# Approximate local influence in generalized linear mixed models

Sergio Alexánder Gómez Noguera

## THESIS PRESENTED TO INSTITUTE OF MATEMATHICS AND STATISTICS OF UNIVERSITY OF SÃO PAULO TO OBTAIN THE TITLE OF DOCTOR OF SCIENCE

## Graduate Program: Statistics Advisor: Prof. Dr. Gilberto Alvarenga Paula

During the development of this work the author received financial support from CAPES/CNPq

São Paulo, December 2019

## Influência local aproximada em modelos lineares generalizados mistos

Sergio Alexánder Gómez Noguera

## TESE APRESENTADA AO INSTITUTO DE MATEMÁTICA E ESTATÍSTICA DA UNIVERSIDADE DE SÃO PAULO PARA OBTENÇÃO DO TÍTULO DE DOUTOR EM CIÊNCIAS

## Programa: Estatística Orientador: Prof. Dr. Gilberto Alvarenga Paula

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro de CAPES e CNPq

São Paulo, Dezembro de 2019

## Influência local aproximada em modelos lineares generalizados mistos

Este exemplar corresponde à redação final da tese devidamente corrigida e defendida por Sergio Alexánder Gómez Noguera e aprovada pela Banca Examinadora.

Banca Examinadora:

- Prof. Dr. Gilberto Alvarenga Paula (orientador) IME USP
- Prof. Dr. Julio da Motta Singer IME USP
- Profa. Dra. Clarice Garcia Borges Demétrio ESALQ USP
- Profa. Dra. Cibele Maria Russo Noveli ICMC USP
- Profa. Dra. Mariana Rodrigues Motta UNICAMP

# Acknowledgements

Agradeço a Deus porque sei que sem Ele nada podemos fazer, mas com ele somos mais que vencedores.

À minha mãe Marina Noguera e meu pai Hernando Gómez pelo apoio, pelo amor e as orações feitas no tempo que estive aqui no Brasil.

Aos meus irmãos Hernán e Diego Gómez pelo apoio fornecido neste processo. À minha sobrinha Mariana Gómez porque, sem ela saber, ela me apoiava todos os dias.

A minha tia Teresa Noguera e sua familia por que sempre que voltei para a Colômbia me senti em casa.

A toda minha família Gómez e Noguera pelo apoio constante e incondicional na minha formação profissional. Obrigado Família.

Ao meu orientador Gilberto Alvarenga Paula pela confiança, paciência e as valiosas sugestões feitas no desenvolvimento deste trabalho.

Ao professor John H. Castillo Gómez que com seu exemplo me motivou a desenvolver meus estudos de pós-graduação. Professor muito obrigado.

Thanks professor Daya K. Nagar for trusting me and giving me another academic opportunity in my life. Thank you teacher.

A Lina pelo apoio, a paciência e os bons momentos.

À minha amiga Estefanía por todos os bons momentos, pelo apoio e todos os dias que compartilhamos no Brasil. Amiga foram anos maravilhosos.

Aos meus amigos Alex Sierra, Sebastián Herrera e Wilson Mutis pelas conversas e os momentos compartilhados.

Aos meus amigos do futebol, obrigado por todos os dias que a gente jogou. Esses jogos foram fundamentais nos dias difíceis do doutorado. Obrigado pessoal.

Aos meus amigos na Colômbia, Luis Miguel Delgado, Nelcy Alomía, Juanito, Catalina Rua e Lucy Aguilera, por me animar nos tempos difíceis.

Aos professores do Instituto de Matemática e Estatística da Universidade de São Paulo, que ajudaram na minha formação acadêmica. Aos professores da banca examinadora pelas suggestões e correções feitas.

Ao Brasil e à USP, pela formação acadêmica e pelos bons tempos vividos.

À CAPES e CNPq, pelo apoio financeiro.

Dedico este trabalho a meus pais Marina Noguera e Hernando Gómez.

iv

# Abstract

GÓMEZ NOGUERA, S. A. **Approximate local influence in generalized linear mixed models.** 2019. Tese (Doutorado) - Institute of Mathematics and Statistics, University of São Paulo, São Paulo, 2019.

Non-Gaussian correlated data are frequent in longitudinal and repeated measure studies. Generalized linear mixed models (GLMMs) are a powerful tool for the analysis and treatment of this kind of data. Residual and sensitivity analysis are useful diagnostic procedures to verify the assumptions made on these models and the adequacy to the data. Among the techniques included in the sensitivity analysis is the local influence, which allows to discriminate observations with a undue weight in the parameter estimates of any statistical model. In this work we present approximated analytical structures for local influence measurements in generalized linear mixed models. These structures were obtained through Laplace approximations for usual perturbation schemes in order to discriminate observations and subjects with excessive influence on the parameter estimates. These measures, which are presented in closed forms for the generalized linear mixed models, have a relatively low computational cost and have been shown to be effective in detection of influential observations and subjects as evidenced by simulation studies and analyses of three real data sets.

**Keywords**: generalized linear mixed model, Laplace approximation, local influence, perturbation scheme, conformal normal curvature, influential observations, influential subjects, random effects.

vi

# Resumo

GÓMEZ NOGUERA, S. A. Influência local aproximada em modelos lineares generalizados mistos 2019. Tese (Doutorado) - Instituto de Matemáticas e Estatística, Universidade de São Paulo, São Paulo, 2019.

Dados correlacionados não Gaussianos são frequentes em estudos longitudinais e de medidas repetidas. Os modelos lineares generalizados mistos (MLGMs) constituem uma ferramenta poderosa para a análise e tratamento de dados desse tipo. Análise de resíduos e análise de sensibilidade são procedimentos de diagnóstico úteis para verificar as suposições feitas para esses modelos e a adequação aos dados. Entre as técnicas incluídas na análise de sensibilidade está a influência local que permite discriminar observações com um peso desproporcional nas estimativas dos parâmetros de qualquer modelo estatístico. Estruturas analíticas aproximadas para medidas de influência local são apresentadas neste trabalho em modelos lineares generalizados mistos. Essas estruturas foram obtidas através de aproximações e grupos com excessiva influência nas estimativas dos parâmetros. Essas medidas, que são apresentadas em formas fechadas para os MLGs mistos, despendem custo computacional relativamente baixo e tem-se mostrado eficientes na detecção de observações e grupos influentes conforme comprovado através de estudos de simulação e análises de três conjuntos de dados reais.

**Palavras chave**: Modelo linear generalizado misto, aproximação de Laplace, influência local, esquema de perturbação, curvatura normal conformal, observações influentes, indivíduos influentes, efeitos aleatórios.

viii

# Contents

Li	List of symbols and abbreviations			xiii
Li	st of	Figures	ş	xv
Li	st of	Tables		xix
1	Intr	oductio	on	1
	1.1	Object	tives and thesis organization	3
2	Bas	ic Conc	epts	5
	2.1	Introd	luction	5
	2.2	Gener	alized linear mixed models (GLMMs)	5
		2.2.1	The Model	6
		2.2.2	On the estimation methods in GLMMs	7
	2.3	Local	influence: general approach	9
	2.4	On the	e Laplace approximation	11
		2.4.1	Some remarks on notation	11
			2.4.1.1 $F_{xx}$ and $F_{xx}^*$	12
		2.4.2	Consequences of the Laplace approximation	13
	2.5	Concl	uding remarks	16
3	Loc	al influ	ence in GLMMs	17
	3.1	Introd	luction	17
	3.2	Works	on local influence in random effect models	18
	3.3	Local	influence in GLMMs: our strategy	20
	3.4	Score	$L_{\psi}$ and Hessian $L_{\psi\psi}$	21
	3.5	Appro	oximate local influence through the Laplace approximation	24
		3.5.1	Remark on the notation	24
		3.5.2	Approximate computation of the moments of $b_i   y_i \dots \dots \dots$	25
			3.5.2.1 Iterative process for computing $\tilde{b}_i$	28

		3.5.2.2 Computing $\tilde{\boldsymbol{b}}_i$ through gamlss package	29
	3.6	Approximation for $L_{\widehat{\psi} \ \widehat{\psi}}$	32
	3.7	Conformal normal curvature under different perturbation schemes: ap-	
		proximation for $\Delta$	35
		3.7.1 Case weights - subjects	36
		3.7.1.1 Conformal normal curvature - subjects	37
		3.7.2 Case weights - observations	37
		3.7.2.1 Conformal normal curvature - observations	39
	3.8	Approximate local influence in random intercept	
		GLMMs	40
		3.8.1 Approximate Hessian matrix	41
		3.8.2 Case weights - subjects	42
		3.8.2.1 Conformal normal curvature - subjects	43
		3.8.3 Case weights - observations	43
		3.8.3.1 Conformal normal curvature - observations	45
	3.9	Concluding remarks	45
4	Nur	merical studies and applications	47
	4.1	Introduction	47
	4.2	Numerical studies	47
		4.2.1 Poisson - normal GLMM	48
		4.2.2 Gamma - normal GLMM	56
	4.3	Applications	60
		4.3.1 Milk data set	61
		4.3.1.1 Potentially influential points	70
		4.3.1.2 MeRC and MRC	74
		4.3.2 Respiratory data set	76
		4.3.2.1 Potentially influential points	84
		4.3.2.2 MeRC and MRC	87
		4.3.3 Insurance data set	89
		4.3.3.1 Potentially influential points	96
		4.3.3.2 MeRC and MRC	99
	4.4	Concluding remarks	100
5	Con	nclusions	103
	5.1	Final considerations	103
	5.2	Suggestions for future research	104

A	Proc	ofs related to the Laplace approximation	107
	A.1	Proof of theorem 1	107
	A.2	Proof of theorem 2	108
	A.3	Proof of theorem 3	109
B	Con	putations of the quantities related with the curvature	111
	<b>B.</b> 1	Some aspects of notation	111
	B.2	Computation of $\mathbf{L}_{\boldsymbol{b}_i}(\boldsymbol{\psi}; \boldsymbol{y}_i, \boldsymbol{b}_i)$ and $\mathbf{L}_{\boldsymbol{b}_i \boldsymbol{b}_i}(\boldsymbol{\psi}; \boldsymbol{y}_i, \boldsymbol{b}_i)$	116
	B.3	Computation of the score function $\mathbf{L}_{\boldsymbol{\psi}}(\boldsymbol{\psi}; \boldsymbol{y})$	117
	<b>B.4</b>	Computation of the Hessian matrix $\mathbf{L}_{\boldsymbol{\psi}\boldsymbol{\psi}}(\boldsymbol{\psi};\boldsymbol{y})$	118
	<b>B</b> .5	Hessians for approximating L $_{\widehat{\psi}\widehat{\psi}}(\psi; y)$	122
	B.6	$\Delta$ under the first perturbation scheme (subjects)	125
	B.7	$\Delta$ under the second perturbation scheme (observations)	126
		B.7.1 Hessians for $\Delta$ under the second perturbation scheme	128
C	Witl	nin-subject correlations	131
D	Imp	lementation in R software	137
Bi	Bibliography		

### xii CONTENTS

# List of symbols and abbreviations

$N_T$	total number of observations.
Ν	number of subjects.
$n_i$	number of observations in the <i>ith</i> subject.
~	distributed as.
$\otimes$	Kronecker product.
Г	gamma function.
$\oplus$	direct sum.
$\mathbb{R}^m$	m-dimensional euclidean space.
<b>1</b> <i>i</i>	$n_i \times 1$ vector of 1's.
а	bold lowercase letters for denoting vectors.
<b>u</b> <sub><i>i</i></sub>	$N \times 1$ vector with 1 in the <i>i</i> th position and zeros in the
	remaining positions.
<b>u</b> <sub><i>ij</i></sub>	$N_T \times 1$ partitioned vector $\mathbf{u}_{ij} = (\mathbf{d}_1^{\top}, \dots, \mathbf{d}_N^{\top})^{\top}$ , which assumes zero
	everywhere except in the <i>jth</i> position of $\mathbf{d}_i \in \mathbb{R}^{n_i}$ , in which it contains 1.
<b>e</b> <sub>ij</sub>	<i>jth</i> column of the identity matrix $\mathbf{I}_{n_i}$ .
y	$N_T \times 1$ vector of observed values of the response variable.
Α	bold capital letters for denoting matrices.
$\mathbf{A}^{\top}$	transpose of the matrix <b>A</b>
<b>A</b>	determinant of the matrix <b>A</b> .
$\mathbf{A}^{-1}$	inverse matrix of the no singular matrix <b>A</b> .
$\mathbf{A} > 0$	positive definite matrix.
$\mathbf{A} \ge 0$	positive semidefinite matrix
$\mathbf{I}_m$	identity matrix of dimension $m \times m$ .
<b>J</b> <sub><i>i</i></sub>	$n_i \times n_i$ matrix of 1's.
$\mathbf{D}_{a}$	$\mathbf{D}_{a} = \text{diag}\{a_{1}, a_{2}, \dots, a_{m}\}, \text{ where } \mathbf{a} = (a_{1}, a_{2}, \dots, a_{m})^{\top}.$
K <sub>rs</sub>	$rs \times rs$ commutation matrix.
$\mathbf{K}_q$	$\mathbf{K}_q = \mathbf{K}_{qq}.$
$\mathbf{G}_q$	$q^2 \times \frac{q \ (q+1)}{2}$ duplication matrix.

$\mathbf{\Omega}_i$	$n_i^2 \times n_i$ matrix defined by the direct sum of the columns of $\mathbf{I}_{n_i}$ .
$\mathrm{E}\left(x ight)$	expected value of the random vector <i>x</i> .
Var(x)	variance-covariance matrix of the random vector $x$ .
$\operatorname{Corr}(\boldsymbol{x}, \boldsymbol{y})$	correlation matrix of the random vectors $x$ and $y$ .
EF (·; ·)	exponential family of distributions.
$f_{Y_{ij} \boldsymbol{b}_i}(\cdot)$	conditional probability density function of the random variable $Y_{ij}$ given $b_i$ .
$g(\cdot)$	link function corresponding to the model.
β	$p \times 1$ vector of fixed effects parameters, i.e. $(\beta_1, \beta_2, \dots, \beta_p)^{\top}$ .
ξ	q(q+1)/2 vector of variance-covariance parameters.
$\phi^{-1}$	dispersion parameter.
$\psi$	vector of parameters to be estimated, i.e. $\boldsymbol{\psi} = (\boldsymbol{\beta}^{T}, \boldsymbol{\xi}^{T}, \boldsymbol{\phi})^{T}$ .
$\boldsymbol{b}_i$	$q \times 1$ vector of random effects.
$\mathbf{N}_q(\boldsymbol{\mu}; \boldsymbol{\Sigma})$	multivariate normal distribution
Ξ	$q \times q$ variance-covariance matrix of the random effects, which depends on $\xi$ .
$L(\boldsymbol{\psi}; \boldsymbol{y})$	marginal log-likelihood function of $y$ .
$\mathrm{L}_{\mathrm{P}}\left(oldsymbol{eta}$ ; $oldsymbol{b} ight)$	penalized log-likelihood function.
ω	perturbation vector.
$\boldsymbol{\omega}_0$	no perturbation vector.
d	direction vector for computing the curvatures.
$C_d(\cdot)$	normal curvature in the direction of the vector <b>d</b> .
$B_{d}(\cdot)$	conformal normal curvature in the direction of the vector <b>d</b> .
$B_i(\cdot)$	conformal normal curvature in the direction of the vector $\mathbf{u}_i$ .
$\mathbf{B}_{ij}\left(\cdot\right)$	conformal normal curvature in the direction of the vector $\mathbf{u}_{ij}$ .
MeRC	Mean Relative Change.
MRC	Maximum Relative Change.
GLMM	generalized linear mixed model.
NRDMM	nonlinear reproductive dispersion mixed model.
GAMLSS	generalized additive model for location, scale and shape.

# **List of Figures**

Comparison of the conformal normal curvatures with $N_T = 150$ and	
k = 10 from the model (4.1) fitted to the generated data sets	50
Comparison of the conformal normal curvatures with $N_T = 150$ and	
k = 15 from the model (4.1) fitted to the generated data sets	50
Comparison of the conformal normal curvatures with $N_T = 150$ and	
k = 30 from the model (4.1) fitted to the generated data sets	51
Comparison of the conformal normal curvatures with $N_T = 150$ and	
k = 50 from the model (4.1) fitted to the generated data sets	51
Comparison of the conformal normal curvatures with $N_T$ = 275 and $k$ =	
16 for the model (4.1) fitted to the generated data sets under unbalanced	
data	53
Comparison of the conformal normal curvatures with $N_T$ = 275 and $k$ =	
17 for the model (4.1) fitted to the generated data sets under unbalanced	
data	54
Comparison of the conformal normal curvatures with $N_T = 100$ and	
k = 10 for the model (4.2) fitted to the generated data sets	57
Comparison of the conformal normal curvatures with $N_T = 100$ and	
k = 20 for the model (4.2) fitted to the generated data sets	58
Comparison of the conformal normal curvatures with $N_T = 100$ and	
k = 25 for the model (4.2) fitted to the generated data sets	58
Comparison of the conformal normal curvatures with $N_T = 51$ and $k = 10$	
for the model (4.2) fitted to the generated data sets under unbalanced data.	59
Comparison of the conformal normal curvatures with $N_T = 63$ and $k = 9$	
for the model (4.2) fitted to the generated data sets under unbalanced data.	59
Robust boxplots of the protein content values for each diet for all cows	
and for all weeks.	62
Profile of the protein content values for each cow classified for each diet	
across weeks.	63
	Comparison of the conformal normal curvatures with $N_T = 150$ and $k = 10$ from the model (4.1) fitted to the generated data sets Comparison of the conformal normal curvatures with $N_T = 150$ and $k = 15$ from the model (4.1) fitted to the generated data sets Comparison of the conformal normal curvatures with $N_T = 150$ and $k = 30$ from the model (4.1) fitted to the generated data sets Comparison of the conformal normal curvatures with $N_T = 150$ and $k = 50$ from the model (4.1) fitted to the generated data sets Comparison of the conformal normal curvatures with $N_T = 275$ and $k = 50$ from the model (4.1) fitted to the generated data sets

4.14	Profile of the mean protein content values for each diet across weeks	63
4.15	Quantile residual against fitted value for $\mu$ (left) and normal probability	
	plot for the quantile residual (right) from the random intercept gamma-	
	normal model fitted to the Milk data set (4.3).	64
4.16	Worm plot from the random intercept gamma-normal model fitted to	
	the Milk data set (4.3).	66
4.17	Predicted random effects density from the random intercept gamma-	
	normal model fitted to the Milk data set (4.3).	67
4.18	Index plot of the conformal normal curvature $B_i(\beta, \xi, \phi)$ from the ran-	
	dom intercept gamma-normal model fitted to the Milk data set (4.3).	71
4.19	Index plot of the conformal normal curvature $B_{ii}(\beta, \xi, \phi)$ from the ran-	
	dom intercept gamma-normal model fitted to the Milk data set (4.3).	72
4.20	Index plots of the conformal normal curvature $B_{ii}(\beta)$ (left top), against	
	$B_{ii}(\xi)$ (right top) and $B_{ii}(\phi)$ (bottom) from the random intercept gamma-	
	normal model fitted to the Milk data set (4.3).	73
4.21	Robust boxplots of age for each respiratory condition for all patients and	
	for all visits.	78
4.22	Quantile residual against fitted value for $\pi$ (left) and normal probability	
	plot for the quantile residual (right) from the random intercept Bernoulli-	
	normal model fitted to the Respiratory data set (4.6)	79
4.23	Worm plots from the random intercept Bernoulli-normal model fitted to	
	the Respiratory data set (4.6).	81
4.24	Predicted random effects density from the random intercept Bernoulli-	
	normal model fitted to the Respiratory data set (4.6)	82
4.25	Index plot of the conformal normal curvature $B_i(\boldsymbol{\beta}, \boldsymbol{\xi})$ from the random	
	intercept logistic model fitted to the Respiratory data set (4.6).	85
4.26	Index plot of the conformal normal curvature $B_{ij}(\boldsymbol{\beta}, \boldsymbol{\xi})$ from the random	
	intercept logistic model fitted to the Respiratory data set (4.6)	86
4.27	Index plots of the conformal normal curvature $B_{ij}\left(oldsymbol{eta} ight)$ (left) and against	
	$B_{ij}(\xi)$ (right) from the random intercept logistic model fitted to the Res-	
	piratory data set (4.6)	87
4.28	Scatterplot of the number of claims versus logarithm of the number of	
	accidents for all territories with fitted smoothing curve	90
4.29	Quantile residual against fitted value for $\mu$ (left) and normal probability	
	plot for the quantile residual (right) from the random intercept negative	
	binomial model with $\nu$ fixed, fitted to the Insurance data set (4.7)	91

4.30	Predicted random effects density from the random intercept negative	
	binomial model with $\nu$ fixed, fitted to the Insurance data set (4.7)	93
4.31	Worm plots from the random intercept negative binomial model with $\nu$	
	fixed, fitted to the Insurance data set (4.7)	94
4.32	Index plot of the conformal normal curvature from the random intercept	
	negative binomial model fitted to the Insurance data (4.7)	97
4.33	Index plot of the conformal normal curvature $B_{ij}(\boldsymbol{\beta}, \boldsymbol{\xi})$ from the random	
	intercept negative binomial model fitted to the Insurance data set (4.7).	98
4.34	Index plots of the conformal normal curvature $B_{ij}\left(oldsymbol{eta} ight)$ (left) and against	
	$B_{ij}(\xi)$ (right) from the random intercept negative binomial model fitted	
	to the Insurance data set	99

### xviii LIST OF FIGURES

# List of Tables

4.1	Summary of the average percentages of potentially influential subjects	
	detected in the 300 replicates by our proposal for each scenario	54
4.2	Summary of the average percentages of potentially influential subjects	
	detected in the 300 replicates by our proposal for each scenario.	60
4.3	Summary of the quantile residuals from the random intercept gamma-	
	normal model fitted to the Milk data set (4.3)	65
4.4	Parameter estimates and their approximated standard errors from the	
	random intercept gamma-normal model fitted to the Milk data set (4.3).	67
4.5	Parameter estimates and their approximated standard errors from the	
	random intercept gamma-normal model fitted to the Milk data set (4.3)	
	dropping the potentially influential subjects and observations on $\widehat{\psi}$ =	
	$(\widehat{\boldsymbol{\beta}}^{T},\widehat{\boldsymbol{\xi}},\widehat{\boldsymbol{\phi}})^{T}$	69
4.6	Necessary matrices for computing the conformal normal curvature in	
	gamma-normal models with logarithmic link	70
4.7	Comparison of the measures MeRC and MRC for potentially influential	
	subjects from the random intercept gamma-normal model fitted to the	
	Milk data set (4.3)	75
4.8	Comparison of the measures MeRC and MRC for potentially influential	
	observations from the random intercept gamma-normal model fitted to	
	the Milk data set (4.3)	76
4.9	Patients with poor respiratory condition in each visit, classified according	
	to the two treatments and gender	77
4.10	Patients with poor respiratory condition in each visit, classified according	
	to the two treatments and age	77
4.11	Parameter estimates and their approximated standard errors from the	
	random intercept logistic model fitted to the Respiratory data set	82

4.12	Parameter estimates and their approximated standard errors from the	
	random intercept logistic model fitted to the Respiratory data set, after	
	dropping potentially influential observations	83
4.13	Necessary matrices for computing the conformal normal curvature in	
	the logistic model	84
4.14	Comparison of the measures MeRC and MRC for potentially influential	
	subjects from the random intercept logistic model fitted to the Respira-	
	tory data set (4.6)	88
4.15	Parameter estimates and their approximated standard errors from the	
	random intercept negative binomial model fitted to the Insurance data	
	set (4.7)	95
4.16	Parameters estimates from the negative binomial model fitted to the	
	Insurance data (4.7), after dropping the influential territories and areas.	95
4.17	Necessary matrices for computing the conformal normal curvature for	
	the negative binomial model with $\nu$ fixed and logarithmic link	96
4.18	Comparison of the measures MeRC and MRC for potentially influential	
	subjects from the random intercept negative binomial model fitted to the	
	Insurance data set (4.7).	99
4.19	Comparison of the measures MeRC and MRC for influential areas from	
	the fit of the negative binomial model to the Insurance data set (4.7).	100
B.1	Some useful results.	111
B.2	Some useful differentials.	115

# Chapter 1 Introduction

Generalized linear mixed models (GLMMs) may be interpreted as an extension of the generalized linear models (McCullagh and Nelder, 1989), by including fixed effects and non-observable random variables, called random effects in the linear predictor. The aim of the random effects is to deal with correlated observations, for example, from longitudinal or repeated measure studies. GLMMs are likely the class of models with random effects most used in the analysis of non-Gaussian correlated data (Rakhmawati et al., 2017). Frequently, the estimation processes in these models require to manipulate intractable high-dimensional integrals, since the likelihood functions associated with GLMMs rarely have closed forms (Nelder et al., 2006). For a complete review of GLMMs and the estimation methods related with these models the reader may refer to Breslow and Clayton (1993), McGilchrist (1994), McCulloch (1997), Aitkin (1999), McCulloch and Searle (2000), Rabe-Hesketh et al. (2002), McCulloch and Neuhaus (2005), Jiang (2007), Wu (2009), Fong et al. (2010), Stroup (2012) and Demidenko (2013), among others. Breslow and Clayton (1993) used Laplace approximation for estimating the parameters in GLMMs. This approach will be extensively used in this work. The reader may refer, for example, to Vonesh and Chinchilli (1996) and Tuerlinckx et al. (2006) and the references therein for more details on the Laplace approximation.

In general, after postulating and fitting a model, it is necessary to assess the existence of

#### 2 INTRODUCTION

observations with undue impact on the parameter estimates which may lead to wrong conclusions. Such observations are called influential observations. For this purpose, according to Lesaffre and Verbeke (1998), deletion diagnostics are the most popular tool to assess the effect of dropping observations on the parameter estimates. This technique was proposed by Cook (1977) for normal linear models and subsequently extended to more general models (see, for instance, Cook and Weisberg, 1982). Tan *et al.* (2001) generalized this methodology for longitudinal mixed - effects models. Xiang *et al.* (2002) studied the Cook distance in GLMMs. Xu *et al.* (2006) developed deletion diagnostics for GLMMs. Tang *et al.* (2006) developed influence diagnostics in nonlinear reproductive dispersion mixed models, which are more general than GLMMs. Recently, Pinho *et al.* (2015) extended the idea of Xiang *et al.* (2002) to allow the separate identification of influential observations in the parameter estimates and prediction of random effects in GLMMs.

An alternative approach based on small perturbations in the model or data was proposed by Cook (1986). Basically, the surface formed by an appropriate influence measure and the perturbation vector is approximated by the normal curvature in some arbitrary direction. Various diagnostic graphs may be derived from the normal curvature. This diagnostic approach has been applied for different kinds of models. For instance, local influence for mixed linear models with normally distributed errors was discussed by Lesaffre and Verbeke (1998). In the context of GLMMs, Ouwens *et al.* (2001) investigated local influence in Poisson and binomial mixed models using Bayesian approach and computational tools. Xiang *et al.* (2003) studied local influence in GLMMs with a particular structure in the variance - covariance matrix of the random effects. That study focused on assessing potentially influential clusters. Zhu and Lee (2003) proposed to develop local influence in GLMMs treating the random effects as missing observations, which allows to identify influential observations through an alternative influence measure (Zhu and Lee, 2001). Recently, Rakhmawati *et al.* (2017) extended the idea of the seminal paper of Cook (1986) in some GLMMs by means of a mixture of analytical and computational techniques. Specifically, the authors focused in highlighting influential subjects rather than observations. However, their influence measures do not have an analytical form that can be applied to any GLMM. More recently, following the approach of Zhu and Lee (2003), Tapia *et al.* (2019a) and Tapia *et al.* (2019b) proposed a sensitivity analysis for Poisson and binomial mixed models. It is worth mentioning that, except Zhu and Lee (2003), the aforementioned papers only perform local influence in particular cases of GLMMs. Also, none of these works supplies general structures for the practical use of this diagnostic tool.

### **1.1** Objectives and thesis organization

Although there are several works on local influence in GLMMs, as far as we know no one of them provides general analytical expressions for the identification of potentially influential observations and subjects. Thus, based on the original idea of Cook (1986), the aim of this work is to supply general analytical forms that allow the detection of subjects and observations with undue weight in the parameter estimates of GLMMs. By general analytical forms, we refer to closed structures for the influence measures that rely on general features of GLMMs, like the variance function, the weight matrix, design matrices for fixed and random effects, among others (McCulloch and Searle, 2000). From such structures, we intend to implement useful functions in the R software, with inexpensive computational routines that allow the use of the proposed approaches in a variety of situations, whose treatments require the use of GLMMs. To achieve this goal we organize the dissertation as follows.

In chapter 2, we describe briefly GLMMs, pointing out some aspects of the estimation methods and presenting a review on the local influence approach. We also propose multivariate extensions of the results presented in section 7.7.1 of Demidenko (2013) on the Laplace approximation. In chapter 3, we derive approximated local influence measures for GLMMs for discriminating potentially influential subjects and observations.

#### 4 INTRODUCTION

We dedicate a section of that chapter for the treatment of local influence in random intercept GLMMs. However, it does not mean that our results are applicable only to these models. Also, we derive the score function and the observed information matrix for such models. Numerical studies and applications with several real data sets are described in chapter 4. In chapter 5 we present some conclusions and suggestions for future research.

In addition, this work is supported by a supplementary material section divided into three appendices. In appendix A we describe the proofs related with the multivariate extensions of the Laplace approximations. Details on the computation of the local influence measures are discussed in appendix B. In addition, in appendix C, we discuss some aspects on the intraclass correlation in GLMMs. Finally, in appendix D, we illustrate the implementation of our proposal in the R software (R Core Team, 2018) for a variety of GLMMs with random slope and intercept.

This work is computationally implemented in the R software (R Core Team, 2018) available in http://www.R-project.org/. Also, this work was written using the typeset system LATEX available in https://www.latex-project.org/ and using the online editor, overleaf available in https://www.overleaf.com/.

# Chapter 2

# **Basic Concepts**

### 2.1 Introduction

This chapter is divided into three parts. In the first one we specify the GLMMs mentioning some special features of these models that are fundamental in the development of our work. Also, we briefly refer to the estimation methods related with the GLMMs and references are suggested for a deeper study of these topics. In the second part we describe the local influence approach (Cook, 1986) justifying the use of the conformal normal curvature (Poon and Poon, 1999) to accomplish sensitivity studies in regression models. In the last part, we focus in clarifying some fundamental aspects of the notation and proposing some multivariate extensions of the Laplace approximation (Demidenko, 2013).

### 2.2 Generalized linear mixed models (GLMMs)

Generalized linear mixed models (GLMMs) may be considered as an extension of generalized linear models, where the linear predictor is formed by fixed effects and non-observable random effects. The random effects are incorporated into the model in order to deal with correlated observations.

#### 2.2.1 The Model

Let  $Y_{ij}$  be the *jth* observation from the *ith* subject, where  $1 \le j \le n_i$ ,  $1 \le i \le N$  and  $N_T = \sum_{i=1}^N n_i$ . Suppose that the probability density function of  $Y_{ij}$  relies on a *q*-dimensional vector  $b_i$  of random effects, via the parameter  $\theta_{ij} \in \mathbb{R}$ . In addition, assume that given  $b_i$ ,  $Y_{ij}$  follows a distribution in the exponential family i. e.,

$$\mathbf{f}_{y_{ij}|\boldsymbol{b}_i}\left(y_{ij};\theta_{ij},\phi\right) = \exp\left[\phi\left\{y_{ij}\,\theta_{ij} - a(\theta_{ij})\right\} + c(y_{ij};\phi)\right],\tag{2.1}$$

where *a*(.) and *c*(.; .) are known functions and  $\phi$  denotes the precision parameter. Under certain regularity conditions (McCulloch and Searle, 2000) it follows that  $\mu_{ij} = E(Y_{ij} | \boldsymbol{b}_i) = \frac{d a(\theta_{ij})}{d \theta_{ij}}$  and  $Var(Y_{ij} | \boldsymbol{b}_i) = \phi^{-1} V_{ij}$ , where  $V_{ij} = V(\mu_{ij}) = \frac{d^2 a(\theta_{ij})}{d \theta_{ij}^2}$  is named the variance function.

The parameter  $\theta_{ij}$  relates  $\mu_{ij}$  to the linear predictor  $\eta_{ij} = \mathbf{x}_{ij}^{\mathsf{T}} \boldsymbol{\beta} + \mathbf{z}_{ij}^{\mathsf{T}} \boldsymbol{b}_i$  through the relationship

$$g(\mu_{ij}) = \eta_{ij} = \boldsymbol{x}_{ij}^{\mathsf{T}} \boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\mathsf{T}} \boldsymbol{b}_{i}, \qquad (2.2)$$

where  $x_{ij}$  is a  $p \times 1$  vector of explanatory variable values,  $\beta$  is a  $p \times 1$  vector of unknown parameters,  $z_{ij}$  is a  $q \times 1$  vector of non-stochastic variables and g(.) is a monotonic and differentiable function. Furthermore, we will assume that  $b_i \sim N_q(0, \Xi)$ , where  $\Xi$ is a symmetric and positive definite variance-covariance matrix, which depends on a  $q(q + 1)/2 \times 1$  vector of parameters  $\xi$ . The generalized linear mixed models are specified jointly by (2.1), (2.2) and the aforementioned distributional assumption on  $b_i$ .

Essentially, in this work, we consider the generalized linear mixed model (GLMM) specified as follows:

$$Y_{ij} | \boldsymbol{b}_i \stackrel{\text{ind}}{\sim} \text{EF} (\mu_{ij}; \phi) ,$$

$$g(\mu_{ij}) = \boldsymbol{x}_{ij}^{\mathsf{T}} \boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\mathsf{T}} \boldsymbol{b}_i \quad \text{and} \quad (2.3)$$

$$\boldsymbol{b}_i \stackrel{\text{iid}}{\sim} N_q(0; \boldsymbol{\Xi}) ,$$

where  $j = 1, ..., n_i$  and i = 1, ..., N. The notation  $EF(\cdot; \cdot)$  refers to the exponential family. That is, we assume that given  $b_i$ ,  $Y_{ij}$  follows a distribution of the form (2.1).

Furthermore, letting  $\boldsymbol{\eta}_i = (\eta_{i1}, \dots, \eta_{in_i})^{\top}$ ,  $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{in_i})^{\top}$  and  $\mathbf{g}(\boldsymbol{\mu}_i) = (\mathbf{g}(\mu_{i1}), \dots, \mathbf{g}(\mu_{in_i}))^{\top}$ , we obtain

$$\mathbf{g}\left(\boldsymbol{\mu}_{i}\right) = \mathbf{X}_{i}\boldsymbol{\beta} + \mathbf{Z}_{i}\boldsymbol{b}_{i},\tag{2.4}$$

where  $\mathbf{X}_{i} = (\mathbf{x}_{i1}, ..., \mathbf{x}_{in_{i}})^{\top}$  and  $\mathbf{Z}_{i} = (\mathbf{z}_{i1}, ..., \mathbf{z}_{in_{i}})^{\top}$ .

Finally, writing  $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^{\mathsf{T}}, \dots, \boldsymbol{\mu}_N^{\mathsf{T}})^{\mathsf{T}}$ ,  $\boldsymbol{\eta} = (\boldsymbol{\eta}_1^{\mathsf{T}}, \dots, \boldsymbol{\eta}_N^{\mathsf{T}})^{\mathsf{T}}$ ,  $\boldsymbol{b} = (\boldsymbol{b}_1^{\mathsf{T}}, \dots, \boldsymbol{b}_N^{\mathsf{T}})^{\mathsf{T}}$  and  $\mathbf{g}(\boldsymbol{\mu}) = (\mathbf{g}(\boldsymbol{\mu}_1)^{\mathsf{T}}, \dots, \mathbf{g}(\boldsymbol{\mu}_N)^{\mathsf{T}})^{\mathsf{T}}$  we can express the relationship (2.4) as

$$\mathbf{g}(\boldsymbol{\mu}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{b},\tag{2.5}$$

where  $\mathbf{X} = (\mathbf{X}_1^{\mathsf{T}}, \dots, \mathbf{X}_N^{\mathsf{T}})^{\mathsf{T}}$  and  $\mathbf{Z} = \bigoplus_{i=1}^N \mathbf{Z}_i$ , with  $\bigoplus$  denoting the direct sum of matrices.

### 2.2.2 On the estimation methods in GLMMs

Denote  $\boldsymbol{\psi} = (\boldsymbol{\beta}^{\mathsf{T}}, \boldsymbol{\xi}^{\mathsf{T}}, \boldsymbol{\phi})^{\mathsf{T}}$  the vector of parameters in the aforementioned model to be estimated. For an exhaustive review of different methods for estimating  $\boldsymbol{\psi}$  the reader may refer to McCulloch and Searle (2000), Breslow and Clayton (1993) and Demidenko (2013), among others.

Let  $\boldsymbol{y} = (\boldsymbol{y}_1^{\mathsf{T}}, \dots, \boldsymbol{y}_N^{\mathsf{T}})^{\mathsf{T}}$  be the  $N_T \times 1$  observation vector, where  $\boldsymbol{y}_i = (y_{i1}, \dots, y_{in_i})^{\mathsf{T}}$  is the observation vector for the *ith* subject. Hence, using similar arguments as Tang *et al.* (2006) the joint log-likelihood function of  $\boldsymbol{y}_i$  and  $\boldsymbol{b}_i$ , for  $i = 1, \dots, N$ , is given by

$$L(\boldsymbol{\psi}; \boldsymbol{y}_{i}, \boldsymbol{b}_{i}) = -\frac{q}{2} \log(2\pi) - \frac{1}{2} \log|\boldsymbol{\Xi}| + \sum_{j=1}^{n_{i}} \left[ \phi \left\{ y_{ij} \,\theta_{ij} - a(\theta_{ij}) \right\} + c(y_{ij}; \phi) \right] - \frac{1}{2} \boldsymbol{b}_{i}^{\top} \boldsymbol{\Xi}^{-1} \boldsymbol{b}_{i}.$$
(2.6)

#### 8 BASIC CONCEPTS

Thus, the marginal log-likelihood functions of  $y_i$  and y are, respectively,

$$L(\boldsymbol{\psi};\boldsymbol{y}_{i}) = \log\left[\int_{\mathbb{R}^{q}} \exp\left\{L\left(\boldsymbol{\psi};\boldsymbol{y}_{i},\boldsymbol{b}_{i}\right)\right\} d\boldsymbol{b}_{i}\right]$$
(2.7)

and

$$L(\boldsymbol{\psi};\boldsymbol{y}) = \sum_{i=1}^{N} L(\boldsymbol{\psi};\boldsymbol{y}_{i}). \qquad (2.8)$$

Unfortunately, in most cases the expression (2.7) can not be evaluated in closed form. However, Tuerlinckx *et al.* (2006) presented a complete review of methods for approximating that expression.

The maximum likelihood approach requires the treatment of (2.8) and hence the use of numerical methods, as the Gauss-Hermite quadrature, becomes a powerful tool to lead with the *q*-dimensional required integration (see, for instance, McCulloch and Searle, 2000). Also, it is possible to apply the Laplace approximation in (2.7) to avoid the multidimensional integration. The use of such approximation is directly associated with a penalized log-likelihood function which leads to "maximum penalized likelihood estimates" for  $\beta$  and *b*. The penalized log-likelihood function assumes the form

$$L_{P}(\boldsymbol{\beta}, \boldsymbol{b}) = \sum_{i=1}^{N} \sum_{j=1}^{n_{1}} \log \left\{ f_{y_{ij}|\boldsymbol{b}_{i}}(y_{ij}; \boldsymbol{\beta}, \boldsymbol{\phi}) \right\} - \frac{1}{2} \sum_{i=1}^{N} \boldsymbol{b}_{i}^{\top} \boldsymbol{\Xi}^{-1} \boldsymbol{b}_{i},$$
(2.9)

which is obtained via the Laplace approximation in (2.7) and by ignoring some terms of the resulting expression. This facilitates the estimation of  $\beta$  and prediction of b, renders the estimation of the variance-covariance parameters  $\xi$  difficult (Breslow and Clayton, 1993).

Alternative strategies of estimation based directly on the study of (2.7) are presented in McGilchrist (1994) and Nelder *et al.* (2006). Bayesian estimation tools for these models are presented, for instance, in Zhao *et al.* (2006) and Zeger and Karim (1991). In addition to the estimation of the fixed effects  $\beta$ , the variance-covariance parameters  $\xi$  and the precision parameter  $\phi$ , GLMMs allow estimation of the within subject dependence structure. In this sense, the intraclass correlation coefficient, Corr ( $Y_{ir}$ ,  $Y_{is}$ ), is extensively derived for particular cases of GLMMs in appendix C. The intraclass correlation for these models is induced by the random- effects distribution (McCulloch and Neuhaus, 2005).

It is worth noting that this work focuses on diagnostic methods for GLMMs and neither on parameter estimation nor on prediction of the random effects. The procedures developed in next sections will assume that the estimated values for  $\psi$  were obtained via some of the estimation methods mentioned earlier. In practice, even though the different estimation methods may lead to different values, the diagnostic procedures to be developed may be applied to each of them.

## 2.3 Local influence: general approach

With the goal of investigating the sensitivity of the estimate of a unknown parameter vector  $(\psi)$  under small perturbations in some components of the model or data, Cook (1986) proposed the concept of local influence. Essentially, the proposal is based on measuring such sensitivity through the behavior of the likelihood displacement  $LD_{\omega} = 2 \{L(\widehat{\psi}; y) - L(\widehat{\psi}_{\omega}; y)\}$ , where  $\widehat{\psi}_{\omega}$  denotes the maximum likelihood estimate under  $L(\psi; y | \omega)$ , which denotes the adopted form by the log-likelihood function of the postulated model  $L(\psi; y)$  when minor modifications are introduced into the model or data through the perturbation vector  $\omega \in \Theta \subseteq \mathbb{R}^l$  (l does not necessarily represent the sample size). The way of introducing the small modifications is not unique and depends on the objective of the investigation. The different possibilities for perturbing a model are called perturbation schemes. However, investigating characteristics of  $LD_{\omega}$ for all  $\omega \in \Theta$  could be unfeasible. Then, Cook (1986) proposed to study the behavior of  $LD_{\omega}$  in a neighborhood of the no perturbation vector  $\omega_0$ , that is,  $\omega_0$  is such that L  $(\psi; y | \omega_0) = L(\psi; y)$ . According to Cook (1986) this study must be made by means of the computation of the normal curvature  $C_d(\psi)$ , in the direction of some unitary vector **d**, of the graph  $LD_{\omega_0+a \, \mathbf{d}}$  against *a*, where  $a \in \mathbb{R}$ . Cook (1986) showed that  $C_d$  may be expressed as

$$C_{\mathbf{d}}(\boldsymbol{\psi}) = 2 \left| \mathbf{d}^{\mathsf{T}} \boldsymbol{\Delta}^{\mathsf{T}} \mathbf{L}_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1} \boldsymbol{\Delta} \mathbf{d} \right|, \qquad (2.10)$$

where

$$\mathbf{L}_{\psi\psi} = \frac{\partial^2 \mathbf{L}(\psi; \boldsymbol{y})}{\partial \psi \partial \psi^{\top}} \quad \text{and} \quad \boldsymbol{\Delta} = \frac{\partial^2 \mathbf{L}(\psi; \boldsymbol{y} \mid \boldsymbol{\omega})}{\partial \psi \partial \boldsymbol{\omega}^{\top}},$$

and these quantities are evaluated at  $\omega = \omega_0$  and  $\psi = \widehat{\psi}$ .

Also, if the vector of parameters is partitioned into  $\boldsymbol{\psi} = (\boldsymbol{\psi}_1^{\mathsf{T}}, \boldsymbol{\psi}_2^{\mathsf{T}})^{\mathsf{T}}$  and we are interested, for instance, in the local influence of the  $r \times 1$  vector  $\boldsymbol{\psi}_1$ , then the normal curvature  $C_d(\boldsymbol{\psi}_1)$  in the direction of the unitary direction **d** takes the form

$$C_{\mathbf{d}}\left(\boldsymbol{\psi}_{1}\right) = \left|\mathbf{d}^{\top}\boldsymbol{\Delta}^{\top}\left(\mathbf{L}_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1} - \mathbf{L}_{22}\right)\boldsymbol{\Delta}\mathbf{d}\right|, \qquad (2.11)$$

where

$$\mathbf{L}_{22} = \begin{bmatrix} \mathbf{0}_{r \times r} & \mathbf{0}_{r \times (s-r)} \\ \mathbf{0}_{(s-r) \times r} & \mathbf{L}_{\psi_2 \psi_2}^{-1} \end{bmatrix} \quad \text{and} \quad \mathbf{L}_{\psi_2 \psi_2} = \frac{\partial^2 \mathbf{L} (\psi; \boldsymbol{y})}{\partial \psi_2 \partial \psi_2^{\top}}.$$

Although the normal curvature (Cook, 1986) has been extensively used, it is not invariant under a uniform change of scale and could take any value. So, it is difficult to judge its size objectively. To reduce these complexities, Poon and Poon (1999) proposed the conformal normal curvature which has one to one correspondence with the normal curvature  $C_d$  and takes values in [0, 1]. On the basis of equations (2.10) and (2.11) the conformal normal curvature (Poon and Poon, 1999) is defined, respectively, by

$$B_{d}(\boldsymbol{\psi}) = \frac{-\left\{ \mathbf{d}^{\top} \boldsymbol{\Delta}^{\top} \left( \mathbf{L}_{\boldsymbol{\psi}\boldsymbol{\psi}} \right)^{-1} \boldsymbol{\Delta} \mathbf{d} \right\}}{\sqrt{\operatorname{tr}\left\{ \left( \boldsymbol{\Delta}^{\top} \mathbf{L}_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1} \boldsymbol{\Delta} \right)^{2} \right\}}}$$
(2.12)

and

$$B_{\mathbf{d}}\left(\boldsymbol{\psi}_{1}\right) = \frac{-\left\{\mathbf{d}^{\top} \boldsymbol{\Delta}^{\top} \left(\mathbf{L}_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1} - \mathbf{L}_{22}\right) \boldsymbol{\Delta} \mathbf{d}\right\}}{\sqrt{\operatorname{tr}\left[\left\{\boldsymbol{\Delta}^{\top} \left(\mathbf{L}_{\boldsymbol{\psi}\boldsymbol{\psi}}^{-1} - \mathbf{L}_{22}\right) \boldsymbol{\Delta}\right\}^{2}\right]}}.$$
(2.13)

This implies that  $0 \le B_d(\psi) \le 1$  and  $0 \le B_d(\psi_1) \le 1$ . All the quantities are evaluated at  $\omega = \omega_0$  and  $\psi = \widehat{\psi}$ .

Even though the conformal normal curvature  $B_d$  may be computed in any direction d, we may have particular interest in some of them and we will refer to such directions in later sections. We will use the conformal normal curvature as a tool to the assessment of local influence. Thus, we will direct our efforts to compute such curvature under GLMMs.

## 2.4 On the Laplace approximation

In this section we extend some results on Laplace approximations presented in Demidenko (2013). Essentially, our results are multivariate extensions of the expressions presented in chapter 7 of that book. However, before describing these extensions we mention some notation aspects that are essential for the understanding and development of our work. The proofs and details of each result in this section are reported in appendix A.

### 2.4.1 Some remarks on notation

In this section we mention briefly some relevant facts that will be used throughout. We begin with a general definition of derivatives for matrix valued functions based on the vectorization of such functions. This definition is very important because it allows an easy and systematic treatment of the Hessian matrix functions, which under this definition may be computed employing differential theory presented in Magnus and Neudecker (1999) or Abadir and Magnus (2005). Furthermore, we mention some key aspects on the notation that are helpful in understanding and obtaining

#### 12 BASIC CONCEPTS

the later results.

**Definition 1.** Let  $\mathbf{F} : \mathbf{S} \subseteq \mathbb{R}^q \longrightarrow \mathbb{R}^{r \times s}$  be defined on an open set  $\mathbf{S}$  in  $\mathbb{R}^q$  and suppose  $\mathbf{x}_0 \in \mathbf{S}$ . We define the Hessian matrix  $\mathbf{F}_{\mathbf{x}_0 \mathbf{x}_0}$  of  $\mathbf{F}$  at  $\mathbf{x}_0$  by

$$\mathbf{F}_{x_0 x_0} = \frac{\partial}{\partial \mathbf{x}^{\mathsf{T}}} \left( \operatorname{vec} \left[ \left\{ \frac{\partial \operatorname{vec} \left( \mathbf{F} \right)}{\partial \mathbf{x}^{\mathsf{T}}} \right\}^{\mathsf{T}} \right] \right) \Big|_{\mathbf{x} = x_0}.$$
(2.14)

Note that **F** could be partitioned as  $\mathbf{F} = (\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_s)$ , where  $\mathbf{F}_j = (\mathbf{F}_{1j}, \mathbf{F}_{2j}, \dots, \mathbf{F}_{rj})^{\top}$ and  $\mathbf{F}_{ij} : \mathbf{S} \subseteq \mathbb{R}^q \longrightarrow \mathbb{R}$  for  $1 \le i \le r$  and  $1 \le j \le s$ . Thus, in terms of the common partial derivative expression (2.14) assumes the form

$$\mathbf{F}_{\mathbf{x}_0 \mathbf{x}_0} = \left( \mathbf{F}_{1_{\mathbf{x}_0 \mathbf{x}_0}}^{\mathsf{T}}, \dots, \mathbf{F}_{s_{\mathbf{x}_0 \mathbf{x}_0}}^{\mathsf{T}} \right)^{\mathsf{T}}, \text{ where } \mathbf{F}_{j_{\mathbf{x}_0 \mathbf{x}_0}} = \left( \partial^2 \mathbf{F}_{1j} / \partial \mathbf{x} \partial \mathbf{x}^{\mathsf{T}}, \dots, \partial^2 \mathbf{F}_{rj} / \partial \mathbf{x} \partial \mathbf{x}^{\mathsf{T}} \right)^{\mathsf{T}} \Big|_{\mathbf{x}=\mathbf{x}_0}.$$

Frequently, computing derivatives of matrix valued functions using common techniques, like the strategies in Harville (1997), is somewhat complicated. For this reason, we will use the differential theory described by Magnus and Neudecker (1999) or Abadir and Magnus (2005), which enables to obtain the derivatives of interest in a simpler way.

#### **2.4.1.1** $F_{xx}$ and $F_{xx}^*$

In this work we distinguish between the Hessian  $r q s \times q$  matrix  $\mathbf{F}_{x_0 x_0}$  defined above and the  $r q \times s q$  matrix of Hessians  $\mathbf{F}_{x_0 x_0}^* = (\mathbf{F}_{1x_0 x_0}, \dots, \mathbf{F}_{s x_0 x_0})$ , which contains the same components that the Hessian matrix, but in a different pattern. Hence, it may be obtained by means of reordering the components of the Hessian matrix  $\mathbf{F}_{x_0 x_0}$ .

Furthermore, it is possible to deduce that for vector valued functions, that is s = 1,  $\mathbf{F}_{x_0 x_0} = \mathbf{F}_{x_0 x_0}^*$ . Consequently, for real valued functions, that is r = s = 1, such equivalence is also satisfied.

