and two step ahead) with the real ones, considering the UNC-NLMEC, AR(1)-NLMEC and AR(2)-NLMEC models, using the discrepancy measures MAPE and MSPE.

Table 5 shows, as expected, that in the two censoring levels, the AR(2)-NLMEC model generates better predictive results than, the UNC-NLMEC and AR(1)-NLMEC

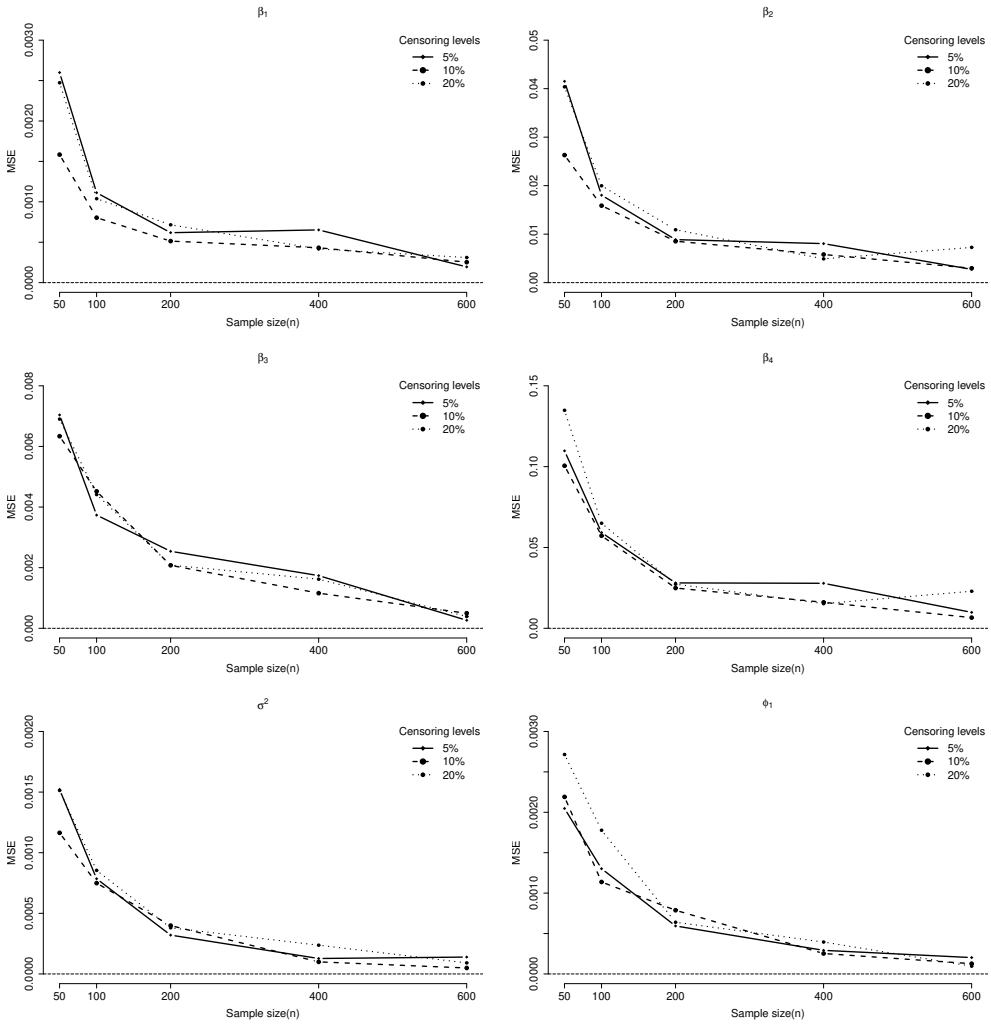

Figure 8 – Average MSE of parameter estimates of the AR(p)-NLMEC model, considering different sample sizes and censoring levels.

model.

Table 5 – Mean absolute prediction error $MAPE$ and mean square prediction error $MSPE$ for different correlation structures

	$l\% = 5\%$					
	MAPE			MSPE		
	UNC	AR(1)	AR(2)	UNC	AR(1)	AR(2)
One step	0.65	0.60	**0.59**	0.75	0.71	**0.69**
Two step	0.70	0.66	**0.62**	0.80	0.65	**0.59**
	$l\% = 20\%$					
	MAPE			MSPE		
	UNC	AR(1)	AR(2)	UNC	AR(1)	AR(2)
One step	0.62	0.55	**0.51**	0.58	0.47	**0.41**
Two step	0.65	0.61	**0.56**	0.61	0.55	**0.49**

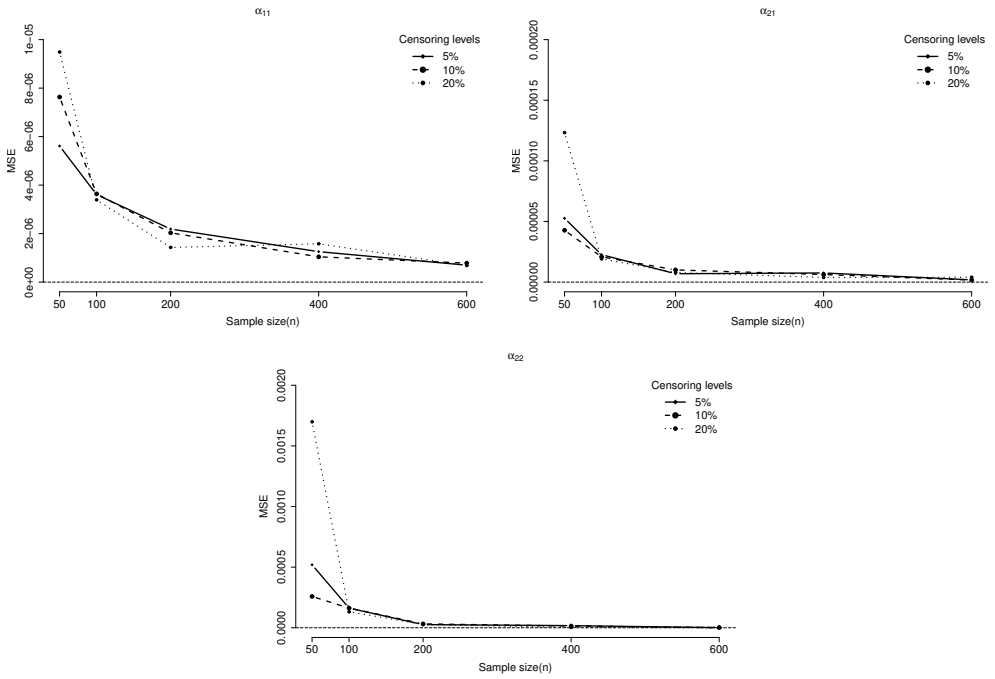

Figure 9 – Average MSE of random effect estimates of the AR(p)-NLMEC model, considering different sample sizes and censoring levels.

2.8 APPLICATION

For the application we consider the analysis of the dataset "*ACTG-315*" described in Section (1.2). For this the dataset, as presented in Matos, Castro e Lachos (2016), we consider the nonlinear mixed effects model given by

$$\begin{aligned} y_{ij} &= \log_{10}(\exp(\lambda_1)+\exp(\lambda_2))+\epsilon_{ij}, \\ \lambda_1 &= \beta_1+b_{1i}-(\beta_2+b_{2i}t_{ij}), \\ \lambda_2 &= \beta_3+b_{3i}-(\beta_4+\beta_5 CD4_{ij}+b_{4i})t_{ij}, \end{aligned}$$

where y_{ij} is the log_{10}-transformation of the viral load for subject "i" at time "t_{ij}", $CD4_{ij}$ indicates the observed CD4 values up to time t_{ij}, $b_i=(b_{1i},\ldots,b_{4i})$ is the random effects vector, for subject "i" and ϵ_{ij} represents the within-individual random error, where $i=1,\ldots,45$ and $j=1,\ldots,n_i$.

We applied the AR(p)-NLMEC model defined in (2.15)-(2.4), considering four cases of correlation structure namely (a) the uncorrelated (UNC) structure, (b) the continuous-time AR(1) structure, (c) the continuous-time AR(2) structure and (d) the continuous-time AR(3) structure. Table 6 shows the ML estimates of the parameters, under the AR(p)-NLMEC models, together with their corresponding standard errors. We observe that, in general, the ML estimates $\widehat{\beta}$ are quite similar, for all scenarios. Table 7

presents some model selection criteria, together, with the values of the log-likelihood. We can see from this table, that the AR(3)-NLMEC model produces more accurate estimates.

In order to evaluate the prediction performance of our approach, we used the same strategy developed by Garay *et al.* (2017a). Thus, we consider the subjects that had at least five measures (45 subjects), at least eights measures (31 subjects) and at least nine measures (14 subjects), and the last two measures were predicted. As in the simulation study (Section 2.7), we computed the MAPE and MSPE measures of the predicted values under different models (UCN, AR(1)-AR(3)), for comparing purposes. Once again the AR(3)-NLMEC models have the best performance in term of prediction. This results are presents in Table 8.

Figure 10 depicts the prediction performance of four randomly selected subjects (#7, #18, #30, #40), with at least eight measures, under the scenarios that produces more accurate estimates. We can see from this figure that the AR(3)-NLMEC model generates predictive values close to the real ones, as expected.

Table 6 – ACTG-315 dataset. ML estimates (Est) and standard errors (SE).

	Correlation structures							
	UNC		AR(1)		AR(2)		AR(3)	
Parameters	Est	SE	Est	SE	Est	SE	Est	SE
β_1	11.670	0.191	11.630	0.280	11.604	0.281	11.590	0.251
β_2	32.209	0.092	33.398	0.012	33.463	0.011	31.619	0.086
β_3	6.653	0.393	6.867	0.337	6.871	0.390	6.770	0.424
β_4	-0.358	0.697	-0.306	0.824	-0.229	0.950	-0.733	0.780
β_5	0.378	0.218	0.391	0.156	0.372	0.167	0.514	0.172
σ^2	0.126	0.015	0.195	0.023	0.207	0.028	0.198	0.020
α_{11}	1.046	0.112	0.587	0.201	0.250	0.513	0.178	0.537
α_{12}	-5.228	0.137	-1.493	0.121	-1.105	0.424	-0.857	0.501
α_{22}	43.631	0.168	7.183	0.355	6.194	0.630	4.317	0.525
α_{13}	0.656	0.041	0.600	0.645	0.149	0.782	0.039	0.893
α_{23}	3.369	0.015	1.178	0.193	1.679	0.212	1.557	0.214
α_{33}	3.489	0.008	1.436	0.517	0.876	0.609	0.850	0.745
α_{14}	-0.534	0.032	-0.396	0.460	-0.528	0.510	-0.472	0.563
α_{24}	13.676	0.322	9.850	0.469	9.855	0.920	8.154	0.744
α_{34}	4.127	0.054	2.119	0.088	2.066	0.093	2.040	0.085
α_{44}	11.209	0.019	9.141	0.073	9.376	0.081	8.274	0.093
ϕ_1	- -	- -	0.525	0.118	0.579	0.127	0.540	0.087
ϕ_2	- -	- -	- -	- -	0.079	0.122	-0.038	0.126
ϕ_3	- -	- -	- -	- -	- -	- -	0.271	0.194

Table 7 – ACTG-315 dataset. Comparison between the AR(p)-NLMEC models, considering different orders of correlation structures.

	UNC	AR(1)	AR(2)	AR(3)
loglik	-281.704	-279.613	-280.421	**-276.253**
AIC	595.408	593.225	596.842	**590.505**
BIC	657.452	659.147	666.641	**664.182**
AICc	597.008	595.030	598.866	**592.760**

Table 8 – ACTG-315 dataset. Evaluation of the prediction accuracy for the AR(p)-NLMEC models, with different correlations structures.

Forecast	For individuals with $n_i \geq 5$ (45 individuals)							
	MAPE				MSPE			
	UNC	AR(1)	AR(2)	AR(3)	UNC	AR(1)	AR(2)	AR(3)
One step	0.8475	0.8907	0.8818	**0.8441**	1.1542	1.2652	1.2406	**1.1411**
Two step	0.7146	0.7073	0.7045	**0.6955**	0.9088	0.8665	0.8547	**0.8597**
Forecast	For individuals with $n_i \geq 8$ (31 individuals)							
	MAPE				MSPE			
	UNC	AR(1)	AR(2)	AR(3)	UNC	AR(1)	AR(2)	AR(3)
One step	0.8308	0.8948	0.8882	**0.8214**	1.1200	1.2915	1.2660	**1.1085**
Two step	0.6600	0.6696	0.6709	**0.6394**	0.7986	0.7961	0.7970	**0.7556**
Forecast	For individuals with $n_i \geq 9$ (14 individuals)							
	MAPE				MSPE			
	UNC	AR(1)	AR(2)	AR(3)	UNC	AR(1)	AR(2)	AR(3)
One step	0.8182	0.8103	0.8095	**0.8058**	1.1930	1.1403	1.1356	**1.1151**
Two step	0.6287	0.6415	0.6410	**0.6220**	0.7315	0.7291	0.7269	**0.6987**

2.9 CONCLUSIONS

This chapter describes a likelihood-based approach to perform inference and prediction in censored mixed effects models with autoregressive errors . We use the EMC algorithm to obtain the ML estimates of model parameters.

For practical demonstration the method is applied to the analysis of a real HIV data set whose measures are subject to the detection limit of the recording assays. We also use simulation to investigate the performance in terms of predictions, parameter recovering and the robustness of the EMC algorithm. In the simulation study comparisons are made between inferences based on the proposed censored data with different correlation structures.

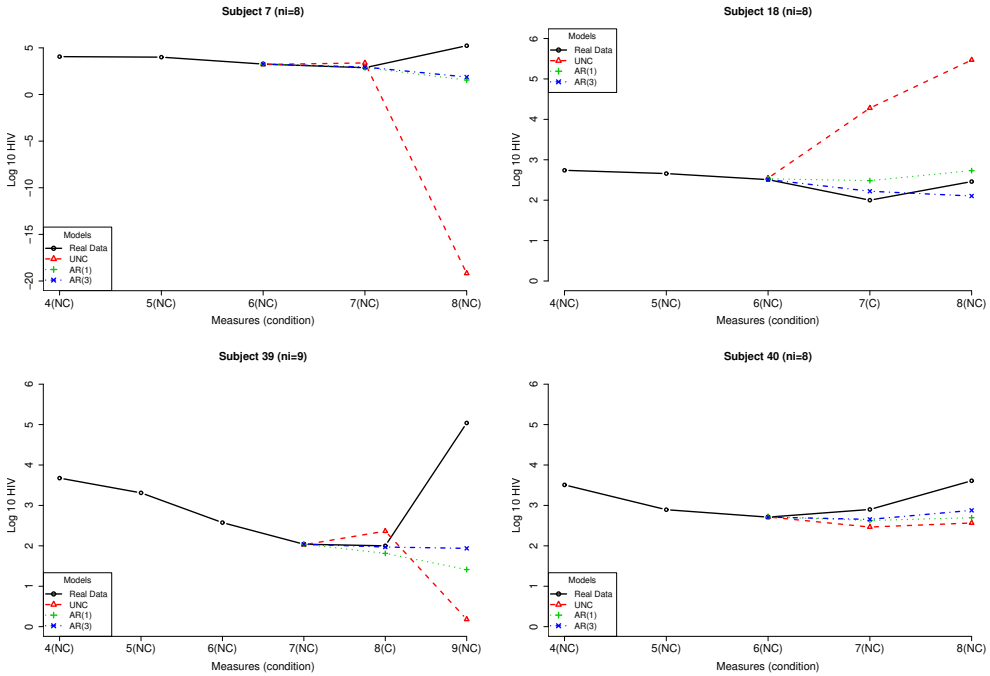

Figure 10 – ACTG-315 dataset. Evaluation of the prediction performance for four randomly selected subjects.

3 PACKAGE "ARPLMEC"

Neste capitulo será apresentado o pacote **ARpLMEC** (Fitting Autoregressive Censored Linear Mixed-Effects models) disponível no CRAN do software R (OLIVARI; GARAY; LACHOS, 2019; R Core Team, 2018). Descreveremos brevemente as funções contidas no pacote e as características de cada função.

3.1 INTRODUCTION

Com objetivo de disponibilizar e facilitar a aplicação da metodologia desenvolvida para diferentes contextos, foi elaborado, em linguagem R, o pacote **ARpLMEC**. Este pacote contém duas funções principais:

i) **ARpMMEC.sim**: utilizada para simular uma base de dados censurada, com efeitos mistos e erros autorregressivos

ii) **ARpMMEC.est**: utilizada para estimar os parâmetros de um modelo linear censurado, com efeitos mistos e erros autorregressivos de ordem "p".