To illustrate the difference between  $\mathbf{F}_{xx}$  and  $\mathbf{F}_{xx}^*$  we consider the special case  $\mathbf{F}$ :  $S \subseteq \mathbb{R}^q \longrightarrow \mathbb{R}^{2\times 3}$ . Then,  $\mathbf{F}$  may be expressed in terms of its real valued components  $F_{ij}: S \subseteq \mathbb{R}^q \longrightarrow \mathbb{R}$  for  $1 \le i \le 2$  and  $1 \le j \le 3$ . Thus, we have

$$\mathbf{F} = \begin{bmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \end{bmatrix}; \quad \mathbf{F}_{x_0 x_0} = \begin{bmatrix} \mathbf{U}_{11} & \mathbf{U}_{21} & \mathbf{U}_{12} & \mathbf{U}_{22} & \mathbf{U}_{13} & \mathbf{U}_{23} \end{bmatrix}^{\mathsf{T}}; \quad \mathbf{F}_{x_0 x_0}^* = \begin{bmatrix} \mathbf{U}_{11} & \mathbf{U}_{12} & \mathbf{U}_{13} \\ \mathbf{U}_{21} & \mathbf{U}_{22} & \mathbf{U}_{23} \end{bmatrix}$$

where  $\mathbf{U}_{ij} = \partial^2 \mathbf{F}_{ij} / \partial \mathbf{x} \partial \mathbf{x}^{\top}$  for  $1 \le i \le 2$  and  $1 \le j \le 3$ .

Although  $F_{x_0x_0}^*$  appears to be a more natural definition for the Hessian of a matrix function **F**, Magnus and Neudecker (1999) discussed the reasons and advantages of the Definition 1.

The reader may be asking why distinguishing between  $\mathbf{F}_{x_0 x_0}$  and  $\mathbf{F}_{x_0 x_0}^*$  and the answer is simple: the strategies exposed by Magnus and Neudecker (1999) and Abadir and Magnus (2005) allow the direct and systematic computation of  $\mathbf{F}_{x_0 x_0}$  but not of  $\mathbf{F}_{x_0 x_0}^*$ . However, for a general and suitable presentation of the results exposed in this work, we require the structure  $\mathbf{F}_{x_0 x_0}^*$ . This can be seen in the next section, in which we present multivariate consequences of the Laplace approximation.

Actually,  $\mathbf{F}_{x_0x_0} \neq \mathbf{F}_{x_0x_0}^*$  only when **F** is a matrix function (that is  $s \neq 1$ ) and in these cases  $\mathbf{F}_{x_0x_0}$  is just the road for obtaining  $\mathbf{F}_{x_0x_0}^*$ . In general, the direct computation of  $\mathbf{F}_{x_0x_0}^*$  may be performed computing the Hessians of the components  $\mathbf{F}_{ij}$  one by one, but this procedure is impractical. Thus, we prefer to compute  $\mathbf{F}_{x_0x_0}$  and reordering its elements to achieve  $\mathbf{F}_{x_0x_0}^*$ .

It is worth mentioning that for a real valued function g the notation  $\mathbf{g}_{xx}$  given in Definition 1 agrees with the usual definition of Hessian for this type of function. That is,  $\mathbf{g}_{xx} = \frac{\partial^2 g}{\partial x \partial x^{\top}}$ . In addition, we write  $\mathbf{g}_x$  for denoting the partial derivative  $\frac{\partial g}{\partial x}$ .

### 2.4.2 Consequences of the Laplace approximation

The Laplace approximation may be used to avoid multidimensional integration. Specifically, it is suitable for obtaining approximations for the integral

$$\int_{\mathbb{R}^q} \exp\left\{h\left(x\right)\right\} dx,\tag{2.15}$$
where h is a unimodal function.

According to Demidenko (2013), if h is an unimodal function with continuous partial derivatives up to second order, we have

$$\int_{\mathbb{R}^{q}} \exp\{\mathbf{h}(\mathbf{x})\} d\mathbf{x} \approx (2\pi)^{q/2} \exp\{\mathbf{h}(\widetilde{\mathbf{x}})\} \left|-\mathbf{h}_{\widetilde{\mathbf{x}}\widetilde{\mathbf{x}}}\right|^{-1/2},$$
(2.16)

where  $\mathbf{h}_{\tilde{x}\tilde{x}}$  is the Hessian matrix of h evaluated at the maximum point of h, denoted by  $\tilde{x}$ .

Demidenko (2013) argued that to check if a function h is unimodal, it is enough to prove that the negative Hessian matrix is positive definite for all  $x \in \mathbb{R}^{q}$ . Some authors, such as Breslow and Clayton (1993) based their estimation procedures for GLMMs on this approximation. Other authors, such as Vonesh and Chinchilli (1996) and Pinheiro and Bates (2000), give special relevance for this approximation as a powerful tool in estimation processes for mixed effect models, where h is some function related with the likelihood function corresponding to the involved model.

In our approach, we want to approximate ratios of integrals such as

$$\frac{\int\limits_{\mathbb{R}^{q}} v(x) \exp\left\{h\left(x\right)\right\} dx}{\int\limits_{\mathbb{R}^{q}} \exp\left\{h\left(x\right)\right\} dx},$$
(2.17)

where  $v : S \subseteq \mathbb{R}^q \longrightarrow \mathbb{R}$  is a continuous function with continuous partial derivatives up to second order. Next we propose some theorems for getting such approximations.

**Theorem 1.** Let  $h, v : S \subseteq \mathbb{R}^q \longrightarrow \mathbb{R}$  be continuous functions with continuous partial derivatives up to second order in a neighborhood of the point  $\tilde{x}$  where h is maximum. Thus, we have that

$$\int_{\mathbb{R}^{q}} v(\mathbf{x}) \exp\{\mathbf{h}(\mathbf{x})\} d\mathbf{x} \approx (2\pi)^{q/2} \exp\{\mathbf{h}(\widetilde{\mathbf{x}})\} \left|-\mathbf{h}_{\widetilde{\mathbf{x}}\widetilde{\mathbf{x}}}\right|^{-1/2} \left[v(\widetilde{\mathbf{x}}) - \frac{1}{2} \operatorname{tr}\left\{(\mathbf{h}_{\widetilde{\mathbf{x}}\widetilde{\mathbf{x}}})^{-1} (v_{\widetilde{\mathbf{x}}\widetilde{\mathbf{x}}})\right\}\right],$$

where  $\mathbf{h}_{\widetilde{x}\widetilde{x}} = \left. \frac{\partial \mathbf{h}(x)}{\partial x \partial x^{\top}} \right|_{x=\widetilde{x}}$  and  $v_{\widetilde{x}\widetilde{x}} = \left. \frac{\partial v(x)}{\partial x \partial x^{\top}} \right|_{x=\widetilde{x}}$ .

Proof. See appendix A.1.

**Theorem 2.** Under the same conditions of Theorem 1, we have

$$\frac{\int\limits_{\mathbb{R}^{q}} \upsilon\left(\mathbf{x}\right) \exp\left\{\mathbf{h}\left(\mathbf{x}\right)\right\} d\mathbf{x}}{\int\limits_{\mathbb{R}^{q}} \exp\left\{\mathbf{h}\left(\mathbf{x}\right)\right\} d\mathbf{x}} \approx \upsilon\left(\widetilde{\mathbf{x}}\right) - \frac{1}{2} \operatorname{tr}\left\{\left(\mathbf{h}_{\widetilde{\mathbf{x}}\,\widetilde{\mathbf{x}}}\right)^{-1} \left(\boldsymbol{\upsilon}_{\widetilde{\mathbf{x}}\,\widetilde{\mathbf{x}}}\right)\right\}$$
$$= \upsilon\left(\widetilde{\mathbf{x}}\right) + \frac{1}{2} \operatorname{vec}\left(\mathbf{I}_{q}\right)^{\top} \left(\boldsymbol{\upsilon}_{\widetilde{\mathbf{x}}\,\widetilde{\mathbf{x}}} \otimes \mathbf{I}_{q}\right) \operatorname{vec}\left(-\mathbf{h}_{\widetilde{\mathbf{x}}\,\widetilde{\mathbf{x}}}^{-1}\right),$$

where  $\mathbf{I}_q$  denotes the identity matrix of order q and  $\otimes$  represents the Kronecker product.

*Proof.* See appendix A.2.

These theorems are essential to approximate moments of an unknown conditional density and to generalize results given in Demidenko (2013) on Laplacian approximations.

One consequence of the previous result is given in the following theorem.

**Theorem 3.** Let  $\Upsilon : S \subseteq \mathbb{R}^q \longrightarrow \mathbb{R}^{r \times s}$  be a matrix valued function with real valued components  $v_{ij} : S \subseteq \mathbb{R}^q \longrightarrow \mathbb{R}$  for  $1 \le i \le r$  and  $1 \le j \le s$ . Also, let  $\mathbf{A} = (a_{ij})$  be a  $r \times s$  matrix, where

$$a_{ij} = \frac{\int\limits_{\mathbb{R}^q} v_{ij}(\mathbf{x}) \exp\left\{h\left(\mathbf{x}\right)\right\} d\mathbf{x}}{\int\limits_{\mathbb{R}^q} \exp\left\{h\left(\mathbf{x}\right)\right\} d\mathbf{x}},$$

and h is a function with the previous conditions. Then,

$$\mathbf{A} \approx \mathbf{\Upsilon} \left( \widetilde{\mathbf{x}} \right) + \left\{ \mathbf{I}_r \otimes \operatorname{vec} \left( \mathbf{I}_q \right)^{\top} \right\} \left( \mathbf{\Upsilon}_{\widetilde{\mathbf{x}} \, \widetilde{\mathbf{x}}}^* \otimes \mathbf{I}_q \right) \left\{ \mathbf{I}_s \otimes \operatorname{vec} \left( -\mathbf{h}_{\widetilde{\mathbf{x}} \, \widetilde{\mathbf{x}}}^{-1} \right) \right\},$$
(2.18)

where  $\Upsilon^*_{\widetilde{x}\widetilde{x}}$  is the matrix of Hessians of  $\Upsilon$ .

*Proof.* See appendix A.3.

Remember that  $\Upsilon_{\tilde{x}\tilde{x}}^*$  may be obtained by reordering the elements of the Hessian matrix  $\Upsilon_{\tilde{x}\tilde{x}}$ , which was considered in Definition 1 of section 2.4.1 (for more details, see section 2.4.1.1).

This theorem reflects the importance of employing the matrix of Hessians  $\mathbf{F}_{x_0x_0}^*$  for a matrix function **F**. This will be applied extensively throughout this work.

## 2.5 Concluding remarks

In this chapter we defined GLMMs and briefly described the local influence methodology for sensitivity analysis. Some aspects of notation were introduced, specifically the difference between the Hessian matrix  $F_{xx}$  and the matrix of Hessians  $F_{xx}^*$ . Also, employing this notation we proposed some generalizations of the results presented in Demidenko (2013) on the Laplace approximation. Such generalizations are one of the key tools for the development of our work.

## Chapter 3

# Local influence in GLMMs

## 3.1 Introduction

In this chapter we adapt the local influence approach presented in section 2.3 to GLMMs. Firstly, we mention some known works in sensitivity analysis for random effect models and we expose our methodology in detail. After that, we justify the impossibility of obtaining exact local influence measures for GLMMs and therefore we propose approximate structures of those measures by means of the results in section 2.4.2. To approximate the local influence measures it is necessary to compute, at least in an approximated way, moments of a random variable  $b_i | y_i$ . This is the aim of a specific section in this chapter.

On the other hand, according to Ouwens *et al.* (2001) it is important to detect influential subjects and influential observations. Following these authors, influential observations may be distributed across the different subjects which are not necessarily discriminated as potentially influential. Studying local influence at the subject and observation levels, we can distinguish between the subject specific influence and the influence due to features of observations on a specific subject. Thus, we perform our sensitivity analysis both at the subject and at the observation levels.

## 3.2 Works on local influence in random effect models

Several papers have been written under the framework of local influence for random effect models. For example, Lesaffre and Verbeke (1998) studied local influence in linear mixed models, which are a particular case of GLMMs. This idea is fully based in the original proposal of Cook (1986). The same ideas were used by Rakhmawati *et al.* (2017) in some special cases of GLMMs via one special perturbation scheme that only allows to discriminate potentially influential subjects but not potentially influential observations. Also that study focused exclusively in the computation of the  $\Delta$  matrix given in (2.10) and the building of  $L_{\psi\psi}$  is performed through fully numerical procedures in special situations without providing a general structure. Using a Bayesian approach, Ouwens *et al.* (2001) studied local influence in Poisson mixed models and in negative binomial mixed models.

In a more general context, Tang *et al.* (2006) studied influence diagnostics in nonlinear reproductive dispersion mixed models (NRDMM) which emcompass generalized Cook distance and local influence tools. Although the NRDMMs are more general than the GLMMs, the study of sensitivity by means of local influence performed by these authors focused only on the fixed effects  $\beta$ . In fact, if we consider that approach for GLMMs, it is possible to deduce that their local influence analysis is equivalent to developing the curvatures (2.12) and (2.13) based on the penalized log-likelihood function (2.9) considered by Breslow and Clayton (1993). This penalized log-likelihood function, which is an approximation of the marginal log-likelihood function L ( $\psi$ ; y), is obtained via Laplace approximation (2.16) in (2.7) by ignoring some terms of the resulting expression. The ignored terms according to Breslow and Clayton (1993) and Tang *et al.* (2006) do not affect the estimation procedures of the fixed effects. This facilitates the estimation of the fixed effects  $\beta$  and the prediction of  $b_i$ , but prevents the estimation of the variance-covariance components  $\xi$ , because among the ignored terms there are quantities with a strong dependence on  $\xi$ . Consequently, the study of local influence with respect to  $\xi$  based in penalized log-likelihood function is not accurate. In fact, Breslow and Clayton (1993) suggested an alternative process motivated by the estimation of the variance components in linear mixed models via restricted maximum likelihood (REML). This is described in chapter 17 of Pawitan (2001). Indeed, this is the reason for which the local influence analysis of Tang *et al.* (2006) is considered only for  $\beta$ . Consequently, since our aim is to study the local influence on  $\psi$ , the penalized log-likelihood function presented in (2.9) is not an option.

As it will be seen later, our work allows to identify observations and subjects with an undue weight on the estimates for  $\boldsymbol{\psi} = (\boldsymbol{\beta}^{\mathsf{T}}, \boldsymbol{\xi}^{\mathsf{T}}, \boldsymbol{\phi})^{\mathsf{T}}$ .

A different approach for the study of GLMMs, and in general for random effect models, is to treat the random effects as missing observations. This inspired the use of a new measure of influence, different from the likelihood displacement  $(LD_{\omega})$  defined previously, but with similar properties. Such a measure is called *Q*-displacement and has been extensively used by Zhu and Lee (2001) in models with missing observations and extended to sensitivity analysis for nonlinear structural equation models with ignorable missing data by Lee *et al.* (2006), for nonlinear mixed effect models by Lee and Xu (2004) and for GLMMs by Zhu and Lee (2003). The use of this measure of influence is based on techniques such as MCMC, Gibbs sampling among others, which could be computationally intensive (Nelder *et al.*, 2006) and difficult in the construction of useful procedures in practical problems. On this respect, Tapia *et al.* (2019b) and Tapia *et al.* (2019a) performed studies of local influence in logistic and Poisson random effect models and argued for this approach to avoid the multidimensional integration emerging from (2.7).

We prefer to use the ideas of the seminal paper by Cook (1986), the proposal of conformal normal curvature given in Poon and Poon (1999) and perturbation schemes for discriminating potentially influential subjects and observations. These considerations will be treated with more detail in the next sections.

## 3.3 Local influence in GLMMs: our strategy

The diagnostic methods proposed in this work has the expressions (2.7) and (2.8) and derived expressions of them as starting point. We will use the Laplace approximation, specifically Theorems 1, 2 and 3, presented in section 2.4.2, to avoid the difficulties of integration coming from the intractable high-dimensional integrals that the log-likelihood function associated with the GLMMs, given in (2.8), involves.

In the context of GLMMs we are interested in identifying influential subjects and observations. In the former, we use the ideas of Lesaffre and Verbeke (1998). Thus, for computing the conformal normal curvature defined in (2.12) and (2.13), we consider the perturbation vector  $\boldsymbol{\omega} \in \mathbb{R}^N$ , the direction defined by  $\mathbf{d} = \mathbf{u}_i \in \mathbb{R}^N$ , where  $\mathbf{u}_i$  denotes a vector with 1 in the ith position and zeros in the remaining positions and we regard the ith subject as likely influential if  $B_{u_i} = B_i > 2\overline{B}$ , where  $B = \{B_i : i = 1, ..., N\}$ . In the latter case, we compute  $B_d$  using  $\omega \in \mathbb{R}^{N_T}$  with direction defined by the partitioned vector  $\mathbf{d} = (\mathbf{d}_1^{\mathsf{T}}, \ldots, \mathbf{d}_N^{\mathsf{T}})^{\mathsf{T}} = \mathbf{u}_{ij} \in \mathbb{R}^{N_T}$ , which assumes zero everywhere, except in the jth position of  $\mathbf{d}_i \in \mathbb{R}^{n_i}$ , in which it contains 1. Following Lee *et al.* (2006) we will say that the jth observation of the ith subject is potentially influential if  $B_{u_{ii}}$  =  $B_{ij} > k_1 \overline{B} + k_2 SD(B)$ , where  $B = \{B_{ij} : j = 1, \dots, n_i \text{ and } i = 1, \dots, N\}$  and  $k_1$  and  $k_2$  are appropriately selected constants. In both cases, the index plots of B<sub>i</sub> and B<sub>ij</sub> may reveal individually influential subjects and observations, respectively. Also, as suggested by Cook (1986), we study the normal curvature  $B_d$  in the direction of the eigenvector of  $-\Delta\,L_{\psi\,\psi}^{-1}\,\Delta$  corresponding to the largest eigenvalue. This vector will be denoted by  $d_{\text{max}}$ and maximizes  $B_d$ . It is worth mentioning that in the former  $d_{max} \in \mathbb{R}^N$  and in the latter  $\mathbf{d}_{\max} \in \mathbb{R}^{N_T}$ . Inspect the index plot of  $|\mathbf{d}_{\max}|$  may reveal simultaneously influential subjects and observations.

For obtaining exact local influence measures like those derived by Lesaffre and Verbeke (1998), Espinheira *et al.* (2008) and Paula (2013a) for linear mixed models, beta regression models and double generalized linear models, respectively, we need a closed form

for the log-likelihood function of the postulated model. Nevertheless, this is not satisfied in GLMMs, since in general L ( $\psi$ ; y) does not have closed form (see, section 2.2.1). Therefore, we will compute approximate local influence measures for these models employing the results given in section 2.4.2 on Laplace approximation jointly with the integral-based approach proposed by Rakhmawati *et al.* (2017). This approach basically consists of using the integral expression for L ( $\psi$ ; y) given in (2.8) and the property of interchangeability of integration and derivation under mild regularity conditions. We want to emphasize that even though the Laplace approximation is the basis of the results exhibited in section 2.4.2, we do not intend to approximate the marginal loglikelihood L ( $\psi$ ; y) and from such approximation to compute the curvatures defined in (2.12). We develop the analysis of sensitivity from the full expression for L ( $\psi$ ; y) and since the computed curvatures do not have a closed form, approximations will be obtained by using the results of section 2.4.2.

From here, the priority will be to compute general structures, at least approximated, for the components of the normal conformal curvature defined in (2.12), namely  $\Delta$  and  $L_{\psi\psi}$ . Thus, in the next section we will begin with the computation of  $L_{\psi\psi}$  and consequently of  $L_{\psi}$ , which do not vary with the perturbation scheme.

## 3.4 Score $L_{\psi}$ and Hessian $L_{\psi\psi}$

In this section, we provide the score  $\mathbf{L}_{\psi} = \partial \mathbf{L}(\psi; y) / \partial \psi$  and the Hessian matrix  $\mathbf{L}_{\psi\psi}$ . In appendices B.3 and B.4 are presented the details on the derivation of these expressions.

To obtain the structures for the score function and the Hessian matrix, we use fundamental concepts of linear algebra as the  $rs \times rs$  commutation matrix  $\mathbf{K}_{rs}$  defined implicitly by the equation  $\mathbf{K}_{rs} \operatorname{vec}(\mathbf{A}) = \operatorname{vec}(\mathbf{A}^{\top})$  for any  $r \times s$  matrix  $\mathbf{A}$  (Abadir and Magnus , 2005). This commutation matrix  $\mathbf{K}_{rs}$  will be denoted by  $\mathbf{K}_r$  when s = r. The explicit definition of  $\mathbf{K}_{rs}$  may be reviewed in Harville (1997). Also, we will employ the  $q^2 \times q(q + 1)/2$  duplication matrix  $\mathbf{G}_q$  defined implicitly by the expression  $\mathbf{G}_q \operatorname{vech}(\mathbf{A}) = \operatorname{vec}(\mathbf{A})$  for any square  $q \times q$  matrix (Abadir and Magnus, 2005). Again, the explicit definition of  $\mathbf{G}_q$  may be reviewed in Harville (1997). Remember that  $\operatorname{vech}(\mathbf{A})$  denotes the q(q + 1)/2 vector that is obtained from  $\operatorname{vec}(\mathbf{A})$  by deleting all supradiagonal elements of A. For convenience we represent the  $r^2 \times r^2$  matrix  $\mathbf{I}_{r^2} + \mathbf{K}_r$  by  $\mathbf{N}_r$ . The duplication matrix  $\mathbf{G}_q$  and the commutation matrix  $\mathbf{K}_{rs}$  are available in the matrixcalc package of the R- software (R Core Team, 2018).

In addition, the following definitions are necessary for an adequate and simplified presentation of the results.

**Definition 2.** Let **A** represent an  $m \times n$  matrix and **B** a  $p \times q$  matrix. Then, we define the matrix valued functions  $\Lambda_q : \mathbb{R}^{m \times n} \longrightarrow \mathbb{R}^{m n q \times q}$  and  $\Pi_m : \mathbb{R}^{p \times q} \longrightarrow \mathbb{R}^{p q m \times m}$  by

$$\mathbf{\Lambda}_{q}\left(\mathbf{A}\right) = \left(\mathbf{I}_{n} \otimes \mathbf{K}_{qm}\right) \left(\operatorname{vec}\left(\mathbf{A}\right) \otimes \mathbf{I}_{q}\right) \quad and \quad \mathbf{\Pi}_{m}\left(\mathbf{B}\right) = \left(\mathbf{K}_{qm} \otimes \mathbf{I}_{p}\right) \left(\mathbf{I}_{m} \otimes \operatorname{vec}\left(\mathbf{B}\right)\right). \quad (3.1)$$

**Definition 3.** Let  $\mathbf{I}_{n_i}$  be the  $n_i \times n_i$  identity matrix and  $\mathbf{e}_{ij}$  its jth column. Then, the direct sum of the  $\mathbf{e}_{ij}$ 's is denoted by  $\mathbf{\Omega}_i$ . That is,  $\mathbf{\Omega}_i = \bigoplus_{i=1}^{n_i} \mathbf{e}_{ij}$ . Note that  $\mathbf{\Omega}_i$  is an  $n_i^2 \times n_i$  matrix.

Now, given (2.8), the computation of  $\mathbf{L}_{\psi}$  and  $\mathbf{L}_{\psi\psi}$  will be obtained from the contribution of the *ith* subject,  $\mathbf{L}^{i} = \mathbf{L}(\psi; y_{i})$  to the marginal log-likelihood function  $\mathbf{L} = \mathbf{L}(\psi; y)$ .

Let

•  $\mathbf{V}_i = \bigoplus_{j=1}^{n_i} \mathbf{V}_{ij}$  and  $\mathbf{W}_i = \bigoplus_{j=1}^{n_i} \mathbf{W}_{ij}$  be the weight matrix (McCulloch and Searle , 2000), where  $W_{ij} = \left( d \, \mu_{ij} / d \, \eta_{ij} \right)^2 / \mathbf{V}_{ij}$ ,

• 
$$\boldsymbol{\tau}_i = \mathbf{W}_i^{1/2} \mathbf{V}_i^{-1/2} (\boldsymbol{y}_i - \boldsymbol{\mu}_i), \boldsymbol{\lambda}_i = \boldsymbol{b}_i \otimes \boldsymbol{b}_i, \boldsymbol{\nu}_i = \sum_{j=1}^{n_i} \{ y_{ij} \theta_{ij} - a(\theta_{ij}) + c_{\phi} (y_{ij}; \phi) \},$$

• 
$$\mathbf{A}_i = \mathrm{d}\boldsymbol{\mu}_i/\mathrm{d}\boldsymbol{\eta}_i^{\mathsf{T}} = \bigoplus_{j=1}^{n_i} \mathrm{d}\boldsymbol{\mu}_{ij}/\mathrm{d}\boldsymbol{\eta}_{ij}, \ \boldsymbol{\rho}_i = (\rho_{i1}, \rho_{i2}, \dots, \rho_{in_i})^{\mathsf{T}},$$

• 
$$\rho_{ij} = \frac{\mathrm{d}}{\mathrm{d}\eta_{ij}} \left( \frac{\mathrm{d}\theta_{ij}}{\mathrm{d}\mu_{ij}} \frac{\mathrm{d}\mu_{ij}}{\mathrm{d}\eta_{ij}} \right) = \frac{\mathrm{d}^2 \theta_{ij}}{\mathrm{d}\mu_{ij}^2} \left( \frac{\mathrm{d}\mu_{ij}}{\mathrm{d}\eta_{ij}} \right)^2 + \frac{\mathrm{d}\theta_{ij}}{\mathrm{d}\mu_{ij}} \frac{\mathrm{d}^2 \mu_{ij}}{\mathrm{d}\eta_{ij}^2} \text{ and}$$

•  $\mathbf{M}_i = \mathbf{D}_{y_i - \mu_i} \mathbf{D}_{\rho_i} - \mathbf{W}_i^{1/2} \mathbf{V}_i^{-1/2} \mathbf{A}_i$ , where for a  $r \times 1$  vector  $\boldsymbol{\alpha}$ ,  $\mathbf{D}_{\boldsymbol{\alpha}} = \bigoplus_{j=1}^r \alpha_j$  and  $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_r)^{\mathsf{T}}$ .

Taking into account that

$$\mathbf{L}_{\psi}^{i} = \left( \left( \mathbf{L}_{\beta}^{i} \right)^{\mathsf{T}}, \left( \mathbf{L}_{\xi}^{i} \right)^{\mathsf{T}}, \mathbf{L}_{\phi}^{i} \right)^{\mathsf{T}} \quad \text{and} \quad \mathbf{L}_{\psi\psi}^{i} = \begin{bmatrix} \mathbf{L}_{\beta\beta}^{i} & \mathbf{L}_{\beta\xi}^{i} & \mathbf{L}_{\beta\phi}^{i} \\ \mathbf{L}_{\xi\beta}^{i} & \mathbf{L}_{\xi\xi}^{i} & \mathbf{L}_{\xi\phi}^{i} \\ \mathbf{L}_{\phi\beta}^{i} & \mathbf{L}_{\phi\xi}^{i} & \mathbf{L}_{\phi\phi}^{i} \end{bmatrix}, \quad (3.2)$$

it is possible to prove that (see appendices B.3 and B.4)

$$\mathbf{L}_{\boldsymbol{\beta}}^{i} = \boldsymbol{\phi} \operatorname{E} \left( \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i} \middle| \boldsymbol{y}_{i} \right), \quad \mathbf{L}_{\boldsymbol{\xi}}^{i} = -\frac{1}{2} \mathbf{G}_{q}^{\mathsf{T}} \left\{ \operatorname{vec} \left( \boldsymbol{\Xi}^{-1} \right) - \left( \boldsymbol{\Xi} \otimes \boldsymbol{\Xi} \right)^{-1} \operatorname{E} \left( \boldsymbol{\lambda}_{i} \middle| \boldsymbol{y}_{i} \right) \right\}, \quad \mathbf{L}_{\boldsymbol{\phi}}^{i} = \operatorname{E} \left( \boldsymbol{\nu}_{i} \middle| \boldsymbol{y}_{i} \right)$$
(3.3)

and

$$\begin{split} \mathbf{L}_{\beta\beta}^{i} &= \phi^{2} \operatorname{Var} \left( \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \middle| \boldsymbol{y}_{i} \right) + \phi \operatorname{E} \left( \mathbf{X}_{i}^{\top} \mathbf{M}_{i} \mathbf{X}_{i} \middle| \boldsymbol{y}_{i} \right), \\ \mathbf{L}_{\phi\phi}^{i} &= \sum_{j=1}^{n_{i}} c_{\phi\phi} \left( \boldsymbol{y}_{ij}; \phi \right) + \operatorname{Var} \left( \boldsymbol{v}_{i} \middle| \boldsymbol{y}_{i} \right), \\ \mathbf{L}_{\beta\xi}^{i} &= \frac{\phi}{2} \operatorname{Cov} \left( \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i}, \boldsymbol{\lambda}_{i} \middle| \boldsymbol{y}_{i} \right) \left( \Xi \otimes \Xi \right)^{-1} \mathbf{G}_{q}, \\ \mathbf{L}_{\beta\phi}^{i} &= \operatorname{E} \left( \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \middle| \boldsymbol{y}_{i} \right) + \phi \operatorname{Cov} \left( \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i}, \boldsymbol{v}_{i} \middle| \boldsymbol{y}_{i} \right), \\ \mathbf{L}_{\xi\phi}^{i} &= \frac{1}{2} \mathbf{G}_{q}^{\top} \left( \Xi \otimes \Xi \right)^{-1} \operatorname{Cov} \left( \boldsymbol{\lambda}_{i}, \boldsymbol{v}_{i} \middle| \boldsymbol{y}_{i} \right) \quad \text{and} \\ \mathbf{L}_{\xi\xi}^{i} &= \frac{1}{2} \mathbf{G}_{q}^{\top} \left[ \mathbf{I}_{q^{2}} - \left\{ \operatorname{E} \left( \boldsymbol{\lambda}_{i} \middle| \boldsymbol{y}_{i} \right)^{\top} \otimes \mathbf{I}_{q^{2}} \right\} \mathbf{P} + \frac{1}{2} \left( \Xi \otimes \Xi \right)^{-1} \operatorname{Var} \left( \boldsymbol{\lambda}_{i} \middle| \boldsymbol{y}_{i} \right) \right] \left( \Xi \otimes \Xi \right)^{-1} \mathbf{G}_{q}, \end{split}$$

where  $\mathbf{P} = \{\mathbf{I}_{q} \otimes \mathbf{\Pi}_{q} (\boldsymbol{\Xi}^{-1})\} + \{\mathbf{\Lambda}_{q} (\boldsymbol{\Xi}^{-1}) \otimes \mathbf{I}_{q}\}.$  Also,  $\mathbf{L}_{\xi\beta}^{i} = (\mathbf{L}_{\beta\xi}^{i})^{\mathsf{T}}, \mathbf{L}_{\phi\beta}^{i} = (\mathbf{L}_{\beta\phi}^{i})^{\mathsf{T}}$  and  $\mathbf{L}_{\phi\xi}^{i} = (\mathbf{L}_{\xi\phi}^{i})^{\mathsf{T}}.$ 

The computations above are valid for i = 1, ..., N. Thus, it is enough to take into account that  $\mathbf{L}_{\psi\psi} = \sum_{i=1}^{N} \mathbf{L}_{\psi\psi}^{i}$  and  $\mathbf{L}_{\psi} = \sum_{i=1}^{N} \mathbf{L}_{\psi}^{i}$  for obtaining the score  $\mathbf{L}_{\psi}$  and the Hessian  $\mathbf{L}_{\psi\psi}$ .

Indeed, the expected values and the variances considered above should be expressed as  $E_{\mathbf{b}_i | \mathbf{y}_i} (\cdot | \mathbf{y}_i)$  and  $\operatorname{Var}_{\mathbf{b}_i | \mathbf{y}_i} (\cdot | \mathbf{y}_i)$ . However, because the simplicity in the notation and since this must not give rise to doubt we prefer using  $E(\cdot | \mathbf{y}_i)$  and  $Var(\cdot | \mathbf{y}_i)$ , respectively.

# 3.5 Approximate local influence through the Laplace approximation

In this section we will use the results obtained in section 2.4.2 on Laplace approximation with the aim of achieving approximated expressions for the components of the conformal normal curvature defined in (2.12) and (2.13).

#### 3.5.1 Remark on the notation

We distinguish between the symbols "  $\widehat{}$  " and "  $\widehat{}$  ". The first one refers to the estimates of the parameters involved in a GLMM ( $\psi$ ), and the latter is related with the predicted values for the random effects b. Unlike generalized linear models (GLMs) some of the expressions defined in section 3.4, such as  $W_i$ ,  $M_i$ ,  $\tau_i$  and  $\nu_i$  rely on  $\psi$  and  $b_i$  jointly and throughout we need to discriminate when these expressions are evaluated at  $\hat{\psi}$  only or at  $\hat{\psi}$  and  $\tilde{b}_i$  jointly. Therefore, an accurate notation is necessary.

For example, the notation  $\widehat{\mathbf{W}}_i$  means that the weight matrix  $\mathbf{W}_i$  is evaluated at  $\widehat{\boldsymbol{\psi}}$  but not at  $\widetilde{\boldsymbol{b}}_i$  and consequently the matrix  $\widehat{\mathbf{W}}_i$  depends on  $\boldsymbol{b}_i$  only. The natural choice for representing  $\mathbf{W}_i$  evaluated at  $\widehat{\boldsymbol{\psi}}$  and  $\widetilde{\boldsymbol{b}}_i$  jointly lead to a tedious notation and we prefer to use  $\widetilde{\mathbf{W}}_i$  to denote this indicate. In summary

$$\widehat{\mathbf{W}}_i = \mathbf{W}_i \Big|_{\widehat{\psi}}$$
 and  $\widetilde{\mathbf{W}}_i = \widehat{\mathbf{W}}_i \Big|_{\widetilde{b}_i} = \mathbf{W}_i \Big|_{\widehat{\psi}, \widetilde{b}_i}$ , for  $i = 1, \dots, N$ .

As a consequence, for other expressions that rely on  $b_i$  only, such as  $\lambda_i$ , the notation  $\lambda_i$  represents  $\lambda_i|_{\widetilde{h}_i}$  as is usual.

## **3.5.2** Approximate computation of the moments of $b_i | y_i$

The expressions in (3.3) and (3.4) require the computation of conditional expectations of the form  $E_{b_i|y_i} \{\kappa(b_i) | y_i\}$ , for some real valued function  $\kappa$ . Particularly, in the simplest case we need to compute  $E_{b_i|y_i}(b_i|y_i)$ , defined in McCulloch and Searle (2000) as the best predictor for  $b_i$ . But, to compute exactly that expression the knowledge of the distribution of  $b_i|y_i$  is required, which is explicitly available only for conjugate distributions (see for instance, Nelder *et al.*, 2006).

Unfortunately, for the majority of GLMMs such distribution does not have a closed form; hence, accurate expressions are unfeasible. Thus, computational tools such as the Gibbs sampling, Markov-Chain Monte-Carlo methods or numerical methods such as Gauss-Hermite integration become important to evaluate those quantities. Nevertheless, according to Nelder *et al.* (2006) the use of such approximation tools is computationally intensive and heavier as the dimension of  $b_i$  increases. For instance, Tapia *et al.* (2019b) and Tapia *et al.* (2019a) used those techniques in a sensitivity analysis via a local influence approach for longitudinal count responses and for mixed effect logistic regression models, respectively. Although such proposals are very useful, they do not allow an accurate generalization for each GLMM.

Instead, we will evaluate the conditional expectations involved in (3.3) and (3.4) by means of Laplace approximations studied in section 2.4.2. Note that any of such expectations evaluated at  $\widehat{\psi}$  may be expressed in the form of the Theorem 3 (see section 2.4.2) where h ( $b_i$ ) = L ( $\widehat{\psi}$ ;  $y_i$ ,  $b_i$ ) and  $v_{ij}$  varies according to the corresponding expected value. In what follows, our focus is on approximate computation for the moments of the random variable  $b_i | y_i$ , for i = 1, 2, ..., N.

Remember that the use of the results in section 2.4.2 entails the existence and search of a maximum of the function h and in our case, a maximum in  $b_i$  of L ( $\hat{\psi}$ ;  $y_i$ ,  $b_i$ ), for i = 1, ..., N, given in (2.6). In this case, we assume that the vector of estimated parameters  $\hat{\psi}$  was obtained by means of some known procedure for this purpose as penalized maximum likelihood, maximum likelihood, restricted maximum likelihood or Bayesian tools such as maximum a posteriori (MAP) estimation (see, for instance, Demidenko, 2013; McCulloch and Searle, 2000). Thus, we want to maximize (in  $b_i$ ) the function  $h(b_i) = L(\widehat{\psi}; y_i, b_i)$  to achieve  $\widetilde{b}_i$  and this may be implemented via an iterative process or in practice, for example, via the gamlss package (Stasinopoulos *et al.*, 2017) in the software R (R Core Team, 2018). The details on the computation of  $\widetilde{b}_i$  will be treated in the next sections. Once  $\widetilde{b}_i$  is obtained, the Theorem 3 in section 2.4.2 provides an analytical way for approximating moments of the random variable  $b_i | y_i$ . Thus, for i = 1, ..., N, we have  $h(b_i) = L(\widehat{\psi}; y_i, b_i)$  and

$$-\mathbf{h}_{\widetilde{b}_{i}\widetilde{b}_{i}} = -\mathbf{L}_{\widetilde{b}_{i}\widetilde{b}_{i}}\left(\widehat{\psi}; y_{i}, b_{i}\right) = \left(\widehat{\phi} \mathbf{Z}_{i}^{\top}\widehat{\mathbf{W}}_{i}\mathbf{Z}_{i} + \widehat{\Xi}^{-1}\right)\Big|_{b_{i}=\widetilde{b}_{i}} = \widehat{\phi} \mathbf{Z}_{i}^{\top}\widetilde{\mathbf{W}}_{i}\mathbf{Z}_{i} + \widehat{\Xi}^{-1},$$

which is positive definite (see, appendix B.2) and hence  $h(\boldsymbol{b}_i) = L(\widehat{\boldsymbol{\psi}}, \boldsymbol{y}_i, \boldsymbol{b}_i)$  is unimodal (Demidenko, 2013).

Although the unimodality is not guaranteed in the non-canonical cases, we will use the outcomes given in section 2.4.2, replacing  $-\mathbf{h}_{\tilde{b}_i\tilde{b}_i} = -\mathbf{L}_{\tilde{b}_i\tilde{b}_i}(\widehat{\psi}; y_i, b_i)$  by  $\mathbf{E}_{y_i|b_i} \{-\mathbf{L}_{b_ib_i}(\widehat{\psi}; y_i, b_i)\}|_{b_i=\widetilde{b}_i}$ . This fact is commonly used when the Laplace approximation is required as in the approximate inference proposed for GLMMs by Breslow and Clayton (1993). In that sense, we can deduce (see appendix B.2) that

$$\mathbf{E}_{\boldsymbol{y}_{i}|\boldsymbol{b}_{i}}\left(-\mathbf{h}_{\boldsymbol{b}_{i}\boldsymbol{b}_{i}}\right)\big|_{\boldsymbol{b}_{i}=\widetilde{\boldsymbol{b}}_{i}}=\mathbf{E}_{\boldsymbol{y}_{i}|\boldsymbol{b}_{i}}\left\{-\mathbf{L}_{\boldsymbol{b}_{i}\boldsymbol{b}_{i}}\left(\widehat{\boldsymbol{\psi}};\,\boldsymbol{y}_{i},\,\boldsymbol{b}_{i}\right)\right\}\big|_{\boldsymbol{b}_{i}=\widetilde{\boldsymbol{b}}_{i}}=\left.\left(\widehat{\boldsymbol{\phi}}\,\mathbf{Z}_{i}^{\top}\,\widehat{\mathbf{W}}_{i}\,\mathbf{Z}_{i}+\widehat{\boldsymbol{\Xi}}^{-1}\right)\big|_{\boldsymbol{b}_{i}=\widetilde{\boldsymbol{b}}_{i}}=\widehat{\boldsymbol{\phi}}\,\mathbf{Z}_{i}^{\top}\,\widetilde{\mathbf{W}}_{i}\,\mathbf{Z}_{i}+\widehat{\boldsymbol{\Xi}}^{-1},$$

where  $\widehat{\mathbf{W}}_i \Big|_{b_i = \widetilde{b}_i} = \widetilde{\mathbf{W}}_i$ .

Anyway, if  $\mathbf{C} = \mathbf{C}(\mathbf{b}_i)$  is an  $r \times s$  matrix, we can rewrite (2.18) as

$$\mathbf{E}_{\boldsymbol{b}_{i}|\boldsymbol{y}_{i}}\left(\mathbf{C}|\boldsymbol{y}_{i}\right)\approx\widetilde{\mathbf{C}}+\left\{\mathbf{I}_{r}\otimes\operatorname{vec}\left(\mathbf{I}_{q}\right)^{\mathsf{T}}\right\}\left(\mathbf{C}_{\widetilde{\boldsymbol{b}}_{i}\widetilde{\boldsymbol{b}}_{i}}^{*}\otimes\mathbf{I}_{q}\right)\left[\mathbf{I}_{s}\otimes\operatorname{vec}\left\{\left(\widehat{\boldsymbol{\phi}}\;\mathbf{Z}_{i}^{\mathsf{T}}\;\widetilde{\mathbf{W}}_{i}\;\mathbf{Z}_{i}+\widehat{\boldsymbol{\Xi}}^{-1}\right)^{-1}\right\}\right],\quad(3.5)$$

where  $\tilde{b}_i$  maximizes  $L(\hat{\psi}; y_i, b_i)$ . This expression is the key tool for approximating equations (3.3) and (3.4) and therefore the conformal normal curvature (2.12). Thus, it will be used extensively throughout.

Note that (3.5) may be reduced if C = c is a vector (s = 1). So, we have

$$\mathbf{E}_{\boldsymbol{b}_{i}|\boldsymbol{y}_{i}}\left(\mathbf{c}|\boldsymbol{y}_{i}\right)\approx\widetilde{\mathbf{c}}+\left\{\mathbf{I}_{r}\otimes\operatorname{vec}\left(\mathbf{I}_{q}\right)^{\mathsf{T}}\right\}\left(\mathbf{c}_{\widetilde{\boldsymbol{b}}_{i}\widetilde{\boldsymbol{b}}_{i}}\otimes\mathbf{I}_{q}\right)\operatorname{vec}\left\{\left(\widehat{\phi}\,\mathbf{Z}_{i}^{\mathsf{T}}\,\widetilde{\mathbf{W}}_{i}\,\mathbf{Z}_{i}+\widehat{\boldsymbol{\Xi}}^{-1}\right)^{-1}\right\},\qquad(3.6)$$

because from the results shown in section 2.4.1.1, we have  $\mathbf{c}_{b_i b_i} = \mathbf{c}^*_{b_i b_i}$ .

In addition, when q = 1 and  $\mathbf{Z}_i = \mathbf{1}_{n_i}$  (random intercept GLMMs) the equation (3.5) may be expressed as

$$\mathbf{E}_{b_i|\boldsymbol{y}_i}\left(\mathbf{C}\,|\,\boldsymbol{y}_i\right) \approx \widetilde{\mathbf{C}} + \mathbf{C}^*_{\widetilde{b}_i\widetilde{b}_i} \times \left\{\widehat{\phi}\operatorname{tr}\left(\widetilde{\mathbf{W}}_i\right) + \widehat{\xi}^{-1}\right\}^{-1}.$$
(3.7)

Motivated by the usefulness of equation (3.5) in later results and for an accurate presentation of these, we propose the following definitions.

**Definition 4.** Let  $\mathbf{C} : \mathbf{S} \subseteq \mathbb{R}^q \longrightarrow \mathbb{R}^{r \times s}$  be a matrix function and  $\mathbf{b}_i$  and  $\mathbf{y}_i$  as defined previously for i = 1, ..., N. Then, we define the matrix function  $\mathbf{R}_1$  as

$$\mathbf{R}_{1}\left(\mathbf{C};\widetilde{\boldsymbol{b}}_{i}\right) = \mathbf{C} + \left\{\mathbf{I}_{r}\otimes\operatorname{vec}\left(\mathbf{I}_{q}\right)^{\mathsf{T}}\right\}\left(\mathbf{C}_{\widetilde{\boldsymbol{b}}_{i}}^{*}\widetilde{\boldsymbol{b}}_{i}\otimes\mathbf{I}_{q}\right)\left[\mathbf{I}_{s}\otimes\operatorname{vec}\left\{\left(\widehat{\boldsymbol{\phi}}\,\mathbf{Z}_{i}^{\mathsf{T}}\,\widetilde{\mathbf{W}}_{i}\,\mathbf{Z}_{i}+\widehat{\boldsymbol{\Xi}}^{-1}\right)^{-1}\right\}\right].$$

**Definition 5.** Let **C** an  $r \times s$  matrix, **B** a  $t \times s$  matrix and  $b_i$ , for i = 1, 2, ..., N. We define the matrix valued function  $\mathbf{R}_2$  as

$$\mathbf{R}_{2}\left(\mathbf{C}, \mathbf{B}; \widetilde{\boldsymbol{b}}_{i}\right) = \mathbf{R}_{1}\left(\mathbf{C}\mathbf{B}^{\mathsf{T}}; \widetilde{\boldsymbol{b}}_{i}\right) - \mathbf{R}_{1}\left(\mathbf{C}; \widetilde{\boldsymbol{b}}_{i}\right) \mathbf{R}_{1}\left(\mathbf{B}; \widetilde{\boldsymbol{b}}_{i}\right)^{\mathsf{T}}.$$

Note that the application of the functions  $\mathbf{R}_1$  and  $\mathbf{R}_2$  requires the computation of  $(\mathbf{C})_{b_i b_i'}^* (\mathbf{C} \mathbf{B}^\top)_{b_i b_i}^*$  and  $(\mathbf{B})_{b_i b_i}^*$ . Here theory of matrix differentiation described in Magnus and Neudecker (1999) or Abadir and Magnus (2005) is important, because it allows the computation of  $(\mathbf{C})_{b_i b_i'} (\mathbf{C} \mathbf{B}^\top)_{b_i b_i}$  and  $(\mathbf{B})_{b_i b_i}$ . Once these Hessian matrices are computed we can reorder their elements, according the guidelines of section 2.4.1.1, to obtain  $(\mathbf{C})_{b_i b_{i'}}^* (\mathbf{C} \mathbf{B}^\top)_{b_i b_i}^*$  and  $(\mathbf{B})_{b_i b_i}^*$ . Following these definitions, equation (3.5) assumes the form

$$\mathrm{E}\left(\mathbf{C} \mid \boldsymbol{y}_{i}\right) \approx \mathbf{R}_{1}\left(\mathbf{C}; \, \widetilde{\boldsymbol{b}}_{i}\right), \quad \text{for} \quad i = 1, \dots, N.$$

Now, for the computation of  $\tilde{b}_i$ , we can use an optimization algorithm such as Newton-Raphson or Fisher scoring. Also, the gamlss package in the R software (R Core Team, 2018) could provide  $\tilde{b}_i$  directly. The details on the computation of  $\tilde{b}_i$ for i = 1, ..., N, are given in the next two sections.

## 3.5.2.1 Iterative process for computing $\tilde{b}_i$

The Newton-Raphson iteration for getting the maximum of L  $(\widehat{\psi}; y_i, b_i)$ , denoted by  $\widetilde{b}_i$ , for i = 1, ..., N, is

$$\boldsymbol{b}_{i}^{(s+1)} = \boldsymbol{b}_{i}^{(s)} + \left\{-\mathbf{L}_{\boldsymbol{b}_{i}\boldsymbol{b}_{i}}\left(\widehat{\boldsymbol{\psi}}; \boldsymbol{y}_{i}, \boldsymbol{b}_{i}^{(s)}\right)\right\}^{-1} \left\{\mathbf{L}_{\boldsymbol{b}_{i}}\left(\widehat{\boldsymbol{\psi}}; \boldsymbol{y}_{i}, \boldsymbol{b}_{i}^{(s)}\right)\right\}, \quad s = 0, 1, \dots, \quad (3.8)$$

where s denotes the iteration index. It is possible to deduce (see appendix B.2) that

$$\mathbf{L}_{\boldsymbol{b}_{i}\boldsymbol{b}_{i}}\left(\widehat{\boldsymbol{\psi}};\boldsymbol{y}_{i},\boldsymbol{b}_{i}\right)=\widehat{\boldsymbol{\phi}}\,\mathbf{Z}_{i}^{\top}\,\widehat{\mathbf{M}}_{i}\,\mathbf{Z}_{i}-\widehat{\boldsymbol{\Xi}}^{-1}\quad\text{and}\quad\mathbf{L}_{\boldsymbol{b}_{i}}\left(\widehat{\boldsymbol{\psi}};\boldsymbol{y}_{i},\boldsymbol{b}_{i}\right)=\widehat{\boldsymbol{\phi}}\,\mathbf{Z}_{i}^{\top}\,\widehat{\boldsymbol{\tau}}_{i}-\widehat{\boldsymbol{\Xi}}^{-1}\,\boldsymbol{b}_{i}.$$

Also, it can be shown that in the canonical exponential family ( $\theta_{ij} = \eta_{ij}$ ), the above expressions take the forms

$$\mathbf{L}_{\boldsymbol{b}_{i}\boldsymbol{b}_{i}}\left(\widehat{\boldsymbol{\psi}};\boldsymbol{y}_{i},\boldsymbol{b}_{i}\right)=-\widehat{\boldsymbol{\phi}}\,\mathbf{Z}_{i}^{\top}\widehat{\mathbf{W}}_{i}\mathbf{Z}_{i}-\widehat{\boldsymbol{\Xi}}^{-1}\quad\text{and}\quad\mathbf{L}_{\boldsymbol{b}_{i}}\left(\widehat{\boldsymbol{\psi}};\boldsymbol{y}_{i},\boldsymbol{b}_{i}\right)=\widehat{\boldsymbol{\phi}}\,\mathbf{Z}_{i}^{\top}\left(\boldsymbol{y}_{i}-\widehat{\boldsymbol{\mu}}_{i}\right)-\widehat{\boldsymbol{\Xi}}^{-1}\boldsymbol{b}_{i}.$$

In fact, this is very important because  $-\mathbf{L}_{b_i b_i} \left( \widehat{\boldsymbol{\psi}}; \boldsymbol{y}_i, \boldsymbol{b}_i \right) = \widehat{\boldsymbol{\phi}} \mathbf{Z}_i^{\mathsf{T}} \widehat{\mathbf{W}}_i \mathbf{Z}_i + \widehat{\boldsymbol{\Xi}}^{-1}$  is positive definite (see appendix B.2), and therefore the iterative process (3.8) converges in a finite number of steps.

In the non-canonical cases, it is not always possible to prove that  $-\mathbf{L}_{b_i b_i}(\widehat{\psi}; y_i, b_i)$  is positive definite and the iterative process (3.8) may not converge. Thus, we use the

Fisher scoring rather than Newton-Raphson iteration procedure, which is obtained from (3.8), replacing  $-\mathbf{L}_{b_i b_i}(\widehat{\boldsymbol{\psi}}; \boldsymbol{y}_i, \boldsymbol{b}_i)$  by  $\mathbf{E}_{\boldsymbol{y}_i | \boldsymbol{b}_i} \{-\mathbf{L}_{b_i b_i}(\widehat{\boldsymbol{\psi}}; \boldsymbol{y}_i, \boldsymbol{b}_i)\} = \widehat{\boldsymbol{\phi}} \mathbf{Z}_i^{\mathsf{T}} \widehat{\mathbf{W}}_i \mathbf{Z}_i + \widehat{\boldsymbol{\Xi}}^{-1}$ , that is positive definite (see appendix B.2).

Among the non-canonical cases we have a special situation, where (2.1) is the gamma probability density function and the link function g(.) is log(.). It is possible to prove, as in the canonical cases, that  $-\mathbf{L}_{b_i b_i}(\widehat{\psi}; y_i, b_i)$  is positive definite and therefore, under this model, we can use the results given in section 2.4.2 directly. The gamma normal model is very useful in practice and later we will employ it in one of our applications.

### 3.5.2.2 Computing $\tilde{b}_i$ through gamlss package

For fixed  $\boldsymbol{\psi} = (\boldsymbol{\beta}^{\mathsf{T}}, \boldsymbol{\xi}^{\mathsf{T}}, \boldsymbol{\phi})^{\mathsf{T}}$ , maximizing the joint log-likelihood function of  $\boldsymbol{y}_i$  and  $\boldsymbol{b}_i$ , given in (2.6) to obtain  $\tilde{\boldsymbol{b}}_i$  is equivalent to maximizing the penalized log-likelihood function (see, for instance Demidenko, 2013; McCulloch and Searle, 2000), namely

$$L_{P}\left(\boldsymbol{\beta},\boldsymbol{b}_{i}\right)=\sum_{j=1}^{n_{i}}\log\left\{f_{y_{ij}\mid\boldsymbol{b}_{i}}\left(y_{ij};\boldsymbol{\beta},\boldsymbol{\phi}\right)\right\}-\frac{1}{2}\boldsymbol{b}_{i}^{\top}\boldsymbol{\Xi}^{-1}\boldsymbol{b}_{i},\quad i=1,\ldots,N,$$

or in the more general form

$$L_{P}\left(\boldsymbol{\beta},\boldsymbol{b}\right) = \sum_{i=1}^{N} \sum_{j=1}^{n_{i}} \log\left\{f_{y_{ij}|\boldsymbol{b}_{i}}\left(y_{ij};\boldsymbol{\beta},\boldsymbol{\phi}\right)\right\} - \frac{1}{2} \sum_{i=1}^{N} \boldsymbol{b}_{i}^{\top} \boldsymbol{\Xi}^{-1} \boldsymbol{b}_{i}, \qquad (3.9)$$

where  $f_{y_{ij}|b_i}(y_{ij}; \psi)$  is defined in (2.1) and  $b = (b_1^{\mathsf{T}}, b_2^{\mathsf{T}}, \dots, b_N^{\mathsf{T}})^{\mathsf{T}}$ . It is precisely the above expression that motivates studying GLMMs as special cases of generalized additive models for location, scale and shape (GAMLSS) with random effects (Stasinopoulos *et al.*, 2017).

Random effect models in GAMLSS allow modeling the location, scale and shape parameters (see, for instance, Stasinopoulos *et al.*, 2017) and the estimation and prediction of the fixed and random effects, respectively, are performed through the maximization, on the random and fixed effects jointly, of an accurate penalized log-likelihood function (Rigby and Stasinopoulos, 2005). When the scale and shape parameters are assumed to be constant (as in the GLMMs), the log-likelihood penalized function associated to the GAMLSS model takes the simplified form given by (3.9). Thus, it makes sense to consider GLMMs as a special case of the random effect models in GAMLSS. In the GAMLSS framework, a GLMM can be specified from (2.1) jointly with (2.5) assuming that  $\boldsymbol{b} \sim N_{Nq}$  ( $\mathbf{0}; \bigoplus_{i=1}^{N} \Xi$ ).

By studying GLMMs as a special case of the random effects models in GAMLSS, we can use the tools offered by the gamlss package in the R software (R Core Team, 2018) for fitting GLMMs. According to Stasinopoulos *et al.* (2017) this package allows, among its multiple utilities, to fit GLMMs by means of the gamlss() function jointly with the random() and re() functions. The random() function is used exclusively for dealing with random intercept GLMMs and the re() function focuses on more general GLMMs. Another option in the gamlss package is the gamlssNP() function, which hast he fitting of random intercept GLMMs among its functionalities. For more details, the reader is referred to Stasinopoulos *et al.* (2017). In the random effects models in GAMLSS, the re() and random() functions allow the inclusions of random effects into the predictors associated with different parameters as scale or shape. However, this is not the aim of this work.

Actually, in the context of GLMMs, the two basic algorithms in the gamlss package called CG (Cole and Green) and RS (Rigby and Stasinopoulos), described in Rigby and Stasinopoulos (2005), allow the maximization of (3.9) with respect to *b* and  $\beta$  to obtain the maximum penalized likelihood estimates of the fixed and random effects. However, this is not the first time that (3.9) was used with that purpose. In fact, this had been considered by Breslow and Clayton (1993) with the aim of obtaining the previously mentioned estimates and its maximization is performed using a Fisher scoring algorithm. In that sense, according to Rigby and Stasinopoulos (2005) the re() and random() functions of the gamlss package have their theoretical justification in the work of Breslow and Clayton (1993). This is because although the gamlss package has the general algorithms CG and RS, these are focused on the maximization of a general version of (3.9). Nevertheless, in the spacial case of a GLMM, the estimates of the variance components  $\xi$  can not be obtained from (3.9) and therefore additional tools are required. Such tools are provided by the approximate inference proposed by Breslow and Clayton (1993) for GLMMs. Specifically, the estimates of  $\xi$  are obtained using restricted maximum likelihood (REML).

So far, we assumed that  $\psi$  was estimated and we focused on getting  $\tilde{b}_i$  which maximizes  $L(\hat{\psi}; y_i, b_i)$  or equivalently  $L_P(\beta, b)$  given in (3.9). That is, our procedures are valid once  $\hat{\psi}$  is obtained. In practice, maximum likelihood, MAP (maximum a posteriori) or maximum penalized likelihood estimates are slightly different. Thus, regardless of the estimation procedure used to fit a GLMM, our diagnostic proposal may be used.

Precisely, as mentioned above, in the gamlss package the penalized likelihood (3.9) is maximized in b and  $\beta$  to produce "maximum penalized likelihood estimates" for  $\beta$  and  $b_i$ . The latter is of special interest in our diagnostic proposal. Therefore, it is reasonable the use of the GAMLSS framework for getting  $\hat{\psi}$  and  $\tilde{b}$  directly, without programming the iterative procedure (3.8) explicitly. In fact, the maximization algorithm of the penalized log-likelihood associated to the random effects GAMLSS described in Rigby and Stasinopoulos (2005) is based on Newton-Raphson iterations.

On the other hand, our diagnostic tools are not restricted to the gamlss package, since given the iterative process (3.8), we can obtain  $\tilde{b}$  from other functions to fit GLMMs, as glmer() of the lme4 package and glmmPQL() of the MASS package (Venables and Ripley, 2002). In that sense, although all the available packages for fitting GLMMs provide predicted values for the random effects, these do not necessarily correspond to  $\tilde{b}_i$ , for i = 1, ..., N, that we require. Some packages produce best linear unbiased predicted (BLUP) (see for instance, McCulloch and Searle, 2000), which is an approximation for E ( $b_i | y_i$ ) that agrees with  $\tilde{b}_i$ , for example, in the normal linear mixed model, but not in general (Pawitan, 2001). However, the BLUP may be considered as starting value  $b_i^{(0)}$  in the iterative process (3.8) to obtain  $\tilde{b}_i$ . In our case, we prefer to use the GAMLSS framework.

# **3.6** Approximation for $L_{\widehat{\psi} \ \widehat{\psi}}$

Now, denoting  $L^i = L(\psi; y_i)$  and taking into account the definitions of  $\mathbf{R}_1$  and  $\mathbf{R}_2$  given in section 3.5.2, we can obtain the approximated structures for (3.4) given by

$$\mathbf{L}_{\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}}^{i} \approx \widehat{\phi}^{2} \mathbf{R}_{2} \left( \mathbf{X}_{i}^{\top} \widehat{\boldsymbol{\tau}}_{i}, \mathbf{X}_{i}^{\top} \widehat{\boldsymbol{\tau}}_{i}; \widetilde{\boldsymbol{b}}_{i} \right) + \widehat{\phi} \mathbf{R}_{1} \left( \mathbf{X}_{i}^{\top} \widehat{\mathbf{M}}_{i} \mathbf{X}_{i}; \widetilde{\boldsymbol{b}}_{i} \right), \\
\mathbf{L}_{\hat{\boldsymbol{\phi}}\hat{\boldsymbol{\phi}}}^{i} \approx \sum_{j=1}^{n_{i}} c_{\widehat{\boldsymbol{\phi}}} \left( y_{ij}; \phi \right) + \mathbf{R}_{2} \left( \widehat{\boldsymbol{\nu}}_{i}, \widehat{\boldsymbol{\nu}}_{i}; \widetilde{\boldsymbol{b}}_{i} \right), \\
\mathbf{L}_{\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\beta}}}^{i} \approx \frac{\widehat{\phi}}{2} \left\{ \mathbf{R}_{2} \left( \mathbf{X}_{i}^{\top} \widehat{\boldsymbol{\tau}}_{i}, \boldsymbol{\lambda}_{i}; \widetilde{\boldsymbol{b}}_{i} \right) \right\} \left( \widehat{\Xi} \otimes \widehat{\Xi} \right)^{-1} \mathbf{G}_{q}, \qquad (3.10) \\
\mathbf{L}_{\hat{\boldsymbol{\beta}}\hat{\boldsymbol{\phi}}}^{i} \approx \mathbf{R}_{1} \left( \mathbf{X}_{i}^{\top} \widehat{\boldsymbol{\tau}}_{i}; \widetilde{\boldsymbol{b}}_{i} \right) + \widehat{\phi} \mathbf{R}_{2} \left( \mathbf{X}_{i}^{\top} \widehat{\boldsymbol{\tau}}_{i}, \widehat{\boldsymbol{\nu}}_{i}; \widetilde{\boldsymbol{b}}_{i} \right), \\
\mathbf{L}_{\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\phi}}}^{i} \approx \frac{1}{2} \mathbf{G}_{q}^{\top} \left( \widehat{\Xi} \otimes \widehat{\Xi} \right)^{-1} \mathbf{R}_{2} \left( \boldsymbol{\lambda}_{i}, \widehat{\boldsymbol{\nu}}_{i}; \widetilde{\boldsymbol{b}}_{i} \right) \quad \text{and} \\
\mathbf{L}_{\hat{\boldsymbol{\xi}}\hat{\boldsymbol{\xi}}}^{i} \approx \frac{1}{2} \mathbf{G}_{q}^{\top} \left[ \mathbf{I}_{q^{2}} - \left\{ \mathbf{R}_{1} \left( \boldsymbol{\lambda}_{i}; \widetilde{\boldsymbol{b}}_{i} \right)^{\top} \otimes \mathbf{I}_{q^{2}} \right\} \widehat{\mathbf{P}} + \frac{1}{2} \left( \widehat{\Xi} \otimes \widehat{\Xi} \right)^{-1} \mathbf{R}_{2} \left( \boldsymbol{\lambda}_{i}, \boldsymbol{\lambda}_{i}; \widetilde{\boldsymbol{b}}_{i} \right) \right] \left( \Xi \otimes \Xi \right)^{-1} \mathbf{G}_{q}, \end{aligned}$$

where  $\widehat{\mathbf{P}} = \left\{ \mathbf{I}_q \otimes \mathbf{\Pi}_q \left( \widehat{\mathbf{\Xi}}^{-1} \right) \right\} + \left\{ \mathbf{\Lambda}_q \left( \widehat{\mathbf{\Xi}}^{-1} \right) \otimes \mathbf{I}_q \right\}$ . Finally, letting

• 
$$\dot{\boldsymbol{p}}_{i} = (\dot{\rho}_{i1}, \dot{\rho}_{i2}, \dots, \dot{\rho}_{in_{i}})^{\mathsf{T}}$$
, where  $\dot{\rho}_{ij} = \frac{d\rho_{ij}}{d\eta_{ij}}$ ,  
•  $\dot{\mathbf{M}}_{i} = \mathbf{D}_{y_{i}-\mu_{i}} \mathbf{D}_{\dot{\rho}_{i}} - \mathbf{W}_{i}^{1/2} \mathbf{V}_{i}^{-1/2} \dot{\mathbf{A}}_{i}$  with  $\dot{\mathbf{A}}_{i} = \bigoplus_{j=1}^{n_{i}} d^{2} \mu_{ij} / d\eta_{ij}^{2}$  and

• 
$$\ddot{\mathbf{M}}_i = \mathbf{D}_{y_i - \mu_i} \mathbf{D}_{\dot{\rho}_i} - \mathbf{W}_i^{1/2} \mathbf{V}_i^{-1/2} \ddot{\mathbf{A}}_i$$
 with  $\ddot{\mathbf{A}}_i = \bigoplus_{j=1}^{n_i} d^3 \mu_{ij} / d \eta_{ij}^3$ ,

and according to the Definitions 4 and 5 given in section 3.5.2, it is necessary taking into account the following expressions extensively developed in appendix B.5:

$$\left(\mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i}\right)_{\boldsymbol{b}_{i} \boldsymbol{b}_{i}} = \left(\mathbf{X}_{i}^{\top} \otimes \mathbf{Z}_{i}^{\top}\right) \, \boldsymbol{\Omega}_{i} \, \left(\dot{\mathbf{M}}_{i} - 2 \, \mathbf{D}_{\boldsymbol{\rho}_{i}} \, \mathbf{A}_{i}\right) \, \mathbf{Z}_{i} ,$$

$$\left( \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i} \boldsymbol{\tau}_{i}^{\mathsf{T}} \mathbf{X}_{i} \right)_{\boldsymbol{b}_{i} \boldsymbol{b}_{i}} = \left( \mathbf{N}_{p} \otimes Z_{i}^{\mathsf{T}} \right) \left\{ \left( \mathbf{X}_{i}^{\mathsf{T}} \otimes \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i} \otimes \mathbf{I}_{n_{i}} \right) \mathbf{\Omega}_{i} \left( \dot{\mathbf{M}}_{i} - 2\mathbf{D}_{\rho_{i}} \mathbf{A}_{i} \right) + \left( \mathbf{I}_{p^{2}} \otimes \mathbf{M}_{i} \mathbf{X}_{i} \right) \mathbf{\Lambda}_{p} \left( \mathbf{I}_{p} \right) \mathbf{X}_{i}^{\mathsf{T}} \mathbf{M}_{i} \right\} \mathbf{Z}_{i} ,$$

$$\left(\mathbf{X}_{i}^{\top} \mathbf{M}_{i} \mathbf{X}_{i}\right)_{b, b_{i}} = \left\{ \left(\mathbf{X}_{i}^{\top} \otimes \mathbf{X}_{i}^{\top}\right) \mathbf{\Omega}_{i} \otimes \mathbf{Z}_{i}^{\top} \right\} \mathbf{\Omega}_{i} \left\{ \ddot{\mathbf{M}}_{i} - 3 \left(\mathbf{D}_{\boldsymbol{\rho}_{i}} \dot{\mathbf{A}}_{i} + \mathbf{D}_{\dot{\boldsymbol{\rho}}_{i}} \mathbf{A}_{i} \right) \right\} \mathbf{Z}_{i},$$

$$(v_i^2)_{\boldsymbol{h},\boldsymbol{h}_i} = 2 \mathbf{Z}_i^{\top} (v_i \mathbf{M}_i + \tau_i \tau_i^{\top}) \mathbf{Z}_i \text{ and } (v_i)_{\boldsymbol{b},\boldsymbol{b}_i} = \mathbf{Z}_i^{\top} \mathbf{M}_i \mathbf{Z}_i,$$

$$(\lambda_i)_{b_i b_i} = (\mathbf{N}_q \otimes \mathbf{I}_q) \ \Lambda_q \ (\mathbf{I}_q) \ , \tag{3.11}$$

$$\begin{aligned} \left( \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i} \boldsymbol{\lambda}_{i}^{\mathsf{T}} \right)_{\boldsymbol{b}_{i} \boldsymbol{b}_{i}} &= \left( \mathbf{I}_{q^{2}} \otimes \mathbf{X}_{i}^{\mathsf{T}} \otimes \mathbf{Z}_{i}^{\mathsf{T}} \right) \left[ \left( \boldsymbol{\lambda}_{i} \otimes \boldsymbol{\Omega}_{i} \right) \left( \dot{\mathbf{M}}_{i} - 2 \mathbf{D}_{\rho_{i}} \mathbf{A}_{i} \right) \mathbf{Z}_{i} + \left\{ \mathbf{N}_{q} \otimes \operatorname{vec} \left( \mathbf{M}_{i} \right) \right\} \left( \mathbf{I}_{q} \otimes \boldsymbol{b}_{i} \right) \right] \\ &+ \left\{ \mathbf{N}_{q} \otimes \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i} \otimes \mathbf{I}_{q} \right\} \mathbf{\Lambda}_{q} \left( \mathbf{I}_{q} \right) + \mathbf{\Lambda}_{p} \left\{ \left( \mathbf{I}_{q} \otimes \boldsymbol{b}_{i}^{\mathsf{T}} \right) \mathbf{N}_{q} \right\} \mathbf{X}_{i}^{\mathsf{T}} \mathbf{M}_{i} \mathbf{Z}_{i} , \end{aligned}$$

$$\left(\mathbf{X}_{i}^{\mathsf{T}}\boldsymbol{\tau}_{i}\boldsymbol{\nu}_{i}\right)_{\boldsymbol{b}_{i}\boldsymbol{b}_{i}}=\left(\mathbf{X}_{i}\otimes\mathbf{Z}_{i}\right)^{\mathsf{T}}\left\{\boldsymbol{\nu}_{i}\boldsymbol{\Omega}_{i}\left(\dot{\mathbf{M}}_{i}-2\,\mathbf{D}_{\boldsymbol{\rho}_{i}}\,\mathbf{A}_{i}\right)+\operatorname{vec}\left(\mathbf{M}_{i}\right)\,\boldsymbol{\tau}_{i}^{\mathsf{T}}+\mathbf{N}_{n_{i}}\left(\mathbf{M}_{i}\otimes\boldsymbol{\tau}_{i}\right)\right\}\,\mathbf{Z}_{i}\,,$$

$$\left(\boldsymbol{\lambda}_{i}\,\boldsymbol{\nu}_{i}\right)_{\boldsymbol{b}_{i}\,\boldsymbol{b}_{i}} = \left(\mathbf{N}_{q}\otimes\mathbf{I}_{q}\right)\,\boldsymbol{\Lambda}_{q}\,\left(\mathbf{I}_{q}\right)\,\left(\mathbf{I}_{q}\,\boldsymbol{\nu}_{i}+\boldsymbol{b}_{i}\,\boldsymbol{\tau}_{i}^{\top}\,\mathbf{Z}_{i}\right) + \left(\mathbf{I}_{q^{2}}\otimes\mathbf{Z}_{i}^{\top}\right)\left[\left(\boldsymbol{\lambda}_{i}\otimes\mathbf{M}_{i}\,\mathbf{Z}_{i}\right) + \left\{\mathbf{N}_{q}\,\left(\mathbf{I}_{q}\otimes\boldsymbol{b}_{i}\right)\otimes\boldsymbol{\tau}_{i}\right\}\right] \quad \text{and}$$

$$\left(\boldsymbol{\lambda}_{i} \, \boldsymbol{\lambda}_{i}^{\mathsf{T}}\right)_{\boldsymbol{b}_{i} \, \boldsymbol{b}_{i}} = \left(\mathbf{N}_{q^{2}} \otimes \mathbf{I}_{q}\right) \left( \left[\mathbf{N}_{q} \left(\mathbf{I}_{q} \otimes \boldsymbol{b}_{i}\right) \otimes \operatorname{vec}\left\{\left(\mathbf{I}_{q} \otimes \boldsymbol{b}_{i}\right)^{\mathsf{T}} \, \mathbf{N}_{q}\right\}\right] + \left\{\mathbf{\Lambda}_{q^{2}} \left(\boldsymbol{\lambda}_{i}^{\mathsf{T}}\right) \, \mathbf{N}_{q} \otimes \mathbf{I}_{q}\right\} \, \mathbf{\Lambda}_{q} \left(\mathbf{I}_{q}\right) \right) \, \mathbf{A}_{q} \left(\mathbf{I}_{q}\right) \, \mathbf$$

In fact, the quantities  $\tau_i$ ,  $\mathbf{D}_{\rho_i}$ ,  $\mathbf{A}_i$ ,  $\mathbf{\dot{A}}_i$ ,  $\mathbf{M}_i$ ,  $\mathbf{\dot{M}}_i$ ,  $\lambda_i$  and  $\nu_i$  in the right side of the above expressions should have the symbol " $\sim$ ", but we omit it by the simplicity. Also,  $\phi$  should be replaced by  $\hat{\phi}$ .

In addition, from section 2.4.1.1 we have that  $\mathbf{F}_{b_i b_i} = \mathbf{F}_{b_i b_i}^*$  when **F** is a vector function, but not in general. However, this is not a problem, because it is enough to reorder the elements of  $\mathbf{F}_{b_i b_i}$  to obtain  $\mathbf{F}_{b_i b_i}^*$ , according to the results of the same section. This procedure may be computationally implemented without major difficulties.

Hence, replacing suitably expressions (3.11) in (3.10), we obtain an approximate analytical expression for  $\mathbf{L}_{\psi\psi}^{i}$  given in (3.2), and consequently for the Hessian matrix  $\mathbf{L}_{\psi\psi} = \sum_{i=1}^{N} \mathbf{L}_{\psi\psi}^{i}$ . For details of these computations, the reader may be refer to appendix B.5. Furthermore, it is very important to mention that these procedures are valid for any GLMM since in general depend of the weight matrix  $\mathbf{W}_{i}$ , the variance function matrix  $\mathbf{V}_{i}$  and the design matrices  $\mathbf{X}_{i}$  and  $\mathbf{Z}_{i}$ , among others. The commutation and duplication matrices  $\mathbf{K}_{mn}$  and  $\mathbf{G}_{q}$ , respectively, are implemented in the matrixcalc package of the

R software (R Core Team, 2018).

Although these expressions seem complex, great simplifications are possible when the GLMM of interest belongs to the canonical exponential family and/or correspond to random intercept GLMMs. Some of these reductions are presented below.

### Canonical exponential family

In the canonical cases ( $\theta_{ij} = \eta_{ij}$ ), we have the following simplifications:

$$\dot{\boldsymbol{\rho}}_{i} = \boldsymbol{\rho}_{i} = \boldsymbol{0}_{n_{i} \times 1}, \qquad \boldsymbol{\tau}_{i} = \boldsymbol{y}_{i} - \boldsymbol{\mu}_{i}, \qquad -\mathbf{M}_{i} = \mathbf{W}_{i} = \mathbf{V}_{i} = \mathbf{A}_{i}$$

$$\mathbf{D}_{\dot{\boldsymbol{\rho}}_{i}} = \mathbf{D}_{\boldsymbol{\rho}_{i}} = \boldsymbol{0}_{n_{i} \times n_{i}}, \qquad \dot{\mathbf{M}}_{i} = -\dot{\mathbf{A}}_{i} \qquad \text{and} \qquad \ddot{\mathbf{M}}_{i} = -\ddot{\mathbf{A}}_{i}.$$
(3.12)

This leads to notorious reductions in expressions given in (3.11) and consequently in the approximated structure for (3.2) from the structures (3.10). To illustrate, we show some of the equations in (3.11) in the canonical form. Thus, we have

$$\left( \mathbf{X}_{i}^{\mathsf{T}} \, \boldsymbol{\tau}_{i} \right)_{\boldsymbol{b}_{i} \boldsymbol{b}_{i}} = \left\{ \mathbf{X}_{i}^{\mathsf{T}} \, \left( \boldsymbol{y}_{i} - \boldsymbol{\mu}_{i} \right) \right\}_{\boldsymbol{b}_{i} \boldsymbol{b}_{i}} = - \left( \mathbf{X}_{i}^{\mathsf{T}} \otimes \mathbf{Z}_{i}^{\mathsf{T}} \right) \, \boldsymbol{\Omega}_{i} \, \dot{\mathbf{A}}_{i} \, \mathbf{Z}_{i} \quad \text{and}$$
$$\left( \mathbf{X}_{i}^{\mathsf{T}} \, \mathbf{M}_{i} \, \mathbf{X}_{i} \right)_{\boldsymbol{b}_{i} \boldsymbol{b}_{i}} = - \left\{ \left( \mathbf{X}_{i}^{\mathsf{T}} \otimes \mathbf{X}_{i}^{\mathsf{T}} \right) \, \boldsymbol{\Omega}_{i} \otimes \mathbf{Z}_{i}^{\mathsf{T}} \right\} \, \boldsymbol{\Omega}_{i} \, \ddot{\mathbf{A}}_{i} \, \mathbf{Z}_{i}.$$

This procedure may be performed for each of the formulas in (3.11).

#### **Random intercept GLMMs**

On the other hand, for random intercept GLMMs, that is, when q = 1 and  $\mathbf{Z}_i = \mathbf{1}_{n_i}$ , we obtain

$$I_q = I_{q^2} = 1$$
,  $K_q = 1$ ,  $G_q = 1$  and  $N_q = N_{q^2} = 2$ . (3.13)

In addition, for two matrices **A** and **B**, if q = 1 we obtain from Definition 2 that  $\Lambda_1(\mathbf{A}) = \text{vec}(\mathbf{A})$  and  $\Pi_1(\mathbf{B}) = \text{vec}(\mathbf{B})$ . To exemplify, some of the expressions in (3.11)

assume the following structures:

$$(v_i)_{b_i b_i} = \operatorname{tr}(\mathbf{M}_i) \text{ and } (v_i^2)_{b_i b_i} = 2v_i \operatorname{tr}(\mathbf{M}_i) + 2 \operatorname{tr}(\mathbf{T}_i)^2,$$

where  $\mathbf{T}_i = \mathbf{W}_i^{1/2} \mathbf{V}_i^{-1/2} \mathbf{D}_{y_i - \mu_i}$ . Analogously, it is possible to get simplified forms for each of the expressions given in (3.11).

Furthermore, Definitions 4 and 5 in section 3.5.2, which allow approximated computations for the moments of  $b_i | y_i$ , assume simple structures that are consequences of (3.7). It should be clear for the reader that for random intercept GLMMs in the canonical exponential family the simplifications are even more drastic. The random intercept logistic model and the random intercept Poisson model with logarithmic link are some examples with these features. In that sense, the logarithmic link is special, because the matrices  $\mathbf{A}_i$ ,  $\dot{\mathbf{A}}_i$  and  $\ddot{\mathbf{A}}_i$  (which are related with the link function g) are the same when this link function is considered. In this sense, we refer in detail to the random intercept GLMMs in section 3.8.

In the next section, equations in (3.11) will be used to approximate the matrix  $\Delta$ , defined in (2.10), for different perturbation schemes.

# 3.7 Conformal normal curvature under different perturbation schemes: approximation for $\Delta$

According to the mentioned criteria in section 3.3, we intend to discriminate possibly influential subjects and observations. Thus, we focus on two perturbation schemes, namely: case weights - subjects and case weights - observations.

The perturbation scheme affects the normal conformal curvature (2.12) or (2.13) through  $\Delta = \partial L(\psi; y | \omega) / \partial \psi \partial \omega^{\top}|_{\omega_0, \hat{\psi}}$ . Consequently, it is necessary to derive  $\Delta$  for each perturbation scheme. These computations may be reviewed in appendices B.6 and B.7. The outcomes are displayed below.

#### 3.7.1 Case weights - subjects

Lesaffre and Verbeke (1998) considered this approach in linear mixed models for identifying influential subjects and we apply it in GLMMs. Hence, we use the perturbation vector  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_N)^{\mathsf{T}}$  and we assume that  $L(\boldsymbol{\psi}; \boldsymbol{y} | \boldsymbol{\omega}) = \sum_{i=1}^N \omega_i L(\boldsymbol{\psi}; \boldsymbol{y}_i)$ . Consequently, the no perturbation vector is  $\boldsymbol{\omega}_0 = \mathbf{1}_N$ .

Under this perturbation scheme it is possible to prove (see appendix B.6) that  $\Delta = (\delta_1, \delta_2, ..., \delta_N)$ , where  $\delta_i = \mathbf{L}_{\psi}^i = \mathbf{L}_{\psi}(\psi; y_i)$  evaluated at  $\psi = \widehat{\psi}$  and  $\mathbf{L}_{\psi}^i$  displayed in equation (3.3). However, as it was mentioned previously, in general, it is not viable to compute exactly this expression. Therefore, from the results of section 3.5.2, we propose the following approximate structure for  $\delta_i$ :

$$\boldsymbol{\delta}_{i} = \begin{bmatrix} \mathbf{L}_{\widehat{\boldsymbol{\beta}}}^{i} \\ \mathbf{L}_{\widehat{\boldsymbol{\xi}}}^{i} \\ \mathbf{L}_{\widehat{\boldsymbol{\phi}}}^{i} \end{bmatrix} \approx \begin{bmatrix} \widehat{\boldsymbol{\phi}} \, \mathbf{R}_{1} \left( \mathbf{X}_{i}^{\top} \, \widehat{\boldsymbol{\tau}}_{i} ; \, \widetilde{\boldsymbol{b}}_{i} \right) \\ -\frac{1}{2} \, \mathbf{G}_{q}^{\top} \left\{ \operatorname{vec} \left( \widehat{\boldsymbol{\Xi}}^{-1} \right) - \left( \widehat{\boldsymbol{\Xi}} \otimes \widehat{\boldsymbol{\Xi}} \right)^{-1} \mathbf{R}_{1} \left( \boldsymbol{\lambda}_{i} ; \, \widetilde{\boldsymbol{b}}_{i} \right) \right\} \\ \mathbf{R}_{1} \left( \widehat{\boldsymbol{\nu}}_{i} ; \, \widetilde{\boldsymbol{b}}_{i} \right) \end{bmatrix}, \quad \text{for} \quad i = 1, \dots, N.$$

$$(3.14)$$

Thus, according to Definition 4 in section 3.5.2, we must replace accurately the expressions  $(\mathbf{X}_i^{\mathsf{T}} \boldsymbol{\tau}_i)_{b_i b_i}$ ,  $(\lambda_i)_{b_i b_i}$  and  $(\nu_i)_{b_i b_i}$  given in (3.11), in expression (3.14) to achieve an analytical structure for  $\boldsymbol{\delta}_i$  and hence for  $\boldsymbol{\Delta}$ . Some special cases are discussed below.

#### Canonical exponential family

For the canonical exponential family, using (3.12), we have the following abbreviated formulas:

$$\mathbf{R}_{1}\left(\mathbf{X}_{i}^{\top}\,\widehat{\boldsymbol{\tau}}_{i}\,;\,\widetilde{\boldsymbol{b}}_{i}\right) = \mathbf{X}_{i}^{\top}\left(\boldsymbol{y}_{i}-\widetilde{\boldsymbol{\mu}}_{i}\right) - \left\{\mathbf{I}_{p}\otimes\operatorname{vec}\left(\mathbf{I}_{q}\right)\right\}^{\top}\left[\left\{\left(\mathbf{X}_{i}^{\top}\otimes\mathbf{Z}_{i}^{\top}\right)\,\mathbf{\Omega}_{i}\,\widetilde{\mathbf{V}}_{i}\mathbf{Z}_{i}\right\}\otimes\mathbf{I}_{q}\right]\operatorname{vec}\left\{\left(\widehat{\boldsymbol{\phi}}\,\mathbf{Z}_{i}^{\top}\,\widetilde{\mathbf{V}}_{i}\,\mathbf{Z}_{i}+\widehat{\boldsymbol{\Xi}}^{-1}\right)^{-1}\right\},\\ \mathbf{R}_{1}\left(\boldsymbol{\lambda}_{i}\,;\,\widetilde{\boldsymbol{b}}_{i}\right) = \widetilde{\boldsymbol{\lambda}}_{i} + \left\{\mathbf{I}_{q^{2}}\otimes\operatorname{vec}\left(\mathbf{I}_{q}\right)\right\}^{\top}\left[\left\{\left(\mathbf{N}_{q}\otimes\mathbf{I}_{q}\right)\,\mathbf{\Lambda}_{q}\left(\mathbf{I}_{q}\right)\right\}\otimes\mathbf{I}_{q}\right]\operatorname{vec}\left\{\left(\widehat{\boldsymbol{\phi}}\,\mathbf{Z}_{i}^{\top}\,\widetilde{\mathbf{V}}_{i}\,\mathbf{Z}_{i}+\widehat{\boldsymbol{\Xi}}^{-1}\right)^{-1}\right\}\quad\text{and}\\ \mathbf{R}_{1}\left(\widehat{\boldsymbol{\nu}}_{i}\,;\,\widetilde{\boldsymbol{b}}_{i}\right) = \widetilde{\boldsymbol{\nu}}_{i} - \operatorname{vec}\left(\mathbf{I}_{q}\right)^{\top}\left\{\left(\mathbf{Z}_{i}^{\top}\widetilde{\mathbf{V}}_{i}\mathbf{Z}_{i}\right)\otimes\mathbf{I}_{q}\right\}\operatorname{vec}\left\{\left(\widehat{\boldsymbol{\phi}}\,\mathbf{Z}_{i}^{\top}\,\widetilde{\mathbf{V}}_{i}\,\mathbf{Z}_{i}+\widehat{\boldsymbol{\Xi}}^{-1}\right)^{-1}\right\}.$$

Replacing the terms above in (3.14), we obtain the explicit analytical expression for  $\delta_i$ .

#### **Random intercept GLMM**

By way of example, in a random intercept GLMM of the canonical exponential family, we obtain

$$\boldsymbol{\delta}_{i} = \begin{bmatrix} \mathbf{L}_{\widehat{\boldsymbol{\beta}}}^{i} \\ \mathbf{L}_{\widehat{\boldsymbol{\xi}}}^{i} \\ \mathbf{L}_{\widehat{\boldsymbol{\phi}}}^{i} \end{bmatrix} \approx \begin{bmatrix} \widehat{\boldsymbol{\phi}} \, \mathbf{X}_{i}^{\mathsf{T}} \left( \boldsymbol{y}_{i} - \widetilde{\boldsymbol{\mu}}_{i} \right) - \widehat{\boldsymbol{\phi}} \, \mathbf{X}_{i}^{\mathsf{T}} \, \widetilde{\mathbf{V}}_{i} \, \mathbf{1}_{i} \left\{ \widehat{\boldsymbol{\phi}} \operatorname{tr} \left( \widetilde{\mathbf{V}}_{i} \right) + \widehat{\boldsymbol{\xi}}^{-1} \right\}^{-1} \\ -\frac{1}{2 \, \widehat{\boldsymbol{\xi}}} + \frac{1}{2 \, \widehat{\boldsymbol{\xi}}^{2}} \left[ \widetilde{\boldsymbol{b}}_{i}^{2} + 2 \left\{ \widehat{\boldsymbol{\phi}} \operatorname{tr} \left( \widetilde{\mathbf{V}}_{i} \right) + \widehat{\boldsymbol{\xi}}^{-1} \right\}^{-1} \right] \\ \widetilde{\boldsymbol{\nu}}_{i} - \operatorname{tr} \left( \widetilde{\mathbf{V}}_{i} \right) \left\{ \widehat{\boldsymbol{\phi}} \operatorname{tr} \left( \widetilde{\mathbf{V}}_{i} \right) + \widehat{\boldsymbol{\xi}}^{-1} \right\}^{-1} \end{bmatrix}, \quad \text{for} \quad i = 1, \dots, N.$$

#### 3.7.1.1 Conformal normal curvature - subjects

The perturbation scheme developed in section 3.7.1 allows to discriminate potentially influential subjects. As it was mentioned in section 3.3, for obtaining a curvature measure associated to each subject, we consider the direction vector  $\mathbf{d} = \mathbf{u}_i \in \mathbb{R}^N$  with i = 1, ..., N. The vector  $\mathbf{u}_i$  represents an  $N \times 1$  vector that assumes zero everywhere, except in the *ith* position, which equals 1. Hence, from (2.12), for i = 1, 2, ..., N we have that the conformal normal curvature associated to the *ith* subject is given by

$$B_{i}(\boldsymbol{\psi}) = B_{\mathbf{u}_{i}}(\boldsymbol{\psi}) \approx \frac{\mathbf{u}_{i}^{\top} \boldsymbol{\Delta}^{\top} \left(-\mathbf{L}_{\widehat{\boldsymbol{\psi}} \widehat{\boldsymbol{\psi}}}\right)^{-1} \boldsymbol{\Delta} \mathbf{u}_{i}}{\sqrt{\mathrm{tr}\left[\left\{\boldsymbol{\Delta}^{\top} \left(\mathbf{L}_{\widehat{\boldsymbol{\psi}} \widehat{\boldsymbol{\psi}}}\right)^{-1} \boldsymbol{\Delta}\right\}^{2}\right]}} = \frac{\boldsymbol{\delta}_{i}^{\top} \left(-\mathbf{L}_{\widehat{\boldsymbol{\psi}} \widehat{\boldsymbol{\psi}}}\right)^{-1} \boldsymbol{\delta}_{i}}{\sqrt{\mathrm{tr}\left[\left\{\boldsymbol{\Delta}^{\top} \left(\mathbf{L}_{\widehat{\boldsymbol{\psi}} \widehat{\boldsymbol{\psi}}}\right)^{-1} \boldsymbol{\Delta}\right\}^{2}\right]}}, \quad (3.15)$$

where  $\Delta = (\delta_1, \delta_2, ..., \delta_N)$ ,  $\delta_i$  is given in (3.14) and  $L_{\widehat{\psi}\widehat{\psi}}$  is approximately obtained from (3.10). This curvature enables the assessment of potentially influential subjects in any GLMM and is one of the main results of this work. Special cases of this curvature are treated in section 3.8.

#### 3.7.2 Case weights - observations

In order to detect influential observations we aggregate weights to each observation via the  $N_T \times 1$  perturbation vector  $\boldsymbol{\omega} = (\boldsymbol{\omega}_1^{\mathsf{T}}, \boldsymbol{\omega}_2^{\mathsf{T}}, \dots, \boldsymbol{\omega}_N^{\mathsf{T}})^{\mathsf{T}}$ , where  $\boldsymbol{\omega}_i = (\omega_{i1}, \omega_{i2}, \dots, \omega_{in_i})^{\mathsf{T}}$  and  $N_T = \sum_{i=1}^N n_i$ . Under this perturbation scheme we consider

$$\mathbf{L}\left(\boldsymbol{\psi};\boldsymbol{y} \,|\, \boldsymbol{\omega}\right) = \sum_{i=1}^{N} \mathbf{L}\left(\boldsymbol{\psi};\boldsymbol{y}_{i} \,|\, \boldsymbol{\omega}_{i}\right) \,,$$

where

$$L(\boldsymbol{\psi};\boldsymbol{y}_{i}|\boldsymbol{\omega}_{i}) = \log \int \exp \left\{ L(\boldsymbol{\psi};\boldsymbol{y}_{i},\boldsymbol{b}_{i}|\boldsymbol{\omega}_{i}) \right\} d\boldsymbol{b}_{i}, \quad i = 1,\ldots,N, \quad (3.16)$$

and

$$L(\boldsymbol{\psi}; \boldsymbol{y}_i, \boldsymbol{b}_i | \boldsymbol{\omega}_i) = -\frac{q}{2} \log(2\pi) - \frac{1}{2} \log|\boldsymbol{\Xi}| + \boldsymbol{\omega}_i^{\top} \mathbf{f}_i - \frac{1}{2} \boldsymbol{b}_i^{\top} \boldsymbol{\Xi}^{-1} \boldsymbol{b}_i$$

with  $\mathbf{f}_{i} = (\mathbf{f}_{i1}, \mathbf{f}_{i2}, \dots, \mathbf{f}_{in_{i}})^{\mathsf{T}}$ ,  $\mathbf{f}_{ij} = \log \{f_{y_{ij}|b_{i}}(y_{ij}; \theta_{ij}, \phi)\}$  and the no perturbation vector given by  $\boldsymbol{\omega}_{0} = (\boldsymbol{\omega}_{0_{1}}^{\mathsf{T}}, \boldsymbol{\omega}_{0_{2}}^{\mathsf{T}}, \dots, \boldsymbol{\omega}_{0_{N}}^{\mathsf{T}})^{\mathsf{T}} = \mathbf{1}_{N_{T}}$ . It can be shown that (see appendix B.7)  $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_{1}, \boldsymbol{\Delta}_{2}, \dots, \boldsymbol{\Delta}_{N})$ , where  $\boldsymbol{\Delta}_{i} = \partial L(\boldsymbol{\psi}; \boldsymbol{y}_{i} | \boldsymbol{\omega}_{i}) / \partial \boldsymbol{\psi} \boldsymbol{\omega}_{i}^{\mathsf{T}} \big|_{\boldsymbol{\omega}_{0}, \hat{\boldsymbol{\psi}}}$  and letting  $\mathbf{T}_{i} = \mathbf{W}_{i}^{1/2} \mathbf{V}_{i}^{-1/2} \mathbf{D}_{\boldsymbol{y}_{i}-\boldsymbol{\mu}_{i}}$  we obtain

$$\Delta_{i} = \frac{\partial L\left(\psi; \boldsymbol{y}_{i} \mid \boldsymbol{\omega}_{i}\right)}{\partial \psi \partial \boldsymbol{\omega}_{i}^{\top}} \bigg|_{\boldsymbol{\omega}_{0}, \widehat{\psi}} = \begin{bmatrix} \frac{\partial L\left(\psi; \boldsymbol{y}_{i} \mid \boldsymbol{\omega}_{i}\right)}{\partial \beta \partial \boldsymbol{\omega}_{i}^{\top}} \\ \frac{\partial L\left(\psi; \boldsymbol{y}_{i} \mid \boldsymbol{\omega}_{i}\right)}{\partial \xi \partial \boldsymbol{\omega}_{i}^{\top}} \\ \frac{\partial L\left(\psi; \boldsymbol{y}_{i} \mid \boldsymbol{\omega}_{i}\right)}{\partial \phi \partial \boldsymbol{\omega}_{i}^{\top}} \end{bmatrix}_{\boldsymbol{\omega}_{0}, \widehat{\psi}} = \begin{bmatrix} \widehat{\phi} \operatorname{Cov}\left(\mathbf{X}_{i}^{\top} \widehat{\boldsymbol{\tau}}_{i}, \widehat{\mathbf{f}}_{i} \middle| \boldsymbol{y}_{i}\right) + \widehat{\phi} \operatorname{E}\left(\mathbf{X}_{i}^{\top} \widehat{\boldsymbol{T}}_{i} \middle| \boldsymbol{y}_{i}\right) \\ \frac{1}{2} \mathbf{G}_{q}^{\top} \left(\widehat{\Xi} \otimes \widehat{\Xi}\right)^{-1} \operatorname{Cov}\left(\boldsymbol{\lambda}_{i}, \widehat{\mathbf{f}}_{i} \middle| \boldsymbol{y}_{i}\right) \\ \operatorname{Cov}\left(\widehat{\nu}_{i}, \widehat{\mathbf{f}}_{i} \middle| \boldsymbol{y}_{i}\right) + \operatorname{E}\left(\widehat{\mathbf{f}}_{i} \middle| \boldsymbol{y}_{i}\right)^{\top} \end{bmatrix},$$
(3.17)

where  $\dot{\mathbf{f}}_i = \partial \mathbf{f}_i / \partial \phi$ . Again, computing exactly the structures in  $\Delta_i$  requires at least knowledge of the probability density function of  $b_i | y_i$ , and for reasons mentioned before, that is not our case. Then, taking  $\mathbf{C}_i = \mathbf{1}_{n_i} \mathbf{f}_i^{\mathsf{T}} + \mathbf{I}_{n_i}$  and supported by the results from section 2.4.2 we obtain, for i = 1, ..., N, the following approximate expressions:

$$\Delta_{i} = \begin{bmatrix} \frac{\partial L(\psi; \boldsymbol{y}_{i} | \boldsymbol{\omega}_{i})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}_{i}^{\mathsf{T}}} \\ \frac{\partial L(\psi; \boldsymbol{y}_{i} | \boldsymbol{\omega}_{i})}{\partial \boldsymbol{\xi} \partial \boldsymbol{\omega}_{i}^{\mathsf{T}}} \\ \frac{\partial L(\psi; \boldsymbol{y}_{i} | \boldsymbol{\omega}_{i})}{\partial \boldsymbol{\xi} \partial \boldsymbol{\omega}_{i}^{\mathsf{T}}} \end{bmatrix}_{\boldsymbol{\omega}_{0}, \hat{\boldsymbol{\psi}}} \approx \begin{bmatrix} \widehat{\boldsymbol{\phi}} \mathbf{R}_{1} \left( \mathbf{X}_{i}^{\mathsf{T}} \widehat{\mathbf{T}}_{i} \widehat{\mathbf{C}}_{i}; \widetilde{\boldsymbol{b}}_{i} \right) - \widehat{\boldsymbol{\phi}} \mathbf{R}_{1} \left( \mathbf{X}_{i}^{\mathsf{T}} \widehat{\boldsymbol{\tau}}_{i}, \widetilde{\boldsymbol{b}}_{i} \right) \mathbf{R}_{1} \left( \widehat{\mathbf{f}}_{i}, \widetilde{\boldsymbol{b}}_{i} \right)^{\mathsf{T}} \\ \frac{1}{2} \mathbf{G}_{q}^{\mathsf{T}} \left( \widehat{\mathbf{\Xi}} \otimes \widehat{\mathbf{\Xi}} \right)^{-1} \mathbf{R}_{2} \left( \lambda_{i}, \widehat{\mathbf{f}}_{i}; \widetilde{\boldsymbol{b}}_{i} \right) \\ \mathbf{R}_{1} \left( \widehat{\mathbf{f}}_{i}^{\mathsf{T}} \widehat{\mathbf{C}}_{i}; \widetilde{\boldsymbol{b}}_{i} \right) - \mathbf{R}_{1} \left( \widehat{\boldsymbol{\nu}}_{i}; \boldsymbol{b}_{i} \right) \mathbf{R}_{1} \left( \widehat{\mathbf{f}}_{i}; \widetilde{\boldsymbol{b}}_{i} \right)^{\mathsf{T}} \end{bmatrix} .$$

$$(3.18)$$

Again, for using the **R**<sub>1</sub> and **R**<sub>2</sub> functions defined in section 3.5.2 it is necessary to take into account  $(\mathbf{X}_i^{\mathsf{T}} \tau_i)_{b,b_i}$ ,  $(\lambda_i)_{b_i b_i}$  and  $(\nu_i)_{b_i b_i}$  presented in (3.11) as well as

$$\begin{aligned} \left( \mathbf{X}_{i}^{\mathsf{T}}\mathbf{T}_{i}\mathbf{C}_{i} \right)_{\boldsymbol{b}_{i}\boldsymbol{b}_{i}} &= \phi \left( \mathbf{I}_{n_{i}p} \otimes \mathbf{Z}_{i}^{\mathsf{T}} \right) \left[ \phi^{-1} \left\{ \left( \mathbf{C}_{i} \otimes \mathbf{X}_{i} \right)^{\mathsf{T}} \mathbf{\Omega}_{i} \otimes \mathbf{I}_{n_{i}} \right\} \mathbf{\Omega}_{i} \left( \dot{\mathbf{M}}_{i} - 2\mathbf{D}_{\rho_{i}}\mathbf{A}_{i} \right) + \left( \mathbf{I}_{n_{i}p} \otimes \mathbf{M}_{i}\mathbf{\Omega}_{i}^{\mathsf{T}} \right) \times \\ \left( \mathbf{T}_{i} \otimes \mathbf{\Pi}_{n_{i}} \left( \mathbf{X}_{i} \right) \mathbf{1}_{n_{i}} \right) + \left( \mathbf{I}_{n_{i}} \otimes \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i} \otimes \mathbf{I}_{n_{i}} \right) \mathbf{\Omega}_{i} \mathbf{M}_{i} \right] \mathbf{Z}_{i} + \phi \mathbf{\Lambda}_{p} \left( \mathbf{Z}_{i}^{\mathsf{T}} \mathbf{T}_{i} \right) \mathbf{X}_{i}^{\mathsf{T}} \mathbf{M}_{i} \mathbf{Z}_{i} , \end{aligned}$$

$$\begin{split} \left(\boldsymbol{\lambda}_{i}\,\mathbf{f}_{i}^{\mathsf{T}}\right)_{\boldsymbol{b}_{i}\,\boldsymbol{b}_{i}} &= \left\{ \left(\mathbf{f}_{i}\otimes\mathbf{I}_{q^{2}}\right)\,\mathbf{N}_{q}\otimes\mathbf{I}_{q} \right\}\,\boldsymbol{\Lambda}_{q}\left(\mathbf{I}_{q}\right) + \phi\,\left\{\mathbf{I}_{n_{i}\,q^{2}}\otimes\left(\mathbf{I}_{q}\otimes\boldsymbol{b}_{i}^{\mathsf{T}}\right)\,\mathbf{N}_{q} \right\}\,\left\{\mathbf{T}_{i}\,\mathbf{Z}_{i}\otimes\operatorname{vec}\left(\mathbf{I}_{q^{2}}\right)\right\} \\ &+ \phi\,\left(\mathbf{I}_{n_{i}}\otimes\boldsymbol{\lambda}_{i}\otimes\mathbf{Z}_{i}^{\mathsf{T}}\right)\,\boldsymbol{\Omega}_{i}\mathbf{M}_{i}\mathbf{Z}_{i} + \phi\boldsymbol{\Lambda}_{q^{2}}\left(\mathbf{Z}_{i}^{\mathsf{T}}\mathbf{T}_{i}\right)\,\mathbf{N}_{q}\left(\mathbf{I}_{q}\otimes\boldsymbol{b}_{i}\right)\,, \end{split}$$

$$\left(\dot{\mathbf{f}}_{i}^{\mathsf{T}}\mathbf{C}_{i}\right)_{\mathbf{b}_{i},\mathbf{b}_{i}} = \phi\left(\mathbf{I}_{n_{i}}\otimes\mathbf{Z}_{i}^{\mathsf{T}}\right)\left\{\left(\mathbf{T}_{i}\otimes\boldsymbol{\tau}_{i}\right) + \phi^{-1}\left(\mathbf{C}_{i}^{\mathsf{T}}\otimes\mathbf{I}_{n_{i}}\right)\boldsymbol{\Omega}_{i}\mathbf{M}_{i} + \operatorname{vec}\left(\mathbf{T}_{i}\right)\boldsymbol{\tau}_{i}^{\mathsf{T}} + \left(\boldsymbol{\nu}_{i}\otimes\boldsymbol{\Omega}_{i}\,\mathbf{M}_{i}\right)\right\}\mathbf{Z}_{i} \text{ and }$$

$$(\mathbf{f}_i)_{\boldsymbol{b}_i \boldsymbol{b}_i} = \phi \ \left( \mathbf{I}_{n_i} \otimes \mathbf{Z}_i^\top \right) \ \mathbf{\Omega}_i \ \mathbf{M}_i \ \mathbf{Z}_i \quad \text{and} \quad \left( \dot{\mathbf{f}}_i \right)_{\boldsymbol{b}_i \ \boldsymbol{b}_i} = \phi^{-1} \ (\mathbf{f}_i)_{\boldsymbol{b}_i \ \boldsymbol{b}_i}$$

These Hessian matrices are computed in appendix B.7.1.