Descreveremos detalhadamente a seguir ambas as funções.

3.2 DESCRIPTION

- **ARpLMEC.sim: Generating Censored Autoregressive Dataset with Linear Mixed Effects.**

  Esta função gera um vetor de variáveis respostas censuradas com efeitos mistos, com erros autorregressivos de ordem "p". Retorna o vetor de censura e o vetor de resposta censurada.

  Uso

R code

```
ARpMMEC.sim(m,x=NULL,z=NULL,nj,beta,sigmae,D1,phi,p.cens=0,
cens.type="left")
```

  Argumentos

m	Número de indivíduos

x Matriz de desenho para os efeitos fixos de ordem $n \times s$, correspondendo ao vetor de efeitos fixos.

z Matriz de desenho para os efeitos aleatórios de ordem $n \times b$, correspondendo ao vetor de efeitos aleatórios.

nj Vetor de ordem $1 \times m$ com o número de observações por indivíduo, em que m representa o número total de indivíduos.

beta Vetor de valores para os efeitos fixos.

sigmae Valor do σ^2.

D1 Matriz de covariância dos efeitos aleatórios.

phi Vetor de ordem $p \times 1$ dos parâmetros autorregressivos do modelo Ar(p).

p.cens Proporção de censura do processo.

cens.type $left$ se for censura à esquerda, $right$ se for censura à direita e $interval$ se for censura intervalar.

- **ARpLMEC.est: Autoregressive Censored Linear Mixed Effects Models**

Esta função estima os parâmetros de um modelo linear com censura à direita, à esquerda ou intervalar, com efeitos mistos e erros autorregressivos de ordem p, por meio do algoritmo tipo EM, obtendo como resultado as estimativas dos parâmetros, os erros padrão e as predições.

Uso

R code

```
ARpMMEC.est(y,x,z,cc,nj,Arp=1,beta0=NULL,sigma0=NULL,D0=NULL,
pi0=NULL,cens.type="left",LI=NULL,LS=NULL,MaxIter=200,
error=1e-04,Prev=FALSE,step=NULL,isubj=NULL,
xpre=NULL,zpre=NULL)
```

Argumentos

y Vetor de respostas censuradas de tamanho n, em que n é a soma do número de observações de cada indivíduo.

cc	Vetor indicador de censurar de tamanho n, em que n é o numero total de observações. Para cada observação será: 0 se não for censurado, 1 se for censurado.
Arp	Ordem do processo autorregressivo. Deve ser um valor inteiro positivo. Para considerar um modelo não correlacionado, utilizar UNC.
beta0	Valor inicial para o vetor de efeitos fixos.
sigma0	Valor inicial para σ^2.
D0	Valor inicial para a matriz de covariância dos efeitos aleatórios.
pi0	Valor inicial para o vetor de coeficientes autorregressivos $\gamma's$.
cens.type	$left$ se for censura à esquerda, $right$ se for censura à direita e $interval$ se for censura intervalar.
LI	Vetor indicador do limite inferior de censura de tamanho n. Para cada observação será: 0 se não for censurado, $-inf$ se for censurado. É indicado apenas quando *cens.type* for $Intervalar$.
LS	Vetor indicador do limite superior de censura de tamanho n. Para cada observação será: 0 se não for censurado, inf se for censurado. É indicado apenas quando *cens.type* for $Intervalar$.
MaxIter	Número máximo de iterações para o algoritmo EM.
error	O erro de convergência máximo.
Prev	Indicador do processo de predição.
step	Número de passos para a predição.
isubj	Vetor indicador dos indivíduos incluídos no processo de predição.
xpre	Matriz de desenho dos efeitos fixos a serem previstos.
zpre	Matriz de desenho dos efeitos aleatórios a serem previstos.

3.3 SEQUENCE TO USE THE PACKAGE

Para comprender melhor a utilização do pacote ARpLMEC, descreveremos a sequência de passos a seguir:

- **Passo 1**: Suponha que queremos estimar os parâmetros de um modelo de efeitos mistos lineares, com um nível de censura de 0.1 (AR(p)-LMEC), definido por:

$$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i + \epsilon_i, \quad \epsilon_i \overset{ind.}{\sim} \mathrm{N}_{n_i}(\mathbf{0_{n_i}}, \sigma^2 M_{n_i}(\phi)), \quad \mathbf{b}_i \overset{iid}{\sim} \mathrm{N}_q(\mathbf{0_q}, \mathbf{D}).$$

  e

$$\mathbf{y}_{obs_i} = \begin{cases} V_i & se \quad y_i \leq V_i; \\ y_i & se \quad y_i > V_i, \end{cases}$$

  para todo $i = 1, \ldots, N$.

- **Passo 2**: Considerando o modelo definido no passo anterior, com diferentes números de medições para cada indivíduo, geraremos uma amostra aleatória para $\mathbf{y}_i$ com censura à esquerda. Os valores para X e Z foram gerados a partir de uma distribuição uniforme, no intervalo (-1,1). Assumimos como valores vedadeiros dos parâmetros: $\beta = (1,2,1)^\top$, $\phi = (0.48, -0.2)^\top$, $\sigma^2 = 0.3$ e

$$D = \begin{bmatrix} 0.049 & 0.001 \\ 0.001 & 0.002 \end{bmatrix}.$$

  Para gerar esta amostra utilizamos a seguinte sequência de comandos no R:

R code

```
##Nivel de censura
p.cens = 0.1
##tamanho de amostra
m = 50
##Valores verdadeiros dos parametros
D = matrix(c(0.049,0.001,0.001,0.002),2,2)
sigma2 = 0.30
phi = c(0.48,-0.2)
beta = c(1,2,1)
##Considerando diferentes tamanhos de observacoes para
##cada individuo i
nj = rep(c(6,5,6,8,5,7,8,9,10,12),5)
##Gerando X e Z
x = matrix(runif(sum(nj)*length(beta),-1,1),
sum(nj),length(beta))
z = matrix(runif(sum(nj)*dim(D)[1],-1,1),
sum(nj),dim(D)[1])
```

```
##Gerandor da amostra censurada para yi
data = ARpLMEC.sim(m,x,z,nj,beta,sigma2,D,phi,p.cens)
##Dados censurados
y_cc = data$y_cc
##Vetor de censura
cc = data$cc
```

- **Passo 3**: Com a amostra gerada do modelo AR(p)-LMEC, estimamos os parâmetros do modelo, sob uma estrutura de correlação AR(2), com a seguinte sequência de comandos:

R code

```
##Estrutura de correlacao
Arp = 2
#Estimacao sem Previsao
teste1 = ARpLMEC.est(y_cc,x,z,cc,nj,Arp,MaxIter = 10)
##Matriz X e Z para previsao
xx=matrix(runif(6*length(beta),-1,1),6,length(beta))
zz=matrix(runif(6*dim(D)[1],-1,1),6,dim(D)[1])
##Numero de individuos para previsao
isubj = c(1,4,5)
#Estimacao com Previsao
teste2 = ARpLMEC.est(y_cc,x,z,cc,nj,Arp,MaxIter=10,Prev=TRUE,
step=2,isubj=isubj,xpre=xx,zpre=zz)
```

- **Passo 4**: Obtendo como resultado:

R code

```
##Estimacao dos parametros
------------------------------------------------
Autoregressive Censored Linear Mixed-Effects Models
------------------------------------------------
Autoregressive order = 2
Subjects = 50 ; Observations = 380
-----------
Estimates
-----------

- Fixed effects
```

```
   Est    SE        IConf(95%)
beta 1 0.983 0.056 < 0.873 , 1.093 >
beta 2 2.061 0.054 < 1.955 , 2.167 >
beta 3 1.040 0.045 < 0.952 , 1.128 >

- Sigma^2
         Est   SE        IConf(95%)
Sigma^2 0.325 0.03 < 0.266 , 0.384 >

- Autoregressives parameters
   Est    SE         IConf(95%)
Phi 1  0.471 0.079  < 0.316 , 0.626 >
Phi 2 -0.146 0.077 < -0.297 , 0.005 >

- Random effects
  Est    SE       IConf(95%)
Alpha 11  0.029 0.043    < 0 , 0.113 >
Alpha 12 -0.015 0.018 < -0.05 , 0.02 >
Alpha 22  0.011 0.023    < 0 , 0.056 >

------------------------
Model selection criteria
------------------------
   Loglik     AIC     BIC  AICc
Value -323.807 665.614 701.075 666.1
-------
Details
-------
Convergence reached? = FALSE
Iterations = 200 / 200
Processing time = 10.88875 mins

##Previsao
teste2$Prev
  subj     step 1     step 2
1    1 -1.0076274  1.9613775
2    4 -0.3282003  1.0041068
3    5 -3.3651604 -0.7217525
```

- **Passo 5**: Finalmente na tabela 9 são apresentadas as estimativas obtidas pelo pacote ARpLMEC com os valores reais assumidos para os parâmetros do modelo.

Table 9 – Estimates obtained by the ARpLMEC package.

Parâmetro	Valor Real	Valor Estimado
β_0	1.000	0.983
β_1	2.000	2.061
β_2	1.000	1.040
σ^2	0.300	0.325
ϕ_1	0.480	0.471
ϕ_2	-0.200	-0.146
α_{11}	0.049	0.029
α_{12}	0.001	-0.015
α_{22}	0.002	0.010

4 CONCLUSION

- Nesta dissertação desenvolvemos uma metodologia para a estimação dos parâmetros baseada na maximização da função de verossimilhança para os modelos de efeitos mistos censurados com erros autorregressivos, alem de uma estrategia de previsão. Utilizamos o algoritmo ECM para obter as estimativas de ML dos parâmetros do modelo. Embora algumas soluções tenham sido propostas na literatura, para lidar com variáveis respostas censuradas, autocorrelacionadas e medidas em intervalos iregulares de tempo, nosso trabalho apresenta uma estrutura de correlacão mais geral que as estudadas anteriormente.
- Para demonstração prática da adequação da metodologia, foi realizado um estudo referente a pacientes com HIV, cujas medidas estão sujeitas aos limites, superiores ou inferiores, de detecção. Também utilizamos estudos de simulação para investigar o desempenho da metodologia em termos da robustez do algoritmo EMC, a estimação dos parâmetros do modelo e previsões de observações futuras.
- Finalmente os métodos propostos também podem ser facilmente aplicados a outras áreas, em que os dados analisados possuem observações censuradas, por meio da implementação do pacote AR(p)-LMEC, apresentado no Capitulo 3, oferecendo assim uma ferramenta adequada para posterior aplicação em diversos conjuntos de dados com as características apresentadas neste trabalho.

4.1 SCIENTIFIC PRODUCTION

- Pacote **AR(p)-LMEC** no software R o qual esta disponível no CRAN (OLIVARI; GARAY; LACHOS, 2019) (vide Apêndice A).
- Artigo intitulado **Autoregressive mixed-effects models for censored data**, o qual foi submetido para publicação (vide Apêndice B).

4.2 FUTURE WORKS

Nesta subsessão são propostas algumas possíveis extensões dos resultados obtidos neste trabalho, os quais são listados a seguir:

- O desenvolvimento de análisis de residuos e técnicas de diagnóstico para o modelo AR(p)-LMEC.

- A extensão dos métodos propostos a conjuntos de dados com observações faltantes e censuradas, utilizando procedimentos de amostragem Bayesianos (Wang e Fan, 2012).
- A implememtação de outras distribuições para os termos aleatórios, que consideram características de não normalidade como assimetria ou presença de caudas pesadas. Como apresentaram Lachos, Bandyopadhyay e Dey (2011), os estudos de cargas virais do HIV com observações censuradas, são baseados sob suposições de normalidade para os termos aleatórios. No entanto, esses estudos podem não fornecer inferências robustas quando as suposições de normalidade são questionáveis, como ocorre em estudos de cargas virais, dado que até mesmo transformações logarítmicas para as respostas não seguem distribuição normal.

REFERENCES

AKAIKE, H. A new look at the statistical model identification. **IEEE transactions on automatic control**, Ieee, v. 19, n. 6, p. 716–723, 1974.

BARNDORFF-NIELSEN, O.; SCHOU, G. On the parametrization of autoregressive models by partial autocorrelations. **Journal of Multivariate Analysis**, Elsevier, v. 3, n. 4, p. 408–419, 1973.

BOX, G. E.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. **Time series analysis: Forecasting and control**. [S.l.]: John Wiley & Sons, 2015.

DEMPSTER, A.; LAIRD, N.; RUBIN, D. Maximum likelihood from incomplete data via the EM algorithm. **Journal of the Royal Statistical Society. Series B, Statistical Methodology**, v. 39, n. 1, p. 1–38, 1977.

EFRON, B. The two sample problem with censored data. In: **Proceedings of the fifth Berkeley symposium on mathematical statistics and probability**. [S.l.: s.n.], 1967. v. 4, n. University of California Press, Berkeley, CA, p. 831–853.

GARAY, A. M.; CASTRO, L. M.; LESKOW, J.; LACHOS, V. H. Censored linear regression models for irregularly observed longitudinal data using the multivariate-t distribution. **Statistical Methods in Medical Research**, SAGE Publications Sage UK: London, England, v. 26, n. 2, p. 542–566, 2017.

GARAY, A. M.; LACHOS, V. H.; BOLFARINE, H.; CABRAL, C. R. B. Linear censored regression models with scale mixtures of normal distributions. **Statistical Papers**, v. 58, n. 1, p. 247–278, 2017.

HUGHES, J. Mixed effects models with censored data with application to HIV RNA levels. **Biometrics**, John Wiley & Sons, v. 55, p. 625–629, 1999.

HURVICH, C. M.; TSAI, C.-L. Regression and time series model selection in small samples. **Biometrika**, Oxford University Press, v. 76, n. 2, p. 297–307, 1989.

LACHOS, V.; BANDYOPADHYAY, D.; DEY, D. Linear and nonlinear mixed-effects models for censored HIV viral loads using normal/independent distributions. **Biometrics**, Wiley Online Library, v. 55, p. 1304–1318, 2011.

LAIRD, N. M.; WARE, J. H. Random-effects models for longitudinal data. **Biometrics**, JSTOR, v. 38, n. 4, p. 963–974, 1982.

LIN, T.; LEE, J. Bayesian analysis of hierarchical linear mixed modeling using the multivariate t distribution. **Journal of Statistical Planning and Inference**, Elsevier, v. 137, n. 2, p. 484–495, 2007.

LIN, T.-I. Robust mixture modeling using multivariate skew t distributions. **Statistics and Computing**, Springer, v. 20, n. 3, p. 343–356, 2010.

LIN, T. I.; WANG, W. L. Multivariate skew-normal linear mixed models for multi-outcome longitudinal data. **Stat Modelling**, v. 13, n. 3, p. 199–221, 2013.

LIU, C. Bayesian robust multivariate linear regression with incomplete data. **Journal of the American Statistical Association**, v. 91, n. 435, p. 1219–1227, 1996.

LOUIS, T. A. Finding the observed information matrix when using the EM algorithm. **Journal of the Royal Statistical Society. Series B, Statistical Methodology.**, JSTOR, v. 44, n. 2, p. 226–233, 1982.

MATOS, L. A.; BANDYOPADHYAY, D.; CASTRO, L. M.; LACHOS, V. H. Influence assessment in censored mixed-effects models using the multivariate Student′s-t distribution. **Journal of Multivariate Analysis**, Elsevier, v. 141, p. 104–117, 2015.

MATOS, L. A.; CASTRO, L. M.; LACHOS, V. H. Censored mixed-effects models for irregularly observed repeated measures with applications to HIV viral loads. **Test**, Springer, v. 25, n. 4, p. 627–653, 2016.

MATOS, L. A.; LACHOS, V. H.; BALAKRISHNAN, N.; LABRA, F. Influence diagnostics in linear and nonlinear mixed-effects models with censored data. **Computational Statistics & Data Analysis**, v. 57, n. 1, p. 450 – 464, 2013.

MATOS, L. A.; PRATES, M. O.; CHEN, M. H.; LACHOS, V. H. Likelihood-based inference for mixed-effects models with censored response using the multivariate-t distribution. **Statistica Sinica**, v. 23, p. 1323–1342, 2013.

MEILIJSON, I. A fast improvement to the EM algorithm on its own terms. **Journal of the Royal Statistical Society. Series B, Statistical Methodology.**, JSTOR, v. 51, n. 1, p. 127–138, 1989.

MENG, X.-L.; RUBIN, D. B. Maximum likelihood estimation via the ECM algorithm: A general framework. **Biometrika**, Biometrika Trust, v. 80, n. 2, p. 267–278, 1993.

MUñOZ, A.; CAREY, V.; SCHOUTEN, J. P.; SEGAL, M.; ROSNER, B. A parametric family of correlation structures for the analysis of longitudinal data. **Biometrics**, JSTOR, v. 48, p. 733–742, 1992.