Again, the quantities  $\mathbf{T}_i$ ,  $\mathbf{T}_i$ ,  $\mathbf{D}_{\rho_i}$ ,  $\mathbf{A}_i$ ,  $\mathbf{C}_i$ ,  $\mathbf{M}_i$ ,  $\dot{\mathbf{M}}_i$ ,  $\lambda_i$  and  $\nu_i$  in the right side of the expressions above should have the symbol " $\sim$ ", but we omit it by the simplicity in the writing. Also,  $\phi$  should be replaced by  $\hat{\phi}$ 

In addition, from section 2.4.1.1 we have that  $\mathbf{F}_{b_i b_i} = \mathbf{F}^*_{b_i b_i}$ , when **F** is a vector function, but not in general. However, this is not a problem, because it is enough to reordering the elements of  $\mathbf{F}_{b_i b_i}$  for obtaining  $\mathbf{F}^*_{b_i b_i}$ , according to the exposed in section 2.4.1.1. This procedure may be computationally implemented without difficulties.

Simplifications of these expressions and therefore of  $\Delta_i$ , may be obtained, from (3.12) for the canonical exponential family. Also, in practice random intercept GLMMs in the canonical exponential family are frequently used and applying jointly (3.12) and (3.13) in the expressions above, we may obtain simple structures for  $\Delta_i$  and  $\Delta$ .

#### 3.7.2.1 Conformal normal curvature - observations

Motivated by the work of Zhu and Lee (2003), we propose this perturbation scheme in order to spotlight potentially influential observations. To define the direction in which the conformal normal curvature will be computed, we consider a partition of the vector  $\mathbf{d} \in \mathbb{R}^{N_T}$ , given by  $\mathbf{d} = (\mathbf{d}_1^{\top}, \mathbf{d}_2^{\top}, \dots, \mathbf{d}_N^{\top})^{\top}$ , where  $\mathbf{d}_i \in \mathbb{R}^{n_i}$ . Thus, as was described in section 3.3, we regard the direction vector  $\mathbf{d} = \mathbf{u}_{ij}$  with  $1 \le j \le n_i$ ,  $1 \le i \le N$  and  $N_T = \sum_{i=1}^N n_i$ . The  $N_T \times 1$  vector  $\mathbf{u}_{ij}$  is a vector with 1 in the *jth* position of  $\mathbf{d}_i \in \mathbb{R}^{n_i}$  and zeros in the remaining positions. In this way, from (2.12) , the conformal normal curvature corresponding to the *jth* observation of the *ith* subject, denoted by  $B_{ij}(\boldsymbol{\psi})$ , is given by

$$\mathbf{B}_{ij}(\boldsymbol{\psi}) = \mathbf{B}_{\mathbf{u}_{ij}}(\boldsymbol{\psi}) = \frac{\mathbf{u}_{ij}^{\top} \boldsymbol{\Delta} (-\mathbf{L}_{\widehat{\boldsymbol{\psi}} \, \widehat{\boldsymbol{\psi}}})^{-1} \boldsymbol{\Delta} \mathbf{u}_{ij}}{\operatorname{tr} \left[ \left\{ \boldsymbol{\Delta}^{\top} \left( \mathbf{L}_{\widehat{\boldsymbol{\psi}} \, \widehat{\boldsymbol{\psi}}} \right)^{-1} \boldsymbol{\Delta} \right\}^{2} \right]},$$
(3.19)

for  $1 \le j \le n_i$  and  $1 \le i \le N$ . In this case, we have  $\Delta = (\Delta_1, \Delta_2, ..., \Delta_N)$ , where  $\Delta_i$  is given in (3.18) and  $L_{\widehat{\psi}\widehat{\psi}}$  is approximately computed by means of (3.10). This curvature enables the assessment of potentially influential observations in any GLMM and is one of the main results of this work.

A special case of the GLMMs are the random intercept GLMMs. These models are frequently used in the practice and their specification is simpler than the general GLMMs. Thus, the conformal normal curvature measures, which allows the assessment of the local influence, present important simplifications that will be the purpose of the next section.

# 3.8 Approximate local influence in random intercept GLMMs

Now we focus on a particular GLMM with one random effect. The main aim of this section is to particularize the results so far to random intercept GLMMs. These models

are specified as:

$$Y_{ij} | b_i \stackrel{\text{ind}}{\sim} \text{EF}(\mu_{ij}; \phi),$$

$$g(\mu_{ij}) = \mathbf{x}_{ij}^{\mathsf{T}} \boldsymbol{\beta} + b_i \quad \text{and} \quad (3.20)$$

$$b_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0; \xi),$$

where  $j = 1, ..., n_i$  and i = 1, ..., N. In addition, the notation  $EF(\cdot; \cdot)$  refers to the exponential family. That is, we assume that given  $b_i$ ,  $Y_{ij}$  follows a distribution in the form (2.1).

These models are special cases of the models defined in section 2.2.1 in which for  $j = 1, ..., n_i$  and i = 1, ..., N we have that  $z_{ij}^{\top} = 1$  or equivalently  $Z_i = \mathbf{1}_i$  for i = 1, ..., N, where  $\mathbf{1}_i$  is a  $n_i \times 1$  vector of 1's.

Under this model the conformal normal curvature defined in (2.12), and computed for different perturbation schemes in (3.15) and (3.19), suffers significant reduction. Essentially, the  $\Delta$  matrix for each perturbation scheme (see sections 3.7.1 and 3.7.2) and the Hessian matrix  $\mathbf{L}_{\widehat{\psi}\widehat{\psi}}$  assume simpler forms. These structures and consequently the conformal normal curvature measures are displayed in the remainder of this chapter.

#### **3.8.1** Approximate Hessian matrix

To obtain the approximated structure of  $\mathbf{L}_{\widehat{\psi}\widehat{\psi}}(\psi; y)$  corresponding to the model (3.20) it is necessary to perform some algebraic operations on (3.10), taking into account that  $\mathbf{Z}_i = \mathbf{1}_i$  and  $\boldsymbol{\Xi}$  becomes a positive real number  $\boldsymbol{\xi}$ . Thus, expressions in (3.10) assume the simplified form

$$\begin{split} \mathbf{L}_{\widehat{\beta}\,\widehat{\beta}}^{i} &\approx \phi^{2} \left(1-t_{i}^{-1}\right) \mathbf{X}_{i}^{\top} \left\{\mathbf{T}_{i} \mathbf{J}_{i} \mathbf{H}_{i}+\mathbf{H}_{i} \mathbf{J}_{i} \mathbf{T}_{i}\right\} \mathbf{X}_{i}+\phi^{2} t_{i}^{-2} \mathbf{X}_{i}^{\top} \mathbf{H}_{i} \mathbf{J}_{i} \mathbf{H}_{i} \mathbf{X}_{i} \\ &+ \phi \mathbf{X}_{i}^{\top} \left[\mathbf{M}_{i}+\left\{\ddot{\mathbf{M}}_{i}-3\left(\mathbf{D}_{\rho_{i}}\dot{\mathbf{A}}_{i}+\mathbf{D}_{\dot{\rho}_{i}}\mathbf{A}_{i}\right)\right\} t_{i}^{-1}\right] \mathbf{X}_{i}, \\ \mathbf{L}_{\widehat{\beta}\,\widehat{\xi}}^{i} &\approx \frac{\phi}{\xi^{2}} \mathbf{X}_{i}^{\top} \left\{2 b_{i} \mathbf{M}_{i} t_{i}^{-1}-\mathbf{H}_{i} t_{i}^{-2}\right\} \mathbf{1}_{i}, \\ \mathbf{L}_{\widehat{\beta}\,\widehat{\phi}}^{i} &\approx \mathbf{X}_{i}^{\top} \left[\mathbf{T}_{i}+\mathbf{H}_{i} \left\{1+\phi \operatorname{tr}\left(\mathbf{M}_{i}\right) t_{i}^{-1}\right\} t_{i}^{-1}+2\phi \operatorname{tr}\left(\mathbf{T}_{i}\right) \mathbf{M}_{i} t_{i}^{-1}\right] \mathbf{1}_{i}, \end{split}$$
(3.21)

$$\begin{split} & \mathcal{L}_{\hat{\xi}\,\hat{\xi}}^{i} \approx \frac{1}{2\xi^{2}} \left\{ 1 - 2\,\xi^{-1}\,b_{i}^{2} - \left(4\xi^{-1} + 4\,b_{i}^{2}\,\xi^{-2}\right)\,t_{i}^{-1} - 2\,\xi^{-2}\,t_{i}^{-2} \right\} ,\\ & \mathcal{L}_{\hat{\xi}\,\hat{\phi}}^{i} \approx \frac{1}{\xi^{2}} \left\{ 4\,b_{i}\,\mathrm{tr}\left(\mathbf{T}_{i}\right)\,t_{i}^{-1} - 2\,\mathrm{tr}\left(\mathbf{M}_{i}\right)t_{i}^{-2} \right\} \quad \mathrm{and} \\ & \mathcal{L}_{\hat{\phi}\,\hat{\phi}}^{i} \approx \sum_{j=1}^{n_{i}}c_{\hat{\phi}\,\hat{\phi}}\left(y_{ij}\,;\,\phi\right) + 2\,\mathrm{tr}\left(\mathbf{T}_{i}\right)^{2}\,t_{i}^{-1} - \mathrm{tr}\left(\mathbf{M}_{i}\right)^{2}\,t_{i}^{-2} , \end{split}$$

where  $\mathbf{H}_{i} = \dot{\mathbf{M}}_{i} - 2 \mathbf{D}_{\rho_{i}} \mathbf{A}_{i}$ ,  $\mathbf{T}_{i} = \mathbf{W}_{i}^{1/2} \mathbf{V}_{i}^{-1/2} \mathbf{D}_{y_{i}-\mu_{i}}$ ,  $\mathbf{M}_{i} = \mathbf{D}_{y_{i}-\mu_{i}} \mathbf{D}_{\rho_{i}} - \mathbf{W}_{i}^{1/2} \mathbf{V}_{i}^{-1/2} \mathbf{A}_{i}$ ,  $t_{i} = \phi \operatorname{tr}(\mathbf{W}_{i}) + \xi^{-1}$ ,  $\dot{\mathbf{M}}_{i} = \mathbf{D}_{y_{i}-\mu_{i}} \mathbf{D}_{\dot{\rho}_{i}} - \mathbf{W}_{i}^{1/2} \mathbf{V}_{i}^{-1/2} \dot{\mathbf{A}}_{i}$ ,  $\ddot{\mathbf{M}}_{i} = \mathbf{D}_{y_{i}-\mu_{i}} \mathbf{D}_{\dot{\rho}_{i}} - \mathbf{W}_{i}^{1/2} \mathbf{V}_{i}^{-1/2} \ddot{\mathbf{A}}_{i}$ ,  $\mathbf{A}_{i} = \bigoplus_{j=1}^{n_{i}} d\mu_{ij} / d\eta_{ij}$ ,  $\dot{\mathbf{A}}_{i} = \bigoplus_{j=1}^{n_{i}} d^{2}\mu_{ij} / d\eta_{ij}^{2}$  and  $\ddot{\mathbf{A}}_{i} = \bigoplus_{j=1}^{n_{i}} d^{3}\mu_{ij} / d\eta_{ij}^{3}$ .

From (3.2), we have an approximated structure for  $\mathbf{L}_{\widehat{\psi} \widehat{\psi}}(\psi; y_i)$  and consequently for  $\mathbf{L}_{\widehat{\psi} \widehat{\psi}}(\psi; y) = \sum_{i=1}^{N} \mathbf{L}_{\widehat{\psi} \widehat{\psi}}(\psi; y_i)$ .

Note that other simplifications are possible by considering canonical links. These simplifications may be obtained from (3.12). The next sections focuses on the simplified form assumed by the  $\Delta$  matrix, for each perturbation scheme, under model (3.20).

#### 3.8.2 Case weights - subjects

The  $\Delta$  matrix associated to this perturbation scheme was developed in section 3.7.1. This matrix was obtained from perturbed log-likelihood function L ( $\psi$ ;  $\boldsymbol{y} | \boldsymbol{\omega}$ ) =  $\sum_{i=1}^{N} \omega_i \operatorname{L}(\psi; \boldsymbol{y}_i)$ , where  $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_N)^{\top} \in \mathbb{R}^N$  is the perturbation vector. Also,  $\boldsymbol{\omega}_0 = \mathbf{1}_N$  is the no perturbation vector.

In appendix B.6 we show that  $\Delta = \frac{\partial L(\psi; y | \omega)}{\partial \psi \partial \omega^{\top}}\Big|_{\widehat{\psi}, \omega_0} = (\delta_1, \dots, \delta_N)$ , where, under the model (3.20),  $\delta_i$  assumes the form

$$\boldsymbol{\delta}_{i} = \begin{bmatrix} \mathbf{L}_{\widehat{\boldsymbol{\beta}}}^{i} \\ \mathbf{L}_{\widehat{\boldsymbol{\xi}}}^{i} \\ \mathbf{L}_{\widehat{\boldsymbol{\phi}}}^{i} \end{bmatrix} \approx \begin{bmatrix} \widehat{\boldsymbol{\phi}} \mathbf{X}_{i}^{\mathsf{T}} \left( \widetilde{\mathbf{W}}_{i}^{1/2} \widetilde{\mathbf{V}}_{i}^{-1/2} \mathbf{D}_{\boldsymbol{y}_{i} - \widetilde{\boldsymbol{\mu}}_{i}} + \widetilde{\boldsymbol{t}}_{i}^{-1} \widetilde{\mathbf{H}}_{i} \right) \mathbf{1}_{i} \\ -\frac{1}{2\widehat{\boldsymbol{\xi}}} + \frac{1}{2\widehat{\boldsymbol{\xi}}^{2}} \left( \widetilde{\boldsymbol{b}}_{i}^{2} + 2 \widetilde{\boldsymbol{t}}_{i}^{-1} \right) \\ \widetilde{\boldsymbol{\nu}}_{i} + \operatorname{tr} \left( \widetilde{\mathbf{M}}_{i} \right) \widetilde{\boldsymbol{t}}_{i}^{-1} \end{bmatrix}, \quad \text{for} \quad i = 1, \dots, N, \quad (3.22)$$

with  $v_i = \sum_{j=1}^{n_i} \{ y_{ij} \theta_{ij} - a(\theta_{ij}) + c_{\phi}(y_{ij}; \phi) \}$ ,  $t_i = \phi \operatorname{tr}(\mathbf{W}_i) + \xi^{-1}$ ,  $\mathbf{H}_i = \dot{\mathbf{M}}_i - 2 \mathbf{D}_{\rho_i} \mathbf{A}_i$ ,  $\mathbf{M}_i = \mathbf{D}_{y_i - \mu_i} \mathbf{D}_{\rho_i} - \mathbf{W}_i^{1/2} \mathbf{V}_i^{-1/2} \mathbf{A}_i$ ,  $\dot{\mathbf{M}}_i = \mathbf{D}_{y_i - \mu_i} \mathbf{D}_{\rho_i} - \mathbf{W}_i^{1/2} \mathbf{V}_i^{-1/2} \dot{\mathbf{A}}_i$ ,  $\mathbf{A}_i = \bigoplus_{j=1}^{n_i} d\mu_{ij} / d\eta_{ij}$  and  $\dot{\mathbf{A}}_i = \bigoplus_{j=1}^{n_i} d^2 \mu_{ij} / d \eta_{ij}^2$ . Thus, under this perturbation scheme and for model (3.20), the  $\boldsymbol{\Delta}$  matrix is expressed by

$$\Delta = \begin{bmatrix} \frac{\partial L(\psi; y | \omega)}{\partial \beta \partial \omega^{\top}} \\ \frac{\partial L(\psi; y | \omega)}{\partial \xi \partial \omega^{\top}} \\ \frac{\partial L(\psi; y | \omega)}{\partial \xi \partial \omega^{\top}} \end{bmatrix}_{\widehat{\psi}, \omega_{0}} \approx \begin{bmatrix} \widehat{\phi} \mathbf{X}^{\top} \left( \widetilde{\mathbf{W}}^{1/2} \widetilde{\mathbf{V}}^{-1/2} \mathbf{D}_{y-\widetilde{\mu}} \mathbf{1} + \widetilde{\mathbf{H}} \mathbf{1} \mathbf{D}_{\widetilde{t}}^{-1} \right) \\ \mathbf{1}_{N}^{\top} \left( -\frac{1}{2\widehat{\xi}} \mathbf{I}_{N} + \frac{1}{2\widehat{\xi}^{2}} \mathbf{D}_{\widetilde{b}}^{2} + \frac{1}{\widehat{\xi}^{2}} \mathbf{D}_{\widetilde{t}}^{-1} \right) \\ \mathbf{1}_{N}^{\top} \left[ \mathbf{D}_{\widetilde{\nu}} + \left\{ \bigoplus_{i=1}^{N} \operatorname{tr} \left( \widetilde{\mathbf{M}}_{i} \right) \right\} \mathbf{D}_{\widetilde{t}}^{-1} \right] \end{bmatrix}, \quad (3.23)$$

with  $\mathbf{t} = (t_1, t_2, ..., t_N)^{\top}, \mathbf{v} = (v_1, v_2, ..., v_N)^{\top}, \mathbf{b} = (b_1, b_2, ..., b_N)^{\top}, \mathbf{W} = \bigoplus_{i=1}^N \mathbf{W}_i,$  $\mathbf{W} = \bigoplus_{i=1}^N \mathbf{V}_i, \mathbf{H} = \bigoplus_{i=1}^N \mathbf{H}_i \text{ and } \mathbf{1} = \bigoplus_{i=1}^N \mathbf{1}_i.$ 

From (3.12) it is possible to obtain other simplifications when the model of interest indicates the canonical link.

#### 3.8.2.1 Conformal normal curvature - subjects

From (3.15), we obtain that the conformal normal curvature associated to the *ith* subject is as follows

$$\mathbf{B}_{i}\left(\boldsymbol{\psi}\right) = \frac{\boldsymbol{\delta}_{i}^{\top} \left(-\mathbf{L}_{\widehat{\boldsymbol{\psi}}\widehat{\boldsymbol{\psi}}}\right)^{-1} \boldsymbol{\delta}_{i}}{\sqrt{\mathrm{tr}\left[\left\{\boldsymbol{\Delta}^{\top} \left(\mathbf{L}_{\widehat{\boldsymbol{\psi}}\widehat{\boldsymbol{\psi}}}\right)^{-1} \boldsymbol{\Delta}\right\}^{2}\right]}},$$

with  $L_{\hat{\psi}\hat{\psi}}$  built by means of (3.21),  $\delta_i$  and  $\Delta$  given in (3.22) and (3.23), respectively. This curvature is useful to detect potentially influential subjects.

#### 3.8.3 Case weights - observations

The  $\boldsymbol{\Delta}$  matrix associated to this perturbation scheme was developed in section 3.7.2 from the perturbed log-likelihood function  $L(\boldsymbol{\psi}; \boldsymbol{y} | \boldsymbol{\omega}) = \sum_{i=1}^{N} L(\boldsymbol{\psi}; \boldsymbol{y}_i | \boldsymbol{\omega}_i)$ , where  $\boldsymbol{\omega}$  is the perturbation vector given by  $\boldsymbol{\omega} = (\boldsymbol{\omega}_1^{\mathsf{T}}, \boldsymbol{\omega}_2^{\mathsf{T}} \dots, \boldsymbol{\omega}_N^{\mathsf{T}})^{\mathsf{T}} \in \mathbb{R}^{N_T}$ , with  $\boldsymbol{\omega}_i = (\omega_{i1}, \omega_{i2}, \omega_{in_i})^{\mathsf{T}} \in \mathbb{R}^{n_i}$  and  $L(\boldsymbol{\psi}; \boldsymbol{y}_i | \boldsymbol{\omega}_i)$  is defined in (3.16). Also, the no perturbation vector is given by  $\boldsymbol{\omega}_0 = (\boldsymbol{\omega}_{0_1}^{\mathsf{T}}, \boldsymbol{\omega}_{0_2}^{\mathsf{T}}, \dots, \boldsymbol{\omega}_{0_N}^{\mathsf{T}})^{\mathsf{T}} = \mathbf{1}_{N_T}$ . In appendix B.7 we proved that under this perturbation scheme  $\Delta = (\Delta_1, \Delta_2, ..., \Delta_N)$  and under model (3.20) it is possible to prove that for i = 1, ..., N,  $\Delta_i$  assumes approximately the form

$$\boldsymbol{\Delta}_{i} = \begin{bmatrix} \frac{\partial \operatorname{L}(\boldsymbol{\psi}; \boldsymbol{y}_{i} | \boldsymbol{\omega}_{i})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}_{i}^{\mathsf{T}}} \\ \frac{\partial \operatorname{L}(\boldsymbol{\psi}; \boldsymbol{y}_{i} | \boldsymbol{\omega}_{i})}{\partial \boldsymbol{\xi} \partial \boldsymbol{\omega}_{i}^{\mathsf{T}}} \\ \frac{\partial \operatorname{L}(\boldsymbol{\psi}; \boldsymbol{y}_{i} | \boldsymbol{\omega}_{i})}{\partial \boldsymbol{\xi} \partial \boldsymbol{\omega}_{i}^{\mathsf{T}}} \end{bmatrix}_{\boldsymbol{\widehat{\psi}}, \boldsymbol{\omega}_{0_{i}}} \approx \begin{bmatrix} \widehat{\boldsymbol{\phi}} \mathbf{X}_{i}^{\mathsf{T}} \{ \widetilde{\mathbf{T}}_{i} + \widetilde{\mathbf{H}}_{i} \, \widetilde{t}_{i}^{-1} + 2 \, \boldsymbol{\phi} \, \widetilde{\mathbf{M}}_{i} \, \mathbf{J}_{i} \, \widetilde{\mathbf{T}}_{i} \, \widetilde{t}_{i}^{-1} + \boldsymbol{\phi} \, \widetilde{\mathbf{H}}_{i} \, \mathbf{J}_{i} \, \widetilde{\mathbf{M}}_{i} \, \widetilde{t}_{i}^{-2} \} \\ \frac{\widehat{\boldsymbol{\phi}}}{\widehat{\boldsymbol{\xi}}^{2}} \{ \mathbf{1}_{i}^{\mathsf{T}} (2 \, \widetilde{b}_{i} \, \widetilde{\mathbf{T}}_{i} \, \widetilde{t}_{i}^{-1} - \widetilde{\mathbf{M}}_{i} \, \widetilde{t}_{i}^{-2}) \} \\ \begin{bmatrix} \widetilde{\mathbf{f}}_{i}^{\mathsf{T}} + \mathbf{1}_{i}^{\mathsf{T}} \left[ \widetilde{k}_{i} \, \widetilde{\mathbf{M}}_{i} \, \widetilde{t}_{i}^{-1} + 2 \, \widehat{\boldsymbol{\phi}} \, \mathrm{tr} \, (\widetilde{\mathbf{T}}_{i}) \, \widetilde{\mathbf{T}}_{i} \, \widetilde{t}_{i}^{-1} - \widehat{\boldsymbol{\phi}} \, \mathrm{tr} \, (\widetilde{\mathbf{M}}_{i}) \, \widetilde{\mathbf{M}}_{i} \, \widetilde{t}_{i}^{-2} ] \end{bmatrix} \right),$$

where  $k_i = (1 - \phi) v_i + 1$ ,  $\mathbf{J}_i$  denotes an  $n_i \times n_i$  matrix of 1's and  $\dot{\mathbf{f}}_i = (\dot{\mathbf{f}}_{i1}, \dots, \dot{\mathbf{f}}_{in_i})^{\top}$ with  $\dot{\mathbf{f}}_{ij} = \partial \log \{ f_{y_{ij}|\mathbf{b}_i} (y_{ij}; \theta_{ij}, \phi) \} / \partial \phi$ . Also, as was previously mentioned,  $\mathbf{H}_i = \dot{\mathbf{M}}_i - 2 \mathbf{D}_{\rho_i} \mathbf{A}_i$ ,  $\mathbf{T}_i = \mathbf{W}_i^{1/2} \mathbf{V}_i^{-1/2} \mathbf{D}_{y_i - \mu_i}$ ,  $\mathbf{M}_i = \mathbf{D}_{y_i - \mu_i} \mathbf{D}_{\rho_i} - \mathbf{W}_i^{1/2} \mathbf{V}_i^{-1/2} \mathbf{A}_i$ ,  $\dot{\mathbf{M}}_i = \mathbf{D}_{y_i - \mu_i} \mathbf{D}_{\dot{\rho}_i} - \mathbf{W}_i^{1/2} \mathbf{V}_i^{-1/2} \dot{\mathbf{A}}_i$ ,  $\mathbf{A}_i = \bigoplus_{j=1}^{n_i} d\mu_{ij} / d\eta_{ij}$  and  $\dot{\mathbf{A}}_i = \bigoplus_{j=1}^{n_i} d^2 \mu_{ij} / d\eta_{ij}^2$ . Consequently, one approximation for  $\mathbf{\Delta} = \partial \mathbf{L} (\psi; y | \omega) / \partial \psi \partial \omega^{\top} |_{\hat{\psi}, \omega_0}$  is given by

$$\boldsymbol{\Delta} \approx \begin{bmatrix} \widehat{\boldsymbol{\phi}} \mathbf{X}^{\mathsf{T}} \left\{ \widetilde{\mathbf{T}} + \widetilde{\mathbf{H}} \left( \bigoplus_{i=1}^{N} \mathbf{I}_{n_{i}} \widetilde{t}_{i}^{-1} \right) + 2 \widehat{\boldsymbol{\phi}} \widetilde{\mathbf{M}} \mathbf{J} \widetilde{\mathbf{T}} \left( \bigoplus_{i=1}^{N} \mathbf{I}_{n_{i}} \widetilde{t}_{i}^{-1} \right) + \boldsymbol{\phi} \widetilde{\mathbf{H}} \mathbf{J} \widetilde{\mathbf{M}} \left( \bigoplus_{i=1}^{N} \mathbf{I}_{n_{i}} \widetilde{t}_{i}^{-2} \right) \right\} \\ \frac{\widehat{\boldsymbol{\phi}}}{\widehat{\xi}^{2}} \left[ \mathbf{1}_{N_{T}}^{\mathsf{T}} \left\{ 2 \widetilde{\mathbf{T}} \left( \bigoplus_{i=1}^{N} \widetilde{b}_{i} \widetilde{t}_{i}^{-1} \mathbf{I}_{n_{i}} \right) - \widetilde{\mathbf{M}} \left( \bigoplus_{i=1}^{N} \widetilde{t}_{i}^{-2} \mathbf{I}_{n_{i}} \right) \right\} \right] \\ \widetilde{\mathbf{f}}^{\mathsf{T}} + \mathbf{1}_{N_{T}}^{\mathsf{T}} \left[ \widetilde{\mathbf{M}} \left( \bigoplus_{i=1}^{N} \widetilde{k}_{i} \widetilde{t}_{i}^{-1} \mathbf{I}_{n_{i}} \right) + 2 \widehat{\boldsymbol{\phi}} \widetilde{\mathbf{T}} \left\{ \bigoplus_{i=1}^{N} \operatorname{tr} \left( \widetilde{t}_{i}^{-1} \widetilde{\mathbf{T}}_{i} \right) \mathbf{I}_{n_{i}} \right\} - \widehat{\boldsymbol{\phi}} \widetilde{\mathbf{M}} \left\{ \bigoplus_{i=1}^{N} \operatorname{tr} \left( \widetilde{\mathbf{M}}_{i} \widetilde{t}_{i}^{-2} \right) \mathbf{I}_{n_{i}} \right\} \right] \end{bmatrix}$$

where  $t_i = \phi \operatorname{tr}(\mathbf{W}_i) + \xi^{-1}$ ,  $\mathbf{T} = \bigoplus_{i=1}^N \mathbf{T}_i$ ,  $\mathbf{M} = \bigoplus_{i=1}^N \mathbf{M}_i$ ,  $\mathbf{J} = \bigoplus_{i=1}^N \mathbf{J}_i$ ,  $\mathbf{H} = \bigoplus_{i=1}^N \mathbf{H}_i$ ,  $\mathbf{I}_{n_i}$  is the  $n_i \times n_i$  identity matrix, and  $\dot{\mathbf{f}} = (\dot{\mathbf{f}}_1^\top, \dot{\mathbf{f}}_2^\top, \dots, \dot{\mathbf{f}}_N^\top)^\top$  with  $\dot{\mathbf{f}}_i = \partial \mathbf{f}_i / \partial \phi$ .

Expression (3.12) enables to achieve simplified forms for  $\Delta$  when the considered model assumes canonical link.

#### 3.8.3.1 Conformal normal curvature - observations

From expression (3.19) we have that the conformal normal curvature corresponding to the *jth* observation of the *ith* subject is given by

$$\mathbf{B}_{ij}(\boldsymbol{\psi}) = \mathbf{B}_{\mathbf{u}_{ij}}(\boldsymbol{\psi}) = \frac{\mathbf{u}_{ij}^{\top} \boldsymbol{\Delta} (-\mathbf{L}_{\widehat{\boldsymbol{\psi}} \widehat{\boldsymbol{\psi}}})^{-1} \boldsymbol{\Delta} \mathbf{u}_{ij}}{\operatorname{tr} \left[ \left\{ \boldsymbol{\Delta}^{\top} (\mathbf{L}_{\boldsymbol{\psi} \boldsymbol{\psi}})^{-1} \boldsymbol{\Delta} \right\}^{2} \right]},$$

where  $\Delta_i$  and  $\Delta$  were previously obtained and  $L_{\hat{\psi}\hat{\psi}}$  may be achieved using the expressions (3.21). This curvature may be used to identify potentially influential observations in random intercept GLMMs.

## 3.9 Concluding remarks

In this chapter we studied an approximation for local influence in GLMMs. The local influence approach that we considered was proposed by Cook (1986). We mentioned briefly some works related with local influence in random effect models and we described our strategy for achieving approximate measures to discriminate potentially influential points. In that sense, we obtained analytical structures  $-L_{\hat{\psi}\hat{\psi}}(\psi; y)$  and  $\Delta$  under different perturbation schemes and consequently for conformal normal curvature. This allowed the simple computation of the conformal normal curvatures defined in (2.12) and (2.13). In addition, we dedicated some pages to the local influence in random intercept GLMMs. Also, we justified the use of the gamlss framework like an accurate tool that provides the necessary elements for the implementation of our approximation of local influence. However, it is possible to use other packages, through of the implementation of a Newton-Raphson or Fisher scoring algorithm.

#### 46 LOCAL INFLUENCE IN GLMMS

## Chapter 4

# Numerical studies and applications

## 4.1 Introduction

The first part of this chapter contains numerical studies for exploring the effectiveness of our proposal. Thus, we consider simulated data from Poisson and gamma mixed models and we develop sensitivity analyses by means of the local influence approach. We consider a comparative study evaluating the components of the conformal normal curvature (2.12), namely the observed Fisher information matrix – L  $_{\hat{\psi}\hat{\psi}}$  and  $\Delta$ , via numerical techniques and using our proposal.

In the last part of this chapter we show the flexibility of our proposal by studying three real data sets, the analyses of which may be accomplished through GLMMs. Also, since the application of our technique requires the previous knowledge of some structures as the weight matrix, the variance function among others, we provide such quantities for different distributions in the exponential family.

## 4.2 Numerical studies

In this section we consider artificial data sets generated from Poisson and gamma random intercept models. Then, we use our proposal and numerical techniques to compute the conformal normal curvature in order to examine if it discriminates influential points.

#### 4.2.1 Poisson - normal GLMM

To exhibit the use of the approximation proposed in this work for the conformal normal curvature (2.12), we study artificial data sets generated from a random intercept GLMM with response Poisson and logarithmic link function. Specifically, the data sets come from the model specified as follows:

$$Y_{ij} | b_i \stackrel{\text{ind}}{\sim} \operatorname{Po}(\mu_{ij}),$$
  
$$g(\mu_{ij}) = \log(\mu_{ij}) = 1 + b_i + 0.5 x_{ij} \text{ and}$$
  
$$b_i \stackrel{\text{iid}}{\sim} N(0; 1),$$
  
$$(4.1)$$

where i = 1, ..., k,  $j = 1, ..., n_i$  and  $N_T = \sum_{i=1}^k n_i$ . k represents the number of units considered in each data set and  $n_i$  is the number of observations per unit. For the explanatory variables  $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{in_i})^T$  we consider two cases: balanced data  $(n_i = n = N_T/k)$  and unbalanced data. In the former,  $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{in})^T$  coincides for each subject and corresponds to a random sample of size n from the normal distribution N (1.5; 0, 5). In the latter case, for each i = 1, 2, ..., k, we assume that  $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{in_i})^T$  is a random sample of size  $n_i$  from the normal distribution N (1.5; 0.5).

The models of interest were fitted by means of the gamlss() function of the gamlss package and the normal conformal curvatures for the units,  $B_i(\psi)$ , are performed using the technique proposed in this work based on the Laplace approximation and alternatively using the Gauss-Hermite quadrature with 20 points of quadrature for computing the integrals involved in such curvatures. The latter, was considered in order to illustrate that our technique is good enough for discriminating potentially influential subjects. It is worth noting that the curvatures via Gauss-Hermite quadrature are achievable because the model in question requires one dimensional integration (q = 1), but in general this procedure may be computationally expensive (Nelder *et al.*, 2006).

We want to emphasize that the aim is not to compare the quality of our proposal with the quality of the numerical integration via Gauss-Hermite quadrature, but to show that the identification of influential subjects with the two methodologies produces similar results. This is an enormous advantage for our proposal because the computational routines of our analytical structures for  $B_i(\psi)$  are inexpensive. In general, what we expect is that the highlighted units coincide in the two methodologies. That is, via Gauss-Hermite quadrature and by means of our proposal.

As illustration, we initially consider four data sets consisting of k = 10, k = 15, k = 30 and k = 50 different units with n = 15, n = 10, n = 5 and n = 3 observations per unit, respectively. Hence, the corresponding models are balanced with  $N_T = 150$  observations each. In addition, 10% of the units were perturbed as will be stated later. A study of sensitivity was made for these models and the results are exposed in Figures 4.1, 4.2, 4.3 and 4.4. Such figures display index plots of  $B_i(\psi)$  obtained through the two methodologies. In the left side of each figure we may see the behavior of the two strategies and should be clear that our approach (solid line) is well behaved in the units with greater curvature, that is, in the potentially influential subjects. This may be confirmed in the right side of the figures in question, because in most cases the highlighted subjects were the same. The dotted and dashed horizontal lines represent twice the average of the curvatures obtained through Gauss-Hermite quadrature and by means of our proposal, respectively. This benchmark was suggested by Lesaffre and Verbeke (1998).


(a)  $B_i(\psi)$  using our proposal (APPROX) and Gauss-Hermite approximation (GH).



**(b)** *Gauss-Hermite approximation.* 



(c) Our proposal.

**Figure 4.1:** Comparison of the conformal normal curvatures with  $N_T = 150$  and k = 10 from the model (4.1) fitted to the generated data sets.



(a)  $B_i(\psi)$  using our proposal (APPROX) and Gauss-Hermite approximation (GH).



**(b)** *Gauss-Hermite approximation.* 



(c) Our proposal.

**Figure 4.2:** Comparison of the conformal normal curvatures with  $N_T = 150$  and k = 15 from the model (4.1) fitted to the generated data sets.



(a)  $B_i(\psi)$  using our proposal (APPROX) and Gauss-Hermite approximation (GH).



**(b)** *Gauss-Hermite approximation.* 



(c) Our proposal.

**Figure 4.3:** Comparison of the conformal normal curvatures with  $N_T = 150$  and k = 30 from the model (4.1) fitted to the generated data sets.



(a)  $B_i(\psi)$  using our proposal (APPROX) and Gauss-Hermite approximation (GH).



**(b)** *Gauss-Hermite approximation.* 





**Figure 4.4:** Comparison of the conformal normal curvatures with  $N_T = 150$  and k = 50 from the model (4.1) fitted to the generated data sets.

Note that Figures 4.1, 4.2, 4.3 and 4.4 cover only four particular cases of data sets coming

from (4.1) and although our proposal reveals good results in these cases, it would be appropriate to study in a more general form the usefulness of our approximation. Thus, to evaluate the effectiveness of our proposal we repeat several times the procedure described previously in different scenarios. Basically, the idea is to build distinct data sets with particular features from (4.1). Hence, we consider different sample sizes for  $N_T$ (total number of observations) and different numbers of units (subjects) k in which the observations are grouped. It is our purpose to investigate the influence of such units. Thus, for each number of observations  $N_T$  we have studied distinct scenarios, each of them defined by the number of subjects k considered. For example, given  $N_T$  = 200 we study four possible scenarios: (*i*) k = 20, (*ii*) k = 25, (*iii*) k = 40 and (*iv*) k = 50. Each scenario was replicated 300 times, thus obtaining 300 data sets from (4.1) in which 10% of the units were randomly perturbed.

The procedure of perturbation for a selected subject *r* is described as follows: we add  $\lceil \overline{y} \rceil$  to each component of the vector of responses  $y_r$ , where  $\overline{y} = (1/N_T) \sum_{i=1}^k \sum_{j=1}^n y_{ij}$  and  $\lceil \cdot \rceil$  is the ceiling function. In each data set potentially influential units were discriminated through our proposal and by means of Gauss-Hermite quadrature with 20 points of quadrature. Finally, given the discriminated units via numerical integration, we established how many of them were detected by our proposal. The average percentage of successes in the 300 replicates by scenario is reported in Table 4.1.

Table 4.1 reveals the good behavior of our approximation, because for different sample sizes grouped in distinct number of subjects, the average rate of detection (in the 300 replicates by scenario) of potentially influential subjects is greater than 90%. This confirms the features exhibited in Figures 4.1, 4.2, 4.3 and 4.4. Also, it shows that such behavior is recurrent in a variety of scenarios. Although the scenarios and the particular structure of the model (4.1) considered here do not correspond to a exhaustive study, these suggest that our proposal is a good tool in the identification of potentially influential units. In addition, the analytical structures obtained in the previous sections for the computation of  $B_i(\psi)$ , for i = 1, ..., k, and the close relationship with the GAMLSS framework, mentioned in section 3.5.2.2, allow sensitivity analyses via local influence, with inexpensive computational routines for any GLMM. On the other hand, studies of local influence via numerical integration or Bayesian methods are restricted to particular GLMMs and with great computational cost, which prevent building computational routines that enable to conduct studies of sensitivity in any practical problem.

Furthermore, unbalanced data were generated from (4.1). Figures 4.5 and 4.6 are special cases of this type of artificial data and these show a good behavior of our methodology. As previously, 300 replicates were performed in different scenarios and the average percentage of successes is reported in the bottom of Table 4.1. The procedure of perturbation is analogous to the mentioned previously. Also, the quantity of observations in each unity was selected randomly for each scenario.



(a)  $B_i(\psi)$  using our proposal (APPROX) and Gauss-Hermite approximation (GH).



**(b)** *Gauss-Hermite approximation.* 



(c) Our proposal.

**Figure 4.5:** Comparison of the conformal normal curvatures with  $N_T = 275$  and k = 16 for the model (4.1) fitted to the generated data sets under unbalanced data.



(c) Our proposal.

**Figure 4.6:** Comparison of the conformal normal curvatures with  $N_T = 275$  and k = 17 for the model (4.1) fitted to the generated data sets under unbalanced data.

Total N <sub>T</sub>	Units size $n_i$	Units	Average percentage
			of detection
	<b>Balanced Case</b> $(n_i = n)$		
	10	k = 10	90%
100	5	k = 20	91.48%
	4	<i>k</i> = 25	91.10%
	6	<i>k</i> = 25	92.73%
150	5	k = 30	93.41%
	3	k = 50	93.48%
	10	k = 20	90%
200	8	<i>k</i> = 25	93.33%
	5	k = 40	93.77%
	4	k = 50	94,38%
	12	k = 25	92.80%
300	10	k = 30	92.7%
	6	k = 50	95.98%

**Table 4.1:** *Summary of the average percentages of potentially influential subjects detected in the 300 replicates by our proposal for each scenario.* 

	5	k = 60	93%
	20	k = 25	90.77%
500	10	k = 50	94.50%
	5	k = 100	96.1%
	4	<i>k</i> = 125	95.5%
	Unbalanced Case		
		k = 25	91.31%
500		k = 45	91.78%
		k = 50	91.21%
		k = 70	90.47%
		<i>k</i> = 71	90.16%
450		k = 51	92.34%
		k = 37	90.97%
		k = 26	91.9%
		k = 26	90.36%
350		k = 51	90%
		<i>k</i> = 37	91.98%
		k = 50	91.96%
275		k = 35	93.87%
		<i>k</i> = 16	91.10%
		k = 17	91.96%
137		k = 25	91.43%
		k = 11	89.70%
		k = 10	89%
110		k = 18	90.15%
		k = 17	91.96%

The Table 4.1 suggests that for balanced and unbalanced data our proposal is well behaved.

The same numerical studies were performed for the gamma normal GLMM. Such studies are the focus of the next section.

### 4.2.2 Gamma - normal GLMM

Again, in order to evaluate our proposal, we consider artificial data sets generated from a random intercept GLMM with response variable gamma and logarithmic link function. Thus, the data sets come from the model specified as follows:

$$Y_{ij} | b_i \stackrel{\text{ind}}{\sim} G(\mu_{ij}; \phi = 1.56),$$

$$g(\mu_{ij}) = \log(\mu_{ij}) = 1 + b_i + 0.5 x_{ij} \text{ and}$$

$$b_i \stackrel{\text{iid}}{\sim} N(0; 1),$$

$$(4.2)$$

where i = 1, ..., k,  $j = 1, ..., n_i$  and  $N_T = \sum_{i=1}^k n_i$ . k represents the number of units considered in each data set and  $n_i$  is the number of observations per unit. For the explanatory variables  $x_i = (x_{i1}, x_{i2}, ..., x_{in_i})^T$  we consider two cases: balanced data  $(n_i = n = N_T/k)$  and unbalanced data. In the former,  $x_i = (x_{i1}, x_{i2}, ..., x_{in_i})^T$  coincides in each subject and corresponds to a random sample of size n from the normal distribution N (1.5; 0, 5). In the latter case, for each i = 1, 2, ..., k, we assume that  $x_i = (x_{i1}, x_{i2}, ..., x_{in_i})^T$  is a random sample of size  $n_i$  from the normal distribution N (1.5; 0.5).

In the balanced case we have data sets with 50, 100, 150, and 200 observations, which are distributed in k subjects. For example, we consider data sets generated from (4.2) with  $N_T = 100$  observations distributed in 10, 20 and 25 subjects. That is, with 100 observations we obtain three scenarios. Each scenario was replicated by 300 times and in each replicate (in each data set) was computed the conformal normal curvature using Gauss-Hermite quadrature and our proposal. Once the influential points were identified via Gauss - Hermite quadrature with 20 quadrature points, we registered how many of them were discriminated by our technique. The average percentage of

detection in the replicates is reported in Table 4.2. The procedure of perturbation in this model was similar to the one performed in section 4.2.1. Unbalanced cases were studied and the average percentage of detection is displayed in the bottom of Table 4.2. The results show an accurate behavior of our technique. It is worth mentioning that the local influence analysis via Gauss-Hermite quadrature is viable because the simple structure of model (4.2), but in general the procedures involved in the computation of conformal normal curvature may be computationally heavy.

Figures 4.7, 4.8 and 4.9 show the comparison between the normal conformal curvatures via Gauss - Hermite quadrature and through our proposal in some particular balanced data sets generated from (4.2). Figures 4.10 and 4.11 display the same comparison but in unbalanced particular data sets. In both cases, the behavior of our proposal is accurate. The results of Table 4.2 confirm the behavior in the aforementioned figures.



(a)  $B_i(\psi)$  using our proposal (APPROX) and Gauss-Hermite approximation (GH).



**(b)** *Gauss-Hermite approximation.* 



(c) Our proposal.

**Figure 4.7:** Comparison of the conformal normal curvatures with  $N_T = 100$  and k = 10 for the model (4.2) fitted to the generated data sets.



(a)  $B_i(\psi)$  using our proposal (APPROX) and Gauss-Hermite approximation (GH).



**(b)** *Gauss-Hermite approximation.* 





**Figure 4.8:** Comparison of the conformal normal curvatures with  $N_T = 100$  and k = 20 for the model (4.2) fitted to the generated data sets.



(c) Our proposal.

Subject

**Figure 4.9:** Comparison of the conformal normal curvatures with  $N_T = 100$  and k = 25 for the model (4.2) fitted to the generated data sets.



(a)  $B_i(\psi)$  using our proposal (APPROX) and Gauss-Hermite approximation (GH).



**(b)** *Gauss-Hermite approximation.* 



(c) Our proposal.

**Figure 4.10:** Comparison of the conformal normal curvatures with  $N_T = 51$  and k = 10 for the model (4.2) fitted to the generated data sets under unbalanced data.





(b) Gauss-Hermite approximation.



(c) Our proposal.



**Figure 4.11:** Comparison of the conformal normal curvatures with  $N_T = 63$  and k = 9 for the model (4.2) fitted to the generated data sets under unbalanced data.

Total N <sub>T</sub>	Units size $n_i$	Units	Average percentage
			of detection
	<b>Balanced Case</b> $(n_i = n)$		
50	5	<i>k</i> = 10	95%
	10	k = 10	93%
100	5	k = 20	94.17%
	4	<i>k</i> = 25	90.96%
	10	k = 15	96.66%
150	6	k = 25	96.10%
	5	k = 30	94.40%
	20	k = 10	96%
200	8	k = 10 k = 25	96.90%
200	5	k = 40	93.54%
	Unbalanced Case		
51	5, 5, 5, 6, 4, 6, 5, 6, 5, 4	<i>k</i> = 10	90%
63	7, 5, 8, 6, 9, 7, 7, 7, 7	<i>k</i> = 9	95%

**Table 4.2:** *Summary of the average percentages of potentially influential subjects detected in the* 300 *replicates by our proposal for each scenario.* 

Thus, our technique appears to be accurate in the discrimination of potentially influential subjects. Later we will use the gamma - normal GLMM in a real context.

In the next section we study a variety of practical problems that shows the practicality and the usefulness of our proposal in the sensitivity analysis for GLMMs.

# 4.3 Applications

In this section we intend to apply the results exhibited in the previous sections in several practical situations under different characteristics, which could be approached

through GLMMs. Analysis of sensitivity by means of local influence will be performed in GLMMs with continuous, dichotomous and count responses. In particular, the gamma, binomial and negative binomial models of random effects will be analyzed in detail. This is possible thanks to the general procedures executed in previous sections for computing the conformal normal curvature. Such procedures enable us to execute a study of influential points (subjects and observations) not only in the cases mentioned in this section, but also in any GLMM with the form given in section 2.2.1.

In each application, once the potentially influential points are highlighted, we measure the effect of such points on the estimates of the parameters of interest. This assessment will be made using the Mean Relative Change (MeRC) and the Maximum Relative Change (MRC), which are described in detail later.

# 4.3.1 Milk data set

First we consider a data set, described and analyzed by Diggle *et al.* (1994), with the aim of illustrating the methodology proposed in this paper. The experiment provides data consisting of protein content of milk samples gathered weekly for 19 weeks from 79 Australian cows, which entered into the study after calving and were fed with one of the three diets: barley, mixtured barley and lupins and lupins alone. The numbers of animals in the diet treatments were 25, 27 and 27 cows, respectively. The milk samples were collected at the first 14 weeks for all the cows and after that the number of animals decreased to 59, 50, 46, 46 and 41 in the weeks 15, 16, 17, 18 and 19, respectively. Thus, the 79 sequences of protein milk measurements do not have the same length. The objective of the milk protein study is to explain the effects of diet on the mean response profile of the protein in milk. Authors like Molenberghs and Kenward (2007) and Diggle *et al.* (1994) have considered several proposals for the analysis of this data set. To analyze this data set and try to achieve the objective of this study we use the GLMMs. Nevertheless, before postulating a model, we perform a preliminary descriptive analysis of the data.

For example, Figure 4.12 presents robust boxplots (Hubert and Vandervieren, 2008)

for the protein for each diet, for all cows and for all weeks. This figure, in general, indicates larger protein content values for the cows fed with barley, whereas the cows treated with lupins and mixture barley/lupins, respectively, present smaller and intermediate values. Also, one may observe some outliers in each group.



Figure 4.12: Robust boxplots of the protein content values for each diet for all cows and for all weeks.

Figure 4.13 presents the profiles of protein for each animal classified according to diet. Among the things that this plot may suggest, we highlight the need of including at least a random intercept in the linear predictor of the proposed model.

In Figure 4.14 the mean protein profiles for each diet along the study weeks are displayed. The means change across weeks and seem to be different for each diet. Also, Figure 4.14 indicates a possible quadratic tendency for the means. This tendency will be considered later in the proposed model. Furthermore, note that the protein content in the milk for each diet appears to be larger in the first weeks. In addition, one may notice indications that the milk produced by the animals treated with barley presents the larger mean values along weeks. In contrast, the animals fed with lupins seem to present smaller mean values, whereas the animals treated with mixture barley/lupins present intermediate mean values.



Figure 4.13: Profile of the protein content values for each cow classified for each diet across weeks.



Figure 4.14: Profile of the mean protein content values for each diet across weeks.

Thus, based on the descriptive graphs we propose the following random intercept gamma-normal model:

$$Y_{ijk} | b_i \stackrel{\text{ind}}{\sim} G(\mu_{ijk}, \phi),$$

$$g(\mu_{ijk}) = \log(\mu_{ijk}) = \beta_0 + \beta_{1(j)} + \text{weeks}_k \beta_2 + \text{weeks}_k^2 \beta_3 + b_i \quad \text{and} \qquad (4.3)$$

$$b_i \stackrel{\text{iid}}{\sim} N(0; \xi),$$

where  $Y_{ijk}$  denotes the protein content of the milk collected for the *ith* animal on the *jth* diet in the *kth* week, for i = 1, 2, ..., 79, j = 1 (barley), 2 (barley and lupins), 3 (lupins) and  $k = 1, 2, ..., n_i$  with  $\sum_{i=1}^{79} n_i = 1337$ .  $\beta_{1(j)}$  denotes the main effect of the *jth* diet and  $b_i$  represents the random effect due to the *ith* animal.

Usually, in the context of the random effect models, the Greek letter  $\xi$  that we use for denoting the variance of the random-effects is replaced by  $\sigma_b^2$ . Thus, in the remainder of this section remember that  $\xi = \sigma_b^2$ .

The density function of a random variable Y with gamma distribution is given by

$$f_{\mathrm{Y}}\left(y;\mu,\phi\right) = \exp\left\{\phi\left(-\frac{y}{\mu} - \log\mu\right) - \log\left\{\Gamma(\phi)\right\} + \phi\log(\phi y) - \log(y)\right\},\$$

where y > 0,  $\phi > 0$ ,  $\mu > 0$  and  $\Gamma(\phi) = \int_0^\infty t^{\phi-1} \exp(-t) dt$ . It should be noted that  $E(Y) = \mu$  and  $Var(Y) = \phi^{-1} \mu^2$ . Thus, the coefficient of variation of Y becomes given by  $\sigma = 1 / \sqrt{\phi}$ .

The adequacy of the model was examined through the gamlss package in R software (R Core Team, 2018), specifically, via the quantile residuals described by Stasinopoulos *et al.* (2017).



**Figure 4.15:** *Quantile residual against fitted value for*  $\mu$  *(left) and normal probability plot for the quantile residual (right) from the random intercept gamma-normal model fitted to the Milk data set (4.3).* 

The main feature of normal quantile residuals is that, if the postulated model is not incorrect, then the residuals will follow asymptotically independent standard normal distributions. So, these residuals enable us to verify the adequacy of a GAMLSS fitted model and in particular of the postulated GLMM.