NETO, P.; SILVA, L. F. da; VIEIRA, N. F.; SOPRANI, M.; CUNHA, C. B.; CABRAL, V. P.; DIETZE, R.; RIBEIRO-RODRIGUES, R. Longitudinal comparison between plasma and seminal hiv-1 viral loads during antiretroviral treatment. **Revista da Sociedade Brasileira de Medicina Tropical**, SciELO Brasil, v. 36, n. 6, p. 689–694, 2003.

OLIVARI, R. C.; GARAY, A. M.; LACHOS, V. H. **ARpLMEC: Fitting Autoregressive Censored Linear Mixed-Effects Models**. [S.l.], 2019. R package version 1.0. Disponível em: <https://CRAN.R-project.org/package=ARpLMEC>.

PINHEIRO, J. C.; BATES, D. M. **Mixed-Effects Models in S and S-PLUS**. New York, NY: Springer, 2000.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2018. Disponível em: <https://www.R-project.org/>.

RAO, C. R. Prediction of future observations in growth curve models. **Statistical Science**, v. 2, p. 434 – 447, 1987.

SCHAFER, J. L.; YUCEL, R. M. Computational strategies for multivariate linear mixed-effects models with missing values. **Journal of computational and Graphical Statistics**, Taylor & Francis, v. 11, n. 2, p. 437–457, 2002.

SCHUMACHER, F. L.; LACHOS, V. H.; DEY, D. K. Censored regression models with autoregressive errors: A likelihood-based perspective. **Canadian Journal of Statistics**, Wiley Online Library, v. 45, n. 4, p. 375–392, 2017.

SCHWARZ, G. *et al.* Estimating the dimension of a model. **The annals of statistics**, Institute of Mathematical Statistics, v. 6, n. 2, p. 461–464, 1978.

TESCHE, F. M.; IANOZ, M.; KARLSSON, T.; KARLSSON, T. **EMC analysis methods and computational models**. [S.l.]: John Wiley & Sons, 1997.

VAIDA, F.; FITZGERALD, A.; DEGRUTTOLA, V. Efficient hybrid EM for linear and nonlinear mixed effects models with censored response. **Computational Statistics & Data Analysis**, Elsevier, v. 51, p. 5718–5730, 2007.

VAIDA, F.; LIU, L. Fast implementation for normal mixed Effects models with censored response. **Journal of Computational and Graphical Statistics**, v. 18, p. 797–817, 2009.

WANG, W.-L. Multivariate t linear mixed models for irregularly observed multiple repeated measures with missing outcomes. **Biometrical Journal**, v. 55, n. 4, p. 554–571, 2013.

WANG, W.-L.; FAN, T.-H. ECM-based maximum likelihood inference for multivariate linear mixed models with autoregressive errors. **Computational Statistics & Data Analysis**, Elsevier, v. 54, n. 5, p. 1328–1341, 2010.

WANG, W. L.; FAN, T. H. Estimation in multivariate t linear mixed models for multivariate longitudinal data. **Statistica Sinica**, v. 21, p. 1857–1880, 2011.

WU, L. A joint model for nonlinear mixed-effects models with censoring and covariates measured with error, with application to AIDS studies. **Journal of the American Statistical Association**, v. 97, n. 460, p. 955–964, 2002.

APPENDIX A – PACKAGE "ARPLMEC"

Package 'ARpLMEC'

January 9, 2019

Type Package

Title Fitting Autoregressive Censored Linear Mixed-Effects Models

Version 1.0

Date 2018-12-13

Author Rommy C. Olivari, Aldo M. Garay and Victor H. Lachos

Maintainer Rommy Camasca Olivari <rco1@de.ufpe.br>

Description It fits left, right or interval censored mixed-effects linear model with autoregressive errors of order p using the EM algorithm. It provides estimates, standard errors of the parameters and prediction of future observations.

Imports Matrix, stats4, gmm, sandwich,mvtnorm, tmvtnorm, numDeriv, utils, graphics, stats, MASS, lmec, mnormt

NeedsCompilation no

License GPL (>= 2)

RoxygenNote 6.1.1

Encoding UTF-8

Repository CRAN

Date/Publication 2019-01-09 17:50:39 UTC

R topics documented:

`ARpLMEC.est` *Autoregressive Censored Linear Mixed Effects Models*

Description

This function fits left, right or interval censored mixed-effects linear model, with autoregressive errors of order `p`, using the EM algorithm. It returns estimates, standard errors and prediction of future observations.

Usage

```
ARpLMEC.est(y, x, z, cc, nj, Arp = 1, beta0 = NULL, sigma0 = NULL,
  D0 = NULL, pi0 = NULL, cens.type = "left", LI = NULL,
  LS = NULL, MaxIter = 200, error = 1e-04, Prev = FALSE,
  step = NULL, isubj = NULL, xpre = NULL, zpre = NULL)
```

Arguments

`y`	Vector `1 x n` of censored responses, where `n` is the sum of the number of observations of each individual.
`x`	Design matrix of the fixed effects of order `n x s`, corresponding to vector of fixed effects.
`z`	Design matrix of the random effects of ordern `x b`, corresponding to vector of random effects.
`cc`	Vector of censoring indicators of length `n`, where `n` is the total of observations. For each observation: `0` if non-censored, `1` if censored.
`nj`	Vector `1 x m` with the number of observations for each subject, where `m` is the total number of individuals.
`Arp`	Order of the autoregressive process. Must be a positive integer value. To consider a model uncorrelated use `UNC`.
`beta0`	Initial values for the vector of fixed effects. If it is not indicated it will be provided automatically. Default is `NULL`.
`sigma0`	Initial values for sigma. If it is not indicated it will be provided automatically. Default is `NULL`.
`D0`	Initial values for the covariance matrix for the random effects. If it is not indicated it will be provided automatically. Default is `NULL`.
`pi0`	Initial values for the vector for autoregressive coefficients pi's. If it is not indicated it will be provided automatically. Default is `NULL`.
`cens.type`	`left` for left censoring, `right` for right censoring and `interval` for interval censoring. Default is `left`.
`LI`	Vector censoring lower limit indicator of length `n`. For each observation: `0` if non-censored, `-inf` if censored. It is only indicated for when `cens.type` is `both`. Default is `NULL`.

LS	Vector censoring upper limit indicator of length n. For each observation: 0 if non-censored, inf if censored.It is only indicated for when cens.type is both. Default is NULL.
MaxIter	The maximum number of iterations of the EM algorithm. Default is 200.
error	The convergence maximum error. Default is 0.0001.
Prev	Indicator of the prediction process. Default is FALSE.
step	Number of steps for prediction. Default is NULL.
isubj	Vector indicator of subject included in the prediction process. Default is NULL.
xpre	Design matrix of the fixed effects to be predicted. Default is NULL.
zpre	Design matrix of the random effects to be predicted. Default is NULL.

Value

returns list of class "ARpMMEC":

FixEffect	Data frame with: estimate, standars erros and confidence intervals of the fixed effects.
Sigma2	Data frame with: estimate, standars erros and confidence intervals of the variance of the white noise process.
Phi	Data frame with: estimate, standars erros and confidence intervals of the autoregressive parameters.
RnEffect	Data frame with: estimate, standars erros and confidence intervals of the random effects.
Est	Vector of parameters estimate (fixed Effects, sigma2, phi, random effects).
SE	Vector of the standard errors of (fixed Effects, sigma2, phi, random effects).
loglik	Log-likelihood value.
AIC	Akaike information criterion.
BIC	Bayesian information criterion.
AICc	Corrected Akaike information criterion.
iter	Number of iterations until convergence.
MI	Information matrix
Prev	Predicted values (if xpre and zpre is not NULL).
time	Processing time.

References

Vaida F, Liu L (2009). Fast implementation for normal mixed Effects models with censored response. Journal of Computational and Graphical Statistics; https://doi.org/10.1198/jcgs.2009.07130

Matos LA, Lachos V, Balakrishnan N, Labra F (2013). Influence diagnostics in linear and nonlinear mixed-effects models with censored data. Computational Statistics & Data Analysis; https://doi.org/10.1016/j.csda.2012.06.021

Schumacher FL, Lachos VH, Dey DK (2017). Censored regression models with autoregressive errors: A likelihood-based perspective. Canadian Journal of Statistics. https://doi.org/10.1002/cjs.11338

Examples

```
## Not run:
p.cens   = 0.1
m           = 10
D = matrix(c(0.049,0.001,0.001,0.002),2,2)
sigma2 = 0.30
phi    = c(0.48,-0.2)
beta   = c(1,2,1)
nj=c(6,5,6,8,5,7,8,6,5,4)
x<-matrix(runif(sum(nj)*length(beta),-1,1),sum(nj),length(beta))
z<-matrix(runif(sum(nj)*dim(D)[1],-1,1),sum(nj),dim(D)[1])
data=ARpLMEC.sim(m,x,z,nj,beta,sigma2,D,phi,p.cens)
attach(data)
Arp    = 2
 ##Estimacao sem Previcao
teste1=ARpLMEC.est(y_cc,x,z,cc,nj,Arp,MaxIter = 10)

##Estimacao com Previcao
xx=matrix(runif(6*length(beta),-1,1),6,length(beta))
zz=matrix(runif(6*dim(D)[1],-1,1),6,dim(D)[1])
isubj=c(1,4,5)
teste2=ARpLMEC.est(y_cc,x,z,cc,nj,Arp,MaxIter=10,Prev=TRUE,step=2,isubj=isubj,xpre=xx,zpre=zz)
teste2$Prev
## End(Not run)
```

`ARpLMEC.sim` *Generating Censored Autoregressive Dataset with Linear Mixed Effects.*

Description

This function simulates a censored response variable with autoregressive errors of order p, with mixed effect and a established censoring rate. This function returns the censoring vector and censored response vector.

Usage

```
ARpLMEC.sim(m, x = NULL, z = NULL, nj, beta, sigmae, D1, phi,
  p.cens = 0, cens.type = "left")
```

Arguments

`m`	Number of individuals
`x`	Design matrix of the fixed effects of order `n x s`, corresponding to vector of fixed effects.

`z`	Design matrix of the random effects of ordern `x b`, corresponding to vector of random effects.
`nj`	Vector `1 x m` with the number of observations for each subject, where `m` is the total number of individuals.
`beta`	Vector of values fixed effects.
`sigmae`	It's the value for sigma.
`D1`	Covariance Matrix for the random effects.
`phi`	Vector of length `Arp`, of values for autoregressive parameters.
`p.cens`	Censoring level for the process. Default is `0`
`cens.type`	`left` for left censoring, `right` for right censoring and `interval` for interval censoring. Default is `left`

Value

returns list:

`cc`	Vector of censoring indicators.
`y_cc`	Vector of responses censoring.

References

Schumacher FL, Lachos VH, Dey DK (2017). Censored regression models with autoregressive errors: A likelihood-based perspective. Canadian Journal of Statistics. https://doi.org/10.1002/cjs.11338

Garay AM, Castro LM, Leskow J, Lachos VH (2017). Censored linear regression models for irregularly observed longitudinal data using the multivariate-t distribution. Statistical Methods in Medical Research. https://doi.org/10.1177/0962280214551191

Examples

```
p.cens   = 0.1
m           = 50
D = matrix(c(0.049,0.001,0.001,0.002),2,2)
sigma2 = 0.30
phi    = c(0.48,-0.2)
beta   = c(1,2,1)
nj=rep(6,m)
x<-matrix(runif(sum(nj)*length(beta),-1,1),sum(nj),length(beta))
z<-matrix(runif(sum(nj)*dim(D)[1],-1,1),sum(nj),dim(D)[1])
data=ARpLMEC.sim(m,x,z,nj,beta,sigma2,D,phi,p.cens)
y<-data$y_cc
cc<-data$cc
```

Index

APPENDIX B – PAPER IN CANADIAN JOURNAL OF STATISTICS

Canadian Journal of Statistics

The Canadian Journal of Statistics

Autoregressive mixed-effects models for censored data

Journal:	*The Canadian Journal of Statistics*
Manuscript ID	Draft
Wiley - Manuscript type:	Research Article
Keywords:	Autoregressive AR(p) models, Censored data, EM algorithm, HIV viral load, Linear/ nonlinear mixed models

Autoregressive mixed-effects models for censored data

Abstract

In AIDS clinical trials, the HIV-1 RNA measurements are often subject to some upper and lower detection limits, depending on the quantification assays. Linear and nonlinear mixed-effects models, with modifications to accommodate censored observations (LMEC/NLMEC), are routinely used to analyze this type of data (Vaida and Liu, 2009). This paper presents a likelihood based approach for fitting LMEC/NLMEC models with autoregressive of order p dependence on the error term. An EM-type algorithm is developed for computing the maximum likelihood estimates, obtaining as a byproduct the standard errors of the fixed effects and the likelihood value. Moreover, the constraints on the parameter space that arise from the stationarity conditions for the autoregressive parameters in the EM algorithm are handled by a reparameterization scheme as discussed in Lin and Lee (2007). To examine the performance of the proposed method, we present some simulation studies and analyze a real AIDS case study. The proposed algorithm and methods are implemented in the new R package **ARpLMEC**.

Keywords: Autoregressive AR(p) models, censored data, EM algorithm, HIV viral load, linear/ nonlinear mixed models.

1. Introduction

The most popular analytic tools for longitudinal data analysis with continuous outcomes are the linear and nonlinear mixed-effects models (LME/NLME). However, in such longitudinal studies, such as those on acquired immunodeficiency syndrome (AIDS) and environmental pollution, some variables may have certain threshold values below or above which the measurements are not quantifiable. For instance, viral load measures the amount of actively replicating virus and, depending on the diagnostic assays used, its measurement over time may be subject to detection limits, below or above which they are not quantifiable. Linear and nonlinear mixed-effects models, with modifications to accommodate censored observations (LMEC/NLMEC), have been proposed to fit this kind of data. In this context, Vaida and Liu (2009) proposed an exact EM-type algorithm for LMEC/NLMEC models that uses closed-form expressions at the E-step as opposed to the Monte Carlo EM algorithm, proposed by Hughes (1999) and Vaida et al. (2007). Matos et al. (2013a) provided additional tools, including influence diagnostics, for LMEC/NLMEC models. In the context of heavy–tailed LMEC/NLMEC, Lachos et al. (2011) advocated the use of the normal/independent (NI) class of distributions, proposed by Liu (1996), and adopted a Bayesian framework to carry out posterior inference. Recently, Matos et al. (2013b) and Matos et al. (2015) proposed a likelihood-based estimation and influence analysis for Student-t LMEC/NLMEC models, respectively. A common feature of these classes of LMEC/NLMEC models is to assume that the correlation structure is induced just by the random effects term. However, in longitudinal studies the repeated measures of each subject are collected over time and hence the random errors tend also to be serially correlated.

Recently, some alternatives for modeling the correlation observation responses and correlations induced by longitudinal data have been proposed in the statistical literature. These proposals consider not only measurements of some variables may have be subjected to certain threshold values below or above which the measurements are not quantifiable. the correlation structure induced by the random effects term, but also by other types of correlation in the

error term. Particularly, Wang (2013) introduced the Student-t LME model for outcome variables recorded on irregular occasions, considering a damping exponential correlation (DEC) structure, as proposed by Muñoz et al. (1992). This correlation structure takes into account the autocorrelation generated by the within-subject dependence among irregular occasions and has as a special case the AR(1) correlation structure. Matos et al. (2016) consider the LMEC/NLMEC model with DEC structure, where an EM-type algorithm is also developed for computing the maximum likelihood (ML) estimates. On the other hand, Wang and Fan (2011) consider the Student-t LME model, with autoregressive dependence structure, of order p (AR(p)), for the within-subject errors.

Even though, some proposal have been made to deal with the problem of serial correlation among the observations in LMEC/NLMEC models, to the best of our knowledge, there are no studies of the LMEC/NLMEC with AR(p) errors. Thus, in this paper we develop a full likelihood-based approach for LMEC/NLMEC modeling with AR(p) errors, hereafter AR(p)-LMEC/NLMEC, including the implementation of a computationally efficient estimation method, via the EM algorithm, with the likelihood function, predictions of unobservable values of the response and the asymptotic standard errors as byproducts. The results developed here are an extension to those presented by Vaida and Liu (2009) and Matos et al. (2016) for the analysis of mixed-effects models with censored responses and HIV data. The proposed algorithm and methods are implemented in the new R package "**ARpLMEC**" (R Development Core Team, 2018).

The rest of the paper is organized as follows, Section 2 introduces the AR(p)-LMEC model and the likelihood function. In Section 3, the related likelihood-based inference is presented, including estimation of the random effects and the expected information matrix. The method for predicting future observations is discussed in Section 4. Section 5 describes the extension to the nonlinear case (AR(p)-NLMEC). The application of the proposed method is presented in Sections 6 and 7 through a simulation study and the analysis of a real AIDS case study, respectively. Finally, Section 8 concludes with a short discussion of issues raised by this study and some possible directions for future research.