In our case, from Figure 4.15 we may notice a plot of the quantile residuals against the fitted value for  $\mu$  indicating a random scatter plot around the horizontal line at 0, which means that the variability seems to be controlled. In addition, the normal probability plot for the quantile residual indicates that the gamma-normal model presents a reasonable fit to the data set. Table 4.3 displays a summary of the quantile residuals in which we have that its mean, variance, coefficient of skewness and coefficient of kurtosis are close to 0, 1, 0 and 3, respectively. This suggests that the residual has approximately standard normal distribution, as it is expected for a suitable model.

**Table 4.3:** *Summary of the quantile residuals from the random intercept gamma-normal model fitted to the Milk data set* (4.3).

Mean	$6.954 \times 10^{-6}$
Variance	1.0007
Coefficient of skewness	-0.0412
Coefficient of kurtosis	3.2988
Filliben correlation coefficient	0.9982

Another diagnostic tool provided by the gamlss package is the worm plot (Buuren and Fredriks, 2001). This is an alternative presentation of the quantile residuals that focuses on deviations of the distribution of these residuals in comparison with the standard normal distribution. The three components of a worm plot are: the worm plot points, a cubic fit of the worm plot points and two elliptical curves that represent point-wise 95% confidence intervals.

In this sense, Figure 4.16 shows the distance between residuals and their expected values represented by the horizontal dotted line. According to Stasinopoulos *et al.* (2017) if the postulated model is not incorrect, we should have at least 95% of the points between the two elliptic dashed curves. Such a feature is revealed in Figure 4.16. In addition, the worm plot may be useful in the detection of outliers, which correspond

to the points that fall outside the elliptic curves.

On the other hand, the shape of the cubic fit of the worm plot points in Figure 4.16 is important because it may reflect problems related to the location, variance, skewness and kurtosis of the fitted distribution (Stasinopoulos *et al.*, 2017, chapter 12). In our case, such cubic fit does not present a specific form. So, it does not seem to reveal inadequate characteristics in the postulated model (4.3).



Figure 4.16: Worm plot from the random intercept gamma-normal model fitted to the Milk data set (4.3).

The density plot of the predicted random effects  $\tilde{b}_i$  is presented in Figure 4.17, and the null hypothesis of normality is not rejected by the Shapiro-Wilk test. In general, as pointed out by McCulloch and Neuhaus (2011), the standard assumption of Gaussian random-effects in GLMMs results in good performance of the predicted random-effects under a wide variety of distributions for the random effects. In other words, the misspecification of the random-effects distribution in GLMMs, in general, does not affect the accuracy of the predicted random-effects  $\tilde{b}_i$ .



**Figure 4.17:** Predicted random effects density from the random intercept gamma-normal model fitted to the Milk data set (4.3).

The estimated parameters and their approximate standard errors were obtained by

means of gamlss package and are given in Table 4.4

**Table 4.4:** *Parameter estimates and their approximated standard errors from the random intercept gamma-normal model fitted to the Milk data set* (4.3).

	Complete data				
Effect	Estimate	S. E.	z-value		
Intercept	1.3260	0.0083	158.3130		
week	-0.0173	0.0020	-8.6340		
week <sup>2</sup>	0.0008	0.0001	8.0550		
Mixtbarleylupins	-0.0270	0.0050	-5.3830		
Lupins	-0.0606	0.0050	-12.0530		
$\sigma \left( \phi^{-1/2} \right)$	0.0744	0.0019	38.4950		
$\xi \left(\sigma_b^2\right)$	0.0024				
ρ	0.2986				

We may notice from Table 4.4 that all effects are marginally significant. One has, for instance, a quadratic effect of time on the mean protein in the milk. The treatment with barley only seems to produce the largest mean, whereas the treatment with lupins only, the smallest one. The estimated coefficient of variation is about 0.0744(7.44%).

In order to evaluate the variance component  $\xi$ , one may use the likelihood ratio test for assessing H<sub>0</sub> :  $\xi = 0$  against H<sub>1</sub> :  $\xi > 0$ . In that sense, according to Zhang and Lin (2008), the asymptotic null distribution of this test in random intercept GLMMs is a mixture of central chi-squared distributions  $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ , where  $\chi_0^2$  represents the degenerate distribution at the origin. In this case, the likelihood ratio statistics equals  $\zeta_{LR} = 282.5$ . The corresponding p-value is  $\frac{1}{2}P(\chi_1^2 \ge \zeta_{LR}) < 0.0001$ , which indicates strong evidence for rejecting H<sub>0</sub>. In other words, there is reason to consider the presence of the random intercept in model (4.3).

To apply this test it was necessary the use of the gamlssNP() function of gamlss.mx package (Stasinopoulos *et al.*, 2017) in the R software (R Core Team, 2018) for fitting the model (4.3). Unlike the gamlss() function, the gamlssNP() function provides maximum likelihood estimates for the parameters in random intercept GLMMs by maximizing the marginal log-likelihood function (2.8). Furthermore, this function allows us to obtain the log-likelihood function evaluated at the maximum likelihood estimates and therefore the application of this test may be easily developed.

In addition, under model (4.3), it is possible to demonstrate (see, appendix C) that the estimated correlation coefficient among two observations of the *ith* subject, induced by the assumption of the random-effects distribution, is

$$\widehat{\rho} = \operatorname{Corr}\left(Y_{ir}, Y_{is}\right) = \frac{\exp\left(\widehat{\xi}\right) - 1}{\left(\widehat{\phi}^{-1} + 1\right)\exp\left(\widehat{\xi}\right) - 1}'$$
(4.4)

where  $1 \le r, s \le n_i$ , i = 1, ..., 79 and  $r \ne s$ . Thus, note that the correlation matrix of the observations for each subject is exchangeable and it is fully specified by the correlation parameter  $\rho$ . The estimate of this parameter is displayed in Table 4.4 and suggests a low correlation. For more details the reader may review appendix C. Note that the intraclass correlation coefficient (4.4), induced by the distributional assumption of the random effects, is always positive.

**Table 4.5:** Parameter estimates and their approximated standard errors from the random intercept gamma-normal model fitted to the Milk data set (4.3) dropping the potentially influential subjects and observations on  $\widehat{\psi} = (\widehat{\beta}^{\top}, \widehat{\xi}, \widehat{\phi})^{\top}$ .

	Dropping influential subjects			Droppin	g influer	ntial obs.
Effect	Estimate	S. E.	z-value	Estimate	S. E.	z-value
Intercept	1.3255	0.0087	152.2140	1.3195	0.0084	156.8910
week	-0.0173	0.0021	-8.2560	-0.0160	0.0020	-7.7850
week <sup>2</sup>	0.0008	0.0001	7.1550	0.0008	0.0001	7.3840
Mixtbarleylupins	-0.0213	0.0051	-4.1690	-0.0278	0.0048	-5.7580
Lupins	-0.0602	0.0053	-11.3490	-0.0628	0.0049	-12.9270
$\sigma \left( \phi^{-1/2} \right)$	0.0718	0.0020	35.6300	0.0715	0.0019	36.7500
$\xi \left(\sigma_{b}^{2}\right)$	0.0024			0.0025		
ρ	0.2986			0.3106		

To use the techniques proposed in this work, it is necessary to note that the fitted model does not belong to the canonical exponential family. So, according to the approximated structures for L  $\hat{\psi}\hat{\psi}$  and  $\Delta$  (related with the specific perturbation scheme) given in sections 3.6 and 3.7, it is necessary to compute some additional matrices along with the variance matrix  $\mathbf{V}_i$ . The structures required for the computation of the conformal normal curvature defined in (2.12) and (2.13) are summarized in Table 4.6. It is worth noting that in the canonical cases the most of such structures are simplified to null matrices and the others have important simplifications as has been mentioned throughout.

**Table 4.6:** Necessary matrices for computing the conformal normal curvature in gamma-normal models with logarithmic link.

$\mathbf{W}_i = \mathbf{I}_{n_i}$	$\mathbf{V}_i = \oplus_{j=1}^{n_i} \mu_{ij}^2$
$\mathbf{D}_{\boldsymbol{\rho}_i} = -\mathbf{V}_i^{-1/2}$	$\mathbf{D}_{\dot{p}_i} = \mathbf{V}_i^{-1}$
$\mathbf{D}_{\ddot{\rho}_i} = -2  \mathbf{V}_i^{-3/2}$	$\mathbf{M}_i = -\mathbf{D}_{y_i} \mathbf{V}_i^{-1/2}$
$\dot{\mathbf{M}}_i = \mathbf{D}_{\mathbf{y}_i} \mathbf{V}_i^{-1} - \mathbf{V}_i^{-1} - \mathbf{I}_{n_i}$	$\ddot{\mathbf{M}}_{i} = -\mathbf{V}_{i}^{-1} \left( \mathbf{V}_{i} + 2 \mathbf{D}_{\mathbf{y}_{i}} \mathbf{V}_{i}^{-1/2} - 2 \mathbf{I}_{n_{i}} \right)$
$\mathbf{A}_i = \dot{\mathbf{A}}_i = \ddot{\mathbf{A}}_i = \mathbf{V}_i^{1/2}$	

Note that these structures are defined in terms of  $V_i$ , which may be built from fitted values for  $\mu$  that may be obtained, for example, from the gamlss package by means of the function fitted(fitmodel, "mu"). The computational cost of these procedures and their subsequent replacement in the analytical expressions for  $L_{\hat{\psi}\hat{\psi}}$  and  $\Delta$  obtained in this work, is inexpensive.

### 4.3.1.1 Potentially influential points

### Potentially influential subjects

Once the approximated structures for  $L_{\widehat{\psi}\widehat{\psi}}$  and  $\Delta$  were computed, we may obtain the conformal normal curvature and highlight the potentially influential points by means of Figures 4.18 and 4.19. In these figures, the most potentially influential subjects or observations correspond to the furthest points of the threshold represented by dashed lines. We suggest to use as cutoff 2  $\overline{B}$  for discriminating influential subjects (Lesaffre and Verbeke, 1998) and  $\overline{B} + 4$  SD (B) for pointing out influential observations.  $\overline{B}$  and SD (B) denote the mean and the standard deviation of B = {B<sub>i</sub> : *i* = 1,...,*N*} or B = {B<sub>ij</sub> : *j* = 1,...,*n<sub>i</sub>* and *i* = 1,...,*N*}, to identify potentially influential subjects or observations, respectively.

In that sense, in Figure 4.18 three subjects are revealed as possible influential subjects: cow 13, fed with a mixture of barley and lupins (corresponding to the animal

number 38) and cows 1 and 6 maintained with lupins alone (correspond to the animals 53 and 62, respectively).

The subjects identified as potentially influential were eliminated in different combinations and the selected model (4.3) was refitted in each case; inferential changes were not observed. To illustrate this, in the left side of Table 4.5 we exhibit the estimated parameters without the subjects potentially influential. Hence, comparing Tables 4.4 and 4.5, we may see that the majority of the parameters changed mildly.



**Figure 4.18:** Index plot of the conformal normal curvature  $B_i(\beta, \xi, \phi)$  from the random intercept gamma-normal model fitted to the Milk data set (4.3).

### Potentially influential observations

On the other hand, in order to detect observations with great impact on the fitting process of the selected model, we use the perturbation scheme developed in section 3.7.2 and by means of the analytical structures exposed in that section, the conformal normal curvatures (2.12) and (2.13) were computed for generating Figures 4.19 and 4.20. Figure 4.19 indicates several potentially influential observations. To label them we have taken into account that the 79 animals have been classified in groups of 25, 27 and 27 according to a specific diet. Hence, we have labeled the observations by (L 02, 1), (L 16, 1) and (BL12, 19). The first two represent the observations collected in the first

week of the cows number 02 and 16 maintained with lupins. The last corresponds to the observation collected in the week 19 in the cow 12 fed with a mixture of barley and lupins. Figure 4.20 shows that the observations mentioned above exercise greater influence on the estimate of the fixed effects  $\hat{\beta}$ . The sample mean of the protein content of the milk collected in the week 19 for the animals under the mixture of barley and lupins diet is 3.40 and the protein content of the observation (BL12, 19) is 2.89, which is well below the mean. Furthermore, the sample mean of the protein measured in the first week is 3.75 for the animals nourished with lupins whereas in observations (L 02, 1) and (L16, 1) the protein content was 4.20 and 4.13, respectively, which are far above the sample mean mentioned. Sample means of content protein in the milk of the animals BL13, L 02 and L 16 were computed and we find that such means are actually different from the protein content of the mentioned observations. Finally, note that according to Figures 4.19 and 4.20, the majority of the observations highlighted as influential appear for the last animals, which are cows fed with lupins. Also, in the left side of the same figures, that is, before of the subject 30 the quantity of potentially influential observations decreases considerably. In other words, most observations corresponding to the cows in the barley diet, do not have a undue weight in the estimation procedure.



**Figure 4.19:** Index plot of the conformal normal curvature  $B_{ij}(\boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\phi})$  from the random intercept gamma-normal model fitted to the Milk data set (4.3).

Again, in order to assess if the detected observations in Figure 4.19 lead to important inferential changes we deleted such points in different combinations and we observed no inferential changes. In the right side of Table 4.5 we present the parameters estimates and their approximated standard errors when the influential observations were completely removed.

Index plots for examining influential observations or subjects separately on  $\hat{\beta}$ ,  $\hat{\xi}$  or  $\hat{\sigma}$  may be obtained following equation (2.13). For instance, Figure 4.20 shows that observations (BL12, 19), (L 02, 1) and (L 16, 1) are the most potentially influential on  $\hat{\beta}$  and (L 02, 1) and (L 16, 1) are the most potentially influential on  $\hat{\xi}$  and  $\hat{\sigma}$ , respectively. Other observations are revealed in this figure. We analysed possible inferential changes and the conclusion was similar to the one made previously.



**Figure 4.20:** Index plots of the conformal normal curvature  $B_{ij}(\beta)$  (left top), against  $B_{ij}(\xi)$  (right top) and  $B_{ij}(\phi)$  (bottom) from the random intercept gamma-normal model fitted to the Milk data set (4.3).

### 4.3.1.2 MeRC and MRC

In order to assess the effect of a set of points in the estimation procedures, we computed the MRC (Maximum Relative Change) proposed by Lee *et al.* (2006) for this purpose. Also, we propose to compute the MeRC (Mean Relative Change) to assess of such effect. These quantities are defined as follows

$$MeRC = \frac{1}{s} \sum_{i=1}^{s} \left| \frac{\widehat{\psi}_{i} - \widehat{\psi}_{i}^{0}}{\widehat{\psi}_{i}} \right| \quad and \quad MRC = \max_{1 \le i \le s} \left| \frac{\widehat{\psi}_{i} - \widehat{\psi}_{i}^{0}}{\widehat{\psi}_{i}} \right|, \tag{4.5}$$

where *s* is the number the parameters,  $\widehat{\psi} = (\widehat{\psi}_1, \dots, \widehat{\psi}_s)^{\top}$  and  $\widehat{\psi}^0 = (\widehat{\psi}_1^0, \dots, \widehat{\psi}_s^0)^{\top}$  are the estimates of  $\psi$  with the complete data set and by dropping the points of interest, respectively. We expect that the impact of the potentially influential points on the estimation processes is greater than the impact of any other set of non-influential points.

To verify this, following Lee *et al.* (2006), given a set of k points discriminated as influential, we develop the procedure described below:

- (1) Compute the MeRC and MRC for the set of potentially influential points.
- (2) Choose randomly several sets of non-influential points of size *k*.
- (3) Compute the MeRC and MRC for each of the sets selected above.
- (4) Compare the MeRC's and MRC's obtained in the items (3) and (1).

Thus, according to Lee *et al.* (2006) we should verify if the MRC for the influential points is greater than the ones for the non-influential points. Similar behavior is expected for the MeRC. This will reflect if the estimation procedure is less sensitive to non-influential points.

# MeRC and MRC for the influential subjects

Since in Figure 4.18 three subjects were highlighted, we selected 10 random samples of non-influential points, each sample with 3 non-influential points and considered the above mentioned method. The results are displayed in Table 4.7 and they reveal what we expected. The greatest MeRC and MRC are associated with the influential subjects.

**Table 4.7:** Comparison of the measures MeRC and MRC for potentially influential subjects from the random intercept gamma-normal model fitted to the Milk data set (4.3).

	Influential	ntial Samples of non-influential subjects						
	subjects	1	2	3	4	5		
MeRC	0.052	0.034	0.018	0.041	0.032	0.008		
MRC	0.208	0.084	0.034	0.149	0.116	0.019		
		Samp	Samples of non-influential subjects					
		6	7	8	9	10		
MeRC		0.020	0.051	0.022	0.040	0.020		
MRC		0.058	0.108	0.05	0.109	0.037		

# MeRC and MRC for the influential observations

Furthermore, we have selected randomly several samples of non-influential observations and we computed the MRC and the MeRC, following the procedure described in section 4.3.1.2. The results are displayed in Table 4.8. As we expect, the greatest impact on the fitted model occurs for the potentially influential observations.

	Influential	Samp	Samples of non-influential observations				
	observations	1	2	3	4	5	
MeRC	0.185	0.038	0.060	0.051	0.084	0.097	
MRC	0.578	0.088	0.166	0.181	0.213	0.177	
		Samp	Samples of non-influential observations				
		6	7	8	9	10	
MeRC		0.081	0.076	0.073	0.072	0.067	
MRC		0.162	0.259	0.153	0.174	0.212	

**Table 4.8:** *Comparison of the measures* MeRC *and* MRC *for potentially influential observations from the random intercept gamma-normal model fitted to the Milk data set* (4.3).

# 4.3.2 Respiratory data set

This data set from the biomedical literature is described by Myers *et al.* (2002) and taken originally from Stokes *et al.* (1995). The aim of the study is to compare two treatments applied to patients with respiratory disease. The data were collected for 56 patients distributed in two groups of 27 and 29 patients, respectively. The first group was submitted to a treatment with active drug and the second was treated with placebo. In the period of the treatment, the respiratory condition of each patient was observed and described as good or poor in each of four visits. Also, other features observed in the patients were Age, Gender and a baseline measure.

A summary of this data set is presented in Tables 4.9 and 4.10. Both tables suggest that the probability of poor respiratory condition is reduced when the patient is treated with the active drug. Also, Table 4.9 indicates that the active drug seems to have different performance in males and females. Essentially, the female patients improved only in the last visit. In addition, from Table 4.10 we have strong indications that the active drug works differently on younger and older patients. In that sense, none of the younger patients treated with the active drug presented poor respiratory condition after the third visit. However, poor respiratory condition in the older patients treated with active drug is presented until the last visit.

Treatment	Gender	Visit 1	Visit 2	Visit 3	Visit 4
Active drug	Female	2/2	2/2	2/2	0/2
	Male	20/25	11/25	3/25	1/25
	Female	5/5	5/5	5/5	4/5
Placebo	Male	15/24	13/24	16/24	11/24

**Table 4.9:** *Patients with poor respiratory condition in each visit, classified according to the two treatments and gender.* 

**Table 4.10:** Patients with poor respiratory condition in each visit, classified according to the two treatments and age.

Treatment	Age	Visit 1	Visit 2	Visit 3	Visit 4
Active drug	(≤30)	13/16	5/16	0/16	0/16
	(> 30)	9/11	8/11	5/11	1/11
D11	(≤30)	9/17	7/17	9/17	8/17
Placedo	(> 30)	11/12	11/12	12/12	7/12

Furthermore, in Figure 4.21 we have robust boxplots (Hubert and Vandervieren, 2008) for the age according to the respiratory condition for all patients, for all visits and the two treatments. Note that the greater ages are related to poor respiratory condition. However, some extreme values are detected in patients with good respiratory condition.



Figure 4.21: Robust boxplots of age for each respiratory condition for all patients and for all visits.

Myers *et al.* (2002) and Stokes *et al.* (1995) proposed to analyze this data set via generalized estimating equations (Liang and Zeger, 1986). Myers *et al.* (2002) focused their analysis by modeling the probability of the respiratory condition poor given the explanatory variables Age, Treatment and Gender and suggested to use a correlation structure AR(1) for the responses of each patient.

In the context of the GLMMs, taking into account that the variable respiratory condition is dichotomous (0 for good, 1 for poor) and following descriptive analyses, we propose to use a random intercept model with response variable Bernoulli with success probability  $\pi_{ij}$ , which represents the probability of poor respiratory condition of the *ith* patient in the *jth* visit given the explanatory variables previously mentioned and the logit link. The model is specified as:

$$\begin{split} \mathbf{Y}_{ij} | b_i \stackrel{\text{ind}}{\sim} & \operatorname{Be}\left(\pi_{ij}\right), \\ \log\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) &= \beta_0 + \beta_1 \operatorname{Age}_i + \beta_2 \operatorname{Treatment}_i + \beta_3 \operatorname{Gender}_i + b_i \quad \text{and} \\ & b_i \stackrel{\text{iid}}{\sim} \operatorname{N}\left(0; \xi\right), \end{split}$$
(4.6)

where  $Y_{ij}$  represents the respiratory condition (1 for poor, 0 for good) of the *i*th patient

in the *jth* visit for i = 1, 2, ..., 56 and j = 1, 2, 3, 4. The explanatory variables are Age in years and the dichotomous variables treatment (0 for active drug, 1 for placebo) and Gender (0 for female, 1 for male) whereas  $b_i$  represents the random effect due to the *ith* patient.



**Figure 4.22:** *Quantile residual against fitted value for*  $\pi$  *(left) and normal probability plot for the quantile residual (right) from the random intercept Bernoulli-normal model fitted to the Respiratory data set* (4.6).

Since the response variable  $Y_{ij}$  is not continuous, Stasinopoulos *et al.* (2017) recommend to use the normalized randomized quantile residuals as a tool for establishing possible departures from the postulated model. Unlike GLMMs with continuous response variable, the quantile residual associated with discrete responses does not have a unique output. This is due to the nature of the cumulative distribution function for a discrete random variable, which is a step function with jumps at the integers and therefore is not a one to one function on [0, 1]. As a consequence, the introduction of random processes for obtaining the quantile residuals, that in this case are called randomized quantile residuals. For more details on the building of these residuals the reader may refer to Stasinopoulos *et al.* (2017) and references therein. Thus, to verify the adequacy of model (4.6) we use several realizations of the randomized quantile residuals. This option is available in the gamlss package by means of the rqres.plot() function.

The right panel of Figure 4.22 displays several realizations of the randomized quantile residuals and their median. Also, for each realization a scatter plot versus the fitted value for  $\pi$  was built, obtaining the graph presented on the left panel of Figure 4.22. Then, we may see that the behavior seems adequate as well as for the normal probability plot on the right panel of Figure 4.22. On the other hand, in order to use the worm plot (see section 4.3.1) as a diagnostic tool, we require quantile residuals that are not unique when the response is not continuous. Hence, following a similar procedure as that considered in Figure 4.22, we use worm plots for several realizations of the normalized randomized quantile residuals which are displayed in Figure 4.23. Since the realizations of the quantile residuals are not the same, the worm plots present different forms, but if the postulated model is adequate we expect suitable forms in each one of them. Thus, note that in each worm plot in Figure 4.23 the worm plot points are located between the elliptical curves and no one of the fitted cubic curves has a specific form. Hence, model 4.6 seems to be adequate.

The correlation matrix associated to the observations in each subject is discussed for some GLMMs in appendix C. The expressions for computing these correlations were achieved for GLMMs with logarithmic link. For other link functions, such as the logit link, it is necessary to use numerical methods for computing the aforementioned correlations.



**Figure 4.23:** Worm plots from the random intercept Bernoulli-normal model fitted to the Respiratory *data set* (4.6).

A plot of predicted random effects density is given in Figure 4.24. As pointed out by McCulloch and Neuhaus (2011), the shape of this density is due to the normality assumption for the random effects in model (4.6) and does not reflect the true density of the random effects. Nevertheless, the knowledge of such density is not necessary since McCulloch and Neuhaus (2011) demonstrated that the predicted values of random effects suffer a slight sensitivity to the shape of the density assumed for the random effects. Thus, the normality assumption for the random effects seems suitable in our



**Figure 4.24:** Predicted random effects density from the random intercept Bernoulli-normal model fitted to the Respiratory data set (4.6).

The parameter estimates and their approximated standard errors are given on the left side of Table 4.11.

**Table 4.11:** *Parameter estimates and their approximated standard errors from the random intercept logistic model fitted to the Respiratory data set.* 

	Complete data			Dropping influential subjects			
Effect	Estimate	S. E.	z-value		Estimate	S. E.	z-value
Intercept	-0.1362	0.8005	-0.1700		-0.0372	0.7952	-0.0470
Age	0.0481	0.0140	3.4360		0.0453	0.0141	3.2180
Placebo	1.0487	0.3062	3.4250		1.0165	0.3082	3.2980
MGender	-1.9807	0.6433	-3.0790		-2.0790	0.6420	-3.2380
$\xi \left(\sigma_b^2\right)$	0.6532				0.4236		

We may notice from Table 4.11 that all effects are marginally significant. As the age increases the probability of poor respiratory condition increases, patients submitted to placebo seems to have a larger probability (estimated odds ratio of  $\hat{\psi} = \exp(1.0486) = 2.85$ ) with respect to the patients submitted to the active drug, whereas female gender

study.

has a larger probability (estimate odds ratio of  $\hat{\psi} = \exp(1.9806) = 7.25$ ), with respect to male gender.

In addition, as mentioned in section 4.3.1, it is possible to assess the variance component  $\xi$  through the likelihood ratio test for evaluating  $H_0: \xi = 0$  against  $H_1: \xi > 0$ , which follows for large  $N_T$  a mixture of central chi-squared distributions  $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ , where  $\chi_0^2$  denotes the degenerate distribution at the origin (Zhang and Lin, 2008). In this case, the likelihood ratio statistic equals  $\zeta_{LR} = 3.99$  and the corresponding p-value is  $\frac{1}{2}P(\chi_1^2 \ge \zeta_{LR}) = 0.0228$ , which indicates evidence against  $H_0$ . In other words, one has evidence for rejecting the absence of patient random effect. As mentioned in section 4.3.1, this test was performed using the gamlssNP() function of the gamlss package in R software (R Core Team, 2018).

Alternatively, one may apply the aforementioned likelihood ratio test using the glmmML() function of the glmmML package in the R software (R Core Team, 2018). This is a function, based on the work of Broström and Holmberg (2011), that provides among other things, the p-value corresponding to the likelihood ratio test for assessing  $H_0: \xi = 0$  against  $H_1: \xi > 0$ . Thus, by means of this function the obtained p-value is 0.02283, which confirms the results achieved through the gamlssNP() function. That is, we have evidence against  $H_0$ . It is worth mentioning that the glmmML() function is only available for random intercept GLMMs with response (conditional to the random effects) Poisson and binomial.

	Dropping observation (18,4)			Droppi	Dropping influential obs.		
Effect	Estimate	S. E.	z-value	Estimate	S. E.	z-value	
Intercept	0.2294	0.8937	0.2570	12.1022	281.9406	0.0430	
Age	0.0515	0.0143	3.6100	0.0546	0.0148	3.7010	
Placebo	0.9873	0.3099	3.1850	0.9825	0.3220	3.0510	
MGender	-2.4159	0.7636	-3.1640	-14.3778	281.9402	-0.0510	
$\xi \left(\sigma_{b}^{2}\right)$	0.6987			0.9170			

**Table 4.12:** *Parameter estimates and their approximated standard errors from the random intercept logistic model fitted to the Respiratory data set, after dropping potentially influential observations.* 

### 84 NUMERICAL STUDIES AND APPLICATIONS

Now, with the purpose of determining points with a undue weight in the estimation processes we will use the approximate local influence approach proposed in sections 3.6 and 3.7. This model belongs to the canonical exponential family and consequently there are important simplifications in the structure of the conformal normal curvature defined in (2.12) and (2.13). Such simplifications are described in (3.12). In addition, we have other important reductions given by (3.13) because model (4.6) is a random intercept model. The required structures are summarized in Table 4.13.

Table 4.13: Necessary matrices for computing the conformal normal curvature in the logistic model.

$\mathbf{A}_{i} = \mathbf{V}_{i} = \mathbf{W}_{i} = \bigoplus_{j=1}^{n_{i}} \pi_{ij} \left( 1 - \pi_{ij} \right)$	$\mathbf{M}_i = -\mathbf{V}_i$
$\dot{\mathbf{M}}_i = -\dot{\mathbf{A}}_i = (2 \mathbf{D}_{\pi_i} - \mathbf{I}_{n_i}) \mathbf{V}_i$	$\phi = 1$
$\ddot{\mathbf{M}}_i = -\ddot{\mathbf{A}}_i = (6 \mathbf{V}_i - \mathbf{I}_{n_i}) \mathbf{V}_i$	

These simple structures enable to compute approximately  $L_{\hat{\psi}\hat{\psi}}$  and  $\Delta$  for each perturbation scheme according to the results in sections 3.6 and 3.7.

#### 4.3.2.1 Potentially influential points

## Potentially influential subjects

Through the perturbation scheme discussed in section 3.7.1 we may compute the conformal normal curvature  $B_i(\psi)$  for each patient and consequently to build Figure 4.25. The dashed line that represents  $2\overline{B}$  (Lesaffre and Verbeke, 1998), as it was mentioned in the previous example allows us to identify at least two potentially influential subjects, the patients 15 and 22. Patient 22 is a 46 year man, which was the only one submitted to active drug that did not present improvement (see, Table 4.10). Patient 15 is the youngest man treated with placebo that presented poor respiratory condition in the four visits. The individual and joint deletion of these patients did not represent changes in the inference. For instance, the right side of Table 4.11 reports the parameter



estimates when the spotlighted subjects are simultaneously deleted.

**Figure 4.25:** Index plot of the conformal normal curvature  $B_i(\beta, \xi)$  from the random intercept logistic model fitted to the Respiratory data set (4.6).

# Potentially influential observations

On the other hand, by means of the perturbation scheme described in section 3.7.2, it is possible to obtain the conformal normal curvature for the *ith* patient in the *jth* visit denoted by  $B_{ij}(\psi) = B_{ij}(\beta, \xi)$  with  $1 \le i \le 56$  and  $1 \le j \le 4$ . Thus, Figure 4.26 shows such curvatures for each patient and the dashed line, that represents  $\overline{B}$ +2 SD (B), enables us to highlight potentially influential observations. In this case,  $\overline{B}$  and SD (B) denote the mean and the standard deviation of  $B = \{B_{ij} : i = 1, ..., 56 \text{ and } j = 1, ..., 4\}$ . We labeled the observations in the form (*patient*, *visit*). Hence, in Figure 4.26 we may see that observations (18, 4), (28, 4) and (53, 4) are discriminated. The observations (18, 4) and (53, 4) correspond to the last visit of the patients 18 and 53. Such patients are the only women treated with active drug that presented good respiratory condition until the last visit (see Table 4.9). Also, observation (28, 4) represents the last visit for the patient 28. Such patient is the only woman treated with placebo that improved in the fourth visit (see Table 4.9).
Myers *et al.* (2002) analyzed these data using the GEE (generalized estimating equations) methodology. From that analysis, Paula (2013b), in section 5.5.2, highlighted potentially influential observations through a generalization of the Cook distance (Cook, 1977) for GEE. Such observations agree exactly with the ones spotlighted in this work.



**Figure 4.26:** Index plot of the conformal normal curvature  $B_{ij}(\boldsymbol{\beta}, \boldsymbol{\xi})$  from the random intercept logistic model fitted to the Respiratory data set (4.6).

The observations were deleted in multiple combinations to evaluate the impact on the parameter estimates and to inspect possible inferential changes. Hence, when such observations are eliminated individually and in groups of two, we achieved similar results to the ones exhibited in the left side of Table 4.12, which correspond to the estimates reached by extracting observation (18, 4). Then, we may see that there are no inferential changes. Note that there is an important shift in the Intercept, but this effect is not significant. However, by extracting jointly the three observations, substantial changes are revealed and exposed in the right side of Table 4.12. The most important occurrence is on the effect of the explanatory variable gender that becomes not significant. Thus, it seems that the observations in question induce the effect of the explanatory variable gender, but when this effect is eliminated from the model, the resulting AIC becomes greater than the one obtained for the model (4.6). Also, the randomized quantile residuals are not well behaved. However, because some standard errors appear very large when compared with the ones from the fit with all observations, one may have indication of singularity with the Fisher information matrix, so the results are not reliable.

All the fits required in this work have been performed by using the gamlss package and eventually compared with the results produced by the glmer function of the lme4 package, getting almost identical outcomes. Nevertheless, when we tried to fit the model (4.6) dropping the observations (18, 4), (28, 4) and (53, 4) jointly, the convergence is not attained. The same problems of divergence occur using the GEE methodology by extracting the mentioned observations.

Furthermore, from Figure 4.27 we may see that the observation (53, 4) is highlighted to be potentially influential on  $\hat{\beta}$  but not on  $\hat{\xi}$ . Two new observations appear spotlighted, (12, 1) and (54, 1), but none of them causes inferential changes.



**Figure 4.27:** Index plots of the conformal normal curvature  $B_{ij}(\beta)$  (left) and against  $B_{ij}(\xi)$  (right) from the random intercept logistic model fitted to the Respiratory data set (4.6).

### 4.3.2.2 MeRC and MRC

#### MeRC and MRC for the influential subjects

Following the procedure described in section 4.3.1.2 we select 10 random samples with two subjects each. Then, we compute the MeRC and the MRC for the highlighted

subjects in Figure 4.25 and for the selected samples on the significant effects. The outcomes are reported in Table 4.14 and present an expected behavior.

	Influential	Samp	les of no	on-influ	ential s	ubjects
	subjects	1	2	3	4	5
MeRC	0.122	0.055	0.066	0.029	0.075	0.098
MRC	0.351	0.117	0.113	0.047	0.216	0.157
		Samp	les of no	on-influ	ential s	ubjects
		6	7	8	9	10
MeRC		0.051	0.066	0.057	0.057	0.077
MRC		0.109	0.090	0.182	0.146	0.122

**Table 4.14:** *Comparison of the measures* MeRC *and* MRC *for potentially influential subjects from the random intercept logistic model fitted to the Respiratory data set* (4.6).

### MeRC and MRC for the influential observations

In order to perform the computations for the MeRC and the MRC, it is necessary to fit the model of interest dropping the potentially influential observations. However, for this data set there is a question if the convergence is attained when the observations of interest are ignored. This doubt arises from the inconsistent results obtained when this model is fitted with distinct packages for the treatment of GLMMs in the R software (R Core Team, 2018). Thus, we prefer to omit the study the MeRC and MRC for the potentially influential observations in this example.

It is worth noting that the doubts are presented in the convergence of the model dropping the discriminated observations, but not in the identification of such observations, which are displayed in Figure 4.26. As it was previously mentioned, the same observations were spotlighted, through the Cook's distance when this problem was treated with GEE by Paula (2013b).

### 4.3.3 Insurance data set

The Third party insurance is a obligatory insurance for vehicle owners in Australia. It insures the drivers if they suffer an accident that causes injuries to other drivers, pedestrian or passengers. The "third party" data set, described by De Jong and Heller (2008), contains the register of third party claims in a period of twelve months between 1984 and 1986 in each of 176 geographical areas (local government areas) in New South Wales, Australia. Such areas are divided into thirteen statistical divisions, that will be called territories. In addition, other features for each local government area are reported in this data set, namely, number of accidents, number of people dead or injured, population density and population.

For the insurers it is important to know the relationship between the third party claims and other variables contained in this data set. This allows to assess the risk of insuring a driver and therefore it becomes an important tool for establishing the insurance premium. Different approaches for analyzing this data set have been exposed in the literature (see, De Jong and Heller, 2008; Stasinopoulos and Rigby, 2007). Although the Poisson distribution is the natural selection by considering as response variable the number of claims, Stasinopoulos and Rigby (2007) and De Jong and Heller (2008) show that this data set presents overdispersion, so a negative binomial distribution may be a more appropriate choice, since it is well known that this distribution accommodates overdispersed data better. Furthermore, De Jong and Heller (2008) suggested to use a quasi-likelihood approach.

A preliminary analyses of this data set reveals important and natural relations among the number of claims in the remaining variables of this data set in any territory. These relations are reviewed in Stasinopoulos and Rigby (2007). In Figure 4.28, we present only a scatterplot of the number of claims versus the logarithm of the number of accidents for all territories and indicates an increase in the number of claims by increasing the number of accidents. Based in this graph, it makes sense to consider the logarithm of the number of accidents as explanatory variable, such as was suggested

### by Pinho et al. (2015).



**Figure 4.28:** *Scatterplot of the number of claims versus logarithm of the number of accidents for all territories with fitted smoothing curve.* 

We adopt the proposed model by Pinho *et al.* (2015) which is random intercept GLMM with negative binomial response specified as follows

$$Y_{ij} | b_i \stackrel{\text{ind}}{\sim} BN(\mu_{ij}, \nu),$$

$$\log(\mu_{ij}) = \beta_0 + \beta_1 \text{LogAccidents}_{ij} + b_i \text{ and} \qquad (4.7)$$

$$b_i \stackrel{\text{iid}}{\sim} N(0; \xi),$$

where  $Y_{ij}$  and  $LogAccidents_{ij}$  denote, respectively, the claims and the logarithm of the number of accidents in the *jth* area of the *ith* territory and  $b_i$  represents the random effect due to the *ith* territory, for i = 1, 2, ..., 13 and  $j = 1, ..., n_i$ , with  $N_T = \sum_{i=1}^{13} n_i = 176$ .

The parametrization that we assume for the probability function of a random variable *Y* with negative binomial distribution is

$$f(y;\mu,\nu) = \frac{\Gamma(\nu+y)}{\Gamma(y+1)\Gamma(\nu)} \left(\frac{\mu}{\mu+\nu}\right)^{y} \left(\frac{\nu}{\mu+\nu}\right)^{\nu}, \qquad (4.8)$$

where  $\mu > 0$ ,  $\nu > 0$  and y = 0, 1, 2, ... We consider  $\nu$  fixed. Hence, it is possible to prove

that (4.8) belongs to the exponential family and therefore the results in this work are applicable.

Pinho *et al.* (2015) proposed an extension of Cook's distance to GLMMs and used it for discriminating influential observations under model (4.7). We will consider the same study, but by means of the approximate local influence approach developed in this work.



**Figure 4.29:** *Quantile residual against fitted value for*  $\mu$  *(left) and normal probability plot for the quantile residual (right) from the random intercept negative binomial model with v fixed, fitted to the Insurance data set (4.7).* 

In light of the fact that the response variable is not continuous, we use several realizations of the normalized randomized quantile residuals which are displayed in probability plots in the right panel of Figure 4.29. These plots appears to be well behaved. Also, for each realization of these residuals was built a scatter plot against the adjusted value for  $\mu$ . One of these scatter plots is presented in the left panel of Figure 4.29 and may suggest that the dispersion should be modeled (the remaining scatter plots present similar features). However, in order to illustrate our methodology, we prefer to use model (4.7) proposed by Pinho *et al.* (2015).

In practice,  $\nu$  may not be assumed to be known and  $\hat{\nu}$  should be computed. We are aware that the negative binomial distribution given in (4.8), for  $\nu$  unknown, can not be written in the form (2.1). That is, the negative binomial distribution does not belong to the exponential family. Nevertheless, for illustrative purposes we obtained, through the gamlss package, the parameter estimates  $\hat{\beta}$ ,  $\hat{\xi}$  and  $\hat{\nu}$  and we focus our

sensitivity analysis on  $\hat{\beta}$  and  $\hat{\xi}$  assuming that  $\nu = \hat{\nu}$ . This is necessary because the diagnostic tool proposed in this work is valid only in the exponential family. A complete sensitivity analysis, through local influence, in model (4.7) requires computation of special curvatures for this model. As far as we know, such curvatures are not available in the statistical literature. Thus, it would be interesting to address this problem in future research.

Unlike the gamma-normal random intercept model (4.3) the correlations between two pairs of observations from *ith* subject, under the model (4.6), are not the same (see, appendix C) and therefore the correlation matrix of the  $n_i$  observations from *ith* subject is not exchangeable and in general may be unstructured. Thus, in appendix C we present an analytical expression for computing the  $n_i (n_i - 1) / 2$  different elements of the correlation matrix corresponding to each subject. In this case, we do not present the correlation matrix associated to the observations of each subject, but this is easily computable from the analytical expressions showed in appendix C.

Figure 4.31 shows worm plots for several realizations of the normalized randomized quantile residuals. These plots reveal a reasonable behavior, although given the specific form of the fitted curve it would be accurate to model the shape parameter v. However, in our case that is not possible, since we consider v fixed.



**Figure 4.30:** *Predicted random effects density from the random intercept negative binomial model with* v *fixed, fitted to the Insurance data set* (4.7).



**Figure 4.31:** Worm plots from the random intercept negative binomial model with v fixed, fitted to the *Insurance data set* (4.7).

A plot of the random effects density is shown in Figure 4.30 and does not reflect the true density of the random effects. However the normality assumption is adequate since the predicted values of the random effects are slightly affected by the misspecification of the random-effects density (McCulloch and Neuhaus, 2011). The parameter estimates obtained from the gamlss() function of the gamlss package are summarized in Table 4.15.

	Complete data				
Effect	Estimate	S. E.	z-value		
Intercept	-2.1157	0.1438	-14.7100		
LogAccidents	1.1893	0.0224	53.0400		
$\xi \left(\sigma_b^2\right)$	0.0023				

**Table 4.15:** *Parameter estimates and their approximated standard errors from the random intercept negative binomial model fitted to the Insurance data set* (4.7).

**Table 4.16:** *Parameters estimates from the negative binomial model fitted to the Insurance data* (4.7), *after dropping the influential territories and areas.* 

	Dropping influential territories			Dropping	g influen	tial areas
Effect	Estimate	S. E.	z-value	Estimate	S. E.	z-value
Intercept	-1.8588	0.1420	-13.0900	-2.0038	0.1378	-14.5400
LogAccidents	1.1526	0.0216	53.1700	1.1699	0.0213	54.7000
$\xi \left(\sigma_{b}^{2}\right)$	0.0012			0.0109		

We may notice from Table 4.15 that all effects are marginally significant. As the number of accidents increases the mean number of claims increases. With respect to the variance component  $\xi$ , one way test  $H_0: \xi = 0$  against  $H_1: \xi > 0$ . As proved by Zhang and Lin (2008) the asymptotic null distribution of this test is a mixture of central chi-squared distributions  $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_0^2$ , where  $\chi_0^2$  denotes the degenerate distribution at the origin. In our case, the likelihood ratio statistic equals  $\zeta_{LR} = 8.720$  and consequently the resulting p-value is  $\frac{1}{2}P(\chi_1^2 \ge 8.720) = 0.0016$ , which reveals a strong evidence against  $H_0$ . That is, the absence of the random effect due to territory is rejected.

The structures required for the implementation of the results of the sections 3.6 and 3.7 are reported in Table 4.17.

**Table 4.17:** Necessary matrices for computing the conformal normal curvature for the negative binomial model with *v* fixed and logarithmic link.

$$\mathbf{V}_{i} = \bigoplus_{j=1}^{n_{i}} \mathbf{V}_{ij}; \mathbf{V}_{ij} = \mu_{ij} \left( 1 + \nu^{-1} \mu_{ij} \right) \qquad \mathbf{W}_{i} = \bigoplus_{j=1}^{n_{i}} \mathbf{W}_{ij}; \mathbf{W}_{ij} = \mu_{ij} \left( 1 + \nu^{-1} \mu_{ij} \right)^{-1}$$
$$\mathbf{D}_{\rho_{i}} = -\nu^{-1} \mathbf{W}_{i} \mathbf{V}_{i} \mathbf{A}_{i} \qquad \mathbf{D}_{\dot{\rho}_{i}} = \mathbf{D}_{\rho_{i}} \mathbf{W}_{i}^{1/2} \left( \mathbf{V}_{i}^{-1/2} - \nu \mathbf{W}_{i}^{1/2} \right)$$
$$\mathbf{M}_{i} = \left( \mathbf{D}_{y_{i}} - \mathbf{A}_{i} \right) \mathbf{D}_{\rho_{i}} - \mathbf{W}_{i} \qquad \dot{\mathbf{M}}_{i} = \left( \mathbf{D}_{y_{i}} - \mathbf{A}_{i} \right) \mathbf{D}_{\dot{\rho}_{i}} - \mathbf{W}_{i} \qquad \dot{\mathbf{A}}_{i} = \dot{\mathbf{A}}_{i} = \mathbf{A}_{i} = \mathbf{D}_{\mu_{i}}$$
$$\mathbf{M}_{i}^{1/2} \left( \mathbf{V}_{i}^{-1/2} - \nu \mathbf{W}_{i}^{1/2} \right) - 2\mathbf{D}_{\rho_{i}} \mathbf{W}_{i} \mathbf{V}_{i}^{-1}$$

We may obtain the conformal normal curvatures  $B_i(\psi)$  and  $B_{ij}(\psi)$  for i = 1, ..., 13and  $j = 1, ..., n_i$ , where  $\psi = (\beta^{\top}, \xi)^{\top}$ . Such curvatures are represented in Figures 4.32 and 4.33. In Figure 4.32, the dashed line represents  $2\overline{B}$  as was suggested in Lesaffre and Verbeke (1998). B is the set of normal conformal curvatures associated to each of the thirteen territories. In Figure 4.33, B is the set of normal conformal curvatures associated to each of the 176 local government areas and the dashed line represents  $\overline{B} + 2 \text{SD}(B)$ .

#### 4.3.3.1 Potentially influential points

#### Potentially influential subjects

Figure 4.32 shows the territories 7 and 13 as potentially influential. These territories have the least and the greatest proportion of claims in relation to the number of accidents, respectively. Particularly, territory 7 had been highlighted as influential in Pinho *et al.* (2015) by means of their proposal of Cook distance. However, territory 13 is special because it contains three geographical areas, being one of them Brokenhill, in which the number of claims is extremely high in relation with the number accidents as well as an unincorporated area in which the number of claims is low in relation with its accidents.



**Figure 4.32:** *Index plot of the conformal normal curvature from the random intercept negative binomial model fitted to the Insurance data* (4.7).

### Potentially influential observations

In Figure 4.33, two local government areas of the territory 12, (12, 4) and (12, 16), one of the territory 13, denoted by (13, 1) and one of the territory 5, represented by (5, 3), were pointed out. We labeled the observations in the form (territory, area). Thus, for example observation (13, 1) corresponds to the first area of the territory 13 and analogously for the remaining highlighted areas. The observation (13, 1) is the Brokenhill area and its influence may be due to the number of claims (912) in relation with the number of accidents(540). The influence of (5, 3), Casino area, may be justified by the same arguments, because the number of claims (182) is high when compared with the number of accidents (208). In fact, in the territory 5 the greatest proportion of claims relative to the number of accidents occur in the Casino area. The remaining areas of this territory has a relatively low proportion of claims. It is worth noting that these areas were identified by Pinho *et al.* (2015) using the Cook distance. Furthermore, we have discriminated the areas (12, 4) and (12, 16), that are Cohargo and Windouran areas. Their influence may be due to have zero claims in spite of occurring accidents.



**Figure 4.33:** Index plot of the conformal normal curvature  $B_{ij}(\beta, \xi)$  from the random intercept negative binomial model fitted to the Insurance data set (4.7).

The territories and areas previously highlighted were deleted in several combinations and model (4.7) was fitted again for studying a possible inferential impact. In Table 4.16 are recorded the most extreme cases, that is when the identified areas and territories are completely deleted. Note that in spite of taking place important changes on  $\hat{\xi}$ (likely related to the elimination of Brokenhill or its territory), there are no inferential changes.

Also, Figure 4.34 displays the same highlighted areas, but it is possible to note that the majority of them have impact only on the fixed effects estimates  $\hat{\beta}$ . Brokenhill is the most influential area on  $\hat{\xi}$  and this was foreseeable because it belongs to the territory 13, that has different features and particularly Brokenhill presents a large number of claims in comparison with other areas in this territory.



**Figure 4.34:** Index plots of the conformal normal curvature  $B_{ij}(\beta)$  (left) and against  $B_{ij}(\xi)$  (right) from the random intercept negative binomial model fitted to the Insurance data set.

### 4.3.3.2 MeRC and MRC

### MeRC and MRC for the potentially influential territories

Using the strategies mentioned in section 4.3.1.2, we accomplish the computations of the MeRC and MRC of the two potentially influential territories and of 10 random samples of two non-influential territories. The results of these procedures are recorded in Table 4.18 and show that the highlighted territories have a greater influence than the other territories.

	Influential	Samp	les of no	on-influ	ential s	ubjects
	subjects	1	2	3	4	5
MeRC	0.367	0.065	0.136	0.144	0.024	0.094
MRC	0.949	0.186	0.352	0.372	0.047	0.264
		Samp	les of no	on-influ	ential s	ubjects
		6	7	8	9	10
MeRC		0.114	0.027	0.100	0.057	0.094
MRC		0.317	0.073	0.270	0.160	0.264

**Table 4.18:** *Comparison of the measures* MeRC *and* MRC *for potentially influential subjects from the random intercept negative binomial model fitted to the Insurance data set* (4.7).

### MeRC and MRC for the potentially influential areas

Table 4.19 reports the MeRC and MRC for the four potentially influential areas displayed in Figure 4.33 and for 10 random samples of non-influential areas. Each sample with four areas. This table shows that the influence of the discriminated areas is greater than the others.

**Table 4.19:** *Comparison of the measures* MeRC *and* MRC *for influential areas from the fit of the negative binomial model to the Insurance data set* (4.7).

	Influential	Samp	Samples of non-influential areas			
	areas	1	2	3	4	5
MeRC	0.171	0.014	0.030	0.037	0.056	0.039
MRC	0.480	0.026	0.090	0.102	0.158	0.110
		Samp	Samples of non-influential areas			
		6	7	8	9	10
MeRC		0.032	0.047	0.018	0.044	0.032
MRC		0.085	0.094	0.047	0.129	0.086

### 4.4 Concluding remarks

In this chapter we accomplished numerical studies to assess the effectiveness of our technique based on the Laplace approximation. Essentially, we compared our proposal with the local influence measures obtained via numerical integration. This comparison was made in random intercept GLMMs with count response variables. The simplicity of that model was necessary because the Gauss-Hermite integration may be computationally heavy when the dimension of the random effects increases. The results of such comparison were satisfactory and consequently suggest that our proposal is suitable for highlighting influential points.

Once the effectiveness of our proposal was guaranteed, we considered three data sets with different features, which carry to the study of a variety of models that are often used in practice. Models with count, continuous and dichotomous response variable were analyzed in our work. This shows the flexibility of our proposal for highlighting potentially influential points. Also, since we built analytical structures for  $L_{\hat{\psi}\hat{\psi}}$  and  $\Delta$ , the computational routines are inexpensive. In addition, for confirming that the detected points are potentially influential we measured the impact of the discriminated points in the parameter estimates by means of the MRC suggested by Lee *et al.* (2006) and the MRC that we proposed. This study suggests what we expected, the identified points have the greatest impact on the fitted models.

Finally, taking advantage that some of the data sets considered in this work have been analyzed by other authors through GEE (Paula, 2013b) or GLMMs (Pinho *et al.*, 2015) and sensitivity analysis were performed for them using Cook's distance, we compared the potentially influential points detected with their proposals and with our proposal. The results are accurate since the detected points by them were also detected by our proposal. Although, in general this does not have to be necessarily true, the points identified with other approaches are confirmed with the approximate local influence that we propose in this work.

### 102 NUMERICAL STUDIES AND APPLICATIONS

# Chapter 5

# Conclusions

### 5.1 Final considerations

Two mathematical tools were essential in the development of this work. First, the Laplace approximation that has been widely used for estimation and diagnostic procedures in regression models. In that sense, we introduced some multivariate extensions of this approximation that facilitate an accurate presentation of the computations for  $L_{\widehat{\psi},\widehat{\psi}}$ ,  $\Delta$  and therefore for the conformal normal curvature  $B_d$ . The second fundamental tool extensively employed in this work was the theory of matrix differentiation (Magnus and Neudecker, 1999), because even though the Laplace approximation avoids the intrinsic multidimensional integration of the GLMMs, this approximation requires the computation of Hessians, and in our case Hessians of matrix functions, which are unfeasible with traditional derivation techniques. For example, Lesaffre and Verbeke (1998) provide local influence measures in linear mixed models, but such measures have complex structures, since were computed using common techniques of derivation. This is an enormous problem, since the implementation of diagnostic procedures may be tedious. However, if such curvatures are computed using the theory of matrix differentiation, it is possible to obtain equivalent results but in a more accurate presentation that allows an easy implementation.

The numerical studies suggested that our proposal is accurate for the identification of potentially influential subjects for balanced and unbalanced data sets. These numerical studies involve the comparison of curvatures obtained via numerical integration and by means of our proposal. The models used to conduct this comparison were chosen to avoid problems with computational time in the numerical integration.

We proposed approximations for the local influence measures in GLMMs. Such approximations have a closed form and therefore can be implemented with inexpensive computational routines. We show that the expressions present important simplifications when the distribution associated to the model belongs to the canonical exponential family. Also, our study is general and may be applied to any GLMM.

In contrast to other diagnostic proposals, our approach enables the distinction between the influence due to features of a subject and the influence related with characteristics of an observation from some subject. In fact, the applications considered in this work showed that even if a subject is not discriminated as potentially influential, some of its observations may be. In some other cases, one subject is discriminated to be influential but none of its observations is not potentially influential. Therefore, in a GLMM, sensitivity analysis is necessary to assess potentially influential subjects and observations. Also in this work we studied the impact of the detected points on the parameter estimates through the mean relative change (MeRC) and the maximum relative change (MRC). Such studies suggest that the discriminated observations and subjects are correctly identified. Furthermore, the structures required for the implementation of our diagnostic procedures in gamma mixed models, negative binomial mixed models and Bernoulli mixed models are provided in Tables 4.17, 4.13 and 4.6 in section 4.

### 5.2 Suggestions for future research

Several research lines directly related with the results of this work may be developed. Thus, we have

• To implement a package in the libraries of GLMMs of the R software (R Core Team, 2018) for discriminating potentially influential subjects and observations through

the local influence approach.

- Tang *et al.* (2006) studied diagnostic techniques for nonlinear reproductive dispersion mixed models, particularly the local influence approach. However, these procedures are valid for highlighting observations with an undue weight only on the fixed effects. Thus, extensions of this sensitivity analysis for the dispersion and covariance parameters are necessary.
- The gamlss framework allows analyses via a great quantity of distributions (Stasinopoulos *et al.*, 2017, chapter 6) that may be used in a regression model according to the response variable of interest. The majority of such distributions does not belong to the exponential family and therefore would be necessary a sensitivity analysis through the local influence approach for these models. In fact, as far as we know, deletion diagnostics such as the Cook distance are not yet available in statistical packages.

In general the gamlss framework allows fitting and selecting a great quantity of regression models with several choices to the distribution of the response variable of interest. In this framework, the analysis of residuals is considered via the quantile residuals and the estimation procedures are performed by maximizing a general penalized likelihood (Rigby and Stasinopoulos, 2005). However, as far as we know, this framework provides nothing on the identification of potentially influential points. This open enormous possibilities for future research. In our case, we are initially interested in count mixed models with excess of zeros, e. g., For example, Poisson mixed models with excess of zeros and negative binomial mixed models with excess of zeros.

• Since the gamlss framework allows the analyses of location, scale, skewness and kurtosis, we intend to extend the local influence approach to such models, beginning with double generalized linear mixed models, that is, models with linear predictors associated to the scale and dispersion parameters. Such predictors in-

### 106 CONCLUSIONS

clude random effects. This work would be an extension of the work performed for double generalized linear models by Paula (2013a).

• In general it would be accurate to develop robust estimation methods in the GLMMs.

# Appendix A

# Proofs related to the Laplace approximation

## A.1 **Proof of theorem 1**

*Proof.* Using the Taylor series expansion of the second order for h and v at  $\tilde{x}$ , we obtain

$$h(\mathbf{x}) \approx h(\widetilde{\mathbf{x}}) + \frac{1}{2} (\mathbf{x} - \widetilde{\mathbf{x}})^{\top} \{ \mathbf{h}_{\widetilde{\mathbf{x}}\widetilde{\mathbf{x}}} \} (\mathbf{x} - \widetilde{\mathbf{x}}) \quad (\text{because } \mathbf{h}_{\widetilde{\mathbf{x}}} = \mathbf{0}_q )$$
$$v(\mathbf{x}) \approx v(\widetilde{\mathbf{x}}) + (\mathbf{x} - \widetilde{\mathbf{x}})^{\top} \mathbf{v}_{\widetilde{\mathbf{x}}} + \frac{1}{2} (\mathbf{x} - \widetilde{\mathbf{x}})^{\top} \{ \mathbf{v}_{\widetilde{\mathbf{x}}\widetilde{\mathbf{x}}} \} (\mathbf{x} - \widetilde{\mathbf{x}}) .$$

Now, using these expressions we can obtain

$$\int_{\mathbb{R}^{q}} v(\mathbf{x}) \exp\{\mathbf{h}(\mathbf{x})\} \approx \int_{\mathbb{R}^{q}} \left\{ v(\tilde{\mathbf{x}}) + (\mathbf{x} - \tilde{\mathbf{x}})^{\top} v_{\tilde{\mathbf{x}}} + \frac{1}{2} (\mathbf{x} - \tilde{\mathbf{x}})^{\top} \{v_{\tilde{\mathbf{x}}\tilde{\mathbf{x}}}\} (\mathbf{x} - \tilde{\mathbf{x}}) \right\} \exp\{\mathbf{h}(\mathbf{x})\} \, \mathrm{d}\,\mathbf{x}.$$
(A.1)

To evaluate this integral we require to perform the computations given below

(a) 
$$\int_{\mathbb{R}^{q}} v(\tilde{x}) \exp\{h(x)\} dx \approx v(\tilde{x}) \int_{\mathbb{R}^{q}} \exp\{h(\tilde{x}) + \frac{1}{2}(x-\tilde{x})^{\top} \{\mathbf{h}_{\tilde{x}\tilde{x}}\}(x-\tilde{x})\} dx$$
$$= v(\tilde{x}) \exp\{h(\tilde{x})\} \int_{\mathbb{R}^{q}} \exp\{\frac{1}{2}(x-\tilde{x})^{\top} [\{\mathbf{h}_{\tilde{x}\tilde{x}}\}^{-1}]^{-1}(x-\tilde{x})\} dx$$
$$= v(\tilde{x}) \exp\{h(\tilde{x})\} \frac{(2\pi)^{q/2} |\{-\mathbf{h}_{\tilde{x}\tilde{x}}\}^{-1}|^{1/2}}{(2\pi)^{q/2} |\{-\mathbf{h}_{\tilde{x}\tilde{x}}\}^{-1}|^{1/2}} \int_{\mathbb{R}^{q}} \exp\{-\frac{1}{2}(x-\tilde{x})^{\top} [\{-\mathbf{h}_{\tilde{x}\tilde{x}}\}^{-1}]^{-1}(x-\tilde{x})\} dx$$
$$= v(\tilde{x}) \exp\{h(\tilde{x})\} (2\pi)^{q/2} |\{-\mathbf{h}_{\tilde{x}\tilde{x}}\}^{-1}|^{1/2}} \int_{\mathbb{R}^{q}} \exp\{-\frac{1}{2}(x-\tilde{x})^{\top} [\{-\mathbf{h}_{\tilde{x}\tilde{x}}\}^{-1}]^{-1}(x-\tilde{x})\} dx$$

$$(\mathbf{b}) \int_{\mathbb{R}^{q}} (\mathbf{x} - \tilde{\mathbf{x}})^{\top} \boldsymbol{v}_{\tilde{\mathbf{x}}} \exp \{\mathbf{h}(\mathbf{x})\} \, d\mathbf{x} \approx \int_{\mathbb{R}^{q}} (\mathbf{x} - \tilde{\mathbf{x}})^{\top} \boldsymbol{v}_{\tilde{\mathbf{x}}} \exp \left\{\mathbf{h}(\tilde{\mathbf{x}}) + \frac{1}{2} (\mathbf{x} - \tilde{\mathbf{x}})^{\top} \{\mathbf{h}_{\tilde{\mathbf{xx}}}\} (\mathbf{x} - \tilde{\mathbf{x}})\right\} \, d\mathbf{x} \\ = \exp \{\mathbf{h}(\tilde{\mathbf{x}})\} \int_{\mathbb{R}^{q}} (\mathbf{x} - \tilde{\mathbf{x}})^{\top} \boldsymbol{v}_{\tilde{\mathbf{x}}} \exp \left\{-\frac{1}{2} (\mathbf{x} - \tilde{\mathbf{x}})^{\top} \{-\mathbf{h}_{\tilde{\mathbf{xx}}}^{-1}\}^{-1} (\mathbf{x} - \tilde{\mathbf{x}})\right\} \, d\mathbf{x} \\ = (2\pi)^{q/2} \left|\{-\mathbf{h}_{\tilde{\mathbf{xx}}}^{*}\}\right|^{-1/2} \exp \{\mathbf{h}(\tilde{\mathbf{x}})\} \, \mathbf{E}_{N_{q}\left(\tilde{\mathbf{x}}: -\mathbf{h}_{\tilde{\mathbf{xx}}}^{-1}\right)} \left\{(\mathbf{x} - \tilde{\mathbf{x}})^{\top} \boldsymbol{v}_{\tilde{\mathbf{x}}}\right\} \\ = (2\pi)^{q/2} \left|\{-\mathbf{h}_{\tilde{\mathbf{xx}}}^{*}\}\right|^{-1/2} \exp \{\mathbf{h}(\tilde{\mathbf{x}})\} \, \mathbf{0}_{q}^{\mathsf{T}} \, \boldsymbol{v}_{\tilde{\mathbf{x}}} = 0. \\ (\mathbf{c}) \int_{\mathbb{R}^{q}} \frac{1}{2} (\mathbf{x} - \tilde{\mathbf{x}})^{\top} \, \boldsymbol{v}_{\tilde{\mathbf{xx}}} (\mathbf{x} - \tilde{\mathbf{x}}) \exp \{\mathbf{h}(\mathbf{x})\} \, d\mathbf{x} \\ \approx \frac{1}{2} \exp \{\mathbf{h}(\tilde{\mathbf{x}})\} \int_{\mathbb{R}^{q}} \frac{1}{2} (\mathbf{x} - \tilde{\mathbf{x}})^{\top} \, \mathbf{v}_{\tilde{\mathbf{xx}}} (\mathbf{x} - \tilde{\mathbf{x}}) \exp \{\mathbf{h}(\mathbf{x})\} \, d\mathbf{x} \\ = \frac{1}{2} (2\pi)^{q/2} \left|\{-\mathbf{h}_{\tilde{\mathbf{xx}}}^{*}\}\right|^{-1/2} \exp \{\mathbf{h}(\tilde{\mathbf{x}})\} \, \mathbf{E}_{N_{q}\left(\tilde{\mathbf{x}}: -\mathbf{h}_{\tilde{\mathbf{x}}}^{-1}\right)} \left\{(\mathbf{x} - \tilde{\mathbf{x}})^{\top} \, \boldsymbol{v}_{\tilde{\mathbf{xx}}} (\mathbf{x} - \tilde{\mathbf{x}})\right\} \, d\mathbf{x} \\ = \frac{1}{2} (2\pi)^{q/2} \left|\{-\mathbf{h}_{\tilde{\mathbf{xx}}}^{*}\}\right|^{-1/2} \exp \{\mathbf{h}(\tilde{\mathbf{x}})\} \, \mathbf{E}_{N_{q}\left(\tilde{\mathbf{x}}: -\mathbf{h}_{\tilde{\mathbf{x}}}^{-1}\right)} \left[\mathbf{tr} \, \{(\mathbf{x} - \tilde{\mathbf{x}}) (\mathbf{x} - \tilde{\mathbf{x}})\} \right] \\ = \frac{1}{2} (2\pi)^{q/2} \left|\{-\mathbf{h}_{\tilde{\mathbf{xx}}}^{*}\}\right|^{-1/2} \exp \{\mathbf{h}(\tilde{\mathbf{x}})\} \, \mathbf{tr} \left[\mathbf{E}_{N_{q}\left(\tilde{\mathbf{x}}: -\mathbf{h}_{\tilde{\mathbf{x}}}^{-1}\right)} \left\{(\mathbf{x} - \tilde{\mathbf{x}}) (\mathbf{x} - \tilde{\mathbf{x}})^{\top}\right\} \, \boldsymbol{v}_{\tilde{\mathbf{xx}}}\right\} \right] \\ = \frac{1}{2} (2\pi)^{q/2} \left|\{-\mathbf{h}_{\tilde{\mathbf{xx}}}^{*}\}\right|^{-1/2} \exp \{\mathbf{h}(\tilde{\mathbf{x}})\} \, \mathbf{tr} \left[\mathbf{E}_{N_{q}\left(\tilde{\mathbf{x}}: -\mathbf{h}_{\tilde{\mathbf{x}}}^{-1}\right)} \left\{(\mathbf{x} - \tilde{\mathbf{x}}) (\mathbf{x} - \tilde{\mathbf{x})^{\top}\right\} \, \boldsymbol{v}_{\tilde{\mathbf{xx}}}\right] \\ = \frac{1}{2} (2\pi)^{q/2} \left|\{-\mathbf{h}_{\tilde{\mathbf{xx}}}^{*}\}\right|^{-1/2} \exp \{\mathbf{h}(\tilde{\mathbf{x})\}} \, \mathbf{tr} \left\{-(\mathbf{h}_{\tilde{\mathbf{x}}}^{*})^{-1} \, \boldsymbol{v}_{\tilde{\mathbf{xx}}}\right\}.$$

Replacing (*a*), (*b*) and (*c*) in (A.1) we obtain the result of theorem 1.

#### 

•

## A.2 Proof of theorem 2

*Proof.* From theorem 1 and by considering v(x) = 1, we obtain the result exposed in (2.16). The first result of this theorem is obtained making the ratio between the outcome of the theorem 1 and the expression (2.16).

Also, we can obtain that

$$\frac{\int v(\mathbf{x}) \exp \{\mathbf{h}(\mathbf{x})\} d\mathbf{x}}{\int \exp \{\mathbf{h}(\mathbf{x})\} d\mathbf{x}} \approx v(\widetilde{\mathbf{x}}) - \frac{1}{2} \operatorname{tr} \{ (\mathbf{h}_{\widetilde{x}\widetilde{x}})^{-1} (\mathbf{v}_{\widetilde{x}\widetilde{x}}) \}$$
  
=  $v(\widetilde{\mathbf{x}}) - \frac{1}{2} \operatorname{tr} \{ \mathbf{I}_q \ \mathbf{I}_q (\mathbf{h}_{\widetilde{x}\widetilde{x}})^{-1} (\mathbf{v}_{\widetilde{x}\widetilde{x}}) \}$   
=  $v(\widetilde{\mathbf{x}}) - \frac{1}{2} \operatorname{vec} (\mathbf{I}_q)^{\top} \operatorname{vec} \{ \mathbf{I}_q (\mathbf{h}_{\widetilde{x}\widetilde{x}})^{-1} (\mathbf{v}_{\widetilde{x}\widetilde{x}}) \}$   
=  $v(\widetilde{\mathbf{x}}) - \frac{1}{2} \operatorname{vec} (\mathbf{I}_q)^{\top} (\mathbf{v}_{\widetilde{x}\widetilde{x}} \otimes \mathbf{I}_q) \operatorname{vec} \{ (\mathbf{h}_{\widetilde{x}\widetilde{x}})^{-1} \}$ 

This equation is important for achieving a more generalized form of this theorem, which is exposed in theorem 3.  $\hfill \Box$ 

# A.3 Proof of theorem 3

*Proof.* The result may be verified by considering in each component  $a_{ij}$  of the matrix **A** its approximated value given in the second result of the theorem 1 and by performing the accurate simplifications.

### 110 APPENDIX A

# Appendix **B**

# Computations of the quantities related with the curvature

### **B.1** Some aspects of notation

In the development of the computations presented below, it is useful to take into account the aspects of notation exposed in Table B.1.

 Table B.1: Some useful results.

Some conventions used in the procedures developed throughout this work  $\boldsymbol{\tau}_i = \mathbf{W}_i^{1/2} \, \overline{\mathbf{V}_i^{-1/2} \, \left( \boldsymbol{y}_i - \boldsymbol{\mu}_i \right)}$  $\boldsymbol{\lambda}_i = \boldsymbol{b}_i \otimes \boldsymbol{b}_i$  $\dot{\mathbf{A}}_i = \bigoplus_{j=1}^{n_i} \mathrm{d}^2 \,\mu_{ij} / \mathrm{d} \,\eta_{ij}^2$  $\mathbf{A}_i = \bigoplus_{i=1}^{n_i} \mathrm{d}\,\mu_{i\,i}/\mathrm{d}\,\eta_{i\,i}$  $v_{i} = \sum_{i=1}^{n_{i}} \left\{ y_{ij} \theta_{ij} - a \left( \theta_{ij} \right) + c_{\phi} \left( y_{ij}; \phi \right) \right\}$  $\ddot{\mathbf{A}}_i = \bigoplus_{j=1}^{n_i} \mathrm{d}^3 \, \mu_{ij} / \mathrm{d} \, \eta_{ij}^3$  $\rho_{ij} = \frac{\mathrm{d}^2 \theta_{ij}}{\mathrm{d}\mu_{ii}^2} \left(\frac{\mathrm{d}\mu_{ij}}{\mathrm{d}\eta_{ii}}\right)^2 + \frac{\mathrm{d}\theta_{ij}}{\mathrm{d}\mu_{ij}} \frac{\mathrm{d}^2 \mu_{ij}}{\mathrm{d}\eta_{ii}^2}$  $\dot{\rho}_{ij} = \frac{\mathrm{d}\,\rho_{ij}}{\mathrm{d}\,n_{ii}}, \quad \ddot{\rho}_{ij} = \frac{\mathrm{d}^2\,\rho_{ij}}{\mathrm{d}\,n_{ii}^2}$  $\boldsymbol{\rho}_i = \left(\rho_{i1}, \ldots, \rho_{in_i}\right)^{\mathsf{T}}$  $\mathbf{D}_{\boldsymbol{\alpha}} = \bigoplus_{i=1}^{n_i} \alpha_i; \ \boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{n_i})^{\top}$  $\dot{\boldsymbol{\rho}}_i = \left(\dot{\rho}_{i1}, \ldots, \dot{\rho}_{in_i}\right)^{\mathsf{T}}$  $\ddot{\boldsymbol{\rho}}_i = \left( \ddot{\rho}_{i1}, \ldots, \ddot{\rho}_{in_i} \right)^{\mathsf{T}}$  $\dot{\mathbf{M}}_i = \mathbf{D}_{\boldsymbol{y}_i - \boldsymbol{\mu}_i} \mathbf{D}_{\dot{\boldsymbol{\rho}}_i} - \mathbf{W}_i^{1/2} \mathbf{V}_i^{-1/2} \dot{\mathbf{A}}_i$  $\mathbf{M}_i = \mathbf{D}_{\boldsymbol{y}_i - \boldsymbol{\mu}_i} \, \mathbf{D}_{\boldsymbol{\rho}_i} - \mathbf{W}_i^{1/2} \, \mathbf{V}_i^{-1/2} \, \mathbf{A}_i$  $\ddot{\mathbf{M}}_i = \mathbf{D}_{\mathbf{y}_i - \boldsymbol{\mu}_i} \mathbf{D}_{\dot{\boldsymbol{p}}_i} - \mathbf{W}_i^{1/2} \mathbf{V}_i^{-1/2} \ddot{\mathbf{A}}_i$  $\mathbf{f}_{i} = (\mathbf{f}_{i1}, \dots, \mathbf{f}_{in})^{\mathsf{T}}; \mathbf{f}_{ii} = \log \{ \mathbf{f}_{u:h}, (y_{ii}; \theta_{ii}, \phi) \}$ 

$$\mathbf{T}_{i} = \mathbf{V}_{i}^{-1/2} \mathbf{W}_{i}^{1/2} \mathbf{D}_{\mathbf{y}_{i} - \boldsymbol{\mu}_{i}} \qquad \qquad \mathbf{C}_{i} = \mathbf{1}_{i} \mathbf{f}_{i}^{\mathsf{T}} + \mathbf{I}_{n_{i}}$$

Also, the following proposition is a powerful tool for an accurate presentation of the later results.

**Proposition 1.** If  $\mathbf{a} = (a_1, a_2, \dots, a_{n_i})^{\top}$  is an  $n_i \times 1$  vector and  $\mathbf{\Omega}_i$  is a matrix as in definition 3, we have that

(a) 
$$(a^{\top} \otimes \mathbf{I}_{n_i}) \ \Omega_i = \mathbf{D}_a$$
, (b)  $(\mathbf{D}_a \otimes \mathbf{I}_{n_i}) \ \Omega_i = \Omega_i \ \mathbf{D}_a$ ,  
(c)  $(\mathbf{I}_{n_i} \otimes a^{\top}) \ \Omega_i = \mathbf{D}_a$  and (d)  $(\mathbf{I}_{n_i} \otimes \mathbf{D}_a) \ \Omega_i = \Omega_i \ \mathbf{D}_a$ .

Proof. (a) :  $(\mathbf{a}^{\top} \otimes \mathbf{I}_{n_{i}}) \mathbf{\Omega}_{i} = (a_{1} \mathbf{I}_{n_{i}} a_{2} \mathbf{I}_{n_{i}} \dots a_{n_{i}} \mathbf{I}_{n_{i}}) \oplus_{j=1}^{n_{i}} \mathbf{e}_{ij}$  $= (a_{1} \mathbf{I}_{n_{i}} a_{2} \mathbf{I}_{n_{i}} \dots a_{n_{i}} \mathbf{I}_{n_{i}}) \left(\mathbf{D}_{e_{i1}} \mathbf{D}_{e_{i2}} \dots \mathbf{D}_{e_{in_{i}}}\right)^{\top}$   $= \sum_{j=1}^{n_{i}} a_{j} \mathbf{D}_{e_{ij}} = \mathbf{D}_{a}.$ (b) :  $(\mathbf{D}_{a} \otimes \mathbf{I}_{n_{i}}) \mathbf{\Omega}_{i} = (\bigoplus_{j=1}^{n_{i}} a_{j} \mathbf{I}_{n_{i}}) (\bigoplus_{j=1}^{n_{i}} \mathbf{e}_{ij}) = \left\{ \left(\mathbf{D}_{e_{i1}} \mathbf{D}_{e_{i2}} \dots \mathbf{D}_{e_{in_{i}}}\right) \left(\bigoplus_{j=1}^{n_{i}} a_{j} \mathbf{I}_{n_{i}}\right) \right\}^{\top}$ 

$$= \left(a_1 \mathbf{D}_{e_{i1}} a_2 \mathbf{D}_{e_{i2}} \dots a_{n_i} \mathbf{D}_{e_{in_i}}\right)^\top$$
$$= \left\{\mathbf{D}_a \left(\mathbf{D}_{e_{i1}} \mathbf{D}_{e_{i2}} \dots \mathbf{D}_{e_{in_i}}\right)\right\}^\top = \mathbf{\Omega}_i \mathbf{D}_a.$$

(c) and (d) : Analogous.

### Some useful differentials

Following Magnus and Neudecker (1999), given a matrix function  $\mathbf{F} : \mathbf{A} \subseteq \mathbb{R}^q \longrightarrow \mathbb{R}^{r\times s}$ , the more suitable form of computing derivatives is through the differentials. Such differentials must be computed by means of general rules of differentiation widely exposed in chapter 13 of Abadir and Magnus (2005) and throughout Magnus and Neudecker (1999). Thus, these rules will be used extensively in what follows and the reader may refer to the books previously mentioned for a detailed review. Among the variety of existing techniques for the treatment with differentials, some of them are essential for obtaining derivatives from the differentials. For example, the identification and uniqueness theorems (Magnus and Neudecker, 1999, chapter 5) allow to associate to each differential one derivative. Other important tool is the Cauchy's rule of invariance (Magnus and Neudecker, 1999, chapter 5), which is the chain rule for differentials and enables to compute differentials of complex matrix functions effectively. In addition, relations between the vec function and the Kronecker product exposed in chapter 10 of Abadir and Magnus (2005) and properties of the duplica-

tion and commutation matrices described in chapter 11 of the same book will be used extensively later.

In what follows we present some important differentials that allow a more effective development of the computations made in later sections. The proposition 1 will be used repeatedly in each computation. In Table B.2 we summarize the results of this section.  $M_i$ :

$$d \operatorname{vec} (\mathbf{M}_{i}) = d \operatorname{vec} \left( \mathbf{D}_{\boldsymbol{y}_{i}-\boldsymbol{\mu}_{i}} \mathbf{D}_{\boldsymbol{\rho}_{i}} - \mathbf{W}_{i}^{1/2} \mathbf{V}_{i}^{-1/2} \mathbf{A}_{i} \right)$$

$$= \left\{ \left( \mathbf{D}_{\boldsymbol{\rho}_{i}} \otimes \mathbf{I}_{n_{i}} \right) \, \boldsymbol{\Omega}_{i} \, \left( -\mathbf{A}_{i} \right) + \left( \mathbf{I}_{n_{i}} \otimes \mathbf{D}_{\boldsymbol{y}_{i}-\boldsymbol{\mu}_{i}} \right) \, \boldsymbol{\Omega}_{i} \, \mathbf{D}_{\boldsymbol{\rho}_{i}} - \left( \mathbf{A}_{i} \otimes \mathbf{I}_{n_{i}} \right) \, \boldsymbol{\Omega}_{i} \, \mathbf{D}_{\boldsymbol{\rho}_{i}} - \left( \mathbf{I}_{n_{i}} \otimes \mathbf{W}_{i}^{1/2} \, \mathbf{V}_{i}^{-1/2} \right) \, \boldsymbol{\Omega}_{i} \, \dot{\mathbf{A}}_{i} \right\} \, \mathbf{Z}_{i} \, \mathrm{d} \, \boldsymbol{b}_{i}$$

$$= \boldsymbol{\Omega}_{i} \, \left( \mathbf{D}_{\boldsymbol{\rho}_{i}} \, \mathbf{A}_{i} + \mathbf{D}_{\boldsymbol{y}_{i}-\boldsymbol{\mu}_{i}} \, \mathbf{D}_{\boldsymbol{\rho}_{i}} - \mathbf{A}_{i} \, \mathbf{D}_{\boldsymbol{\rho}_{i}} - \mathbf{W}_{i}^{1/2} \, \mathbf{V}_{i}^{-1/2} \, \dot{\mathbf{A}}_{i} \right) \, \mathbf{Z}_{i} \, \mathrm{d} \, \boldsymbol{b}_{i} = \boldsymbol{\Omega}_{i} \, \left( \dot{\mathbf{M}}_{i} - 2 \, \mathbf{D}_{\boldsymbol{\rho}_{i}} \, \mathbf{A}_{i} \right) \, \mathbf{Z}_{i} \, \mathrm{d} \, \boldsymbol{b}_{i} \, ,$$

 $\dot{\mathbf{M}}_i$  :

$$\begin{split} \operatorname{d}\operatorname{vec}\left(\dot{\mathbf{M}}_{i}\right) &= \operatorname{d}\operatorname{vec}\left(\mathbf{D}_{\boldsymbol{y}_{i}-\boldsymbol{\mu}_{i}} \mathbf{D}_{\dot{\rho}_{i}} - \mathbf{W}_{i}^{1/2} \mathbf{V}_{i}^{-1/2} \dot{\mathbf{A}}_{i}\right) \\ &= \left\{\left(\mathbf{D}_{\dot{\rho}_{i}} \otimes \mathbf{I}_{n_{i}}\right) \, \boldsymbol{\Omega}_{i} \left(-\mathbf{A}_{i}\right) + \left(\mathbf{I}_{n_{i}} \otimes \mathbf{D}_{\boldsymbol{y}_{i}-\boldsymbol{\mu}_{i}}\right) \boldsymbol{\Omega}_{i} \mathbf{D}_{\dot{\rho}_{i}} - \left(\dot{\mathbf{A}}_{i} \otimes \mathbf{I}_{n_{i}}\right) \boldsymbol{\Omega}_{i} \mathbf{D}_{\rho_{i}} - \left(\mathbf{I}_{n_{i}} \otimes \mathbf{W}^{1/2} \mathbf{V}_{i}^{-1/2}\right) \boldsymbol{\Omega}_{i} \ddot{\mathbf{A}}_{i}\right\} \mathbf{Z}_{i} \operatorname{d} \boldsymbol{b}_{i} \\ &= \boldsymbol{\Omega}_{i} \left\{-\mathbf{D}_{\dot{\rho}_{i}} \mathbf{A}_{i} + \mathbf{D}_{\boldsymbol{y}_{i}-\boldsymbol{\mu}_{i}} \mathbf{D}_{\dot{\rho}_{i}} - \mathbf{D}_{\rho_{i}} \dot{\mathbf{A}}_{i} - \mathbf{W}_{i}^{1/2} \mathbf{V}_{i}^{-1/2} \, \ddot{\mathbf{A}}_{i}\right\} \mathbf{Z}_{i} \operatorname{d} \boldsymbol{b}_{i} \\ &= \boldsymbol{\Omega}_{i} \left(\ddot{\mathbf{M}}_{i} - \mathbf{D}_{\dot{\rho}_{i}} \mathbf{A}_{i} - \mathbf{D}_{\rho_{i}} \, \dot{\mathbf{A}}_{i}\right) \mathbf{Z}_{i} \operatorname{d} \boldsymbol{b}_{i} \,, \end{split}$$

 $\mathbf{X}_i^{\mathsf{T}} \, \boldsymbol{\tau}_i$  and  $\mathbf{Z}_i^{\mathsf{T}} \, \boldsymbol{\tau}_i$  :

d  $(\mathbf{Z}_i^{\top} \boldsymbol{\tau}_i) = \mathbf{Z}_i^{\top} \mathbf{M}_i \mathbf{Z}_i$  (Analogous) and

 $\lambda_i$  :

 $\mathbf{f}_i$  and  $\dot{\mathbf{f}}_i$ : Let  $\mathbf{f}_i = (\mathbf{f}_{i1}, \mathbf{f}_{i2}, \dots, \mathbf{f}_{in_i})^{\top}$ , where  $\mathbf{f}_{ij} = \phi \{y_{ij} \,\theta_{ij} - a \,(\theta_{ij})\} + c \,(y_{ij}; \phi)$ . Note that

$$\frac{\partial \mathbf{f}_i}{\partial \boldsymbol{b}_i^{\top}} = \left(\frac{\partial \mathbf{f}_{i1}}{\partial \boldsymbol{b}_i}, \dots, \frac{\partial \mathbf{f}_{in_i}}{\partial \boldsymbol{b}_i}\right)^{\top} \quad \text{and} \quad \frac{\partial \mathbf{f}_{ij}}{\partial \boldsymbol{b}_i^{\top}} = \phi \left(y_{ij} - \mu_{ij}\right) \frac{\mathrm{d}\,\theta_{ij}}{\mathrm{d}\,\mu_{ij}} \frac{\mathrm{d}\,\mu_{ij}}{\mathrm{d}\,\eta_{ij}} \boldsymbol{z}_{ij}^{\top}.$$

Then,

$$\frac{\partial \mathbf{f}_i}{\partial \boldsymbol{b}_i^{\mathsf{T}}} = \phi \, \mathbf{D}_{\boldsymbol{y}_i - \boldsymbol{\mu}_i} \, \mathbf{W}_i^{1/2} \, \mathbf{V}_i^{-1/2} \, \mathbf{Z}_i = \phi \, \mathbf{T}_i \, \mathbf{Z}_i \qquad \text{and} \qquad \mathrm{d} \, \dot{\mathbf{f}}_i = \mathbf{T}_i \, \mathbf{Z}_i \, \mathrm{d} \, \boldsymbol{b}_i.$$

Also, since  $\dot{\mathbf{f}}_i = (\dot{\mathbf{f}}_{i1}, \dots, \dot{\mathbf{f}}_{in_i})^{\mathsf{T}}$ , where  $\dot{\mathbf{f}}_{ij} = d\dot{\mathbf{f}}_{ij}/d\phi$ , then the procedure for computing  $d\dot{\mathbf{f}}_i$  is as above.

 $v_i$ : Note that  $v_i = \sum_{j=1}^{n_i} \{ y_{ij} \theta_{ij} - a \left( \theta_{ij} + c_{\phi} \left( y_{ij}; \phi \right) \right) \} = \mathbf{1}_{n_i}^{\mathsf{T}} \mathbf{f}_i$ . Hence, we have

 $d(v_i) = d(\mathbf{1}_{n_i}^{\top} \dot{\mathbf{f}}_i) = \mathbf{1}_{n_i}^{\top} d\dot{\mathbf{f}}_i = \mathbf{1}_{n_i}^{\top} \mathbf{T}_i \mathbf{Z}_i = \mathbf{1}_{n_i}^{\top} \mathbf{D}_{\mathbf{y}_i - \boldsymbol{\mu}_i} \mathbf{W}_i^{1/2} \mathbf{V}_i^{-1/2} \mathbf{Z}_i = (\mathbf{y}_i - \boldsymbol{\mu}_i)^{\top} \mathbf{W}_i^{1/2} \mathbf{V}_i^{-1/2} \mathbf{Z}_i = \boldsymbol{\tau}_i^{\top} \mathbf{Z}_i.$ 

 $\mathbf{X}_i^{ op} \mathbf{T}_i$  and  $\mathbf{Z}_i^{ op} \mathbf{T}_i$  :

$$d \operatorname{vec} \left( \mathbf{X}_{i}^{\top} \mathbf{T}_{i} \right) = \left( \mathbf{I}_{n_{i}} \otimes \mathbf{X}_{i}^{\top} \right) d \operatorname{vec} \left( \mathbf{T}_{i} \right)$$
$$= \left( \mathbf{I}_{n_{i}} \otimes \mathbf{X}_{i}^{\top} \right) \left\{ \left( \mathbf{D}_{\boldsymbol{y}_{i} - \boldsymbol{\mu}_{i}} \otimes \mathbf{I}_{n_{i}} \right) d \operatorname{vec} \left( \mathbf{V}_{i}^{-1/2} \mathbf{W}_{i}^{1/2} \right) + \left( \mathbf{I}_{n_{i}} \otimes \mathbf{V}_{i}^{-1/2} \mathbf{W}_{i}^{1/2} \right) d \operatorname{vec} \left( \mathbf{D}_{\boldsymbol{y}_{i} - \boldsymbol{\mu}_{i}} \right) \right\}$$
$$= \left( \mathbf{I}_{n_{i}} \otimes \mathbf{X}_{i}^{\top} \right) \mathbf{\Omega}_{i} \left( \mathbf{D}_{\boldsymbol{y}_{i} - \boldsymbol{\mu}_{i}} \mathbf{D}_{\boldsymbol{\rho}_{i}} - \mathbf{V}_{i}^{-1/2} \mathbf{W}_{i}^{1/2} \mathbf{A}_{i} \right) \mathbf{Z}_{i} d \mathbf{b}_{i} = \left( \mathbf{I}_{n_{i}} \otimes \mathbf{X}_{i}^{\top} \right) \mathbf{\Omega}_{i} \mathbf{M}_{i} \mathbf{Z}_{i} d \mathbf{b}_{i},$$

 $\operatorname{d}\operatorname{vec}\left(\mathbf{Z}_{i}^{\mathsf{T}}\mathbf{T}_{i}
ight)$  (Analogous) and

 $\mathbf{C}_i$  :

d vec ( $\mathbf{C}_i$ ) = d vec ( $\mathbf{1}_i \mathbf{f}_i^{\mathsf{T}} + \mathbf{I}_{n_i}$ ) =  $\phi$  ( $\mathbf{I}_{n_i} \otimes \mathbf{1}_i$ ) d  $\mathbf{f}_i = \phi$  ( $\mathbf{I}_{n_i} \otimes \mathbf{1}_{n_i}$ ) T<sub>i</sub> Z<sub>i</sub> d  $\mathbf{b}_i = \phi$  ( $\mathbf{T}_i \mathbf{Z}_i \otimes \mathbf{1}_{n_i}$ ) d  $\mathbf{b}_i$ A summary of these computations is presented in the table below.

Matrix function (F)	d vec (F)
$\mathbf{M}_i$	$\mathbf{\Omega}_i \left( \dot{\mathbf{M}}_i - 2  \mathbf{D}_{\boldsymbol{\rho}_i}  \mathbf{A}_i \right)  \mathbf{Z}_i  \mathrm{d}  \boldsymbol{b}_i$
$\dot{\mathbf{M}}_i$	$\boldsymbol{\Omega}_i \left( \ddot{\mathbf{M}}_i - \mathbf{D}_{\dot{\boldsymbol{\rho}}_i} \mathbf{A}_i - \mathbf{D}_{\boldsymbol{\rho}_i} \dot{\mathbf{A}}_i \right)  \mathbf{Z}_i  \mathrm{d}  \boldsymbol{b}_i$
$\mathbf{X}_i^{ op}  m{ au}_i$	$\mathbf{X}_i^{ op}  \mathbf{M}_i  \mathbf{Z}_i  \mathrm{d}  m{b}_i$
$\mathbf{Z}_i^{\top}  \boldsymbol{\tau}_i$	$\mathbf{Z}_i^{ op}  \mathbf{M}_i  \mathbf{Z}_i  \mathrm{d}  \boldsymbol{b}_i$
$oldsymbol{\lambda}_i$	$\mathbf{N}_q \; ig( \mathbf{I}_q \otimes oldsymbol{b}_i ig) \; \mathrm{d} \: oldsymbol{b}_i$
${\cal V}_i$	$oldsymbol{ au}_i^{ op}  oldsymbol{Z}_i  \mathrm{d}  oldsymbol{b}_i$
$\dot{\mathbf{f}}_i$	$\mathbf{T}_i \mathbf{Z}_i \mathrm{d} \boldsymbol{b}_i$
$\mathbf{f}_i$	$\phi  {f T}_i  {f Z}_i  {f d}  {f b}_i$
$\mathbf{X}_i^{ op} \mathbf{T}_i$	$\left(\mathbf{I}_{n_i}\otimes\mathbf{X}_i^{ op} ight)~\mathbf{\Omega}_i\mathbf{M}_i\mathbf{Z}_i\mathrm{d}m{b}_i$
$\mathbf{Z}_i^{ op} \mathbf{T}_i$	$\left(\mathbf{I}_{n_i}\otimes\mathbf{Z}_i^{ op} ight)\mathbf{\Omega}_i\mathbf{M}_i\mathbf{Z}_i\mathrm{d}m{b}_i$
<b>C</b> <sub><i>i</i></sub>	$\phi \ (\mathbf{T}_i \mathbf{Z}_i \otimes 1_{n_i}) \ \mathbf{d} \ \boldsymbol{b}_i.$

**Table B.2:** Some useful differentials.

Each differential presented in the previous table allows to obtain first order derivatives with respect to  $b_i$ . Details of this process may be revised in Magnus and Neudecker (1999). We make a brief summary in the next section.

### **Derivatives from differentials**

One of the advantages of working with differentials is that **F** and **dF** have the same dimension, and this allows more comfortable computations. To summarize, given **F** as above we may compute the differential d **F** and the first identification theorem (Magnus and Neudecker, 1999, chapter 5) ensures that, under accurate conditions,  $\operatorname{vec}(d\mathbf{F}) = \operatorname{d}\operatorname{vec}(\mathbf{F}) = \mathbf{C}(x) \operatorname{d} x$ , where  $\mathbf{C}(x) = \partial \operatorname{vec}(\mathbf{F}) / \partial x^{\mathsf{T}}$  and  $x \in A \subseteq \mathbb{R}^q$ . In our case, the expressions in question depend on  $b_i \in \mathbb{R}^q$  and therefore the differentiation procedures are performed on  $b_i \in \mathbb{R}^q$ . For example, the first row of Table B.2 indicates that

$$\operatorname{dvec}(\mathbf{M}_{i}) = \mathbf{\Omega}_{i} \left( \dot{\mathbf{M}}_{i} - 2 \mathbf{D}_{\boldsymbol{\rho}_{i}} \mathbf{A}_{i} \right) \mathbf{Z}_{i} \operatorname{d} \boldsymbol{b}_{i} \Longleftrightarrow \frac{\partial \operatorname{vec}(\mathbf{M}_{i})}{\partial \boldsymbol{b}_{i}^{\top}} = \mathbf{\Omega}_{i} \left( \dot{\mathbf{M}}_{i} - 2 \mathbf{D}_{\boldsymbol{\rho}_{i}} \mathbf{A}_{i} \right) \mathbf{Z}_{i}$$

Note that if **F** is a vector function (s = 1), vec (**F**) = **F** and therefore it is not necessary the vec (.) function for applying the identification theorem. For instance, from the third

row in Table B.2 we may obtain

$$d\left(\mathbf{X}_{i}^{\top}\boldsymbol{\tau}_{i}\right) = \mathbf{X}_{i}^{\top}\mathbf{M}_{i}\mathbf{Z}_{i} \Longleftrightarrow \frac{\partial \mathbf{X}_{i}^{\top}\boldsymbol{\tau}_{i}}{\partial \boldsymbol{b}_{i}^{\top}} = \mathbf{X}_{i}^{\top}\mathbf{M}_{i}\mathbf{Z}_{i}.$$

# **B.2** Computation of $\mathbf{L}_{b_i}(\boldsymbol{\psi}; \boldsymbol{y}_i, \boldsymbol{b}_i)$ and $\mathbf{L}_{b_i b_i}(\boldsymbol{\psi}; \boldsymbol{y}_i, \boldsymbol{b}_i)$

From (2.1) and (2.6)

$$L(\boldsymbol{\psi}; \boldsymbol{y}_{i}, \boldsymbol{b}_{i}) = -\frac{q}{2} \log(2\pi) - \frac{1}{2} \log |\boldsymbol{\Xi}| + \sum_{j=1}^{n_{i}} \log \left\{ f_{y_{ij}|\boldsymbol{b}_{i}}(y_{ij}; \theta_{ij}, \phi) \right\} - \frac{1}{2} \boldsymbol{b}_{i}^{\top} \boldsymbol{\Xi}^{-1} \boldsymbol{b}_{i}.$$

First, note that

$$d\left(\boldsymbol{b}_{i}^{\top}\boldsymbol{\Xi}^{-1}\boldsymbol{b}_{i}\right) = \left(d\,\boldsymbol{b}_{i}^{\top}\right)\,\boldsymbol{\Xi}^{-1}\,\boldsymbol{b}_{i} + \boldsymbol{b}_{i}^{\top}\,\boldsymbol{\Xi}\,d\,\boldsymbol{b}_{i} = 2\,\boldsymbol{b}_{i}^{\top}\,\boldsymbol{\Xi}^{-1}\,d\,\boldsymbol{b}_{i} \Longrightarrow \frac{\partial\left(\boldsymbol{b}_{i}^{\top}\,\boldsymbol{\Xi}^{-1}\,\boldsymbol{b}_{i}\right)}{\partial\,\boldsymbol{b}_{i}^{\top}} = 2\,\boldsymbol{b}_{i}^{\top}\,\boldsymbol{\Xi}^{-1}.$$

Then,

$$\begin{aligned} \mathbf{L}_{\boldsymbol{b}_{i}}\left(\boldsymbol{\psi};\boldsymbol{y}_{i},\boldsymbol{b}_{i}\right) &= \frac{\partial \mathbf{L}\left(\boldsymbol{\psi};\boldsymbol{y}_{i},\boldsymbol{b}_{i}\right)}{\partial \boldsymbol{b}_{i}} = \phi \sum_{j=1}^{n_{i}} \left[ \left\{ y_{ij} - \frac{\mathrm{d}\,a\left(\theta_{ij}\right)}{\mathrm{d}\,\theta_{ij}} \right\} \frac{\mathrm{d}\theta_{ij}}{\mathrm{d}\,\mu_{ij}} \frac{\mathrm{d}\mu_{ij}}{\mathrm{d}\,\eta_{ij}} \frac{\mathrm{d}\eta_{ij}}{\mathrm{d}\,\boldsymbol{b}_{i}} \right] - \boldsymbol{\Xi}^{-1} \boldsymbol{b}_{i} \\ &= \phi \sum_{j=1}^{n_{i}} \left[ \left\{ y_{ij} - \mu_{ij} \right\} \frac{\mathrm{d}\theta_{ij}}{\mathrm{d}\,\mu_{ij}} \frac{\mathrm{d}\mu_{ij}}{\mathrm{d}\,\eta_{ij}} z_{ij} \right] - \boldsymbol{\Xi}^{-1} \boldsymbol{b}_{i} = \phi \mathbf{Z}_{i}^{\top} \mathbf{W}_{i}^{1/2} \mathbf{V}_{i}^{-1/2} \left( \boldsymbol{y}_{i} - \boldsymbol{\mu}_{i} \right) - \boldsymbol{\Xi}^{-1} \boldsymbol{b}_{i} \\ &= \phi \mathbf{Z}_{i}^{\top} \boldsymbol{\tau}_{i} - \boldsymbol{\Xi}^{-1} \boldsymbol{b}_{i} \end{aligned}$$

and

$$\mathbf{L}_{\boldsymbol{b}_{i}\boldsymbol{b}_{i}}\left(\boldsymbol{\psi};\boldsymbol{y}_{i},\boldsymbol{b}_{i}\right) = \frac{\partial}{\partial\boldsymbol{b}_{i}^{\top}}\left(\phi\,\mathbf{Z}_{i}^{\top}\,\boldsymbol{\tau}_{i}-\boldsymbol{\Xi}^{-1}\,\boldsymbol{b}_{i}\right) = \phi\,\mathbf{Z}_{i}^{\top}\,\frac{\partial\boldsymbol{\tau}_{i}}{\partial\boldsymbol{b}_{i}^{\top}} - \boldsymbol{\Xi}^{-1} = \phi\,\mathbf{Z}_{i}^{\top}\,\mathbf{M}_{i}\,\mathbf{Z}_{i}-\boldsymbol{\Xi}^{-1}.$$

Also, from (3.12), in the canonical exponential family we have that

$$\mathbf{L}_{\boldsymbol{b}_{i}}\left(\boldsymbol{\psi};\boldsymbol{y}_{i},\boldsymbol{b}_{i}\right)=\phi\,\mathbf{Z}_{i}^{\top}\left(\boldsymbol{y}_{i}-\boldsymbol{\mu}_{i}\right)-\boldsymbol{\Xi}^{-1}\,\boldsymbol{b}_{i}\quad\text{and}\quad\mathbf{L}_{\boldsymbol{b}_{i}\boldsymbol{b}_{i}}\left(\boldsymbol{\psi};\boldsymbol{y}_{i},\boldsymbol{b}_{i}\right)=-\phi\,\mathbf{Z}_{i}^{\top}\,\mathbf{W}_{i}\,\mathbf{Z}_{i}-\boldsymbol{\Xi}^{-1}.$$

Expected value of  $E_{y_i|b_i} \left\{ - L_{b_i b_i} \left( \widehat{\psi}; y_i, b_i \right) \right\}$ 

Note that

$$\mathbf{E}_{\boldsymbol{y}_i|\boldsymbol{b}_i}\left(\mathbf{D}_{\boldsymbol{y}_i-\boldsymbol{\mu}_i} \mid \boldsymbol{b}_i\right) = \mathbf{E}_{\boldsymbol{y}_i|\boldsymbol{b}_i}\left(\mathbf{D}_{\boldsymbol{y}_i}-\mathbf{D}_{\boldsymbol{\mu}_i} \mid \boldsymbol{b}_i\right) = \mathbf{E}_{\boldsymbol{y}_i|\boldsymbol{b}_i}\left(\mathbf{D}_{\boldsymbol{y}_i} \mid \boldsymbol{b}_i\right) - \mathbf{D}_{\boldsymbol{\mu}_i} = \mathbf{0}_{n_i}$$

Thus, since  $\mathbf{M}_i = \mathbf{D}_{y_i - \mu_i} \mathbf{D}_{\rho_i} - \mathbf{W}_i^{1/2} \mathbf{V}_i^{-1/2} \mathbf{A}_i = \mathbf{D}_{y_i - \mu_i} \mathbf{D}_{\rho_i} - \mathbf{W}_i$ , we have that

$$\mathbf{E}_{\boldsymbol{y}_i|\boldsymbol{b}_i}\left\{-\mathbf{L}_{\boldsymbol{b}_i\boldsymbol{b}_i}\left(\widehat{\boldsymbol{\psi}};\boldsymbol{y}_i,\boldsymbol{b}_i\right)|\boldsymbol{b}_i\right\}=\mathbf{E}_{\boldsymbol{y}_i|\boldsymbol{b}_i}\left(-\phi\,\mathbf{Z}_i^{\top}\,\mathbf{M}_i\,\mathbf{Z}_i+\boldsymbol{\Xi}^{-1}\,|\,\boldsymbol{b}_i\right)=\phi\,\mathbf{Z}_i^{\top}\,\mathbf{W}_i\,\mathbf{Z}_i+\boldsymbol{\Xi}^{-1}$$

This matrix is positive definite because  $\mathbf{W}_i$  is a diagonal matrix with positive elements (Remember that a diagonal matrix is positive definite if and only if its elements are positive). Thus,  $\mathbf{Z}_i^{\top} \mathbf{W}_i \mathbf{Z}_i$  is semidefinite positive. Also  $\Xi^{-1}$  is positive definite because  $\Xi$  is definite positive, therefore  $\mathbf{Z}_i^{\top} \mathbf{W}_i \mathbf{Z}_i + \Xi^{-1}$  is positive definite. The properties used in this argument could be reviewed in chapter 8 of Abadir and Magnus (2005).

# **B.3** Computation of the score function $L_{\psi}(\psi; y)$

Since  $L(\boldsymbol{\psi}; \boldsymbol{y}) = \sum_{i=1}^{N} L(\boldsymbol{\psi}; \boldsymbol{y}_i)$ , we perform the calculations on  $L(\boldsymbol{\psi}; \boldsymbol{y}_i)$ . Hence, we will compute  $L_{\boldsymbol{\psi}}^i = \partial L(\boldsymbol{\psi}; \boldsymbol{y}_i) / \partial \boldsymbol{\psi}$ , for i = 1, ..., N.