2. Model formulation

In the non-censored case, a Gaussian LME model is specified as

$$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i + \varepsilon_i, \tag{1}$$

where $\mathbf{b}_i \overset{iid}{\sim} \mathrm{N}_q(\mathbf{0}_q, \mathbf{D})$ is independent of $\varepsilon_i \overset{ind.}{\sim} \mathrm{N}_{n_i}(\mathbf{0}_{n_i}, \Omega_i)$, $i = 1, \ldots, n$ and $\mathbf{0}_q$ and $\mathbf{0}_{n_i}$ represent the zeros vector of dimension q and n_i respectively. The subscript i represents the subject index; $\mathbf{y}_i = (y_{i1}, \ldots, y_{in_i})^\top$ is an $n_i \times 1$ vector of observed continuous responses for subject i measured at particular time points $\mathbf{t}_i = (t_{i1}, \ldots, t_{n_i})^\top$; $\mathbf{X}_i$ is the $n_i \times l$ design matrix corresponding to the fixed effects, β, of dimension $l \times 1$; $\mathbf{Z}_i$ is the $n_i \times q$ design matrix corresponding to the $q \times 1$ vector of random effects $\mathbf{b}_i$; ε_i, with dimension $(n_i \times 1)$, is the vector of random errors; and the dispersion matrix $\mathbf{D} = \mathbf{D}(\alpha)$ depends on the unknown and reduced parameters α. The correlation structure of the error vector is assumed to be $\Omega_i = \sigma^2 M_{n_i}(\phi)$, where the $n_i \times n_i$ matrix $M_{n_i}(\phi)$ incorporates a time-dependence structure. In this paper, we consider $M_{n_i}(\phi)$ with a structured $AR(p)$ dependence matrix for the within-subject errors, denoted by AR(p)-LME model. Specifically,

$$M_{n_i}(\phi) = \frac{1}{1 - \phi_1\rho_1 - \ldots - \phi_p\rho_p}[\rho_{|r-s|}],$$

where $r, s = 1, \ldots, n_i$ and $\rho_k's$ are implicit functions of autoregressive parameters $\phi = (\phi_1, \ldots, \phi_p)^\top$ and satisfy the Yule–Walker equation (Box et al., 2015), i.e.,

$$\rho_k = \phi_1\rho_{k-1} + \ldots + \phi_p\rho_{k-p},\ \rho_o = 1,\ (k = 1, \ldots, n_i - 1).$$

In addition, the roots of $1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p=0$ must lie outside the unit circle to ensure the stationarity of the model. For the pure AR model, admissible values of ϕ are limited in a p-dimensional hypercube $\mathbb{C}_p$.

In order to simplify the estimation procedure and ensure the admissibility of ϕ, we follow Barndorff-Nielsen and Schou (1973), to reparameterize ϕ as

$$\begin{aligned}\phi_p^{(p)} &= \gamma_p,\\ \phi_v^{(p)} &= \phi_v^{(p-1)}-\gamma_p\phi_{p-v}^{(p-1)},\end{aligned} \tag{2}$$

where $\phi_v^{(p)}$ is the v-th AR parameter in the AR(p) model given by Equation (1), and $\gamma_v=\phi_v^{(v)}$ represents the partial autocorrelation function, of lag v, for $v=1,\ldots,p-1$. With this notation, the matrix $M_{n_i}(\phi)$ can be written as:

$$M_{n_i}(\phi)=\begin{bmatrix}\gamma_0 & \gamma_1 & \cdots & \gamma_{n_i-1}\\ \gamma_1 & \gamma_0 & \cdots & \gamma_{n_i-2}\\ \vdots & \vdots & \ddots & \vdots\\ \gamma_{n_i-1} & \gamma_{n_i-2} & \cdots & \gamma_0\end{bmatrix}, \text{ for } i=1,\ldots,n.$$

The recursion given in Equation (2), can be used to define a transformation

$$\mathscr{B}:\gamma=(\gamma_1,\ldots,\gamma_p)^\top\to\phi=(\phi_1,\ldots,\phi_p)^\top,$$

which is one-to-one, continuous and differentiable inside the admissible region. This parameterization, has the advantage that, in the γ-space the admissible region is simply the p-dimensional cube with boundary surfaces corresponding to ±1, while in the ϕ-space it is very complicated. As an illustration, for $p=2$ the transformation is $\phi_1=\gamma_1(1-\gamma_2)$ and $\phi_2=\gamma_2$. For $p=3$, it can be written as $\phi_1=\gamma_1(1-\gamma_2)-\gamma_2\gamma_3$, $\phi_2=\gamma_2(1+\gamma_1\gamma_3)-\gamma_1\gamma_3$ and $\phi_3=\gamma_3$ (see, Schumacher et al., 2017).

As mentioned above, the proposed model also considers censored observations, *i.e.*, we assume that the response Y_{ij} is not fully observed for all i, j. Let $(\mathbf{V}_i,\mathbf{C}_i)$ be the observed data for the i-th subject, where $\mathbf{V}_i$ represents the vector of uncensored readings or censoring level and $\mathbf{C}_i$ is the vector of censoring indicators, such that:

$$\begin{aligned}y_{ij} &\le V_{ij} \text{ if } C_{ij}=1,\\ y_{ij} &= V_{ij} \text{ if } C_{ij}=0.\end{aligned} \tag{3}$$

Note that, since the observed response y_{ij} is defined over the real line, extensions to right-censored data are straightforward. In fact, the right-censored problem can be represented by a left-censored problem by simultaneously transforming the response y_{ij} and censoring level V_{ij} to $-y_{ij}$ and $-V_{ij}$. The model defined in Equations (1)-(3), is henceforth called the AR(p)-LMEC model.

2.1. The log-likelihood function

Following Vaida and Liu (2009), classic inference on the parameter vector $\theta=(\beta^\top,\sigma^2,\alpha^\top,\phi^\top)^\top$ is based on the marginal distribution of $\mathbf{y}_i$. For complete data, the marginal distribution of the vector $\mathbf{y}_i$, follows a multivariate normal distribution $\mathrm{N}_{n_i}(\mathbf{X}_i\beta,\Sigma_i)$, where $\Sigma_i=\Omega_i+\mathbf{Z}_i\mathbf{D}\mathbf{Z}_i^\top$, for $i=1,\ldots,n$. The strategy to compute the likelihood function associated with the AR(p)-LMEC model, defined by Equations (1)-(3), is to treat separately the observed and censored components of $\mathbf{y}_i$.

Let $\mathbf{y}_i^o$ be the n_i^o-vector of observed outcomes and $\mathbf{y}_i^c$ be the n_i^c-vector of censored observations for subject i $(n_i=n_i^o+n_i^c)$, such that $C_{ij}=0$ for all elements in $\mathbf{y}_i^o$, and $C_{ij}=1$ for all elements in $\mathbf{y}_i^c$. After reordering, $\mathbf{y}_i$, $\mathbf{V}_i$, $\mathbf{X}_i$ and Σ_i can be partitioned as follows:

$$\mathbf{y}_i=vec(\mathbf{y}_i^o,\mathbf{y}_i^c),\ \mathbf{V}_i=vec(\mathbf{V}_i^o,\mathbf{V}_i^c),\ \mathbf{X}_i^\top=(\mathbf{X}_i^o,\mathbf{X}_i^c) \text{ and } \Sigma_i=\begin{pmatrix}\Sigma_i^{oo} & \Sigma_i^{oc}\\ \Sigma_i^{co} & \Sigma_i^{cc}\end{pmatrix}.$$

In this setup, the operator $vec(\cdot)$ denotes the function with stack vectors or matrices of the same number of columns. Consequently, from the marginal-conditional decomposition of the multivariate normal distribution, $\mathbf{y}_i^o \sim \mathrm{N}_{n_i^o}(\mathbf{X}_i^o\beta, \Sigma_i^{oo})$ and $\mathbf{y}_i^c \mid \mathbf{y}_i^o \sim \mathrm{N}_{n_i^c}(\mu_i, \mathbf{S}_i)$, where $\mu_i = \mathbf{X}_i^c\beta + \Sigma_i^{co}(\Sigma_i^{oo})^{-1}(\mathbf{y}_i^o - \mathbf{X}_i^o\beta)$ and $\mathbf{S}_i = \Sigma_i^{cc} - \Sigma_i^{co}(\Sigma_i^{oo})^{-1}\Sigma_i^{oc}$. Now, let $\Phi_{n_i}(\mathbf{u};\mathbf{a},\mathbf{A})$ and $\phi_{n_i}(\mathbf{u};\mathbf{a},\mathbf{A})$ be the *cdf* (left tail) and *pdf*, respectively, of the multivariate normal distribution $\mathrm{N}_{n_i}(\mathbf{a},\mathbf{A})$, computed at vector $\mathbf{u}$. From Vaida and Liu (2009) and Matos et al. (2013a), the likelihood function for subject i (using conditional probability arguments) is given by:

$$\begin{aligned} L_i(\theta) &= f(\mathbf{y}_i^c \le \mathbf{V}_i^c \mid \mathbf{y}_i^o = \mathbf{V}_i^o, \theta) f(\mathbf{y}_i^o = \mathbf{V}_i^o \mid \theta) = f(\mathbf{y}_i^c \le \mathbf{V}_i^c \mid \mathbf{y}_i^o, \theta) f(\mathbf{y}_i^o \mid \theta) \\ &= \Phi_{n_i^c}(\mathbf{V}_i^c; \mu_i, \mathbf{S}_i)\phi_{n_i^o}(\mathbf{y}_i^o; \mathbf{X}_i^o\beta, \Sigma_i^{oo}), \end{aligned} \tag{4}$$

which can be easily evaluated computationally.

The log-likelihood function for the observed data, given by $\ell(\theta|\mathbf{y}) = \sum_{i=1}^n \log L_i(\theta)$, is used to compute different model selection criteria, such as:

$$\mathrm{AIC} = 2m - 2\ell_{max} \text{ and } \mathrm{BIC} = m\log N - 2\ell_{max},$$

where m is the number of the model parameters, $N = \sum_{i=1}^n n_i$ and ℓ_{max} is the maximized log-likelihood value.

3. The EM algorithm

In this section, we describe in detail, how the parameters of the proposed model specified in Equations (1)-(3) can be fitted by using the ECM algorithm (Meng and Rubin, 1993). The EM algorithm, (Dempster et al., 1977) has several appealing features, such as stability of monotone convergence with each iteration, increasing the likelihood and simplicity of implementation. Due to the computational difficulty at the M-step, we use the ECM algorithm (an extension of the EM algorithm), which shares the appealing features of the EM and converges faster than the original algorithm.

Let $\mathbf{y} = (\mathbf{y}_1^\top, \ldots, \mathbf{y}_n^\top)^\top$, $\mathbf{b} = (\mathbf{b}_1^\top, \ldots, \mathbf{b}_n^\top)^\top$, $\mathbf{V} = vec(\mathbf{V}_1, \ldots, \mathbf{V}_n)$ and $\mathbf{C} = vec(\mathbf{C}_1, \ldots, \mathbf{C}_n)$. Considering $\mathbf{b}$ as the hypothetical missing data, the complete data are denoted by:

$$\mathbf{y}_c = (\mathbf{C}^\top, \mathbf{V}^\top, \mathbf{y}^\top, \mathbf{b}^\top)^\top.$$

Hence, the ECM algorithm is applied to the complete data log-likelihood function

$$\begin{aligned} \ell_{c_i}(\theta \mid \mathbf{y}_c) &= -\frac{1}{2}\Big[n_i\log\sigma^2 + \log(|M_{n_i}(\phi)|) + \frac{1}{\sigma^2}(\mathbf{y}_i - \mathbf{X}_i\beta - \mathbf{Z}_i\mathbf{b}_i)^\top M_{n_i}^{-1}(\phi)(\mathbf{y}_i - \mathbf{X}_i\beta - \mathbf{Z}_i\mathbf{b}_i) \\ &+ \log|\mathbf{D}| + \mathbf{b}_i^\top\mathbf{D}^{-1}\mathbf{b}_i\Big] + K, \end{aligned} \tag{5}$$

with K being a constant that does not depend on the parameter vector θ. Given the current estimate $\theta = \widehat{\theta}^{(k)}$, the E-step calculates the conditional expectation of the complete data log-likelihood function, given by

$$Q\left(\theta \mid \widehat{\theta}^{(k)}\right) = E\left(\ell_c(\theta \mid \mathbf{y}_c) \mid \mathbf{V}, \mathbf{C}, \widehat{\theta}^{(k)}\right) = \sum_{i=1}^n Q_{1i}\left(\beta, \sigma^2 \mid \widehat{\theta}^{(k)}\right) + \sum_{i=1}^n Q_{2i}\left(\alpha \mid \widehat{\theta}^{(k)}\right),$$

where

$$\begin{aligned} Q_{1i}\left(\beta, \sigma^2, \phi \mid \widehat{\theta}^{(k)}\right) &= -\frac{n_i}{2}\log\widehat{\sigma^2}^{(k)} - \frac{1}{2}\log(|\widehat{M}_{n_i}^{(k)}(\phi)|) - \frac{1}{2\widehat{\sigma^2}^{(k)}}\Big[\widehat{a}_i^{(k)}(\widehat{\phi}^{(k)}) \\ &- 2\widehat{\beta}^{(k)\top}\mathbf{X}_i^\top\widehat{M}_{n_i}^{-1(k)}(\phi)\left(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{\mathbf{b}}_i^{(k)}\right) + \widehat{\beta}^{(k)\top}\mathbf{X}_i^\top\widehat{M}_{n_i}^{-1(k)}(\phi)\mathbf{X}_i\widehat{\beta}^{(k)}\Big], \end{aligned} \tag{6}$$

$$Q_{2i}\left(\alpha \mid \widehat{\theta}^{(k)}\right) = -\frac{1}{2}\log|\widehat{\mathbf{D}}^{(k)}| - \frac{1}{2}tr\left(\widehat{\mathbf{b}_i\mathbf{b}_i^\top}^{(k)}\widehat{\mathbf{D}}^{-1(k)}\right), \tag{7}$$

with $\widehat{a}_i^{(k)}(\phi) = tr\left(\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} M_{n_i}^{-1}(\phi) - 2\widehat{\mathbf{y}_i\mathbf{b}_i^\top}^{(k)}\mathbf{Z}_i^\top M_{n_i}^{-1}(\phi) + \widehat{\mathbf{b}_i\mathbf{b}_i^\top}^{(k)}\mathbf{Z}_i^\top M_{n_i}^{-1}(\phi)\mathbf{Z}_i\right)$, and

$$\begin{aligned}
\widehat{\mathbf{b}}_i^{(k)} &= E\left(\mathbf{b}_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\theta}^{(k)}\right) = \widehat{\varphi}_i^{(k)}(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{X}_i\widehat{\beta}^{(k)}),\\
\widehat{\mathbf{b}_i\mathbf{b}_i^\top}^{(k)} &= E\left(\mathbf{b}_i\mathbf{b}_i^\top \mid \mathbf{V}_i, \mathbf{C}_i,\ \widehat{\theta}^{(k)}\right)\\
&= \widehat{\Lambda}_i^{(k)} + \widehat{\varphi}_i^{(k)}(\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} - \widehat{\mathbf{y}}_i^{(k)}\widehat{\beta}^{(k)\top}\mathbf{X}_i^\top - \mathbf{X}_i\widehat{\beta}^{(k)}\widehat{\mathbf{y}}_i^{(k)\top} + \mathbf{X}_i\widehat{\beta}^{(k)}\widehat{\beta}^{(k)\top}\mathbf{X}_i^\top)\widehat{\varphi}_i^{(k)\top},\\
\widehat{\mathbf{y}_i\mathbf{b}_i^\top}^{(k)} &= E\left(\mathbf{y}_i\mathbf{b}_i^\top \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\theta}^{(k)}\right) = (\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} - \widehat{\mathbf{y}}_i^{(k)}\widehat{\beta}^{(k)\top}\mathbf{X}_i^\top)\widehat{\varphi}_i^{(k)\top},
\end{aligned}$$

with $\widehat{\Lambda}_i^{(k)} = (\widehat{\mathbf{D}}^{-1(k)} + \mathbf{Z}_i^\top\widehat{M}_{n_i}^{-1(k)}(\phi)\mathbf{Z}_i/\widehat{\sigma^2}^{(k)})^{-1}$, $\widehat{\varphi}_i^{(k)} = \widehat{\Lambda}_i^{(k)}\mathbf{Z}_i^\top\widehat{M}_{n_i}^{-1(k)}(\phi)/\widehat{\sigma^2}^{(k)}$ and $tr(\mathbf{A})$ being the trace function of the matrix $\mathbf{A}$.

It is easy to see, from Equations (6) and (7), that the E-step reduces only to the computation of

$$\widehat{\mathbf{y}_i\mathbf{y}_i^\top}^{(k)} = E\left(\mathbf{y}_i\mathbf{y}_i^\top \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\theta}^{(k)}\right) \text{ and } \widehat{\mathbf{y}}_i^{(k)} = E\left(\mathbf{y}_i \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\theta}^{(k)}\right).$$

These conditional expectations rely on the first and second moments of a multivariate truncated normal distribution and can be determined in closed-form, (for further details on the computation of these moments. see Vaida and Liu (2009) and Matos et al. (2013a)).

The conditional maximization step (CM) conditionally maximizes $Q(\theta|\widehat{\theta}^{(k)})$, with respect to θ, obtaining a new estimate $\widehat{\theta}^{(k+1)}$, as follows:

$$\begin{aligned}
\widehat{\beta}^{(k+1)} &= \left(\sum_{i=1}^n \mathbf{X}_i^\top\widehat{M}_{n_i}^{-1(k)}(\phi)\mathbf{X}_i\right)^{-1}\sum_{i=1}^n \mathbf{X}_i^\top\widehat{M}_{n_i}^{-1(k)}(\phi)\left(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{\mathbf{b}}_i^{(k)}\right),\\
\widehat{\sigma^2}^{(k+1)} &= \frac{1}{N}\sum_{i=1}^n\left[\widehat{a}_i^{(k)} - 2\widehat{\beta}^{(k+1)\top}\mathbf{X}_i^\top\widehat{M}_{n_i}^{-1(k)}(\phi)\left(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{\mathbf{b}}_i^{(k)}\right) + \widehat{\beta}^{(k+1)\top}\mathbf{X}_i^\top\widehat{M}_{n_i}^{-1(k)}(\phi)\mathbf{X}_i\widehat{\beta}^{(k+1)}\right],\\
\widehat{\mathbf{D}}^{(k+1)} &= \frac{1}{n}\sum_{i=1}^n\widehat{\mathbf{b}_i\mathbf{b}_i^\top}^{(k)},\\
\widehat{\gamma}^{(k+1)} &= \underset{\gamma\in(-1,1)^p}{\text{argmax}}\left(-\frac{1}{2}\log(|M_{n_i}(\phi)|) - \frac{1}{2\widehat{\sigma^2}^{(k+1)}}\left[\widehat{a}_i^{(k)}(\phi) + \widehat{\beta}^{(k+1)\top}\mathbf{X}_i^\top M_{n_i}^{-1}(\phi)\mathbf{X}_i\widehat{\beta}^{(k+1)}\right.\right.\\
&\quad \left.\left. - 2\widehat{\beta}^{(k+1)\top}\mathbf{X}_i^\top M_{n_i}^{-1}(\phi)\left(\widehat{\mathbf{y}}_i^{(k)} - \mathbf{Z}_i\widehat{\mathbf{b}}_i^{(k)}\right)\right]\right).
\end{aligned}$$

Finally, $\hat{\phi}^{(k+1)} = \mathscr{B}(\hat{\gamma}^{(k+1)})$, where the transformation $\mathscr{B}: \gamma \to \phi$ was discussed in Section 2. This process is iterated until some distance between two successive parameter estimates, such as $||\theta^{(k+1)} - \theta^{(k)}||$, becomes small enough, where $||\theta||$ denotes the Euclidean norm of the vector θ. The initial values can be calculated by taking the censored values to be observed ones and proceeding as in a usual LME model.

3.1. Estimation of random effects and standard errors

To estimate the random effects, we consider the conditional mean of $\mathbf{b}_i$ given the observed data $\mathbf{V}_i$ and $\mathbf{C}_i$, that is, $E(\mathbf{b}_i \mid \mathbf{V}_i, \mathbf{C}_i)$. Thus, for a given value of $\theta = (\beta^\top, \sigma^2, \alpha^\top, \phi^\top)^\top$, the conditional mean of $\mathbf{b}_i$ given $\mathbf{V}_i$ and $\mathbf{C}_i$ is:

$$\widehat{\mathbf{b}}_i(\theta) = E(\mathbf{b}_i \mid \mathbf{V}_i, \mathbf{C}_i) = \varphi_i(\widehat{\mathbf{y}}_i - \mathbf{X}_i\beta), \tag{8}$$

where $\varphi_i = \Lambda_i\mathbf{Z}_i^\top M_{n_i}^{-1}(\phi)/\sigma^2$ and $\Lambda_i = (\mathbf{D}^{-1} + \mathbf{Z}_i^\top M_{n_i}^{-1}(\phi)\mathbf{Z}_i/\sigma^2)^{-1}$. Note that $\widehat{\mathbf{y}}_i = E(\mathbf{y}_i|\mathbf{Q}_i, \mathbf{C}_i)$ is given by the first moment of a multivariate truncated normal distribution. In practice, the

estimator of $\mathbf{b}_i$, $\widehat{\mathbf{b}}_i$, can be obtained by substituting the ML estimates $\widehat{\theta}$ into Equation (8), leading to $\widehat{\mathbf{b}}_i = \widehat{\mathbf{b}}_i(\widehat{\theta})$. On the other hand, the conditional covariance matrix of $\mathbf{b}_i$, given $\mathbf{V}_i$ and $\mathbf{C}_i$, is

$$Var\left(\mathbf{b}_i \mid \mathbf{Q}_i, \mathbf{C}_i\right) = E\left(\mathbf{b}_i\mathbf{b}_i^\top \mid \mathbf{Q}_i, \mathbf{C}_i\right) - \widehat{\mathbf{b}}_i(\theta)\widehat{\mathbf{b}}_i(\theta)^\top = \Lambda_i + \varphi_i Var\left(\mathbf{y}_i \mid \mathbf{V}_i, \mathbf{C}_i\right)\varphi_i^\top.$$

Note that $Var(\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i)$ can be easily obtained as a byproduct of the proposed ECM algorithm developed previously.

3.2. The empirical information matrix

Following Lin (2010), we compute the asymptotic covariance of the ML estimates, through the empirical information matrix, which is computed as in Meilijson (1989)

$$\mathbf{I}_e(\theta \mid \mathbf{y}) = \sum_{i=1}^{n} \mathbf{s}(\mathbf{y}_i \mid \theta)\mathbf{s}^\top(\mathbf{y}_i \mid \theta) - \frac{1}{n}\mathbf{S}(\mathbf{y}_i \mid \theta)\mathbf{S}^\top(\mathbf{y}_i \mid \theta), \tag{9}$$

where $\mathbf{S}(\mathbf{y}_i \mid \theta) = \sum_{i=1}^{n} \mathbf{s}(\mathbf{y}_i \mid \theta)$ and $\mathbf{s}(\mathbf{y}_i \mid \theta)$ is the empirical score function for subject 'i'. According to Louis (1982), it is possible to relate the score function of the incomplete data log-likelihood, with the conditional expectation of the complete data log-likelihood function. Therefore, the individual score can be determined as:

$$\mathbf{s}(\mathbf{y}_i \mid \theta) = \frac{\partial \log f(\mathbf{y}_i \mid \theta)}{\partial \theta} = E\left(\frac{\partial \ell_{c_i}(\theta \mid \mathbf{y}_c)}{\partial \theta} \mid \mathbf{V}_i, \mathbf{C}_i, \theta\right), \tag{10}$$

where $\ell_i(\theta \mid \mathbf{y}_c)$ is the complete data log-likelihood, formed from the observation 'i'. Using the ML estimates $\widehat{\theta}$, that is, $\mathbf{S}(\mathbf{y}_i \mid \widehat{\theta}) = 0$, it follows that Equation (9), can be approximated by:

$$\mathbf{I}_e(\widehat{\theta} \mid \mathbf{y}) = \sum_{i=1}^{n} \widehat{\mathbf{s}}_i\widehat{\mathbf{s}}_i^\top, \tag{11}$$

where $\widehat{\mathbf{s}}_i = \mathbf{s}(\mathbf{y}_i \mid \widehat{\theta}) = \left(\widehat{\mathbf{s}}_{i,\beta}^\top, \widehat{\mathbf{s}}_{i,\sigma^2}^\top, \widehat{\mathbf{s}}_{i,\alpha}^\top, \widehat{\mathbf{s}}_{i,\phi}^\top\right)^\top$ has elements given by

$$\widehat{\mathbf{s}}_{i,\beta} = E\left(\frac{\partial \ell_{c_i}(\theta \mid \mathbf{y}_c)}{\partial \beta} \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\theta}\right) = \frac{1}{\widehat{\sigma^2}}\left[\mathbf{X}_i^\top \widehat{M}_{n_i}(\phi)^{-1}\left(\widehat{\mathbf{y}}_i - \mathbf{Z}_i\widehat{\mathbf{b}}_i\right) - \mathbf{X}_i^\top \widehat{M}_{n_i}(\phi)^{-1}\mathbf{X}_i\widehat{\beta}\right],$$

$$\begin{aligned}\widehat{\mathbf{s}}_{i,\sigma^2} &= E\left(\frac{\partial \ell_{c_i}(\theta \mid \mathbf{y}_c)}{\partial \sigma^2} \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\theta}\right) \\ &= -\frac{n_i}{2\widehat{\sigma^2}} + \frac{1}{2\widehat{\sigma^4}}\left[\widehat{h}_i - 2\widehat{\beta}^\top \mathbf{X}_i^\top \widehat{M}_{n_i}(\phi)^{-1}\left(\widehat{\mathbf{y}}_i - \mathbf{Z}_i\widehat{\mathbf{b}}_i\right) + \widehat{\beta}^\top \mathbf{X}_i^\top \widehat{M}_{n_i}(\phi)^{-1}\mathbf{X}_i\widehat{\beta}\right],\end{aligned}$$

$$\widehat{\mathbf{s}}_{i,\alpha} = E\left(\frac{\partial \ell_{c_i}(\theta \mid \mathbf{y}_c)}{\partial \alpha} \mid \mathbf{V}_i, \mathbf{C}_i, \widehat{\theta}\right) = -\frac{1}{2}tr\left(\widehat{D}^{-1}\frac{\partial D}{\partial \alpha}\widehat{D}^{-1}(\widehat{D} - \widehat{\mathbf{b}}_i\widehat{\mathbf{b}}_i^\top)\right),$$

with $\widehat{h}_i = tr\left(\widehat{\mathbf{y}_i\mathbf{y}_i^\top}\widehat{M}_{n_i}(\phi)^{-1} - 2\widehat{\mathbf{y}_i\mathbf{b}_i^\top}\mathbf{Z}_i^\top \widehat{M}_{n_i}(\phi)^{-1} + \widehat{\mathbf{b}_i\mathbf{b}_i^\top}\mathbf{Z}_i^\top \widehat{M}_{n_i}(\phi)^{-1}\mathbf{Z}_i\right)$.

Considering the reparameterization (2), the parts of the log-likelihood function that depends on ϕ can be written as $|\mathbf{M}_{n_i}(\phi)| = |\mathbf{M}_{p_i}(\phi)| = \prod_{j=1}^{p}(1-\gamma_j^2)^{-j}$ and $(\mathbf{y}_i - \mathbf{X}_i\beta - \mathbf{Z}_i\mathbf{b}_i)^\top M_{n_i}^{-1}(\phi)(\mathbf{y}_i - \mathbf{X}_i\beta - \mathbf{Z}_i\mathbf{b}_i) = \lambda^\top D(\mathbf{y}, \beta)\lambda$, with $\lambda^\top = (-1, \phi^\top) = (-1, \mathscr{B}(\gamma)^\top)$ and $D(\mathbf{y}, \beta)$ being the $(p+1)\times(p+1)$ matrix with the (r,s)-entry defined by:

$$D_{r,s} = D_{s,r} = \mathbf{d}_{(r,s)} \cdot \mathbf{d}_{(s,r)}, \tag{12}$$

where the vector $\mathbf{d}_{(r,s)} = (e_r, \ldots, e_{n+1-s})$, such that $\mathbf{e} = (\mathbf{y}_i - \mathbf{X}_i^\top\beta - \mathbf{Z}_i\widehat{\mathbf{b}}_i)$ and $A \cdot B$ is the internal product of vector A and B.

Now, to calculate the derivative $\frac{\partial \ell_{c_i}(\boldsymbol{\theta}|\mathbf{y}_c)}{\partial \boldsymbol{\phi}}$, we partition $D = D(\mathbf{y}_c, \beta)$ as:

$$D = \begin{bmatrix} D_{11} & D_{\phi 1}^{\top} \\ D_{\phi 1} & D_{\phi\phi} \end{bmatrix},$$

such that D_{11} is 1×1, $D_{\phi 1}$ is $p \times 1$ and $D_{\phi\phi}$ is $p \times p$. Then, the sum of squares in Equation (12) can be written as:

$$\lambda^{\top} D \lambda = \begin{pmatrix} -1 & \phi^{\top} \end{pmatrix} \begin{bmatrix} D_{11} & D_{\phi 1}^{\top} \\ D_{\phi 1} & D_{\phi\phi} \end{bmatrix} \begin{pmatrix} -1 \\ \phi \end{pmatrix} = D_{11} - 2\phi^{\top} D_{\phi 1} + 2\phi^{\top} D_{\phi\phi}\phi.$$

Therefore, we have:

$$\frac{\partial}{\partial \phi} \lambda^{\top} D \lambda = -2D_{\phi 1} + 2D_{\phi\phi}\phi.$$

Finally, the vector $\widehat{\mathbf{s}}_{i,\phi}$ will be defined as

$$\widehat{\mathbf{s}}_{i,\phi} = E\left(\frac{\partial \ell_{c_i}(\theta \mid \mathbf{y}_c)}{\partial \phi} \mid \mathbf{V}_i, \mathbf{C}_i, \theta \right) = \frac{1}{\widehat{\sigma^2}}(-D_{\phi 1} + D_{\phi\phi}\phi) + \frac{1}{2}\widehat{M}_{p_i}(\phi)^{-1} \frac{\partial M_{p_i}}{\partial \phi}.$$

4. Prediction of future observations

The problem related to the prediction of future values has a great impact on many practical applications. Rao (1987) pointed out that the predictive accuracy of future observations can be taken as an alternative measure of "goodness-of-fit". In order to propose a strategy to generate predicted values from the AR(p)-LMEC model, we use the approach proposed by Wang (2013). Thus, let $\mathbf{y}_{i,obs}$ be an observed response vector of dimension $n_{i,obs} \times 1$ for a new subject i over the first portion of time and $\mathbf{y}_{i,pred}$ be the corresponding $n_{i,pred} \times 1$ response vector over the future portion of time. Let $\bar{\mathbf{X}}_i = (\mathbf{X}_{i,obs}, \mathbf{X}_{i,pred})$ be the $(n_{i,obs} + n_{i,pred}) \times p$ design matrix corresponding to $\bar{\mathbf{y}}_i = (\mathbf{y}_{i,obs}^{\top}, \mathbf{y}_{i,pred}^{\top})$.

To deal with the censored values existing in $\mathbf{y}_{i,obs}$, we use the imputation procedure, by replacing the censored values by $\widehat{\mathbf{y}}_i = E(\mathbf{y}_i|\mathbf{V}_i, \mathbf{C}_i, \widehat{\theta})$ obtained from the EM algorithm. Therefore, when the censored values are imputed, a complete dataset, denoted by $\mathbf{y}_{i,obs^*}$, is obtained. The reason to use the imputation procedure is that it avoids computing truncated conditional expectations of the multivariate normal distribution originated by the censoring scheme. Hence, we have that

$$\bar{\mathbf{y}}_i^* = \left(\mathbf{y}_{i,obs^*}^{\top}, \mathbf{y}_{i,pred}^{\top} \right)^{\top} \sim N_{n_{i,obs}+n_{i,pred}} \left(\mathbf{X}_i\beta, \Sigma_i \right),$$

where the matrix Σ_i, can be represented by $\Sigma_i = \begin{pmatrix} \Sigma_i^{obs^*,obs^*} & \Sigma_i^{obs^*,pred} \\ \Sigma_i^{pred,obs^*} & \Sigma_i^{pred,pred} \end{pmatrix}$. As mentioned in Wang (2013), the best linear predictor of $\mathbf{y}_{i,pred}$ with respect to the minimum mean squared error (MSE) criterion is the conditional expectation of $\mathbf{y}_{i,pred}$ given $\mathbf{y}_{i,obs^*}$, which is given by

$$\widehat{\mathbf{y}}_{i,pred}(\theta) = \mathbf{X}_{i,pred}\beta + \Sigma_i^{pred,obs^*} \Sigma_i^{obs^*,obs^*-1} \left(\mathbf{y}_{i,obs^*} - \mathbf{X}_{i,obs^*}\beta \right). \tag{13}$$

Therefore, $\mathbf{y}_{i,pred}$ can be estimated directly, by substituting $\widehat{\theta}$ into Equation (13), leading to $\widehat{\widehat{\mathbf{y}}_{i,pred}} = \widehat{\mathbf{y}}_{i,pred}(\widehat{\theta})$.