For convenience in the computations, let us consider  $\mathcal{L}_i$  as the joint log-likelihood function of  $\boldsymbol{y}_i$  and  $\boldsymbol{b}_i$  for i = 1, ..., N. That is,  $\mathcal{L}_i = \log \{f_{\boldsymbol{y}_i, \boldsymbol{b}_i}(\boldsymbol{y}_i, \boldsymbol{b}_i; \boldsymbol{\psi})\} = L(\boldsymbol{\psi}; \boldsymbol{y}_i, \boldsymbol{b}_i)$ . Thus, from (2.7) and (2.6), we have that

$$\begin{split} \mathbf{L}_{\boldsymbol{\beta}}^{i} &= \frac{\partial \operatorname{L}\left(\boldsymbol{\psi}; \boldsymbol{y}_{i}\right)}{\partial \boldsymbol{\beta}} = \frac{\partial}{\partial \boldsymbol{\beta}} \log \left\{ \int_{\mathbb{R}^{q}} \exp\left(\mathcal{L}_{i}\right) \, \mathrm{d} \, \boldsymbol{b}_{i} \right\} = \frac{1}{\int_{\mathbb{R}^{q}} \exp\left(\mathcal{L}_{i}\right) \, \mathrm{d} \, \boldsymbol{b}_{i}} \int_{\mathbb{R}^{q}} \frac{\partial}{\partial \boldsymbol{\beta}} \exp\left(\mathcal{L}_{i}\right) \, \mathrm{d} \, \boldsymbol{b}_{i} \\ &= \frac{1}{\int_{\mathbb{R}^{q}} \exp\left(\mathcal{L}_{i}\right) \, \mathrm{d} \, \boldsymbol{b}_{i}} \int_{\mathbb{R}^{q}} \frac{\partial \mathcal{L}_{i}}{\partial \boldsymbol{\beta}} \exp\left(\mathcal{L}_{i}\right) \, \mathrm{d} \, \boldsymbol{b}_{i} \\ &= \frac{\phi}{\int_{\mathbb{R}^{q}} \exp\left(\mathcal{L}_{i}\right) \, \mathrm{d} \, \boldsymbol{b}_{i}} \int_{\mathbb{R}^{q}} \mathbf{X}_{i}^{\top} \, \boldsymbol{\tau}_{i} \exp\left(\mathcal{L}_{i}\right) \, \mathrm{d} \, \boldsymbol{b}_{i} = \phi \operatorname{E}\left(\mathbf{X}_{i}^{\top} \, \boldsymbol{\tau}_{i} \middle| \mathbf{y}_{i}\right). \end{split}$$

Note that, in the previous procedure, we use the fact that

$$L(\boldsymbol{\psi};\boldsymbol{y}_{i}) = \log \left\{ f_{\boldsymbol{y}_{i}}(\boldsymbol{y}_{i};\boldsymbol{\psi}) \right\} = \log \left\{ \int_{\mathbb{R}^{q}} f_{\boldsymbol{y}_{i},\boldsymbol{b}_{i}}(\boldsymbol{y}_{i},\boldsymbol{b}_{i};\boldsymbol{\psi}) d\boldsymbol{b}_{i} \right\} = \log \left[ \int_{\mathbb{R}^{q}} \exp \left\{ L(\boldsymbol{\psi};\boldsymbol{y}_{i},\boldsymbol{b}_{i}) \right\} \right].$$

In addition, we have used the following result:

$$\begin{aligned} \frac{\partial \mathcal{L}_{i}}{\partial \boldsymbol{\beta}} &= \frac{\partial \operatorname{L}\left(\boldsymbol{\psi}; \boldsymbol{y}_{i}, \boldsymbol{b}_{i}\right)}{\partial \boldsymbol{\beta}} = \frac{\partial}{\partial \boldsymbol{\beta}} \left( \sum_{j=1}^{n_{i}} \left[ \phi \left\{ y_{ij} \theta_{ij} - a \left( \theta_{ij} \right) \right\} + c \left( y_{ij}; \phi \right) \right] \right) \\ &= \sum_{j=1}^{n_{i}} \left\{ \phi \left( y_{ij} - \mu_{ij} \right) \frac{\mathrm{d} \, \theta_{ij}}{\mathrm{d} \, \mu_{ij}} \frac{\mathrm{d} \, \mu_{ij}}{\mathrm{d} \, \eta_{ij}} \frac{\mathrm{d} \, \eta_{ij}}{\mathrm{d} \, \beta} \right\} = \sum_{j=1}^{n_{i}} \left\{ \phi \left( y_{ij} - \mu_{ij} \right) \frac{\mathrm{d} \, \theta_{ij}}{\mathrm{d} \, \mu_{ij}} \frac{\mathrm{d} \, \mu_{ij}}{\mathrm{d} \, \eta_{ij}} \mathbf{x}_{ij} \right\} \\ &= \phi \, \mathbf{X}_{i}^{\mathsf{T}} \, \mathbf{W}_{i}^{1/2} \, \mathbf{V}_{i}^{-1/2} \, \left( \mathbf{y}_{i} - \boldsymbol{\mu}_{i} \right) = \phi \, \mathbf{X}_{i}^{\mathsf{T}} \, \boldsymbol{\tau}_{i} \,, \end{aligned}$$

where  $W_i$  is the weight matrix,  $V_i$  is the matrix associated with the variance function and  $\tau_i = \mathbf{W}_i^{1/2} \mathbf{V}_i^{-1/2} (\mathbf{y}_i - \boldsymbol{\mu}_i)$ . Also,  $\mathbf{L}_{\xi}^i = \partial \mathbf{L} (\boldsymbol{\psi}; \mathbf{y}_i) / \partial \xi$  and  $\mathbf{L}_{\phi}^i = \partial \mathbf{L} (\boldsymbol{\psi}; \mathbf{y}_i) / \partial \phi$  in (3.3) may be computed in a

similar way to the procedure exposed previously. It is enough take into account that

$$d\mathcal{L}_{i} = d\left\{L\left(\psi, y_{i}, b_{i}\right)\right\} = d\left(-\frac{1}{2}\log|\Xi| - \frac{1}{2}b_{i}^{\top}\Xi^{-1}b_{i}\right) = -\frac{1}{2}\mathrm{tr}\left(\Xi^{-1}\,\mathrm{d}\,\Xi\right) + \frac{1}{2}b_{i}^{\top}\Xi^{-1}\,\mathrm{d}\,\Xi\Xi^{-1}\,b_{i}$$
$$= -\frac{1}{2}\operatorname{vec}\left(\Xi^{-1}\right)^{\top}\,\mathrm{d}\operatorname{vec}\left(\Xi\right) + \frac{1}{2}\operatorname{tr}\left(\Xi^{-1}\,b_{i}\,b_{i}^{\top}\Xi\,\mathrm{d}\,\Xi\right)$$
$$= -\frac{1}{2}\operatorname{vec}\left(\Xi^{-1}\right)^{\top}\,\mathrm{d}\operatorname{vec}\left(\Xi\right) + \frac{1}{2}\operatorname{vec}\left(\Xi^{-1}\,b_{i}\,b_{i}^{\top}\Xi\right)^{\top}\,\mathrm{d}\operatorname{vec}\left(\Xi\right)$$
$$= -\frac{1}{2}\left\{\operatorname{vec}\left(\Xi^{-1}\right)^{\top} - \left(b_{i}^{\top}\otimes b_{i}^{\top}\right)\left(\Xi\otimes\Xi\right)^{-1}\right\}G_{q}\,\mathrm{d}\,\xi\,,$$

$$\implies \frac{\partial \mathcal{L}_i}{\partial \xi} = \frac{\partial L\left(\psi; y_i, b_i\right)}{\partial \xi} = -\frac{1}{2} \mathbf{G}_q^{\top} \left\{ \operatorname{vec}\left(\boldsymbol{\Xi}^{-1}\right) - \left(\boldsymbol{\Xi} \otimes \boldsymbol{\Xi}\right)^{-1} \lambda_i \right\} .$$
(B.1)

Finally, from (2.6), note that

$$\frac{\partial \mathcal{L}_{i}}{\partial \phi} = \frac{\partial \operatorname{L}\left(\boldsymbol{\psi}; \boldsymbol{y}_{i}, \boldsymbol{b}_{i}\right)}{\partial \phi} = \sum_{j=1}^{n_{i}} \left\{ y_{ij} \, \theta_{ij} - a\left(\theta_{ij}\right) + \frac{\operatorname{d} c\left(y_{ij}; \phi\right)}{\operatorname{d} \phi} \right\} = \nu_{i}.$$

#### Computation of the Hessian matrix $L_{\psi\psi}(\psi; y)$ **B.4**

Again, since  $L(\boldsymbol{\psi}; \boldsymbol{y}) = \sum_{i=1}^{N} L(\boldsymbol{\psi}; \boldsymbol{y}_i)$ , we perform the procedures on  $L(\boldsymbol{\psi}; \boldsymbol{y}_i)$  for  $i = 1, \ldots, N$ .

In order to compute  $\mathbf{L}_{\psi\psi}^{i} = \frac{\partial^{2} \mathbf{L}(\psi; \mathbf{y}_{i})}{\partial \psi \partial \psi^{\top}}$  in (3.2), we will compute each of its components  $\mathbf{L}_{\beta\beta}^{i}$ ,  $\mathbf{L}_{\xi\xi}^{i}$ ,  $\mathbf{L}_{\phi\phi}^{i}$ ,  $\mathbf{L}_{\beta\xi}^{i}$ ,  $\mathbf{L}_{\beta\phi}^{i}$ ,  $\mathbf{L}_{\beta\phi}^{i}$  and we will obtain the results exposed in (3.4). Thus, using some results of Table B.2, we obtain

$$\begin{split} \mathbf{L}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{i} &: \ \mathbf{L}_{\boldsymbol{\beta}\boldsymbol{\beta}}^{i} = \frac{\partial \mathbf{L}_{\boldsymbol{\beta}}^{i}}{\partial \boldsymbol{\beta}^{\mathsf{T}}} = \phi \frac{\partial}{\partial \boldsymbol{\beta}^{\mathsf{T}}} \left\{ \int_{\mathbb{R}^{q}} \mathbf{X}_{i}^{\mathsf{T}} \tau_{i} \exp\left(\mathcal{L}_{i}\right) d \boldsymbol{b}_{i} \left( \int_{\mathbb{R}^{q}} \exp\left(\mathcal{L}_{i}\right) d \boldsymbol{b}_{i} \right)^{-1} \right\} \\ &= \int_{\mathbb{R}^{q}} \left( \frac{\partial \mathbf{X}_{i}^{\mathsf{T}} \tau_{i}}{\partial \boldsymbol{\beta}^{\mathsf{T}}} + \mathbf{X}_{i}^{\mathsf{T}} \tau_{i} \frac{\partial \mathcal{L}_{i}}{\partial \boldsymbol{\beta}^{\mathsf{T}}} \right) \exp\left(\mathcal{L}_{i}\right) d \boldsymbol{b}_{i} \times \left( \int_{\mathbb{R}^{q}} \exp\left(\mathcal{L}_{i}\right) d \boldsymbol{b}_{i} \right)^{-1} \\ &- \phi \int_{\mathbb{R}^{q}} \mathbf{X}_{i}^{\mathsf{T}} \tau_{i} \exp\left(\mathcal{L}_{i}\right) d \boldsymbol{b}_{i} \int_{\mathbb{R}^{q}} \frac{\partial \mathcal{L}_{i}}{\partial \boldsymbol{\beta}^{\mathsf{T}}} \exp\left(\mathcal{L}_{i}\right) d \boldsymbol{b}_{i} \left( \int_{\mathbb{R}^{q}} \exp\left(\mathcal{L}_{i}\right) d \boldsymbol{b}_{i} \right)^{-2} \\ &= \phi \int_{\mathbb{R}^{q}} \left( \mathbf{X}_{i}^{\mathsf{T}} \mathbf{M}_{i} \mathbf{X}_{i} + \phi \mathbf{X}_{i}^{\mathsf{T}} \tau_{i} \tau_{i}^{\mathsf{T}} \mathbf{X}_{i} \right) \exp\left(\mathcal{L}_{i}\right) d \boldsymbol{b}_{i} \left( \int_{\mathbb{R}^{q}} \exp\left(\mathcal{L}_{i}\right) d \boldsymbol{b}_{i} \right)^{-1} \\ &- \phi^{2} \int_{\mathbb{R}^{q}} \mathbf{X}_{i}^{\mathsf{T}} \tau_{i} \exp\left(\mathcal{L}_{i}\right) d \boldsymbol{b}_{i} \int_{\mathbb{R}^{q}} \tau_{i}^{\mathsf{T}} \mathbf{X}_{i} \exp\left(\mathcal{L}_{i}\right) d \boldsymbol{b}_{i} \left( \int_{\mathbb{R}^{q}} \exp\left(\mathcal{L}_{i}\right) d \boldsymbol{b}_{i} \right)^{-2} \\ &= \phi \operatorname{E} \left( \mathbf{X}_{i}^{\mathsf{T}} \mathbf{M}_{i} \mathbf{X}_{i} | \mathbf{y}_{i} \right) + \phi^{2} \operatorname{E} \left\{ \mathbf{X}_{i}^{\mathsf{T}} \tau_{i} \left( \mathbf{X}_{i}^{\mathsf{T}} \tau_{i} \right)^{\mathsf{T}} \right| \mathbf{y}_{i}^{\mathsf{T}} - \phi^{2} \operatorname{E} \left( \mathbf{X}_{i}^{\mathsf{T}} \tau_{i} | \mathbf{y}_{i} \right) \operatorname{E} \left( \mathbf{X}_{i}^{\mathsf{T}} \tau_{i} | \mathbf{y}_{i} \right)^{\mathsf{T}} \\ &= \phi \operatorname{E} \left( \mathbf{X}_{i}^{\mathsf{T}} \mathbf{M}_{i} \mathbf{X}_{i} | \mathbf{y}_{i} \right) + \phi^{2} \operatorname{Var} \left( \mathbf{X}_{i}^{\mathsf{T}} \tau_{i} | \mathbf{y}_{i} \right) . \end{split}$$

 $\mathbf{L}_{\xi\xi}^{i}$  : First, note that

$$\begin{aligned} \frac{\partial}{\partial \xi^{\mathrm{T}}} \left\{ \mathrm{E} \left( \lambda_{i} | y_{i} \right) \right\} &= \frac{\partial}{\partial \xi^{\mathrm{T}}} \left\{ \int_{\mathbb{R}^{q}} \lambda_{i} \exp \left( \mathcal{L}_{i} \right) \mathrm{d} b_{i} \left( \int_{\mathbb{R}^{q}} \exp \left( \mathcal{L}_{i} \right) \mathrm{d} b_{i} \right)^{-1} \right\} \\ &= \int_{\mathbb{R}^{q}} \lambda_{i} \frac{\partial \mathcal{L}_{i}}{\partial \xi^{\mathrm{T}}} \exp \left( \mathcal{L}_{i} \right) \mathrm{d} b_{i} \left( \int_{\mathbb{R}^{q}} \exp \left( \mathcal{L}_{i} \right) \mathrm{d} b_{i} \right)^{-1} \\ &- \int_{\mathbb{R}^{q}} \lambda_{i} \exp \left( \mathcal{L}_{i} \right) \mathrm{d} b_{i} \int_{\mathbb{R}^{q}} \frac{\partial \mathcal{L}_{i}}{\partial \xi^{\mathrm{T}}} \exp \left( \mathcal{L}_{i} \right) \mathrm{d} b_{i} \left( \int_{\mathbb{R}^{q}} \exp \left( \mathcal{L}_{i} \right) \mathrm{d} b_{i} \right)^{-2} \\ &= \mathrm{E} \left( \lambda_{i} \frac{\partial \mathcal{L}_{i}}{\partial \xi^{\mathrm{T}}} \left| y_{i} \right) - \mathrm{E} \left( \lambda_{i} | y_{i} \right) \mathrm{E} \left( \frac{\partial \mathcal{L}_{i}}{\partial \xi} \right| y_{i} \right)^{\mathrm{T}} = \mathrm{Cov} \left( \lambda_{i}, \frac{\partial \mathcal{L}_{i}}{\partial \xi} \right| y_{i} \right) \\ &= \mathrm{Cov} \left[ \lambda_{i}, -\frac{1}{2} \mathbf{G}_{q}^{\mathrm{T}} \left\{ \operatorname{vec} \left( \Xi^{-1} \right) - \left( \Xi^{-1} \otimes \Xi^{-1} \right) \lambda_{i} \right\} \right| y_{i} \right] = \mathrm{Cov} \left\{ \lambda_{i}, \frac{1}{2} \mathbf{G}_{q}^{\mathrm{T}} \left( \Xi^{-1} \otimes \Xi^{-1} \right) \lambda_{i} \right| y_{i} \right\} \\ &= \frac{1}{2} \mathrm{Var} \left( \lambda_{i} | y_{i} \right) (\Xi \otimes \Xi)^{-1} \mathbf{G}_{q}. \end{aligned}$$
(B.2)

Now 
$$\mathbf{L}_{\xi\xi}^{i} = \frac{\partial \mathbf{L}_{\xi}^{i}}{\partial \xi^{\top}} = \frac{\partial}{\partial \xi^{\top}} \left[ -\frac{1}{2} \mathbf{G}_{q}^{\top} \left\{ \operatorname{vec} \left( \boldsymbol{\Xi}^{-1} \right) - \left( \boldsymbol{\Xi} \otimes \boldsymbol{\Xi} \right)^{-1} \operatorname{E} \left( \boldsymbol{\lambda}_{i} | \boldsymbol{y}_{i} \right) \right\} \right].$$

Thus, we may compute

$$d\left[-\frac{1}{2}\mathbf{G}_{q}^{\top}\left\{\operatorname{vec}\left(\boldsymbol{\Xi}^{-1}\right)-\left(\boldsymbol{\Xi}\otimes\boldsymbol{\Xi}\right)^{-1}\mathrm{E}\left(\boldsymbol{\lambda}_{i}|\boldsymbol{y}_{i}\right)\right\}\right]=\frac{1}{2}\mathbf{G}_{q}^{\top}\left[-\operatorname{d}\operatorname{vec}\left(\boldsymbol{\Xi}^{-1}\right)+\operatorname{d}\left\{\left(\boldsymbol{\Xi}\otimes\boldsymbol{\Xi}\right)^{-1}\mathrm{E}\left(\boldsymbol{\lambda}_{i}|\boldsymbol{y}_{i}\right)\right\}\right].$$

Then,

(a) d 
$$\left\{ \operatorname{vec}\left(\Xi^{-1}\right) \right\} = -\operatorname{vec}\left\{\Xi^{-1} \left( \operatorname{d}\Xi \right) \Xi^{-1} \right\} = -\left(\Xi \otimes \Xi\right)^{-1} \mathbf{G}_{q} \operatorname{d} \xi$$
 and

(b) d {
$$(\Xi \otimes \Xi)^{-1} \operatorname{E} (\lambda_i | y_i)$$
} = { $\operatorname{E} (\lambda_i | y_i)^{\top} \otimes \mathbf{I}_{q^2}$ } d vec ( $\Xi^{-1} \otimes \Xi^{-1}$ ) + ( $\Xi^{-1} \otimes \Xi^{-1}$ ) d E ( $\lambda_i | y_i$ )  
= { $\operatorname{E} (\lambda_i | y_i)^{\top} \otimes \mathbf{I}_{q^2}$ } d vec ( $\Xi^{-1} \otimes \Xi^{-1}$ ) + ( $\Xi^{-1} \otimes \Xi^{-1}$ )  $\left[\frac{\partial}{\partial \operatorname{vec}}(\Xi)^{\top}$  {E ( $\lambda_i | y_i$ )} d vec ( $\Xi$ )  
= {E ( $\lambda_i | y_i$ ) <sup>$\top$</sup>   $\otimes \mathbf{I}_{q^2}$ } P d vec ( $\Xi^{-1}$ ) + ( $\Xi^{-1} \otimes \Xi^{-1}$ )  $\left[\frac{\partial}{\partial \operatorname{vec}}(\Xi)^{\top}$  {E ( $\lambda_i | y_i$ )} G<sub>q</sub> d  $\xi$   
= { $\left[ \left( \lambda_i | y_i \right)^{\top} \otimes \mathbf{I}_{q^2} \right]$  P ( $\Xi^{-1} \otimes \Xi^{-1}$ ) G<sub>q</sub> + ( $\Xi^{-1} \otimes \Xi^{-1}$ )  $\left[\frac{\partial}{\partial \xi^{\top}}$  {E ( $\lambda_i | y_i$ )} d  $\xi$ ,

where  $\mathbf{P} = \{\mathbf{I}_q \otimes \mathbf{\Pi}_q (\mathbf{\Xi}^{-1})\} + \{\mathbf{\Lambda}_q (\mathbf{\Xi}^{-1}) \otimes \mathbf{I}_q\}$ . Thus, from (*a*) and (*b*) we have that

$$\mathbf{L}_{\xi\xi}^{i} = \frac{1}{2} \mathbf{G}_{q}^{\top} \left( \mathbf{I}_{q^{2}} - \left\{ \mathbf{E} \left( \boldsymbol{\lambda}_{i} | \boldsymbol{y}_{i} \right) \otimes \mathbf{I}_{q^{2}} \right\} \mathbf{P} \right) \left( \boldsymbol{\Xi} \otimes \boldsymbol{\Xi} \right)^{-1} \mathbf{G}_{q} + \frac{1}{2} \mathbf{G}_{q}^{\top} \left( \boldsymbol{\Xi} \otimes \boldsymbol{\Xi} \right)^{-1} \left[ \frac{\partial}{\partial \xi^{\top}} \left\{ \mathbf{E} \left( \boldsymbol{\lambda}_{i} | \boldsymbol{y}_{i} \right) \right\} \right].$$

The result in (3.4) is obtained replacing (B.2) in the expression above. In addition, we have

$$\begin{split} \mathbf{L}_{\phi\phi}^{i} &: \ \mathbf{L}_{\phi\phi}^{i} = \frac{\partial \mathbf{L}_{\phi}^{i}}{\partial \phi} = \frac{\partial}{\partial \phi} \left\{ \mathbf{E} \left( v_{i} | \boldsymbol{y}_{i} \right) \right\} = \frac{\partial}{\partial \phi} \left\{ \int_{\mathbb{R}^{q}} v_{i} \exp \left( \mathcal{L}_{i} \right) \, \mathrm{d} \, \boldsymbol{b}_{i} \left( \int_{\mathbb{R}^{q}} \exp \left( \mathcal{L} \right) \, \mathrm{d} \, \boldsymbol{b}_{i} \right)^{-1} \right\} \\ &= \int_{\mathbb{R}^{q}} \left( \frac{\mathrm{d} \, v_{i}}{\mathrm{d} \, \phi} + v_{i} \frac{\mathrm{d} \, \mathcal{L}_{i}}{\mathrm{d} \, \phi} \right) \exp \left( \mathcal{L}_{i} \right) \, \mathrm{d} \, \boldsymbol{b}_{i} \left( \int_{\mathbb{R}^{q}} \exp \left( \mathcal{L}_{i} \right) \, \mathrm{d} \, \boldsymbol{b}_{i} \right)^{-1} \\ &- \int_{\mathbb{R}^{q}} v_{i} \exp \left( \mathcal{L}_{i} \right) \, \mathrm{d} \, \boldsymbol{b}_{i} \int_{\mathbb{R}^{q}} \frac{\mathrm{d} \, \mathcal{L}_{i}}{\mathrm{d} \, \phi} \exp \left( \mathcal{L}_{i} \right) \, \mathrm{d} \, \boldsymbol{b}_{i} \left( \int_{\mathbb{R}^{q}} \exp \left( \mathcal{L}_{i} \right) \, \mathrm{d} \, \boldsymbol{b}_{i} \right)^{-2} \\ &= \frac{\mathrm{d} \, v_{i}}{\mathrm{d} \, \phi} + \mathrm{E} \left( v_{i} \frac{\mathrm{d} \, \mathcal{L}_{i}}{\mathrm{d} \, \phi} \middle| \, \boldsymbol{y}_{i} \right) - \mathrm{E} \left( v_{i} | \, \boldsymbol{y}_{i} \right) \mathrm{E} \left( \frac{\mathrm{d} \, \mathcal{L}_{i}}{\mathrm{d} \, \phi} \middle| \, \boldsymbol{y}_{i} \right) = \sum_{j=1}^{n_{i}} \frac{\mathrm{d}^{2} c \left( y_{ij}; \phi \right)}{\mathrm{d} \, \phi^{2}} + \mathrm{Cov} \left( v_{i}, \frac{\partial \mathcal{L}_{i}}{\partial \phi} \middle| \, \boldsymbol{y}_{i} \right) \\ &= \sum_{j=1}^{n_{i}} \frac{\mathrm{d}^{2} c \left( y_{ij}; \phi \right)}{\mathrm{d} \, \phi^{2}} + \mathrm{Var} \left( v_{i} \middle| \, \boldsymbol{y}_{i} \right) \,. \end{split}$$

$$\begin{split} \mathbf{L}_{\beta\xi}^{i} &: \mathbf{L}_{\beta\xi}^{i} = \frac{\partial}{\partial \xi^{\top}} \left\{ \phi \operatorname{E} \left( \mathbf{X}_{i}^{\top} \tau_{i} | y_{i} \right) \right\} \\ &= \phi \frac{\partial}{\partial \xi^{\top}} \left\{ \int_{\mathbb{R}^{q}} \mathbf{X}_{i}^{\top} \tau_{i} \exp \left( \mathcal{L}_{i} \right) \mathrm{d} b_{i} \left( \int_{\mathbb{R}^{q}} \exp \left( \mathcal{L}_{i} \right) \mathrm{d} b_{i} \right)^{-1} \right\} \\ &= \phi \int_{\mathbb{R}^{q}} \mathbf{X}_{i}^{\top} \tau_{i} \frac{\partial \mathcal{L}_{i}}{\partial \xi^{\top}} \exp \left( \mathcal{L}_{i} \right) \mathrm{d} b_{i} \left( \int_{\mathbb{R}^{q}} \exp \left( \mathcal{L}_{i} \right) \mathrm{d} b_{i} \right)^{-1} \\ &- \phi \int_{\mathbb{R}^{q}} \mathbf{X}_{i}^{\top} \tau_{i} \exp \left( \mathcal{L}_{i} \right) \mathrm{d} b_{i} \int_{\mathbb{R}^{q}} \frac{\partial \mathcal{L}_{i}}{\partial \xi^{\top}} \exp \left( \mathcal{L}_{i} \right) \mathrm{d} b_{i} \left( \int_{\mathbb{R}^{q}} \exp \left( \mathcal{L}_{i} \right) \mathrm{d} b_{i} \right)^{-2} \\ &= \phi \operatorname{E} \left( \mathbf{X}_{i}^{\top} \tau_{i} \frac{\partial \mathcal{L}_{i}}{\partial \xi^{\top}} \middle| \mathbf{y}_{i} \right) - \phi \operatorname{E} \left( \mathbf{X}_{i}^{\top} \tau_{i} \middle| \mathbf{y}_{i} \right) \operatorname{E} \left( \frac{\partial \mathcal{L}_{i}}{\partial \xi} \middle| \mathbf{y}_{i} \right)^{\top} = \phi \operatorname{Cov} \left( \mathbf{X}_{i}^{\top} \tau_{i}, \frac{\partial \mathcal{L}_{i}}{\partial \xi} \middle| \mathbf{y}_{i} \right) \\ &= \operatorname{Cov} \left( \mathbf{X}_{i}^{\top} \tau_{i}, -\frac{1}{2} \mathbf{G}_{q}^{\top} \left\{ \operatorname{vec} \left( \mathbf{\Xi}^{-1} \right) - \left( \mathbf{\Xi} \otimes \mathbf{\Xi} \right)^{-1} \lambda_{i} \right\} \middle| \mathbf{y}_{i} \right) \\ &= \phi \operatorname{Cov} \left( \mathbf{X}_{i}^{\top} \tau_{i}, \frac{1}{2} \mathbf{G}_{q}^{\top} \left( \mathbf{\Xi} \otimes \mathbf{\Xi} \right)^{-1} \lambda_{i} \middle| \mathbf{y}_{i} \right) = \frac{\phi}{2} \operatorname{Cov} \left( \mathbf{X}_{i}^{\top} \tau_{i}, \lambda_{i} \middle| \mathbf{y}_{i} \right) \left( \mathbf{\Xi} \otimes \mathbf{\Xi} \right)^{-1} \mathbf{G}_{q}. \end{split}$$

 $\mathbf{L}^{i}_{\boldsymbol{\xi}\phi}$  : First, note that

$$\begin{aligned} \frac{\partial}{\partial \phi} \left\{ \mathbf{E} \left( \lambda_{i} | \mathbf{y}_{i} \right) \right\} &= \frac{\partial}{\partial \phi} \left\{ \int_{\mathbb{R}^{q}} \lambda_{i} \exp \left( \mathcal{L}_{i} \right) \mathrm{d} \mathbf{b}_{i} \left( \int_{\mathbb{R}^{q}} \exp \left( \mathcal{L}_{i} \right) \mathrm{d} \mathbf{b}_{i} \right)^{-1} \right\} \\ &= \int_{\mathbb{R}^{q}} \lambda_{i} \frac{\partial \mathcal{L}_{i}}{\partial \phi} \exp \left( \mathcal{L}_{i} \right) \mathrm{d} \mathbf{b}_{i} \left( \int_{\mathbb{R}^{q}} \exp \left( \mathcal{L}_{i} \right) \mathrm{d} \mathbf{b}_{i} \right)^{-1} \\ &- \int_{\mathbb{R}^{q}} \lambda_{i} \exp \left( \mathcal{L}_{i} \right) \mathrm{d} \mathbf{b}_{i} \int_{\mathbb{R}^{q}} \frac{\partial \mathcal{L}_{i}}{\partial \phi} \exp \left( \mathcal{L}_{i} \right) \mathrm{d} \mathbf{b}_{i} \left( \int_{\mathbb{R}^{q}} \exp \left( \mathcal{L}_{i} \right) \mathrm{d} \mathbf{b}_{i} \right)^{-2} \\ &= \mathbf{E} \left( \lambda_{i} \frac{\partial \mathcal{L}_{i}}{\partial \phi} \middle| \mathbf{y}_{i} \right) - \mathbf{E} \left( \lambda_{i} | \mathbf{y}_{i} \right) \mathbf{E} \left( \frac{\partial \mathcal{L}_{i}}{\partial \phi} \middle| \mathbf{y}_{i} \right) = \mathbf{Cov} \left( \lambda_{i}, \frac{\partial \mathcal{L}_{i}}{\partial \phi} \middle| \mathbf{y}_{i} \right) = \mathbf{Cov} \left( \lambda_{i}, \nu_{i} | \mathbf{y}_{i} \right). \end{aligned}$$

Then,

$$\mathbf{L}_{\xi\phi}^{i} = \frac{\partial \mathbf{L}_{\xi}^{i}}{\partial \phi} = \frac{\partial}{\partial \phi} \left[ -\frac{1}{2} \mathbf{G}_{q}^{\top} \left\{ \operatorname{vec} \left( \boldsymbol{\Xi}^{-1} \right) - \left( \boldsymbol{\Xi} \otimes \boldsymbol{\Xi} \right)^{-1} \mathbf{E} \left( \boldsymbol{\lambda}_{i} | \boldsymbol{y}_{i} \right) \right\} \right] \\ = \frac{1}{2} \mathbf{G}_{q}^{\top} \left( \boldsymbol{\Xi} \otimes \boldsymbol{\Xi} \right)^{-1} \frac{\partial}{\partial \phi} \left\{ \mathbf{E} \left( \boldsymbol{\lambda}_{i} | \boldsymbol{y}_{i} \right) \right\} = \frac{1}{2} \mathbf{G}_{q}^{\top} \left( \boldsymbol{\Xi} \otimes \boldsymbol{\Xi} \right)^{-1} \operatorname{Cov} \left( \boldsymbol{\lambda}_{i}, v_{i} | \boldsymbol{y}_{i} \right).$$

 $\mathbf{L}^{i}_{\pmb{\beta} \phi}$  : First, note that

$$\begin{aligned} \frac{\partial}{\partial \phi} \mathbf{E} \left( \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i} \middle| \mathbf{y}_{i} \right) &= \frac{\partial}{\partial \phi} \left\{ \int_{\mathbb{R}^{q}} \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i} \exp \left( \mathcal{L}_{i} \right) \, \mathrm{d} \, \mathbf{b}_{i} \left( \int_{\mathbb{R}^{q}} \exp \left( \mathcal{L}_{i} \right) \, \mathrm{d} \, \mathbf{b}_{i} \right)^{-1} \right\} \\ &= \int_{\mathbb{R}^{q}} \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i} \frac{\partial \mathcal{L}_{i}}{\partial \phi} \exp \left( \mathcal{L}_{i} \right) \, \mathrm{d} \, \mathbf{b}_{i} \left( \int_{\mathbb{R}^{q}} \exp \left( \mathcal{L}_{i} \right) \, \mathrm{d} \, \mathbf{b}_{i} \right)^{-1} \\ &- \int_{\mathbb{R}^{q}} \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i} \exp \left( \mathcal{L}_{i} \right) \, \mathrm{d} \, \mathbf{b}_{i} \int_{\mathbb{R}^{q}} \frac{\partial \mathcal{L}_{i}}{\partial \phi} \exp \left( \mathcal{L}_{i} \right) \, \mathrm{d} \, \mathbf{b}_{i} \left( \int_{\mathbb{R}^{q}} \exp \left( \mathcal{L}_{i} \right) \, \mathrm{d} \, \mathbf{b}_{i} \right)^{-2} \\ &= \mathbf{E} \left( \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i} \frac{\partial \mathcal{L}_{i}}{\partial \phi} \middle| \mathbf{y}_{i} \right) - \mathbf{E} \left( \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i} \middle| \mathbf{y}_{i} \right) \mathbf{E} \left( \frac{\partial \mathcal{L}_{i}}{\partial \phi} \middle| \mathbf{y}_{i} \right) = \mathbf{Cov} \left( \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i} \frac{\partial \mathcal{L}_{i}}{\partial \phi} \middle| \mathbf{y}_{i} \right) = \mathbf{Cov} \left( \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i}, \frac{\partial \mathcal{L}_{i}}{\partial \phi} \middle| \mathbf{y}_{i} \right) = \mathbf{Cov} \left( \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i}, \frac{\partial \mathcal{L}_{i}}{\partial \phi} \middle| \mathbf{y}_{i} \right) = \mathbf{Cov} \left( \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i}, \frac{\partial \mathcal{L}_{i}}{\partial \phi} \middle| \mathbf{y}_{i} \right) = \mathbf{Cov} \left( \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i}, \frac{\partial \mathcal{L}_{i}}{\partial \phi} \middle| \mathbf{y}_{i} \right) = \mathbf{Cov} \left( \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i}, \frac{\partial \mathcal{L}_{i}}{\partial \phi} \middle| \mathbf{y}_{i} \right) = \mathbf{Cov} \left( \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i}, \frac{\partial \mathcal{L}_{i}}{\partial \phi} \middle| \mathbf{y}_{i} \right) = \mathbf{Cov} \left( \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i}, \frac{\partial \mathcal{L}_{i}}{\partial \phi} \middle| \mathbf{y}_{i} \right) = \mathbf{Cov} \left( \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i}, \frac{\partial \mathcal{L}_{i}}{\partial \phi} \middle| \mathbf{y}_{i} \right) = \mathbf{Cov} \left( \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i}, \frac{\partial \mathcal{L}_{i}}{\partial \phi} \middle| \mathbf{y}_{i} \right) = \mathbf{Cov} \left( \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i}, \frac{\partial \mathcal{L}_{i}}{\partial \phi} \middle| \mathbf{y}_{i} \right) = \mathbf{Cov} \left( \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i}, \frac{\partial \mathcal{L}_{i}}{\partial \phi} \middle| \mathbf{y}_{i} \right) = \mathbf{Cov} \left( \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i}, \frac{\partial \mathcal{L}_{i}}{\partial \phi} \middle| \mathbf{y}_{i} \right) = \mathbf{Cov} \left( \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i}, \frac{\partial \mathcal{L}_{i}}{\partial \phi} \middle| \mathbf{y}_{i} \right) = \mathbf{Cov} \left( \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i}, \frac{\partial \mathcal{L}_{i}}{\partial \phi} \middle| \mathbf{y}_{i} \right) = \mathbf{Cov} \left( \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i}, \frac{\partial \mathcal{L}_{i}}{\partial \phi} \middle| \mathbf{y}_{i} \right) = \mathbf{Cov} \left( \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i}, \frac{\partial \mathcal{L}_{i}}{\partial \phi} \middle| \mathbf{y}_{i} \right) = \mathbf{Cov} \left( \mathbf{X}_{i}^{\mathsf{T}} \boldsymbol{\tau}_{i}, \frac{\partial \mathcal{L}_{i}}{\partial \phi} \right)$$

Then,

$$\mathbf{L}_{\beta\phi}^{i} = \frac{\partial \mathbf{L}_{\beta}^{i}}{\partial \phi} = \frac{\partial}{\partial \phi} \left\{ \phi \operatorname{E} \left( \mathbf{X}_{i}^{\mathsf{T}} \tau_{i} \middle| \mathbf{y}_{i} \right) \right\} = \operatorname{E} \left( \mathbf{X}_{i}^{\mathsf{T}} \tau_{i} \middle| \mathbf{y}_{i} \right) + \phi \frac{\partial}{\partial \phi} \left\{ \operatorname{E} \left( \mathbf{X}_{i}^{\mathsf{T}} \tau_{i} \middle| \mathbf{y}_{i} \right) \right\} \\ = \frac{\partial \mathbf{L}_{\beta}^{i}}{\partial \phi} = \frac{\partial}{\partial \phi} \left\{ \phi \operatorname{E} \left( \mathbf{X}_{i}^{\mathsf{T}} \tau_{i} \middle| \mathbf{y}_{i} \right) \right\} = \operatorname{E} \left( \mathbf{X}_{i}^{\mathsf{T}} \tau_{i} \middle| \mathbf{y}_{i} \right) + \phi \operatorname{Cov} \left( \mathbf{X}_{i}^{\mathsf{T}} \tau_{i}, \nu_{i} \middle| \mathbf{y}_{i} \right).$$
## **B.5** Hessians for approximating $L_{\widehat{\psi}\widehat{\psi}}(\psi; y)$

For this purpose, it is necessary to obtain Hessian matrices from differentials. A exhaustive review of this issue, may be found in Magnus and Neudecker (1999) or Abadir and Magnus (2005). However, we briefly describe the procedure to achieve such Hessians in the next section. After that, we apply this procedure in the cases that we need.

#### Hessian matrices from differentials

Now, as it is suggested in Magnus and Neudecker (1999) or Abadir and Magnus (2005), if the interest is to obtain the Hessian matrix of **F** denoted by  $\mathbf{F}_{b_i b_i}$  and defined in (2.14), we have two options. The former is to use the second identification theorem presented in the chapter 6 of Magnus and Neudecker (1999) and the latter is to apply successively the first identification theorem, which was described briefly above. We choose the last. Thus, following definition 1 we have that

$$\mathbf{F}_{\boldsymbol{b}_{i}\boldsymbol{b}_{i}} = \frac{\partial}{\partial \boldsymbol{b}_{i}^{\mathsf{T}}} \left( \operatorname{vec} \left[ \left\{ \frac{\partial \operatorname{vec} \left( \mathbf{F} \right)}{\partial \boldsymbol{b}_{i}^{\mathsf{T}}} \right\}^{\mathsf{T}} \right] \right).$$

Consequently, for obtaining  $\mathbf{F}_{b_i b_i}$  we should conduct the procedure exposed below:

**Step 1:** To compute d vec (**F**) and hence we obtain  $\partial \text{vec}(\mathbf{F}) / \partial \boldsymbol{b}_i^{\mathsf{T}}$ .

**Step 2:** Transpose the resulting matrix for achieving  $(\partial \text{vec}(\mathbf{F}) / \partial \boldsymbol{b}_i^{\mathsf{T}})^{\mathsf{T}}$ .

**Step 3:** To compute  $d \operatorname{vec} \left\{ \left( \partial \operatorname{vec}(\mathbf{F}) / \partial \boldsymbol{b}_i^{\mathsf{T}} \right)^{\mathsf{T}} \right\}$  and use the first identification theorem for obtaining  $\mathbf{F}_{\boldsymbol{b}_i, \boldsymbol{b}_i}$ .

In order to illustrate this procedure we will compute  $\mathbf{F}_{b_i b_i}$  for  $\mathbf{F} = \mathbf{X}_i^{\top} \boldsymbol{\tau}_i$ . Thus, we have

**Step 1:** From Table B.2 we have that  $d \operatorname{vec} (\mathbf{X}_i^{\mathsf{T}} \boldsymbol{\tau}_i) = d \mathbf{X}_i^{\mathsf{T}} \boldsymbol{\tau}_i = \mathbf{X}_i^{\mathsf{T}} \mathbf{M}_i \mathbf{Z}_i d \boldsymbol{b}_i$ .

**Step 2:**  $\partial \text{vec}(\mathbf{F}) / \partial \boldsymbol{b}_i^{\mathsf{T}} = \mathbf{X}_i^{\mathsf{T}} \mathbf{M}_i \mathbf{Z}_i \text{ and } \left\{ \partial \text{vec}(\mathbf{F}) / \partial \boldsymbol{b}_i^{\mathsf{T}} \right\}^{\mathsf{T}} = \mathbf{X}_i^{\mathsf{T}} \mathbf{M}_i \mathbf{Z}_i$ 

Step 3: Using Table B.2 we obtain

$$\operatorname{d}\operatorname{vec}(\mathbf{Z}_{i}^{\top}\mathbf{M}_{i}\mathbf{X}_{i}) = \left(\mathbf{X}_{i}^{\top}\otimes\mathbf{Z}_{i}^{\top}\right)\operatorname{d}\operatorname{vec}\left(\mathbf{M}_{i}\right) = \left(\mathbf{X}_{i}^{\top}\otimes\mathbf{Z}_{i}^{\top}\right)\mathbf{\Omega}_{i}\left(\dot{\mathbf{M}}_{i}-2\mathbf{D}_{\boldsymbol{\rho}_{i}}\mathbf{A}_{i}\right)\mathbf{Z}_{i}\operatorname{d}\boldsymbol{b}_{i}.$$

Then, the Hessian matrix of **F** on  $b_i$  is

$$\mathbf{F}_{\boldsymbol{b}_i \boldsymbol{b}_i} = \left( \mathbf{X}_i^\top \boldsymbol{\tau}_i \right)_{\boldsymbol{b}_i \boldsymbol{b}_i} = \left( \mathbf{X}_i^\top \otimes \mathbf{Z}_i^\top \right) \boldsymbol{\Omega}_i \left( \dot{\mathbf{M}}_i - 2 \, \mathbf{D}_{\boldsymbol{\rho}_i} \, \mathbf{A}_i \right) \, \mathbf{Z}_i.$$

## Hessians for the approximation of $L^i_{\ \widehat{\psi}}$ and $L^i_{\ \widehat{\psi} \ \widehat{\psi}}$

In this section we present the detailed computations of the Hessian matrices exhibited in (3.11). For this purpose, we will use in each case the procedure exposed in the previous section. The first Hessian in (3.11) was computed in the previous section. Thus, we will deal with the remaining Hessians.

 $\lambda_i$ : from Table B.2, we have that  $d\lambda_i = \mathbf{N}_q (\mathbf{I}_q \otimes \mathbf{b}_i) d\mathbf{b}_i$ . Thus,

$$\operatorname{d}\operatorname{vec}\left\{\left(\mathbf{I}_{q}\otimes\boldsymbol{b}_{i}^{\mathsf{T}}\right)\mathbf{N}_{q}\right\}=\left(\mathbf{N}_{q}\otimes\mathbf{I}_{q}\right)\boldsymbol{\Lambda}_{q}\left(\mathbf{I}_{q}\right)\operatorname{d}\boldsymbol{b}_{i}\Longrightarrow\left(\boldsymbol{\lambda}\right)_{\boldsymbol{b}_{i}\boldsymbol{b}_{i}}=\left(\mathbf{N}_{q}\otimes\mathbf{I}_{q}\right)\boldsymbol{\Lambda}_{q}\left(\mathbf{I}_{q}\right),$$

 $\nu_i$ : from Table B.2, we have that  $d \nu_i = \tau_i^\top \mathbf{Z}_i d \mathbf{b}_i$  and  $d \mathbf{Z}_i^\top \tau_i = \mathbf{Z}_i^\top \mathbf{M}_i \mathbf{Z}_i d \mathbf{b}_i$ . Therefore  $(\nu_i)_{b_i b_i} = \mathbf{Z}_i^\top \mathbf{M}_i \mathbf{Z}_i$ .