5. The nonlinear case

As mentioned in the Introduction, some approximations based on the EM algorithm have been proposed in the statistical literature to deal with NLME models. In this context, we use an approximation of the nonlinear functions mentioned by Vaida and Liu (2009). This approximation (15) was considered by Matos et al. (2013a) in the context of censored nonlinear mixed-effects models. In that paper, simulation studies revealed that the approximation can efficiently estimate the model parameters. Wang (2013) used this approximation to implement an ECM algorithm to carry out ML estimation in Student-t nonlinear mixed-effects models for multi-outcome longitudinal data with missing values. Consequently, we conclude that this approximation is robust, stable, and does not produce any severe consequences in inference when applied to other types of (censored) nonlinear models.

The NLME (without censoring) of Pinheiro and Bates (2000) is defined as:

$$\mathbf{y}_i = \eta(\psi_i, \mathbf{X}_i) + \varepsilon_i, \ \psi_i = \mathbf{A}_i\beta + \mathbf{B}_i\mathbf{b}_i, \ i = 1, \ldots, n, \tag{14}$$

where $\mathbf{b}_i \overset{iid}{\sim} \mathrm{N}_q(\mathbf{0}_q, \mathbf{D})$ and $\varepsilon_i \overset{ind}{\sim} \mathrm{N}_{n_i}(\mathbf{0}_{n_i}, \sigma^2\mathrm{E}_i)$ are independent; $\mathbf{y}_i$ is an $(n_i \times 1)$ vector of observed responses for subject i; η is a nonlinear function of the individual random parameter ψ_i; $\mathbf{A}_i$ and $\mathbf{B}_i$ are known design matrices of dimensions $r \times p$ and $r \times q$, respectively, possibly depending on some covariate values; β is the $(p \times 1)$ vector of fixed effects; and $\mathbf{b}_i$ is the $(q \times 1)$ vector of random effects.

As mentioned by Vaida and Liu (2009), the linearization (L) procedure to obtain the approximate MLE of $\theta = (\beta^\top, \sigma^2, \alpha^\top, \phi^\top)^\top$ involves taking the first-order Taylor expansion of η_i around the current parameter estimate $\widetilde{\beta}$ and the random effect estimates $\widetilde{\mathbf{b}}_i$ (empirical predictors). This procedure is equivalent to iteratively solving the following LME model (L-step):

$$\widetilde{\mathbf{Y}}_i = \widetilde{\mathbf{W}}_i\beta + \widetilde{\mathbf{H}}_i\mathbf{b}_i + \varepsilon_i, \ i = 1, \ldots, n, \tag{15}$$

where $\mathbf{b}_i \overset{iid}{\sim} \mathrm{N}_q(\mathbf{0}_q, \mathbf{D})$ and $\varepsilon_i \overset{ind}{\sim} \mathrm{N}_{n_i}(\mathbf{0}_{n_i}, \sigma^2\mathrm{E}_i)$; and $\widetilde{\mathbf{Y}}_i = \mathbf{Y}_i - \eta(\psi(\widetilde{\beta}, \widetilde{\mathbf{b}}_i), \mathbf{X}_i)$, with

$$\widetilde{\mathbf{H}}_i = \frac{\partial \eta(\mathbf{A}_i\beta + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_i)}{\partial \mathbf{b}_i^\top}\Big|_{\mathbf{b}_i = \widetilde{\mathbf{b}}_i} \quad \text{and} \quad \widetilde{\mathbf{W}}_i = \frac{\partial \eta(\mathbf{A}_i\beta + \mathbf{B}_i\mathbf{b}_i, \mathbf{X}_i)}{\partial \beta_i^\top}\Big|_{\beta_i = \widetilde{\beta}_i}.$$

Thus, in the censored case, the model in (15) is an LME with censored data that can be fitted using the strategy explained in Section 3. The model matrices in (15) depend on the current parameter value, and need to be recalculated at each iteration. The algorithm iterates between the L-, E- and CM-steps until convergence.

6. Simulation studies

In order to examine the performance of our proposed model, we develop three simulation studies, considering the linear and nonlinear cases: (i) the first simulation study shows the asymptotic behavior of the parameter estimates, when the sample size is increasing; (ii) the goal of the second simulation is to study the consistency of the standard errors of the ML estimates and (iii) the third simulation shows the performance of the prediction of future values.

6.1. The linear case

For the linear case, we perform a simulation studies based on the AR(p)-LMEC model defined in (1)-(3), considering a left-censored setting with same number of measurements for each individual i, i.e., $n_i = 6$. As in Schumacher et al. (2017), we assume an autoregressive dependence for the error term of order $p = 2$ and covariate vector for each individual given by $\mathbf{X}_i = (\mathbf{1}_6, \ \mathbf{x}_{1i}, \ \mathbf{x}_{2i})$, where each vector $\mathbf{x}_{ti}$ was generated independently, from a uniform distribution $U(-1, 1)$ and $\mathbf{1}_{n_i}$ represents a $1 \times n_i$ vector of ones. We consider $\mathbf{Z}_i = (\mathbf{1}_6, \ z_{1i})$,

with $z_{1i} = (1,\ 2,\ 3,\ 4,\ 5,\ 6)^\top$, as the covariate vector associated with the random effect. It is important to stress that, all these values were fixed throughout the replications.

For all simulation schemes, the values of the parameters were set at: $\beta = (\beta_1, \beta_2, \beta_3)^\top = (1, 2, 1)^\top$, $\sigma^2 = 0.48$ and $\phi = (\phi_1, \phi_2)^\top = (0.48, -0.2)^\top$. The random effect $\mathbf{b}_i = (b_{1i}, b_{2i})^\top$ was generated, previously, from a bivariate normal distribution $\mathrm{N}_2(\mathbf{0}_2, \mathbf{D})$, with

$$\mathbf{D} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} = \begin{bmatrix} 0.049 & 0.001 \\ 0.001 & 0.002 \end{bmatrix},$$

as presented by Matos et al. (2016). We generated left-censored Monte Carlo data with $l\%$ of censoring level, that is, $l\%$ of the observations, in each dataset, were left censored.

- **Simulation study 1:**

  We generated $R = 100$ datasets for each one of the combinations, of different samples sizes $n = \{50,\ 100,\ 200,\ 400,\ 600\}$ and censoring levels $l\% = \{0\%,\ 5\%,\ 20\%,\ 40\%\}$. Considering our proposed EM algorithm, we compute the bias (Bias) and the mean squared error (MSE) for the parameter estimates $\widehat{\theta}_i$, over the R samples and twenty (20) different situations. These measures are defined by:

$$\text{Bias}\left(\widehat{\theta}_i\right) = \frac{1}{R}\sum_{j=1}^{R}\left(\widehat{\theta}_i^{(j)} - \theta_i\right) \text{ and } \text{MSE}\left(\widehat{\theta}_i\right) = \frac{1}{R}\sum_{j=1}^{R}\left(\widehat{\theta}_i^{(j)} - \theta_i\right)^2,$$

  where $\widehat{\theta}_i^{(j)}$ denotes the ML estimate of parameter θ_i, for the j-th replication, with $j = 1, \ldots, R$.

  The results are shown in Figures 1 and 2. It can observed that, in general, the Bias and MSE tend to zero, when the sample size increases, indicating that the estimates, based on our proposed EM-type algorithm have good asymptotic properties.

- **Simulation study 2:**

  In this second simulation study, we fixed the sample size at $n = 100$ and generated $R = 100$ Monte Carlo samples, with different censoring levels $l\% = \{0\%,\ 5\%,\ 10\%,\ 20\%,\ 40\%\}$. In order to show that the approximated method, suggested in Section 3.2, provides reasonable SE of the ML estimates and has good asymptotic properties, we analyzed the SE of the parameter estimates (MC-SE), the average values of the standard errors computed, using the empirical information matrix (MC-IM-SE), and the percentage of times in the R samples that the 95% asymptotic confidence intervals contained the true parameter values (MC-COV). These measures are defined by:

$$\text{MC-IM-SE}\left(\widehat{\theta}_i\right) = \frac{1}{R}\sum_{j=1}^{R}\widehat{SE\left(\theta_i\right)}^{(j)} \text{ and } \text{MC-SE}\left(\widehat{\theta}_i\right) = \sqrt{\frac{1}{R-1}\left(\sum_{j=1}^{R}\left(\widehat{\theta}_i^{(j)}\right)^2 - \left(\frac{1}{R}\sum_{j=1}^{R}\widehat{\theta}_i^{(j)}\right)^2\right)},$$

  where $\widehat{\theta}_i^{(j)}$ and $\widehat{SE\left(\theta_i\right)}^{(j)}$ represent the ML estimates of parameter θ_i and the SE estimate of $\widehat{\theta}_i^{(j)}$, respectively.

  The results are given in Table 1. This table shows a reasonable MC coverage of the parameters $\beta = (\beta_1, \beta_2, \beta_3)^\top$ and ϕ_2, although, the values of σ^2 and ϕ_1 tend to be lower than the nominal level 95%. Taking into account the moderate sample size ($n = 100$), we consider these results quite satisfactory. Similar results were obtained by Schumacher et al. (2017) in the context of AR(p) censored linear models.

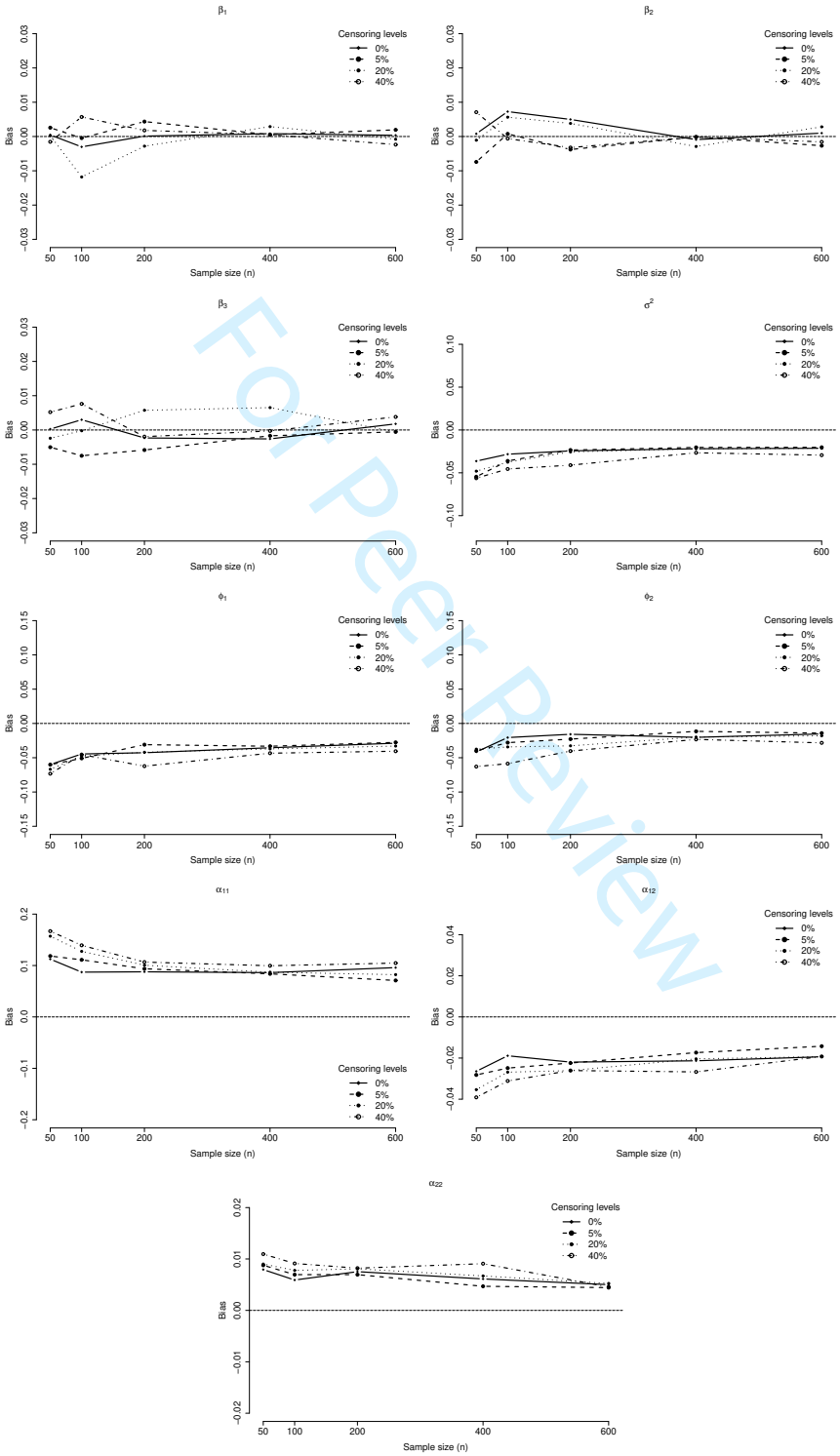

Figure 1: Simulation study - linear case. Average Bias of parameter estimates of the AR(p)-LMEC model, considering different sample sizes "*n*" and censoring levels "*l%*".

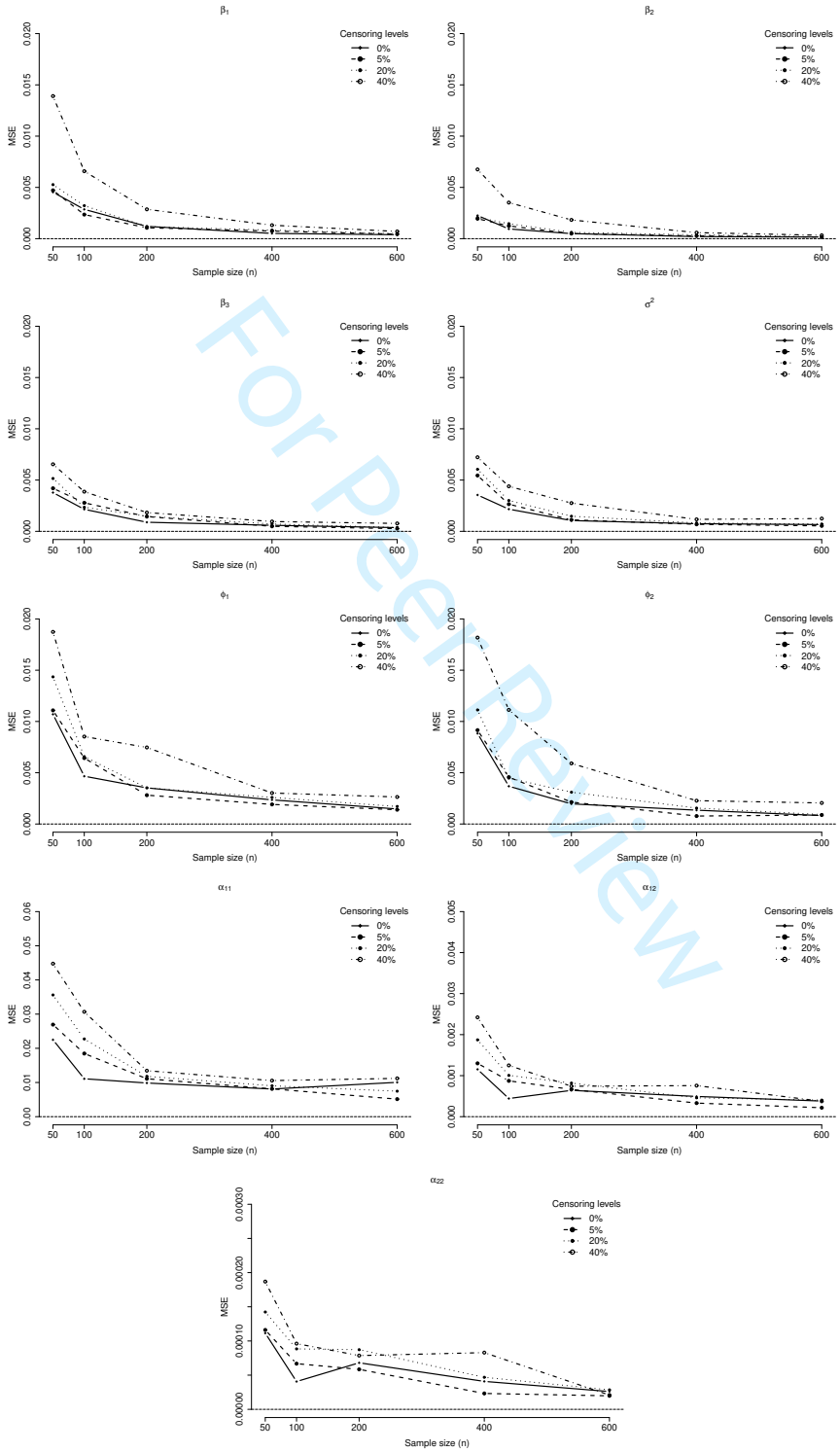

Figure 2: Simulation study - linear case. Average MSE of parameter estimates of the AR(p)-LMEC model, considering different sample sizes "*n*" and censoring levels "*l*%".

Table 1: Simulation study - linear case. Standard errors of parameter estimates (MC-SE), average values of the standard errors (MC-IM-SE) and MC-COV

Censoring levels "l%"		Parameters					
		β_1	β_2	β_3	σ^2	ϕ_1	ϕ_2
0%	MC-IM-SE	0.063	0.030	0.047	0.030	0.050	0.055
	MC-SE	0.048	0.034	0.054	0.039	0.066	0.061
	MC-COV	94%	97%	95%	87%	87%	93%
5%	MC-IM-SE	0.065	0.032	0.047	0.031	0.051	0.056
	MC-SE	0.053	0.032	0.045	0.038	0.066	0.059
	MC-COV	97%	93%	94%	88%	93%	90%
10%	MC-IM-SE	0.065	0.032	0.045	0.031	0.054	0.056
	MC-SE	0.052	0.027	0.051	0.038	0.056	0.073
	MC-COV	95%	96%	96%	86%	93%	92%
20%	MC-IM-SE	0.069	0.035	0.050	0.031	0.057	0.061
	MC-SE	0.052	0.036	0.054	0.036	0.063	0.055
	MC-COV	95%	95%	94%	87%	89%	90%
40%	MC-IM-SE	0.085	0.050	0.057	0.027	0.076	0.073
	MC-SE	0.068	0.048	0.060	0.054	0.087	0.088
	MC-COV	95%	94%	96%	89%	88%	92%

- **Simulation study 3:**

  Here, we fixed the sample size $n = 100$ and compared the predicted values of the AR(p)-LMEC model by using the approach proposed in Section 4. We considered (i) an autoregressive dependence of order $p = \{1, 2\}$, i.e. $AR(1)$ and $AR(2)$ and (ii) an uncorrelated structure (UNC) for the error term. In order to do that, $R = 100$ Monte Carlo samples of size $n = 100$ were generated from an AR(2)-LMEC model with two censoring levels $l\% = \{5\%, 20\%\}$. Then, we obtained the predicted values of the last two observations (one and two step ahead forecast), for each subject. A suggested by Schumacher et al. (2017) and Garay et al. (2017), the comparison was made by using the mean absolute prediction error (MAPE) and mean square prediction error (MSPE), which are defined by:

$$\text{MAPE} = \frac{1}{R \times h} \sum_i \sum_j |y_{ij} - \widehat{y}_{ij}| \quad \text{and} \quad \text{MSPE} = \frac{1}{R \times h} \sum_i \sum_j (y_{ij} - \widehat{y}_{ij})^2,$$

  where $h = \{1,2\}$ denotes the step number ahead. $y_{i,j}$ and $\widehat{y}_{i,j}$ represent the original and predicted value of the j-th observation, for the i-th subject, respectively, with $i = 1,\ldots,n$ and $j = 5,\ 6$.
  Table 2 shown the comparison between the predicted values (one and two step ahead forecast) with the real ones, in the three models and two censoring levels $l\%$. It can be seen that, as expected, the AR(2)-LMEC model provides better predictive results than the UNC-LMEC and AR(1)-LMEC models under both censoring levels $l\% = \{5\%,\ 20\%\}$.

6.2. The nonlinear case

As in the linear case, we performed three simulation studies based in the right-censored AR(p)-NLMEC model with the same number of measurements ($n_i = 10$) for each subject. As in Matos et al. (2016), we considered the following logistic model:

$$y_{ij} = \delta_{1i} + \frac{\delta_2}{1 + \exp\left(\frac{x_{ij} - \delta_3}{\delta_{4i}}\right)} + \varepsilon_{ij},$$

where $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{in_i})^\top$ and $\varepsilon_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{in_i})^\top \sim \mathrm{N}_{n_i}(\mathbf{0}_{n_i}, \sigma^2 M_{n_i}(\phi))$. $M_{n_i}(\phi)$ assumes a first-order autoregressive structure (AR(1)), $\mathbf{x}_{ij}$ is generated from a discrete uniform

Table 2: Simulation study - linear case. Mean absolute prediction error (MAPE) and mean square prediction error (MSPE) for different correlation structures

	$l\% = 5\%$					
	MAPE			MSPE		
	UNC	AR(1)	AR(2)	UNC	AR(1)	AR(2)
One step	0.70	0.61	**0.59**	0.80	0.57	**0.57**
Two step	0.66	0.60	**0.57**	0.69	0.55	**0.53**
	$l\% = 20\%$					
	MAPE			MSPE		
	UNC	AR(1)	AR(2)	UNC	AR(1)	AR(2)
One step	0.69	0.66	**0.65**	0.83	0.70	**0.54**
Two step	0.70	0.64	**0.63**	0.78	0.66	**0.63**

distribution $U_d(0,90)$. To force the parameters to be positive we reparameterized the model to $\delta_{1i} = \exp(\beta_1 + b_{1i})$, $\delta_2 = \exp(\beta_2)$, $\delta_3 = \exp(\beta_3)$ and $\delta_{4i} = \exp(\beta_4 + b_{2i})$.

For all simulation schemes, the parameter values were set at: $\beta = (\beta_1,\ \beta_2,\ \beta_3,\ \beta_4)^\top = (1.6094,\ 0.6931,\ 3.8067,\ 2.3026)^\top$, $\sigma^2 = 0.55$ and $\phi = 0.58$. The random effects $\mathbf{b}_i = (b_{1i},\ b_{2i})^\top$ were generated from a bivariate normal distribution $\mathrm{N}_2(\mathbf{0}_2, \mathbf{D})$, with

$$\mathbf{D} = \begin{bmatrix} \alpha_{11} & \\ \alpha_{21} & \alpha_{22} \end{bmatrix} = \begin{bmatrix} 0.0025 & -0.0010 \\ -0.0010 & 0.0100 \end{bmatrix}.$$

- **Simulation study 1:**

  In this case, we generated $R = 100$ datasets, for each one of the combinations of different samples size $n = \{50,\ 100,\ 200,\ 400,\ 600\}$ and censoring levels $l\% = \{5\%,\ 10\%,\ 20\%\}$. As in the linear case, we estimate the parameter models using the proposed EM algorithm and computed the bias (Bias) and the mean squared error (MSE), defined previously, for $\widehat{\theta}$, over the R samples, where $\theta = (\beta^\top, \sigma^2, \alpha^\top, \phi)^\top$.

  The results are shown in Figures 3 and 4. It can be seen that, as in a linear case, the Bias and MSE tend to zero, when the sample size increases, indicating that the EM estimates have good asymptotic properties. That is, our proposed EM algorithm produces consistent estimators, as expected.

- **Simulation study 2:**

  In this simulation study, we fixed the sample size at $n = 100$ and generated $R = 100$ Monte Carlo samples, with different censoring levels $l\% = \{5\%,\ 10\%,\ 20\%\}$. As in the linear case, we analyzed the MC-SE, MC-IM-SE and MC-COV, for each censoring level. The results are presented in Table 3 which shows a reasonable MC coverage for the all parameters, including σ^2 and ϕ. Thus, taking into account the moderate sample size $(n = 100)$, we consider these results satisfactory, meaning the proposed asymptotic approximation for the SE of the parameters is reliable.

- **Simulation study 3:**

  Here we set the sample size at $n = 100$ to evaluate the predicted performance of the the AR(p)-NLMEC model. We considered: (i) an autoregressive dependence of order $p = \{1,\ 2\}$ and (ii) an uncorrelated structure (UNC) for the error term. Thus, $R = 100$ datasets were generated from an AR(2)-NLMEC model with $\phi = (\phi_1, \phi_2)^\top = (0.58, -0.1)^\top$, and two censoring levels $l\% = \{5\%,\ 20\%\}$. We recorded the predicted values of the last two observations (one and two step ahead forecast) for each subject

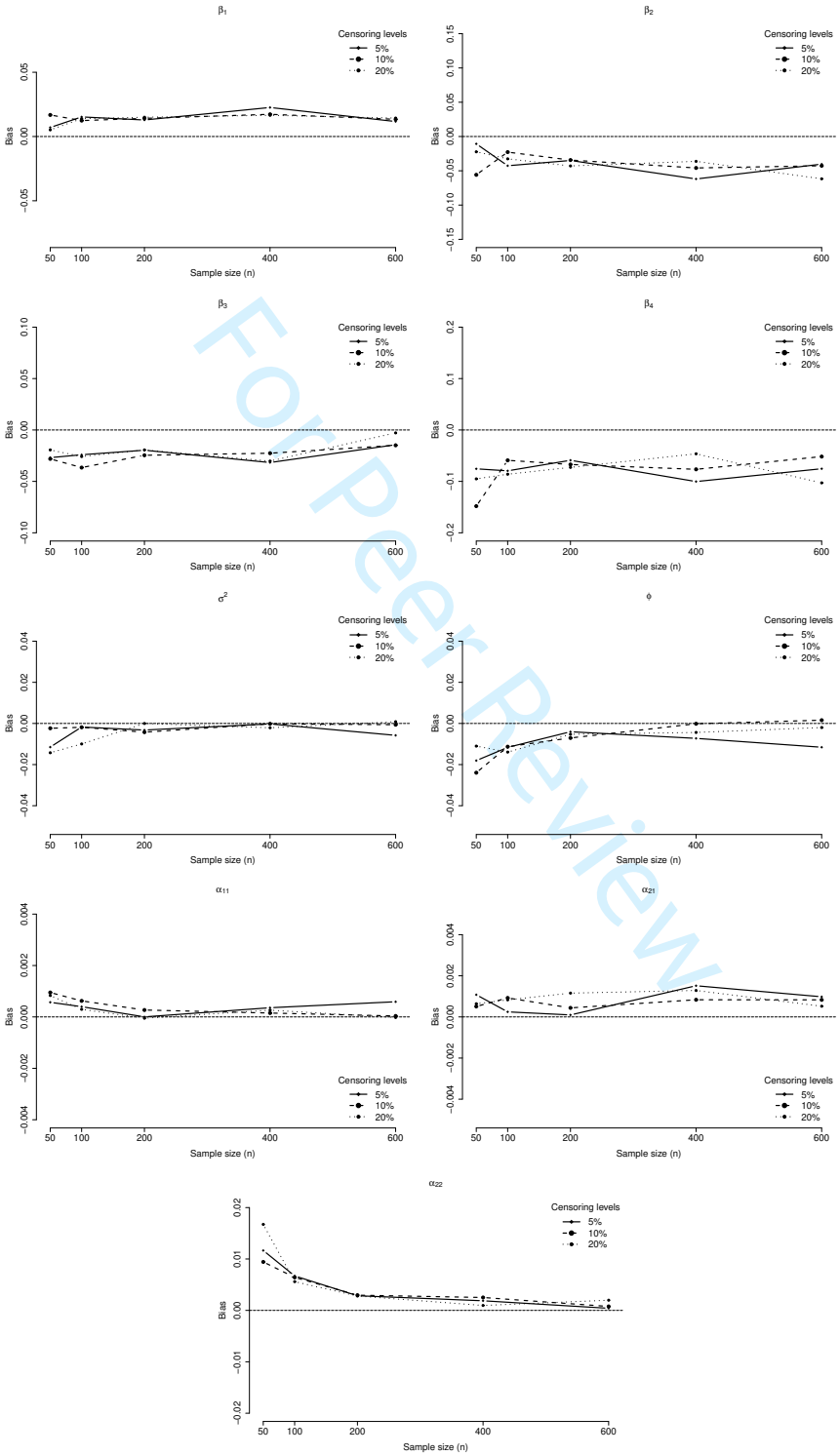

Figure 3: Simulation study - nonlinear case. Average Bias of parameter estimates of the AR(p)-NLMEC model, considering different sample sizes and censoring levels $l\%$.

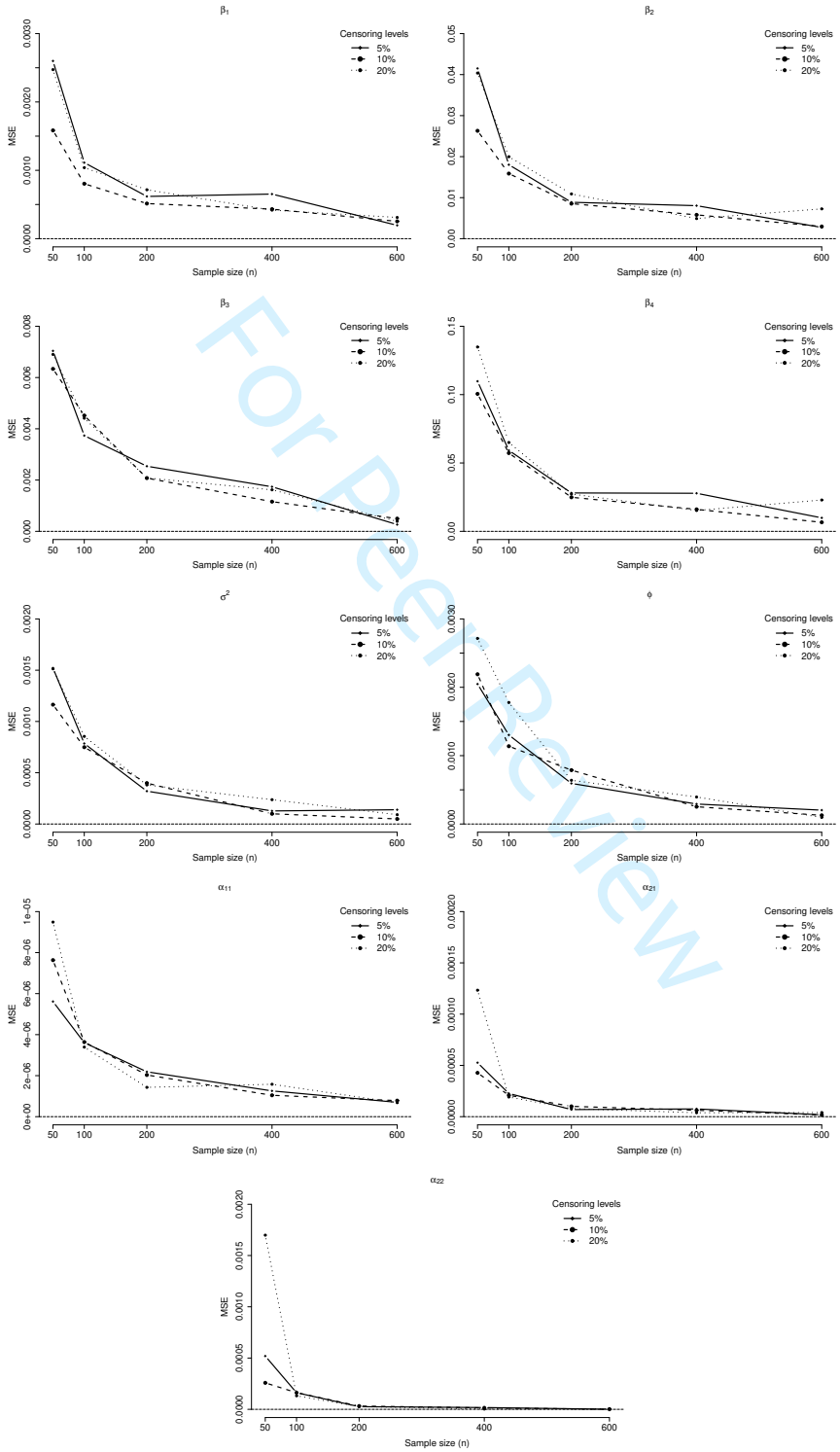

Figure 4: Simulation study - nonlinear case. Average MSE of parameter estimates of the AR(p)-NLMEC model, considering different sample sizes and censoring levels *l*%.

Table 3: Simulation study - nonlinear case. Standard errors of parameter estimates (MC-SE), average values of the standard errors (MC-IM-SE) and MC-COV

Censoring levels "l%"		Parameters					
		β_1	β_2	β_3	β_4	σ^2	ϕ
5%	MC-IM-SE	0.031	0.142	0.059	0.245	0.030	0.003
	MC-SE	0.030	0.128	0.057	0.231	0.028	0.034
	MC-COV	96%	93%	95%	93%	95%	95%
10%	MC-IM-SE	0.031	0.144	0.061	0.248	0.030	0.003
	MC-SE	0.026	0.125	0.057	0.233	0.027	0.032
	MC-COV	91%	95%	90%	92%	92%	95%
20%	MC-IM-SE	0.030	0.137	0.061	0.233	0.029	0.003
	MC-SE	0.029	0.138	0.062	0.241	0.028	0.040
	MC-COV	96%	95%	94%	92%	94%	94%

in order to analyze the predictive performance compared with the real one. As in the linear case, we used the discrepancy measures MAPE and MSPE previously defined. The results are presented in Table 4, which shows that, as expected, for the two censoring levels l% the true AR(2)-NLMEC model generates better predictive results than the UNC-NLMEC and AR(1)-NLMEC models.

Table 4: Simulation study - nonlinear case. Mean absolute prediction error *MAPE* and mean square prediction error *MSPE* for different correlation structures

	$l\% = 5\%$					
	MAPE			MSPE		
	UNC	AR(1)	AR(2)	UNC	AR(1)	AR(2)
One step	0.65	0.60	**0.59**	0.75	0.71	**0.69**
Two step	0.70	0.66	**0.62**	0.80	0.65	**0.59**
	$l\% = 20\%$					
	MAPE			MSPE		
	UNC	AR(1)	AR(2)	UNC	AR(1)	AR(2)
One step	0.62	0.55	**0.51**	0.58	0.47	**0.41**
Two step	0.65	0.61	**0.56**	0.61	0.55	**0.49**

7. Application

In this section, we apply the proposed techniques to the "*ACTG-315*" dataset, previously analyzed by Wu (2002) and Matos et al. (2016). This dataset involves 46 HIV-1 infected patients treated with a potent antiretroviral drug cocktail, based on the protease inhibitor ritonavir and two reverse transcription inhibitor drugs (zidovudine and lamivudine). Before initiating the antiretroviral therapy, all patients discontinued their antiretroviral regimen for five weeks as a "washout" period. The aim of this antiretroviral regimen is to show that immunity can be partially restored in people with moderately advanced HIV disease. The viral load was quantified on days 0, 2, 7, 10, 14, 21, 28, 56, 84, 168 and 196 after starting treatment. The dataset contains 361 observations, but we excluded the subject that had fewer than 4 measurements, in order to obtain a dataset that allows using different autoregressive structure orders. Therefore, the analyzed dataset was reduced to 357 observations of 45 subjects.

An immunologic marker known as CD4+ cell count was also measured along with viral load and 72 out of 357 (20.16%) CD4 values were missing due to a mismatch of the CD4 and the viral load measurement schedules. Viral load measurements below the detectable threshold of 100 copies/mL occurred in 31 out of 357 observations(8.6%).

For this the dataset, following Matos et al. (2016), we considered the nonlinear mixed effects model given by:

$$\begin{aligned} y_{ij} &= \log_{10}(\exp(\lambda_1)+\exp(\lambda_2))+\varepsilon_{ij}, \\ \lambda_1 &= \beta_1+b_{1i}-(\beta_2+b_{2i}t_{ij}), \\ \lambda_2 &= \beta_3+b_{3i}-(\beta_4+\beta_5 CD4_{ij}+b_{4i})t_{ij}, \end{aligned}$$

where y_{ij} is the log_{10}-transformation of the viral load for subject "i" at time "t_{ij}", $CD4_{ij}$ indicates the observed CD4 values up to time t_{ij}, $b_i=(b_{1i},\ldots,b_{4i})$ is the random effects vector, for subject "i" and ε_{ij} represents the within-individual random error, where $i=1,\ldots,45$ and $j=1,\ldots,n_i$.

We applied the AR(p)-NLMEC model defined in (3)-(14), considering four correlation structures, namely: (a) the uncorrelated (UNC) structure; (b) the continuous-time AR(1) structure; (c) the continuous-time AR(2) structure; and (d) the continuous-time AR(3) structure. Table 5 shows the ML estimates of the parameters, in the AR(p)-NLMEC models, together with their corresponding standard errors. In general, the ML estimates $\widehat{\beta}$ are quite similar, for all scenarios. Table 6 presents some model selection criteria, together, with the values of the log-likelihood. It can be seen from this table, that the AR(3)-NLMEC model produces more accurate estimates.

In order to evaluate the prediction performance of our approach, we used the same strategy developed by Garay et al. (2017). Thus, we considered the subjects that had at least five measures (45 subjects), at least eight measures (31 subjects) and at least nine measures (14 subjects), and the last two measures were predicted. As in the simulation study (Section 6) we computed the MAPE and MSPE measures of the predicted values under different models (UCN, AR(1)-AR(3), for comparison. Once again the AR(3)-NLMEC models had the best performance in terms of prediction.

Figure 5 depicts the prediction performance of four randomly selected subjects (#7,#18,#30,#40), with at least eight measures, in the scenarios that produce more accurate estimates. It can seen from this figure that the AR(3)-NLMEC model generates predictive values close to the real ones, as expected.

Table 5: ACTG-315 dataset. ML estimates (Est) and standard errors (SE).

	Correlation structures							
	UNC		AR(1)		AR(2)		AR(3)	
Parameters	Est	SE	Est	SE	Est	SE	Est	SE
β_1	11.670	0.191	11.630	0.280	11.604	0.281	11.590	0.251
β_2	32.209	0.092	33.398	0.012	33.463	0.011	31.619	0.086
β_3	6.653	0.393	6.867	0.337	6.871	0.390	6.770	0.424
β_4	-0.358	0.697	-0.306	0.824	-0.229	0.950	-0.733	0.780
β_5	0.378	0.218	0.391	0.156	0.372	0.167	0.514	0.172
σ^2	0.126	0.015	0.195	0.023	0.207	0.028	0.198	0.020
α_{11}	1.046	0.112	0.587	0.201	0.250	0.513	0.178	0.537
α_{12}	-5.228	0.137	-1.493	0.121	-1.105	0.424	-0.857	0.501
α_{22}	43.631	0.168	7.183	0.355	6.194	0.630	4.317	0.525
α_{13}	0.656	0.041	0.600	0.645	0.149	0.782	0.039	0.893
α_{23}	3.369	0.015	1.178	0.193	1.679	0.212	1.557	0.214
α_{33}	3.489	0.008	1.436	0.517	0.876	0.609	0.850	0.745
α_{14}	-0.534	0.032	-0.396	0.460	-0.528	0.510	-0.472	0.563
α_{24}	13.676	0.322	9.850	0.469	9.855	0.920	8.154	0.744
α_{34}	4.127	0.054	2.119	0.088	2.066	0.093	2.040	0.085
α_{44}	11.209	0.019	9.141	0.073	9.376	0.081	8.274	0.093
ϕ_1	- -	- -	0.525	0.118	0.579	0.127	0.540	0.087
ϕ_2	- -	- -	- -	- -	0.079	0.122	-0.038	0.126
ϕ_3	- -	- -	- -	- -	- -	- -	0.271	0.194

Table 6: ACTG-315 dataset. Comparison between the AR(p)-NLMEC models, considering different orders of correlation structures.

	UNC	AR(1)	AR(2)	AR(3)
loglik	-281.704	-279.613	-280.421	**-276.253**
AIC	595.408	593.225	596.842	**590.505**
BIC	657.452	659.147	666.641	**664.182**
AICc	597.008	595.030	598.866	**592.760**

Table 7: ACTG-315 dataset. Evaluation of the prediction accuracy for the AR(p)-NLMEC models, with different correlations structures.

Forecast	For individuals with $n_i \geq 5$ (45 individuals)							
	MAPE				MSPE			
	UNC	AR(1)	AR(2)	AR(3)	UNC	AR(1)	AR(2)	AR(3)
One step	0.8475	0.8907	0.8818	**0.8441**	1.1542	1.2652	1.2406	**1.1411**
Two step	0.7146	0.7073	0.7045	**0.6955**	0.9088	0.8665	0.8547	**0.8597**
Forecast	For individuals with $n_i \geq 8$ (31 individuals)							
	MAPE				MSPE			
	UNC	AR(1)	AR(2)	AR(3)	UNC	AR(1)	AR(2)	AR(3)
One step	0.8308	0.8948	0.8882	**0.8214**	1.1200	1.2915	1.2660	**1.1085**
Two step	0.6600	0.6696	0.6709	**0.6394**	0.7986	0.7961	0.7970	**0.7556**
Forecast	For individuals with $n_i \geq 9$ (14 individuals)							
	MAPE				MSPE			
	UNC	AR(1)	AR(2)	AR(3)	UNC	AR(1)	AR(2)	AR(3)
One step	0.8182	0.8103	0.8095	**0.8058**	1.1930	1.1403	1.1356	**1.1151**
Two step	0.6287	0.6415	0.6410	**0.6220**	0.7315	0.7291	0.7269	**0.6987**

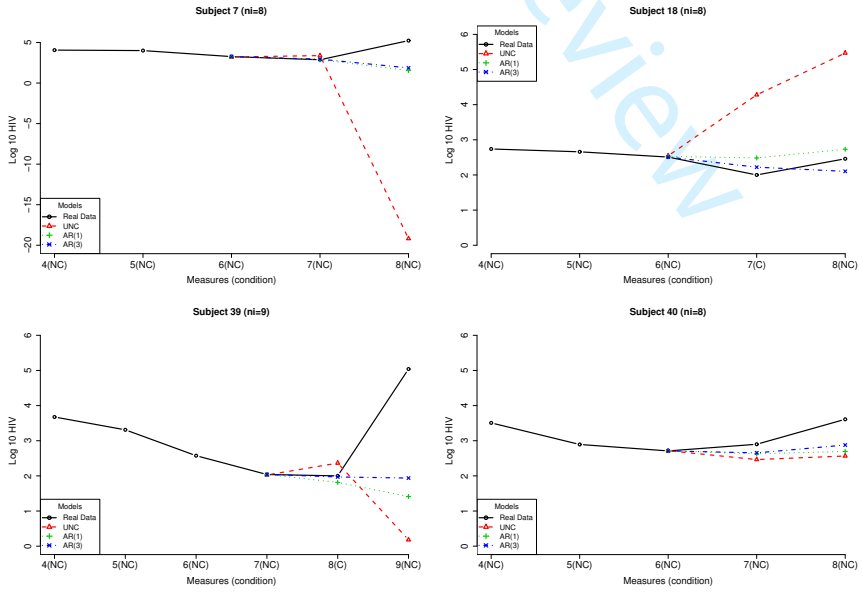

Figure 5: ACTG-315 dataset. Evaluation of the prediction performance for four randomly selected subjects.

8. conclusions

This work describes a likelihood-based approach to perform inference and prediction in censored mixed effects models with autoregressive errors. We used the EM algorithm to

obtain the ML estimates of model parameters. For practical demonstration the method was applied to the analysis of a real HIV dataset whose measures are subject to the detection limit of the recording assays. We also used simulation to investigate the performance in terms of predictions, parameter recovery and the robustness of the EM algorithm. In the simulation study comparisons are made between inferences based on the proposed censored data with different correlation structures. We showed that the differences in inference of the autoregressive structure can be substantial. The proposed methods can be implemented in the new R package "ARpLMEC", providing practitioners with a convenient tool for further applications in their domain.

The proposed methods can also be easily applied to other substantive areas where the data being analyzed have censored observations. In line with this future work includes extending the proposed methods to accommodate missing values in addition to censoring using hybrid Bayesian sampling procedures (Wang and Fan, 2012) and multivariate outcomes (Wang, 2013). Further, the models developed here do not consider skewness and heavy-tails in the responses because typically in HIV-AIDS studies, the responses (censored viral load) are log-transformed to achieve a 'close to normal' shape. However, features of non-normality, such as skewness and thick tails (Lachos et al., 2010), need to be incorporated in the proposed method to come up with a more general framework for censored mixed models. This issue is currently under investigation, and we hope to report these findings in a future paper.

References

Barndorff-Nielsen, O., Schou, G., 1973. On the parametrization of autoregressive models by partial autocorrelations. Journal of Multivariate Analysis 3 (4), 408–419.

Box, G. E., Jenkins, G. M., Reinsel, G. C., Ljung, G. M., 2015. Time series analysis: Forecasting and control. John Wiley & Sons.

Dempster, A., Laird, N., Rubin, D., 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society. Series B, Statistical Methodology 39, 1–38.

Garay, A. M., Castro, L. M., Leskow, J., Lachos, V. H., 2017. Censored linear regression models for irregularly observed longitudinal data using the multivariate-t distribution. Statistical Methods in Medical Research 26 (2), 542–566.

Hughes, J., 1999. Mixed effects models with censored data with application to HIV RNA levels. Biometrics 55, 625–629.

Lachos, V., Bandyopadhyay, D., Dey, D., 2011. Linear and nonlinear mixed-effects models for censored HIV viral loads using normal/independent distributions. Biometrics 55, 1304–1318.

Lachos, V. H., Ghosh, P., Arellano-Valle, R. B., 2010. Likelihood based inference for skew–normal independent linear mixed models. Statistica Sinica 20, 303–322.

Lin, T., Lee, J., 2007. Bayesian analysis of hierarchical linear mixed modeling using the multivariate t distribution. Journal of Statistical Planning and Inference 137 (2), 484–495.

Lin, T.-I., 2010. Robust mixture modeling using multivariate skew t distributions. Statistics and Computing 20 (3), 343–356.

Liu, C., 1996. Bayesian robust multivariate linear regression with incomplete data. Journal of the American Statistical Association 91, 1219–1227.

Louis, T. A., 1982. Finding the observed information matrix when using the EM algorithm. Journal of the Royal Statistical Society. Series B, Statistical Methodology. 44 (2), 226–233.

Matos, L. A., Bandyopadhyay, D., Castro, L. M., Lachos, V. H., 2015. Influence assessment in censored mixed-effects models using the multivariate Student′s-*t* distribution. Journal of Multivariate Analysis 141, 104–117.

Matos, L. A., Castro, L. M., Lachos, V. H., 2016. Censored mixed-effects models for irregularly observed repeated measures with applications to hiv viral loads. Test 25 (4), 627–653.

Matos, L. A., Lachos, V., Balakrishnan, N., Labra, F., 2013a. Influence diagnostics in linear and nonlinear mixed-effects models with censored data. Computational Statistics & Data Analysis 57 (1), 450–464.

Matos, L. A., Prates, M. O., Chen, M. H., Lachos, V. H., 2013b. Likelihood-based inference for mixed-effects models with censored response using the multivariate-t distribution. Statistica Sinica 23, 1323–1342.

Meilijson, I., 1989. A fast improvement to the EM algorithm on its own terms. Journal of the Royal Statistical Society. Series B, Statistical Methodology. 51 (1), 127–138.

Meng, X.-L., Rubin, D. B., 1993. Maximum likelihood estimation via the ECM algorithm: A general framework. Biometrika 80 (2), 267–278.

Muñoz, A., Carey, V., Schouten, J. P., Segal, M., Rosner, B., 1992. A parametric family of correlation structures for the analysis of longitudinal data. Biometrics 48, 733–742.

Pinheiro, J. C., Bates, D. M., 2000. Mixed-Effects Models in S and S-PLUS. Springer, New York, NY.

R Development Core Team, 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0. URL `http://www.R-project.org`

Rao, C. R., 1987. Prediction of future observations in growth curve models. Statistical Science 2, 434–447.

Schumacher, F. L., Lachos, V. H., Dey, D. K., 2017. Censored regression models with autoregressive errors: A likelihood-based perspective. Canadian Journal of Statistics 45 (4), 375–392.

Vaida, F., Fitzgerald, A., DeGruttola, V., 2007. Efficient hybrid EM for linear and nonlinear mixed effects models with censored response. Computational Statistics & Data Analysis 51, 5718–5730.

Vaida, F., Liu, L., 2009. Fast implementation for normal mixed Effects models with censored response. Journal of Computational and Graphical Statistics 18, 797–817.

Wang, W.-L., 2013. Multivariate t linear mixed models for irregularly observed multiple repeated measures with missing outcomes. Biometrical Journal 55 (4), 554–571.

Wang, W. L., Fan, T. H., 2011. Estimation in multivariate t linear mixed models for multivariate longitudinal data. Statistica Sinica 21, 1857–1880.

Wang, W.-L., Fan, T.-H., 2012. Bayesian analysis of multivariate t linear mixed models using a combination of ibf and gibbs samplers. Journal of Multivariate Analysis 105 (1), 300–310.

Wu, L., 2002. A joint model for nonlinear mixed-effects models with censoring and covariates measured with error, with application to AIDS studies. Journal of the American Statistical Association 97 (460), 955–964.