 $\mathbf{X}_i^{\mathsf{T}} \tau_i \tau_i^{\mathsf{T}} \mathbf{X}_i$ : from Table B.2, we have expressions for d  $\mathbf{X}_i^{\mathsf{T}} \tau_i$  and d vec (**M**<sub>*i*</sub>). Thus,

$$d \operatorname{vec} \left( \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \boldsymbol{\tau}_{i}^{\top} \mathbf{X}_{i} \right) = d \operatorname{vec} \left\{ \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \left( \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \right)^{\top} \right\}$$
$$= \left( \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \otimes \mathbf{I}_{p} \right) d \left( \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \right) + \left( \mathbf{I}_{p} \otimes \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \right) d \left( \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \right)$$
$$= \left\{ \mathbf{K}_{p} \left( \mathbf{I}_{p} \otimes \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \right) \mathbf{K}_{p1} + \left( \mathbf{I}_{p} \otimes \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \right) \right\} d \left( \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \right)$$
$$= \left\{ \mathbf{N}_{p} \left( \mathbf{I}_{p} \otimes \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \right) \right\} d \left( \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \right) = \mathbf{N}_{p} \left( \mathbf{I}_{p} \otimes \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \right) \mathbf{X}_{i}^{\top} \mathbf{M}_{i} \mathbf{Z}_{i} d \mathbf{b}_{i} \quad \text{and}$$

$$d \operatorname{vec} \left\{ \mathbf{Z}_{i}^{\top} \mathbf{M}_{i} \mathbf{X}_{i} \left( \mathbf{I}_{p} \otimes \boldsymbol{\tau}_{i}^{\top} \mathbf{X}_{i} \right) \mathbf{N}_{p} \right\}$$

$$= \left( \mathbf{N}_{p} \otimes \mathbf{Z}_{i}^{\top} \mathbf{M}_{i} \mathbf{X}_{i} \right) \Lambda_{p} \left( \mathbf{I}_{p} \right) \mathbf{X}_{i}^{\top} \mathbf{M}_{i} \mathbf{Z}_{i} d \mathbf{b}_{i} + \left[ \left\{ \mathbf{N}_{p} \left( \mathbf{I}_{p} \otimes \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \right) \mathbf{X}_{i}^{\top} \right\} \otimes \mathbf{Z}_{i}^{\top} \right] d \operatorname{vec} \left( \mathbf{M}_{i} \right)$$

$$= \left( \mathbf{N}_{p} \otimes \mathbf{Z}_{i}^{\top} \mathbf{M}_{i} \mathbf{X}_{i} \right) \Lambda_{p} \left( \mathbf{I}_{p} \right) \mathbf{X}_{i}^{\top} \mathbf{M}_{i} \mathbf{Z}_{i} d \mathbf{b}_{i} + \left\{ \mathbf{N}_{p} \left( \mathbf{X}_{i}^{\top} \otimes \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \right) \otimes \mathbf{Z}_{i}^{\top} \right\} \mathbf{\Omega}_{i} \left( \dot{\mathbf{M}}_{i} - 2 \mathbf{D}_{\rho_{i}} \mathbf{A}_{i} \right) \mathbf{Z}_{i} d \mathbf{b}_{i}$$

$$= \left( \mathbf{N}_{p} \otimes \mathbf{Z}_{i}^{\top} \right) \left\{ \left( \mathbf{I}_{p^{2}} \otimes \mathbf{M}_{i} \mathbf{X}_{i} \right) \Lambda_{p} \left( \mathbf{I}_{p} \right) \mathbf{X}_{i}^{\top} \mathbf{M}_{i} + \left( \mathbf{X}_{i}^{\top} \otimes \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \otimes \mathbf{I}_{n_{i}} \right) \mathbf{\Omega}_{i} \left( \dot{\mathbf{M}}_{i} - 2 \mathbf{D}_{\rho_{i}} \mathbf{A}_{i} \right) \right\} \mathbf{Z}_{i},$$

$$\Longrightarrow \left( \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \boldsymbol{\tau}_{i}^{\top} \mathbf{X}_{i} \right)_{\boldsymbol{b}_{i} \boldsymbol{b}_{i}} = \left( \mathbf{N}_{p} \otimes \mathbf{Z}_{i}^{\top} \right) \left\{ \left( \mathbf{I}_{p^{2}} \otimes \mathbf{M}_{i} \mathbf{X}_{i} \right) \Lambda_{p} \left( \mathbf{I}_{p} \right) \mathbf{X}_{i}^{\top} \mathbf{M}_{i} + \left( \mathbf{X}_{i}^{\top} \otimes \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \otimes \mathbf{I}_{n_{i}} \right) \boldsymbol{\Omega}_{i} \left( \dot{\mathbf{M}}_{i} - 2 \mathbf{D}_{\boldsymbol{\rho}_{i}} \mathbf{A}_{i} \right) \right\} \mathbf{Z}_{i}.$$

 $\mathbf{X}_{i}^{\mathsf{T}} \mathbf{M}_{i} \mathbf{X}_{i}$ : from Table B.2 we have expressions for d vec ( $\mathbf{M}_{i}$ ) and d vec ( $\dot{\mathbf{M}}_{i}$ ). Thus,

$$\operatorname{d}\operatorname{vec}\left(\mathbf{X}_{i}^{\top}\mathbf{M}_{i}\mathbf{X}_{i}\right) = \left(\mathbf{X}_{i}^{\top}\otimes\mathbf{X}_{i}^{\top}\right)\operatorname{d}\operatorname{vec}\left(\mathbf{M}_{i}\right) = \left(\mathbf{X}_{i}^{\top}\otimes\mathbf{X}_{i}^{\top}\right)\mathbf{\Omega}_{i}\left(\dot{\mathbf{M}}_{i}-2\mathbf{D}_{\rho_{i}}\mathbf{A}_{i}\right)\mathbf{Z}_{i}\operatorname{d}\boldsymbol{b}_{i} \quad \text{and} \quad \mathbf{D}_{i}$$

$$\begin{aligned} \operatorname{d}\operatorname{vec}\left\{\mathbf{Z}_{i}^{\top}\left(\dot{\mathbf{M}}_{i}-2\,\mathbf{D}_{\rho_{i}}\,\mathbf{A}_{i}\right)\,\mathbf{\Omega}_{i}^{\top}\left(\mathbf{X}_{i}\otimes\mathbf{X}_{i}\right)\right\} &= \left\{\left(\mathbf{X}_{i}^{\top}\otimes\mathbf{X}_{i}^{\top}\right)\,\mathbf{\Omega}_{i}\otimes\mathbf{Z}_{i}^{\top}\right\}\,\operatorname{d}\operatorname{vec}\left(\dot{\mathbf{M}}_{i}-2\,\mathbf{D}_{\rho_{i}}\,\mathbf{A}_{i}\right)\right\} \\ &= \left\{\left(\mathbf{X}_{i}^{\top}\otimes\mathbf{X}_{i}^{\top}\right)\,\mathbf{\Omega}_{i}\otimes\mathbf{Z}_{i}^{\top}\right\}\,\left\{\operatorname{d}\operatorname{vec}\left(\dot{\mathbf{M}}_{i}\right)-2\,\operatorname{d}\operatorname{vec}\left(\mathbf{D}_{\rho_{i}}\,\mathbf{A}_{i}\right)\right\} \\ &= \left\{\left(\mathbf{X}_{i}^{\top}\otimes\mathbf{X}_{i}^{\top}\right)\,\mathbf{\Omega}_{i}\otimes\mathbf{Z}_{i}^{\top}\right\}\,\left\{\operatorname{d}\operatorname{vec}\left(\dot{\mathbf{M}}_{i}\right)-2\,\mathbf{\Omega}_{i}\left(\mathbf{D}_{\rho_{i}}\,\mathbf{A}_{i}+\mathbf{D}_{\rho_{i}}\,\dot{\mathbf{A}}_{i}\right)\,\mathbf{Z}_{i}\,\operatorname{d}\boldsymbol{b}_{i}\right\} \\ &= \left\{\left(\mathbf{X}_{i}^{\top}\otimes\mathbf{X}_{i}^{\top}\right)\,\mathbf{\Omega}_{i}\otimes\mathbf{Z}_{i}^{\top}\right\}\,\mathbf{\Omega}_{i}\,\left\{\ddot{\mathbf{M}}_{i}-3\,\left(\mathbf{D}_{\rho_{i}}\,\mathbf{A}_{i}+\mathbf{D}_{\rho_{i}}\,\dot{\mathbf{A}}_{i}\right)\right\}\,\mathbf{Z}_{i}\,\operatorname{d}\boldsymbol{b}_{i}\,,\\ &\Longrightarrow\\ \left(\mathbf{X}_{i}^{\top}\,\mathbf{M}_{i}\,\mathbf{X}_{i}\right)_{\boldsymbol{b}_{i},\boldsymbol{b}_{i}} = \left\{\left(\mathbf{X}_{i}^{\top}\otimes\mathbf{X}_{i}^{\top}\right)\,\mathbf{\Omega}_{i}\otimes\mathbf{Z}_{i}^{\top}\right\}\,\mathbf{\Omega}_{i}\,\left\{\ddot{\mathbf{M}}_{i}-3\,\left(\mathbf{D}_{\rho_{i}}\,\mathbf{A}_{i}+\mathbf{D}_{\rho_{i}}\,\dot{\mathbf{A}}_{i}\right)\right\}\,\mathbf{Z}_{i}\,.\end{aligned}$$

 $v_i^2$ : from Table B.2 we have the computation for d  $v_i$  and d  $\mathbf{Z}_i^{\mathsf{T}} \tau_i$ . Thus,

$$d(v_i^2) = 2v_i \frac{\partial v_i}{\partial b_i^{\mathsf{T}}} db_i = 2v_i \tau_i^{\mathsf{T}} \mathbf{Z}_i db_i \text{ and}$$

$$2d(\mathbf{Z}_i^{\mathsf{T}} \tau_i v_i) = 2d\{\mathbf{Z}_i \tau_i\} v_i + 2\tau_i d(v_i)$$

$$= 2v_i \mathbf{Z}_i^{\mathsf{T}} \mathbf{M}_i \mathbf{Z}_i db_i + 2\mathbf{Z}_i \tau_i \tau_i^{\mathsf{T}} \mathbf{Z}_i db_i \implies (v_i)_{b_i b_i} = 2\mathbf{Z}_i^{\mathsf{T}} (v_i \mathbf{M}_i + \tau_i \tau_i^{\mathsf{T}}) \mathbf{Z}_i$$

$$= 2\mathbf{Z}_i^{\mathsf{T}} (v_i \mathbf{M}_i + \tau_i \tau_i^{\mathsf{T}}) \mathbf{Z}_i db_i.$$

 $\mathbf{X}_i^{\mathsf{T}} \boldsymbol{\tau}_i \boldsymbol{\lambda}_i^{\mathsf{T}}$ : expressions for d  $\mathbf{X}_i^{\mathsf{T}} \boldsymbol{\tau}_i$ , d  $\boldsymbol{\lambda}_i$  and d vec ( $\mathbf{M}_i$ ) are displayed in Table B.2. Thus, we have

$$d \operatorname{vec} \left( \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \boldsymbol{\lambda}_{i}^{\top} \right) = \left( \mathbf{I}_{q^{2}} \otimes \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \right) d \boldsymbol{\lambda}_{i} + \left( \boldsymbol{\lambda}_{i} \otimes \mathbf{I}_{p} \right) d \left( \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \right)$$
$$= \left\{ \operatorname{N}_{q} \left( \mathbf{I}_{q} \otimes \boldsymbol{b}_{i} \right) \otimes \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \right\} d \boldsymbol{b}_{i} + \left( \boldsymbol{\lambda}_{i} \otimes \mathbf{X}_{i}^{\top} \mathbf{M}_{i} \mathbf{Z}_{i} \right) d \boldsymbol{b}_{i} \text{ and }$$

$$\begin{aligned} \operatorname{d}\operatorname{vec}\left[\left\{\left(\mathbf{I}_{q}\otimes\boldsymbol{b}_{i}^{\mathsf{T}}\right)\,\mathbf{N}_{q}\otimes\boldsymbol{\tau}_{i}^{\mathsf{T}}\,\mathbf{X}_{i}\right\}+\left(\boldsymbol{\lambda}_{i}^{\mathsf{T}}\otimes\mathbf{Z}_{i}^{\mathsf{T}}\,\mathbf{M}_{i}\,\mathbf{X}_{i}\right)\right] =\\ =& \mathbf{\Lambda}_{p}\left\{\left(\mathbf{I}_{q}\otimes\boldsymbol{b}_{i}^{\mathsf{T}}\right)\,\mathbf{N}_{q}\right\}\,\mathbf{X}_{i}^{\mathsf{T}}\,\mathbf{M}_{i}\,\mathbf{Z}_{i}\,\operatorname{d}\boldsymbol{b}_{i}+\left\{\mathbf{N}_{q}\otimes\mathbf{\Pi}_{q}\left(\boldsymbol{\tau}_{i}^{\mathsf{T}}\,\mathbf{X}_{i}\right)\right\}\,\mathbf{\Lambda}_{q}\left(\mathbf{I}_{q}\right)\,\operatorname{d}\boldsymbol{b}_{i}\\ &+\left\{\mathbf{N}_{q}\left(\mathbf{I}_{q}\otimes\boldsymbol{b}_{i}^{\mathsf{T}}\right)\,\mathbf{N}_{q}\right\}\,\mathbf{X}_{i}^{\mathsf{T}}\,\mathbf{M}_{i}\,\mathbf{Z}_{i}\,\operatorname{d}\boldsymbol{b}_{i}+\left\{\mathbf{N}_{q}\left(\mathbf{I}_{q}\otimes\boldsymbol{b}_{i}\right)\otimes\operatorname{vec}\left(\mathbf{Z}_{i}^{\mathsf{T}}\,\mathbf{M}_{i}\,\mathbf{X}_{i}\right)\right\}\,\operatorname{d}\boldsymbol{b}_{i}\\ =& \mathbf{\Lambda}_{p}\left\{\left(\mathbf{I}_{q}\otimes\boldsymbol{b}_{i}^{\mathsf{T}}\right)\,\mathbf{N}_{q}\right\}\,\mathbf{X}_{i}^{\mathsf{T}}\,\mathbf{M}_{i}\,\mathbf{Z}_{i}\,\operatorname{d}\boldsymbol{b}_{i}+\left\{\mathbf{N}_{q}\left(\mathbf{I}_{q}\otimes\boldsymbol{b}_{i}\right)\otimes\operatorname{vec}\left(\mathbf{Z}_{i}^{\mathsf{T}}\,\mathbf{M}_{i}\,\mathbf{X}_{i}\right)\right\}\,\operatorname{d}\boldsymbol{b}_{i}\\ &+\left(\mathbf{N}_{q}\otimes\mathbf{X}_{i}^{\mathsf{T}}\,\boldsymbol{\tau}_{i}\otimes\mathbf{I}_{q}\right)\,\mathbf{\Lambda}_{q}\left(\mathbf{I}_{q}\right)\,\operatorname{d}\boldsymbol{b}_{i}+\left(\boldsymbol{\lambda}_{i}\otimes\mathbf{X}_{i}^{\mathsf{T}}\otimes\mathbf{Z}_{i}^{\mathsf{T}}\right)\,\mathbf{\Omega}_{i}\left(\dot{\mathbf{M}}_{i}-2\,\mathbf{D}_{\rho_{i}}\,\mathbf{A}_{i}\right)\,\mathbf{Z}_{i}\,\operatorname{d}\boldsymbol{b}_{i}\\ &=\left(\mathbf{I}_{q^{2}}\otimes\mathbf{X}_{i}^{\mathsf{T}}\otimes\mathbf{Z}_{i}^{\mathsf{T}}\right)\left[\left\{\mathbf{N}_{q}\left(\mathbf{I}_{q}\otimes\boldsymbol{b}_{i}\right)\otimes\operatorname{vec}\left(\mathbf{M}_{i}\right)\right\}+\left(\boldsymbol{\lambda}_{i}\otimes\mathbf{\Omega}_{i}\right)\left(\dot{\mathbf{M}}_{i}-2\,\mathbf{D}_{\rho_{i}}\,\mathbf{A}_{i}\right)\,\mathbf{Z}_{i}\right]\,\operatorname{d}\boldsymbol{b}_{i}\\ &+\left(\mathbf{N}_{q}\otimes\mathbf{X}_{i}^{\mathsf{T}}\,\boldsymbol{\tau}_{i}\otimes\mathbf{I}_{q}\right)\,\mathbf{\Lambda}_{q}\left(\mathbf{I}_{q}\right)\,\operatorname{d}\boldsymbol{b}_{i}+\mathbf{\Lambda}_{p}\left\{\left(\mathbf{I}_{q}\otimes\boldsymbol{b}_{i}^{\mathsf{T}}\right)\,\mathbf{N}_{q}\right\}\,\mathbf{X}_{i}^{\mathsf{T}}\,\mathbf{M}_{i}\,\mathbf{Z}_{i}\,\operatorname{d}\boldsymbol{b}_{i}\\ &\Longleftrightarrow\left(\mathbf{X}_{i}^{\mathsf{T}}\,\boldsymbol{\tau}_{i}\,\boldsymbol{\lambda}_{q}\right)\right]\,\operatorname{d}\boldsymbol{b}_{i}+\mathbf{\Lambda}_{p}\left\{\left(\mathbf{I}_{q}\otimes\boldsymbol{b}_{i}^{\mathsf{T}}\right)\,\mathbf{N}_{q}\right\}\,\mathbf{X}_{i}^{\mathsf{T}}\,\mathbf{M}_{i}\,\mathbf{Z}_{i}\,\operatorname{d}\boldsymbol{b}_{i}\\ &+\left(\mathbf{N}_{q}\otimes\mathbf{X}_{i}^{\mathsf{T}}\,\boldsymbol{\tau}_{i}\,\mathbf{\lambda}_{q}\right)\,\mathbf{\Lambda}_{q}\left(\mathbf{I}_{q}\otimes\mathbf{b}_{i}\right)\otimes\operatorname{vec}\left(\mathbf{M}_{i}\right)\right\}+\left(\boldsymbol{\lambda}_{i}\otimes\mathbf{\Omega}_{i}\right)\left(\dot{\mathbf{M}_{i}-2\,\mathbf{D}_{\rho_{i}}\,\mathbf{A}_{i}\right)\,\mathbf{Z}_{i}\right]\,\operatorname{d}\boldsymbol{b}_{i}\\ &+\left(\mathbf{N}_{q}\otimes\mathbf{X}_{i}^{\mathsf{T}}\,\boldsymbol{\tau}_{i}\,\mathbf{\lambda}_{q}\right)\left(\mathbf{I}_{q}\left(\mathbf{I}_{q}\otimes\mathbf{b}_{i}\right)\otimes\operatorname{vec}\left(\mathbf{M}_{i}\right)\right)\right\}+\left(\boldsymbol{\lambda}_{i}\otimes\mathbf{\Omega}_{i}\right)\left(\dot{\mathbf{M}_{i}-2\,\mathbf{D}_{\rho_{i}}\,\mathbf{A}_{i}\right)\,\mathbf{Z}_{i}\right)\,\operatorname{d}\boldsymbol{b}_{i}\\ &+\left(\mathbf{N}_{q}\otimes\mathbf{X}_{i}^{\mathsf{T}}\,\boldsymbol{\tau}_{i}\,\mathbf{N}_{q}\right)\left(\mathbf{A}_{i}\left(\mathbf{I}_{q}\otimes\mathbf{b}_{i}\right)\otimes\operatorname{vec}\left(\mathbf{M}_{i}\right)\right)\right\}+\left(\mathbf{A}_{i}\otimes\mathbf{\Omega}_{i}\right)\left(\dot{\mathbf{M}_{i}-2\,\mathbf{D}_{\rho_{i}}\,\mathbf{A}_{i}\right)\,\mathbf{Z}_{i}\right)\,\operatorname{d}\boldsymbol{b}_{i}\\ &+\left(\mathbf{N}_{q}\otimes\mathbf{X}_{i}^{\mathsf{T}}\,\boldsymbol{\tau}_{i}\otimes\mathbf{Z}_{i}\right)\left(\mathbf$$

 $\mathbf{X}_{i}^{\mathsf{T}} \tau_{i} \nu_{i}$ : In Table B.2 we have expressions for d  $\mathbf{X}_{i}^{\mathsf{T}} \tau_{i}$ , d  $\mathbf{Z}_{i}^{\mathsf{T}} \tau_{i}$  and d  $\nu_{i}$ . Thus, we have

$$\operatorname{d}\operatorname{vec}\left(\mathbf{X}_{i}^{\top}\,\boldsymbol{\tau}_{i}\,\boldsymbol{\nu}_{i}\right) = \operatorname{d}\left(\mathbf{X}_{i}^{\top}\,\boldsymbol{\tau}_{i}\,\boldsymbol{\nu}_{i}\right) = \boldsymbol{\nu}_{i}\,\mathbf{X}_{i}^{\top}\,\mathbf{M}_{i}\,\mathbf{Z}_{i}\,\operatorname{d}\boldsymbol{b}_{i} + \mathbf{X}_{i}^{\top}\,\boldsymbol{\tau}_{i}\,\boldsymbol{\tau}_{i}^{\top}\,\mathbf{Z}_{i}\,\operatorname{d}\boldsymbol{b}_{i} \quad \text{and}$$

$$d \operatorname{vec} \left( \mathbf{Z}_{i}^{\top} \tau_{i} \tau_{i}^{\top} \mathbf{X}_{i} + \nu_{i} \mathbf{Z}_{i}^{\top} \mathbf{M}_{i} \mathbf{X}_{i} \right)$$

$$= \left( \mathbf{X}_{i}^{\top} \tau_{i} \otimes \mathbf{I}_{q} \right) \mathbf{Z}_{i}^{\top} \mathbf{M}_{i} \mathbf{Z}_{i} d b_{i} + \left( \mathbf{I}_{p} \otimes \mathbf{Z}_{i}^{\top} \tau_{i} \right) \mathbf{X}_{i}^{\top} \mathbf{M}_{i} \mathbf{Z}_{i} d b_{i}$$

$$+ \left( \mathbf{X}_{i}^{\top} \otimes \mathbf{Z}_{i}^{\top} \right) \operatorname{vec} \left( \mathbf{M}_{i} \right) \tau_{i}^{\top} \mathbf{Z}_{i} d b_{i} + \nu_{i} \left( \mathbf{X}_{i}^{\top} \otimes \mathbf{Z}_{i}^{\top} \right) \mathbf{\Omega}_{i} \left( \dot{\mathbf{M}}_{i} - 2 \mathbf{D}_{\rho_{i}} \mathbf{A}_{i} \right) \mathbf{Z}_{i} d b_{i}$$

$$= \left( \mathbf{X}_{i}^{\top} \otimes \mathbf{Z}_{i}^{\top} \right) \left\{ \left( \tau_{i} \otimes \mathbf{M}_{i} \right) + \left( \mathbf{M}_{i} \otimes \tau_{i} \right) + \operatorname{vec} \left( \mathbf{M}_{i} \right) \tau_{i}^{\top} + \nu_{i} \mathbf{\Omega}_{i} \left( \dot{\mathbf{M}}_{i} - 2 \mathbf{D}_{\rho_{i}} \mathbf{A}_{i} \right) \right\} \mathbf{Z}_{i} d b_{i}$$

$$= \left( \mathbf{X}_{i}^{\top} \otimes \mathbf{Z}_{i}^{\top} \right) \left\{ \mathbf{N}_{n_{i}} \left( \mathbf{M}_{i} \otimes \tau_{i} \right) + \operatorname{vec} \left( \mathbf{M}_{i} \right) \tau_{i}^{\top} + \nu_{i} \mathbf{\Omega}_{i} \left( \dot{\mathbf{M}}_{i} - 2 \mathbf{D}_{\rho_{i}} \mathbf{A}_{i} \right) \right\} \mathbf{Z}_{i} d b_{i}$$

$$\Longrightarrow \left( \mathbf{X}_{i}^{\top} \tau_{i} \nu_{i} \right)_{b_{i} b_{i}} = \left( \mathbf{X}_{i}^{\top} \otimes \mathbf{Z}_{i}^{\top} \right) \left\{ \mathbf{N}_{n_{i}} \left( \mathbf{M}_{i} \otimes \tau_{i} \right) + \operatorname{vec} \left( \mathbf{M}_{i} \right) \tau_{i}^{\top} + \nu_{i} \mathbf{\Omega}_{i} \left( \dot{\mathbf{M}}_{i} - 2 \mathbf{D}_{\rho_{i}} \mathbf{A}_{i} \right) \right\} \mathbf{Z}_{i}.$$

 $\lambda_i v_i$ : In Table B.2 are reported  $d\lambda_i$ ,  $dv_i$  and  $d\mathbf{Z}_i^{\top} \boldsymbol{\tau}_i$ . Using these expressions we

obtain

$$\operatorname{dvec}\left(\lambda_{i} \nu_{i}\right) = \mathbf{N}_{q}\left(\mathbf{I}_{q} \otimes \boldsymbol{b}_{i}\right) \operatorname{d} \boldsymbol{b}_{i} \nu_{i} + \lambda_{i} \tau_{i}^{\top} \mathbf{Z}_{i} \operatorname{d} \boldsymbol{b}_{i} = \left\{\nu_{i} \mathbf{N}_{q}\left(\mathbf{I}_{q} \otimes \boldsymbol{b}_{i}\right) + \lambda_{i} \tau_{i}^{\top} \mathbf{Z}_{i}\right\} \operatorname{d} \boldsymbol{b}_{i} \quad \text{and}$$

$$d \left( \mathbf{Z}_{i}^{\mathsf{T}} \tau_{i} \boldsymbol{\lambda}_{i}^{\mathsf{T}} + \left( \mathbf{I}_{q} \otimes \boldsymbol{b}_{i}^{\mathsf{T}} \right) \mathbf{N}_{q} \boldsymbol{\nu}_{i} \right) = \left( \boldsymbol{\lambda}_{i} \otimes \mathbf{I}_{q} \right) \mathbf{Z}_{i}^{\mathsf{T}} \mathbf{M}_{i} \mathbf{Z}_{i} d \boldsymbol{b}_{i} + \left( \mathbf{I}_{q^{2}} \otimes \mathbf{Z}_{i}^{\mathsf{T}} \tau_{i} \right) \mathbf{N}_{q} \left( \mathbf{I}_{q} \otimes \boldsymbol{b}_{i} \right) d \boldsymbol{b}_{i} + \operatorname{vec} \left\{ \left( \mathbf{I}_{q} \otimes \boldsymbol{b}_{i}^{\mathsf{T}} \right) \mathbf{N}_{q} \right\} \tau_{i}^{\mathsf{T}} \mathbf{Z}_{i} d \boldsymbol{b}_{i} + \boldsymbol{\nu}_{i} \left( \mathbf{N}_{q} \otimes \mathbf{I}_{q} \right) \boldsymbol{\Lambda}_{q} \left( \mathbf{I}_{q} \right) d \boldsymbol{b}_{i} = \left( \boldsymbol{\lambda}_{i} \otimes \mathbf{Z}_{i}^{\mathsf{T}} \mathbf{M}_{i} \mathbf{Z}_{i} \right) d \boldsymbol{b}_{i} + \left( \mathbf{N}_{q} \otimes \mathbf{Z}_{i}^{\mathsf{T}} \tau_{i} \right) \left( \mathbf{I}_{q} \otimes \boldsymbol{b}_{i} \right) d \boldsymbol{b}_{i} + \left( \mathbf{N}_{q} \otimes \mathbf{I}_{q} \right) \boldsymbol{\Lambda}_{q} \left( \mathbf{I}_{q} \right) \boldsymbol{b}_{i} \tau_{i}^{\mathsf{T}} \mathbf{Z}_{i} d \boldsymbol{b}_{i} + \boldsymbol{\nu}_{i} \left( \mathbf{N}_{q} \otimes \mathbf{I}_{q} \right) \boldsymbol{\Lambda}_{q} \left( \mathbf{I}_{q} \right) d \boldsymbol{b}_{i} = \left( \mathbf{N}_{q} \otimes \mathbf{I}_{q} \right) \boldsymbol{\Lambda}_{q} \left( \mathbf{I}_{q} \right) \left\{ \mathbf{I}_{q} \boldsymbol{\nu}_{i} + \boldsymbol{b}_{i} \tau_{i}^{\mathsf{T}} \mathbf{Z}_{i} \right\} d \boldsymbol{b}_{i} + \left( \mathbf{I}_{q^{2}} \otimes \mathbf{Z}_{i}^{\mathsf{T}} \right) \left\{ \left( \boldsymbol{\lambda}_{i} \otimes \mathbf{M}_{i} \mathbf{Z}_{i} \right) + \left( \mathbf{N}_{q} \otimes \tau_{i} \right) \left( \mathbf{I}_{q} \otimes \boldsymbol{b}_{i} \right) \right\} d \boldsymbol{b}_{i} ,$$

$$\implies (\boldsymbol{\lambda}_{i} \, \boldsymbol{\nu}_{i})_{\boldsymbol{b}_{i}\boldsymbol{b}_{i}} = \\ = \left( \mathbf{N}_{q} \otimes \mathbf{I}_{q} \right) \, \boldsymbol{\Lambda}_{q} \left( \mathbf{I}_{q} \right) \left\{ \mathbf{I}_{q} \, \boldsymbol{\nu}_{i} + \boldsymbol{b}_{i} \, \boldsymbol{\tau}_{i}^{\mathsf{T}} \, \mathbf{Z}_{i} \right\} \, \mathrm{d} \, \boldsymbol{b}_{i} + \left( \mathbf{I}_{q^{2}} \otimes \mathbf{Z}_{i}^{\mathsf{T}} \right) \left\{ (\boldsymbol{\lambda}_{i} \otimes \mathbf{M}_{i} \, \mathbf{Z}_{i}) + \left( \mathbf{N}_{q} \otimes \boldsymbol{\tau}_{i} \right) \left( \mathbf{I}_{q} \otimes \boldsymbol{b}_{i} \right) \right\}.$$

 $\lambda_i \lambda_i^{\mathsf{T}}$ : Using d  $\lambda_i$  of Table B.2, we obtain

$$dvec (\lambda_i \lambda_i^{\top}) = (\lambda_i \otimes \mathbf{I}_{q^2}) d\lambda_i + (\mathbf{I}_{q^2} \otimes \lambda_i) d\lambda_i$$
$$= \mathbf{N}_{q^2} (\lambda_i \otimes \mathbf{I}_{q^2}) \mathbf{N}_q (\mathbf{I}_q \otimes b_i) db_i = \mathbf{N}_{q^2} \{\lambda_i \otimes \mathbf{N}_q (\mathbf{I}_q \otimes b_i)\} db_i \text{ and}$$

$$d \operatorname{vec} \left[ \left\{ \lambda_{i}^{\mathsf{T}} \otimes \left( \mathbf{I}_{q} \otimes \boldsymbol{b}_{i}^{\mathsf{T}} \right) \mathbf{N}_{q} \right\} \mathbf{N}_{q^{2}} \right] = d \operatorname{vec} \left\{ \left( \mathbf{I}_{q} \otimes \boldsymbol{b}_{i}^{\mathsf{T}} \right) \left( \lambda_{i}^{\mathsf{T}} \otimes \mathbf{N}_{q} \right) \mathbf{N}_{q^{2}} \right\} \\ = \left( \mathbf{N}_{q^{2}} \otimes \mathbf{I}_{q} \right) d \operatorname{vec} \left\{ \left( \mathbf{I}_{q} \otimes \boldsymbol{b}_{i}^{\mathsf{T}} \right) \left( \lambda_{i}^{\mathsf{T}} \otimes \mathbf{N}_{q} \right) \right\} \\ = \left( \mathbf{N}_{q^{2}} \otimes \mathbf{I}_{q} \right) \left\{ \left( \mathbf{I}_{q^{5}} \otimes \boldsymbol{b}_{i}^{\mathsf{T}} \right) d \operatorname{vec} \left( \lambda_{i}^{\mathsf{T}} \otimes \mathbf{N}_{q} \right) + \left( \lambda_{i} \otimes \mathbf{N}_{q} \otimes \mathbf{I}_{q} \right) \mathbf{\Lambda}_{q} \left( \mathbf{I}_{q} \right) d \mathbf{b}_{i} \right\} \\ = \left( \mathbf{N}_{q^{2}} \otimes \mathbf{I}_{q} \right) \left[ \left( \mathbf{I}_{q^{5}} \otimes \boldsymbol{b}_{i}^{\mathsf{T}} \right) \left\{ \mathbf{I}_{q^{2}} \otimes \operatorname{vec} \left( \mathbf{N}_{q} \right) \right\} d \lambda_{i} + \left( \lambda_{i} \otimes \mathbf{N}_{q} \otimes \mathbf{I}_{q} \right) \mathbf{\Lambda}_{q} \left( \mathbf{I}_{q} \right) d \mathbf{b}_{i} \right] \\ = \left( \mathbf{N}_{q^{2}} \otimes \mathbf{I}_{q} \right) \left[ \left( \mathbf{I}_{q^{5}} \otimes \boldsymbol{b}_{i}^{\mathsf{T}} \right) \left\{ \mathbf{I}_{q^{2}} \otimes \operatorname{vec} \left( \mathbf{N}_{q} \right) \right\} \mathbf{N}_{q} \left( \mathbf{I}_{q} \otimes \mathbf{b}_{i} \right) d \mathbf{b}_{i} + \left( \lambda_{i} \otimes \mathbf{N}_{q} \otimes \mathbf{I}_{q} \right) \mathbf{\Lambda}_{q} \left( \mathbf{I}_{q} \right) d \mathbf{b}_{i} \right] \\ = \left( \mathbf{N}_{q^{2}} \otimes \mathbf{I}_{q} \right) \left[ \left( \mathbf{I}_{q^{5}} \otimes \boldsymbol{b}_{i}^{\mathsf{T}} \right) \left\{ \mathbf{N}_{q} \left( \mathbf{I}_{q} \otimes \mathbf{b}_{i} \right) \otimes \operatorname{vec} \left( \mathbf{N}_{q} \right) \right\} d \mathbf{b}_{i} + \left( \lambda_{i} \otimes \mathbf{N}_{q} \otimes \mathbf{I}_{q} \right) \mathbf{\Lambda}_{q} \left( \mathbf{I}_{q} \right) d \mathbf{b}_{i} \right] \\ = \left( \mathbf{N}_{q^{2}} \otimes \mathbf{I}_{q} \right) \left[ \left\{ \mathbf{N}_{q} \left( \mathbf{I}_{q} \otimes \mathbf{b}_{i} \right) \otimes \left( \mathbf{I}_{q^{3}} \otimes \mathbf{b}_{i}^{\mathsf{T}} \right) \operatorname{vec} \left( \mathbf{N}_{q} \right) \right\} d \mathbf{b}_{i} + \left( \lambda_{i} \otimes \mathbf{N}_{q} \otimes \mathbf{I}_{q} \right) \mathbf{\Lambda}_{q} \left( \mathbf{I}_{q} \right) d \mathbf{b}_{i} \right] \\ = \left( \mathbf{N}_{q^{2}} \otimes \mathbf{I}_{q} \right) \left( \left[ \mathbf{N}_{q} \left( \mathbf{I}_{q} \otimes \mathbf{b}_{i} \right) \otimes \operatorname{vec} \left\{ \left( \mathbf{I}_{q} \otimes \mathbf{b}_{i}^{\mathsf{T}} \right) \mathbf{N}_{q} \right\} \right] + \left( \lambda_{i} \otimes \mathbf{N}_{q} \otimes \mathbf{I}_{q} \right) \mathbf{\Lambda}_{q} \left( \mathbf{I}_{q} \right) d \mathbf{b}_{i} ,$$

## **B.6** $\Delta$ under the first perturbation scheme (subjects)

In section 3.7.1 we considered  $\mathbf{L}(\boldsymbol{\psi}; \boldsymbol{y} | \boldsymbol{\omega}) = \sum_{i=1}^{N} \omega_i \mathbf{L}(\boldsymbol{\psi}; \boldsymbol{y}_i)$ , where  $\boldsymbol{\omega} = (\omega_i, \omega_2, \dots, \omega_N)^{\top}$ . The no perturbation vector is  $\boldsymbol{\omega}_0 = \mathbf{1}_N$ . Under this perturbation scheme  $\boldsymbol{\Delta}$  assumes the form

$$\begin{split} \Delta &= \frac{\partial^2 \operatorname{L}(\psi; y)}{\partial \psi \partial \omega^{\top}} \bigg|_{\widehat{\psi}, \omega_0} = \frac{\partial^2}{\partial \psi \partial \omega^{\top}} \left\{ \sum_{i=1}^N \omega_i \operatorname{L}(\psi; y_i) \right\} \bigg|_{\widehat{\psi}, \omega_0} = \frac{\partial}{\partial \psi} \left[ \frac{\partial}{\partial \omega^{\top}} \left\{ \sum_{i=1}^N \omega_i \operatorname{L}(\psi; y_i) \right\} \right] \bigg|_{\widehat{\psi}, \omega_0} \\ &= \frac{\partial}{\partial \psi} \left( \operatorname{L}(\psi; y_1), \operatorname{L}(\psi; y_2), \dots, \operatorname{L}(\psi; y_N) \right) \bigg|_{\widehat{\psi}} = (\delta_1, \delta_2, \dots, \delta_N), \\ \text{where} \quad \delta_i &= \frac{\partial \operatorname{L}(\psi; y_i)}{\partial \psi} \bigg|_{\widehat{\psi}} = \operatorname{L}_{\widehat{\psi}}(\psi; y_i) = \operatorname{L}_{\widehat{\psi}}^i \text{ for } i = 1, \dots, N. \end{split}$$

 $\mathbf{L}_{\psi}^{i}$  is computed in appendix B.3.

Also, the approximated expression for  $\delta_i$  given in (3.14) is obtained by means of the definition 4 and using the Hessians  $(\mathbf{X}_i^{\mathsf{T}} \tau_i)_{b_i b_i}$ ,  $(\lambda_i)_{b_i b_i}$  and  $(\nu_i)_{b_i b_i}$ , which were computed in appendix B.5.

## B.7 $\Delta$ under the second perturbation scheme (observations)

In section 3.7.2 we considered  $\mathbf{L}(\boldsymbol{\psi}; \boldsymbol{y} | \boldsymbol{\omega}) = \sum_{i=1}^{N} \mathbf{L}(\boldsymbol{\psi}; \boldsymbol{y}_i | \boldsymbol{\omega}_i)$  and  $\mathbf{L}(\boldsymbol{\psi}; \boldsymbol{y}_i | \boldsymbol{\omega}_i)$  given in (3.16). Under this perturbation scheme  $\boldsymbol{\Delta}$  may be obtained as follows:

$$\begin{split} \Delta &= \left. \frac{\partial^{2} \mathcal{L} \left( \psi; \boldsymbol{y} \mid \boldsymbol{\omega} \right)}{\partial \psi \partial \boldsymbol{\omega}^{\top}} \right|_{\widehat{\psi}, \omega_{0}} = \left. \frac{\partial^{2}}{\partial \psi \partial \boldsymbol{\omega}^{\top}} \left\{ \sum_{i=1}^{N} \mathcal{L} \left( \psi; \boldsymbol{y}_{i} \mid \boldsymbol{\omega}_{i} \right) \right\} \right|_{\widehat{\psi}, \omega_{0}} = \left. \frac{\partial}{\partial \psi} \left\{ \sum_{i=1}^{N} \left. \frac{\partial \mathcal{L} \left( \psi; \boldsymbol{y}_{i} \mid \boldsymbol{\omega}_{i} \right)}{\partial \boldsymbol{\omega}^{\top}} \right\} \right|_{\widehat{\psi}, \omega_{0}} = \left. \frac{\partial}{\partial \psi} \left( \frac{\partial \mathcal{L} \left( \psi; \boldsymbol{y}_{1} \mid \boldsymbol{\omega}_{1} \right)}{\partial \boldsymbol{\omega}_{1}^{\top}}, \frac{\partial \mathcal{L} \left( \psi; \boldsymbol{y}_{2} \mid \boldsymbol{\omega}_{2} \right)}{\partial \boldsymbol{\omega}_{2}^{\top}}, \dots, \frac{\partial \mathcal{L} \left( \psi; \boldsymbol{y}_{N} \mid \boldsymbol{\omega}_{N} \right)}{\partial \boldsymbol{\omega}_{N}^{\top}} \right) \right|_{\widehat{\psi}, \omega_{0}} \\ &= \left. \left( \Delta_{1}, \Delta_{2}, \dots, \Delta_{N} \right), \\ \text{where } \Delta_{i} &= \left. \frac{\partial^{2} \mathcal{L} \left( \psi; \boldsymbol{y}_{i} \mid \boldsymbol{\omega}_{i} \right)}{\partial \psi \partial \boldsymbol{\omega}_{i}^{\top}} \right|_{\widehat{\psi}, \omega_{0}}. \end{split}$$

In the computation above we have used the fact that, for i = 1, ..., N,

$$\frac{\partial \operatorname{L}(\boldsymbol{\psi};\boldsymbol{y}_{i} | \boldsymbol{\omega}_{i})}{\partial \boldsymbol{\omega}} = \left(\mathbf{0}_{n_{1}}^{\mathsf{T}}, \dots, \mathbf{0}_{n_{i-1}}^{\mathsf{T}}, \frac{\partial \operatorname{L}(\boldsymbol{\psi};\boldsymbol{y}_{i} | \boldsymbol{\omega}_{i})}{\partial \boldsymbol{\omega}_{i}^{\mathsf{T}}}, \mathbf{0}_{n_{i+1}}^{\mathsf{T}}, \dots, \mathbf{0}_{n_{N}}^{\mathsf{T}}\right)^{\mathsf{T}},$$

where  $\mathbf{0}_{n_i}$  is a  $n_i \times 1$  vector of zeros.

In order to determine the structure of  $\Delta$ , in what follows we will compute, for i = 1, ..., N, the form of  $\Delta_i$  given in (3.17). Denoting  $\mathcal{L}(\omega_i) = L(\psi; y_i, b_i | \omega_i)$  and since  $\psi = (\beta^{\mathsf{T}}, \xi^{\mathsf{T}}, \phi)^{\mathsf{T}}$ , we may see that

$$\frac{\partial \mathcal{L}(\omega_{i})}{\partial \omega_{i}^{\top}} = \frac{\partial L(\psi; y_{i}, b_{i} | \omega_{i})}{\partial \omega_{i}^{\top}} = \frac{\partial}{\partial \omega_{i}^{\top}} (\omega_{i}^{\top} \mathbf{f}_{i}) = \mathbf{f}_{i}^{\top},$$
  
$$\frac{\partial \mathcal{L}(\omega_{i})}{\partial \beta} = \frac{\partial L(\psi; y_{i}, b_{i} | \omega_{i})}{\partial \beta} = \frac{\partial}{\partial \beta} (\mathbf{f}_{i}^{\top} \omega_{i}) = \phi \mathbf{X}_{i}^{\top} \mathbf{W}_{i}^{1/2} \mathbf{V}_{i}^{-1/2} \mathbf{D}_{y_{i} - \mu_{i}} \omega_{i} = \phi \mathbf{X}_{i}^{\top} \mathbf{T}_{i} \omega_{i} \text{ and }$$

$$\frac{\partial \mathbf{f}_i^{\mathsf{T}}}{\partial \boldsymbol{\beta}} = \phi \, \mathbf{X}_i^{\mathsf{T}} \, \mathbf{T}_i.$$

Now, using the expressions above and according to equation (3.16) we have that

$$\begin{split} \frac{\partial^{2} L\left(\psi; \boldsymbol{y}_{i} | \boldsymbol{\omega}_{i}\right)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}_{i}^{\top}} &= \frac{\partial}{\partial \boldsymbol{\beta}} \left( \frac{\partial}{\partial \boldsymbol{\omega}_{i}^{\top}} \log \left[ \int_{\mathbb{R}^{q}} \exp \left\{ \mathcal{L}\left(\boldsymbol{\omega}_{i}\right) \right\} d \boldsymbol{b}_{i} \right] \right) \\ &= \frac{\partial}{\partial \boldsymbol{\beta}} \left( \left[ \int_{\mathbb{R}^{q}} \exp \left\{ \mathcal{L}\left(\boldsymbol{\omega}_{i}\right) \right\} d \boldsymbol{b}_{i} \right]^{-1} \times \left[ \int_{\mathbb{R}^{q}} \exp \left\{ \mathcal{L}\left(\boldsymbol{\omega}_{i}\right) \right\} \frac{\partial \mathcal{L}\left(\boldsymbol{\omega}_{i}\right)}{\partial \boldsymbol{\omega}_{i}^{\top}} d \boldsymbol{b}_{i} \right] \right) \\ &= \frac{\partial}{\partial \boldsymbol{\beta}} \left( \left[ \int_{\mathbb{R}^{q}} \exp \left\{ \mathcal{L}\left(\boldsymbol{\omega}_{i}\right) \right\} d \boldsymbol{b}_{i} \right]^{-1} \times \left[ \int_{\mathbb{R}^{q}} \exp \left\{ \mathcal{L}\left(\boldsymbol{\omega}_{i}\right) \right\} f_{i}^{\top} d \boldsymbol{b}_{i} \right] \right) \\ &= \left[ \int_{\mathbb{R}^{q}} \exp \left\{ \mathcal{L}\left(\boldsymbol{\omega}_{i}\right) \right\} d \boldsymbol{b}_{i} \right]^{-1} \int_{\mathbb{R}^{q}} \left( \frac{\partial \mathcal{L}\left(\boldsymbol{\omega}_{i}\right)}{\partial \boldsymbol{\beta}} f_{i}^{\top} + \frac{\partial f_{i}^{\top}}{\partial \boldsymbol{\beta}} \right) \exp \left\{ \mathcal{L}\left(\boldsymbol{\omega}_{i}\right) \right\} d \boldsymbol{b}_{i} \\ &- \left[ \int_{\mathbb{R}^{q}} \exp \left\{ \mathcal{L}\left(\boldsymbol{\omega}_{i}\right) \right\} d \boldsymbol{b}_{i} \right]^{-2} \int_{\mathbb{R}^{q}} \exp \left\{ \mathcal{L}\left(\boldsymbol{\omega}_{i}\right) \right\} \frac{\partial \mathcal{L}\left(\boldsymbol{\omega}_{i}\right)}{\partial \boldsymbol{\beta}} d \boldsymbol{b}_{i} \int_{\mathbb{R}^{q}} \exp \left\{ \mathcal{L}\left(\boldsymbol{\omega}_{i}\right) \right\} f_{i}^{\top} d \boldsymbol{b}_{i} \\ &= \left[ \int_{\mathbb{R}^{q}} \exp \left\{ \mathcal{L}\left(\boldsymbol{\omega}_{i}\right) \right\} d \boldsymbol{b}_{i} \right]^{-1} \int_{\mathbb{R}^{q}} \left( \phi \mathbf{X}_{i}^{\top} \mathbf{T}_{i} \boldsymbol{\omega}_{i} f_{i}^{\top} + \phi \mathbf{X}_{i}^{\top} \mathbf{T}_{i} \right) \exp \left\{ \mathcal{L}\left(\boldsymbol{\omega}_{i}\right) \right\} d \boldsymbol{b}_{i} \\ &- \left[ \int_{\mathbb{R}^{q}} \exp \left\{ \mathcal{L}\left(\boldsymbol{\omega}_{i}\right) \right\} d \boldsymbol{b}_{i} \right]^{-2} \int_{\mathbb{R}^{q}} \exp \left\{ \mathcal{L}\left(\boldsymbol{\omega}_{i}\right) \right\} \phi \mathbf{X}_{i}^{\top} \mathbf{T}_{i} \boldsymbol{\omega}_{i} d \boldsymbol{b}_{i} \int_{\mathbb{R}^{q}} \exp \left\{ \mathcal{L}\left(\boldsymbol{\omega}_{i}\right) \right\} f_{i}^{\top} d \boldsymbol{b}_{i}. \end{split}$$

Since under this perturbation scheme the no perturbation vector is  $\boldsymbol{\omega}_0 = \mathbf{1}_{N_T}$  and  $\boldsymbol{\omega} = (\boldsymbol{\omega}_1^{\mathsf{T}}, \boldsymbol{\omega}_2^{\mathsf{T}}, \dots, \boldsymbol{\omega}_N^{\mathsf{T}})^{\mathsf{T}}$ , we have that  $\boldsymbol{\omega}_{0_i} = \mathbf{1}_{n_i}$  for  $i = 1, \dots, N$ . Thus,

$$\mathcal{L}(\boldsymbol{\omega}_{i_0}) = L(\boldsymbol{\psi}; \boldsymbol{y}_i, \boldsymbol{b}_i | \boldsymbol{\omega}_{i_0}) = L(\boldsymbol{\psi}; \boldsymbol{y}_i, \boldsymbol{b}_i | \mathbf{1}_{n_i}) = L(\boldsymbol{\psi}; \boldsymbol{y}_i, \boldsymbol{b}_i) = \mathcal{L}_i.$$

Then,

$$\frac{\partial^{2} \mathcal{L}_{\omega_{i}} (\boldsymbol{\psi}; \boldsymbol{y}_{i})}{\partial \boldsymbol{\beta} \partial \omega_{i}^{\top}} \bigg|_{\widehat{\boldsymbol{\psi}}, \omega_{i_{0}}} = \phi \mathcal{E} \left( \mathbf{X}_{i}^{\top} \mathbf{T}_{i} \mathbf{1}_{n_{i}} \mathbf{f}_{i}^{\top} + \mathbf{X}_{i}^{\top} \mathbf{T}_{i} \big| \boldsymbol{y}_{i} \right) + \phi \mathcal{E} \left( \mathbf{X}_{i}^{\top} \mathbf{T}_{i} \mathbf{1}_{n_{i}} \big| \boldsymbol{y}_{i} \right) \mathcal{E} \left( \mathbf{f}_{i} \big| \boldsymbol{y}_{i} \right)^{\top}$$
$$= \phi \mathcal{E} \left( \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \mathbf{f}_{i}^{\top} + \mathbf{X}_{i}^{\top} \mathbf{T}_{i} \big| \boldsymbol{y}_{i} \right) + \phi \mathcal{E} \left( \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \big| \boldsymbol{y}_{i} \right) \mathcal{E} \left( \mathbf{f}_{i} \big| \boldsymbol{y}_{i} \right)^{\top}$$
$$= \hat{\phi} \mathcal{E} \left( \mathbf{X}_{i}^{\top} \mathbf{T}_{i} \big| \boldsymbol{y}_{i} \right) + \hat{\phi} \mathcal{C} \operatorname{ov} \left( \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \big| \boldsymbol{y}_{i} \right).$$

On the other hand  $\frac{\partial^2 L(\psi; y_i | \omega_i)}{\partial \xi \partial \omega_i^{\top}} \Big|_{\widehat{\psi}, \omega_{i_0}}$  and  $\frac{\partial^2 L(\psi; y_i | \omega_i)}{\partial \phi \partial \omega_i^{\top}} \Big|_{\widehat{\psi}, \omega_{i_0}}$  may be computed similarly. It is enough to take into account that

$$\frac{\partial \mathcal{L}(\omega_i)}{\partial \xi} = \frac{\partial \mathcal{L}_i}{\partial \xi} = \frac{\partial \mathcal{L}(\psi; y_i, b_i)}{\partial \xi} = -\frac{1}{2} \mathbf{G}_q^{\mathsf{T}} \left\{ \operatorname{vec} \left( \boldsymbol{\Xi}^{-1} \right) - \left( \boldsymbol{\Xi} \otimes \boldsymbol{\Xi} \right)^{-1} \lambda_i \right\}.$$
 As it was seen in (B.1)

$$\frac{\partial \mathcal{L}(\boldsymbol{\omega}_i)}{\partial \boldsymbol{\phi}} = \frac{\partial \operatorname{L}\left(\boldsymbol{\psi}; \boldsymbol{y}_i, \boldsymbol{b}_i \mid \boldsymbol{\omega}_i\right)}{\partial \boldsymbol{\phi}} = \mathbf{1}_{n_i}^{\mathsf{T}} \frac{\partial \mathbf{f}_i}{\partial \boldsymbol{\phi}} = \mathbf{1}_{n_i}^{\mathsf{T}} \dot{\mathbf{f}}_i = \boldsymbol{\nu}_i.$$

Hence, for each i = 1, ..., N, we have obtained the structure for  $\Delta_i$  and therefore for  $\Delta$ . Finally, since  $\Delta_i$  does not have a closed form, following the definition (4) we require the expressions given in section 3.7.2 and computed below.

#### **B.7.1** Hessians for $\Delta$ under the second perturbation scheme

 $\mathbf{X}_i^{\mathsf{T}} \mathbf{T}_i \mathbf{C}_i$ : Expressions for dvec ( $\mathbf{C}_i$ ), dvec ( $\mathbf{T}_i$ ) and dvec ( $\mathbf{M}_i$ ) are displayed in B.2. Thus, we have

$$d \operatorname{vec} \left( \mathbf{X}_{i}^{\top} \mathbf{T}_{i} \mathbf{C}_{i} \right) = d \operatorname{vec} \left( \mathbf{X}_{i}^{\top} \mathbf{V}_{i}^{-1/2} \mathbf{W}_{i}^{1/2} \mathbf{D}_{y_{i}-\mu_{i}} \mathbf{C}_{i} \right)$$
$$= \left( \mathbf{C}_{i}^{\top} \otimes \mathbf{I}_{p} \right) d \operatorname{vec} \left( \mathbf{X}_{i}^{\top} \mathbf{T}_{i} \right) + \left( \mathbf{I}_{n_{i}} \otimes \mathbf{X}_{i}^{\top} \mathbf{T}_{i} \right) d \operatorname{vec} \left( \mathbf{C}_{i} \right)$$
$$= \left( \mathbf{C}_{i}^{\top} \otimes \mathbf{X}_{i}^{\top} \right) \mathbf{\Omega}_{i} \mathbf{M}_{i} \mathbf{Z}_{i} d \mathbf{b}_{i} + \phi \left( \mathbf{T}_{i} \mathbf{Z}_{i} \otimes \mathbf{X}_{i}^{\top} \mathbf{\tau}_{i} \right) \text{ and }$$

$$d \operatorname{vec} \left\{ \mathbf{Z}_{i}^{\top} \mathbf{M}_{i} \mathbf{\Omega}_{i}^{\top} (\mathbf{C}_{i} \otimes \mathbf{X}_{i}) + \phi \left( \mathbf{Z}_{i}^{\top} \mathbf{T}_{i} \otimes \boldsymbol{\tau}_{i}^{\top} \mathbf{X}_{i} \right) \right\}$$

$$= \left\{ \left( \mathbf{C}_{i}^{\top} \otimes \mathbf{X}_{i}^{\top} \right) \mathbf{\Omega}_{i} \otimes \mathbf{Z}_{i}^{\top} \right\} d \operatorname{vec} \left( \mathbf{M}_{i} \right) + \left( \mathbf{I}_{n_{i}p} \otimes \mathbf{Z}_{i}^{\top} \mathbf{M}_{i} \mathbf{\Omega}_{i}^{\top} \right) d \operatorname{vec} \left( \mathbf{C}_{i} \otimes \mathbf{X}_{i} \right)$$

$$+ \phi \left( \mathbf{I}_{n_{i}} \otimes \Pi_{q} \left( \boldsymbol{\tau}_{i}^{\top} \mathbf{X}_{i} \right) \right) d \operatorname{vec} \left( \mathbf{Z}_{i}^{\top} \mathbf{T}_{i} \right) + \phi \Lambda_{p} \left( \mathbf{Z}_{i}^{\top} \mathbf{T}_{i} \right) d \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i}$$

$$= \left\{ \left( \mathbf{C}_{i}^{\top} \otimes \mathbf{X}_{i}^{\top} \right) \mathbf{\Omega}_{i} \otimes \mathbf{Z}_{i}^{\top} \right\} \mathbf{\Omega}_{i} \left( \dot{\mathbf{M}}_{i} - 2 \mathbf{D}_{\rho_{i}} \mathbf{A}_{i} \right) \mathbf{Z}_{i} d \mathbf{b}_{i} + \phi \left( \mathbf{I}_{n_{i}} \otimes \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \otimes \mathbf{Z}_{i}^{\top} \right) \mathbf{\Omega}_{i} \mathbf{M}_{i} \mathbf{Z}_{i} d \mathbf{b}_{i}$$

$$+ \phi \Lambda_{p} \left( \mathbf{Z}_{i}^{\top} \mathbf{T}_{i} \right) \mathbf{X}_{i}^{\top} \mathbf{M}_{i} \mathbf{Z}_{i} d \mathbf{b}_{i} + \phi \left( \mathbf{I}_{n_{i}p} \otimes \mathbf{Z}_{i}^{\top} \mathbf{M}_{i} \mathbf{\Omega}_{i}^{\top} \right) \left( \mathbf{T}_{i} \mathbf{Z}_{i} \otimes \mathbf{\Pi}_{n_{i}} \left( \mathbf{X}_{i} \right) \mathbf{1}_{n_{i}} \right) d \mathbf{b}_{i}$$

$$= \left( \mathbf{I}_{n_{i}p} \otimes \mathbf{Z}_{i}^{\top} \right) \left[ \left\{ \left( \mathbf{C}_{i}^{\top} \otimes \mathbf{X}_{i}^{\top} \right) \mathbf{\Omega}_{i} \otimes \mathbf{I}_{n_{i}} \right\} \mathbf{\Omega}_{i} \left( \dot{\mathbf{M}}_{i} - 2 \mathbf{D}_{\rho_{i}} \mathbf{A}_{i} \right) + \phi \left( \mathbf{I}_{n_{i}} \otimes \mathbf{X}_{i}^{\top} \boldsymbol{\tau}_{i} \otimes \mathbf{I}_{n_{i}} \right) \mathbf{\Omega}_{i} \mathbf{M}_{i}$$

$$+ \phi \left( \mathbf{I}_{n_{i}p} \otimes \mathbf{M}_{i} \mathbf{\Omega}_{i}^{\top} \right) \left( \mathbf{T}_{i} \otimes \mathbf{H}_{n_{i}} \left( \mathbf{X}_{i} \right) \mathbf{1}_{n_{i}} \right] \mathbf{Z}_{i} d \mathbf{b}_{i} + \phi \mathbf{\Lambda}_{p} \left( \mathbf{Z}_{i}^{\top} \mathbf{T}_{i} \right) \mathbf{X}_{i}^{\top} \mathbf{M}_{i} \mathbf{Z}_{i} d \mathbf{b}_{i} ,$$

$$\Longrightarrow (\mathbf{X}_{i}^{\top} \mathbf{T}_{i} \mathbf{C}_{i})_{\boldsymbol{b}_{i}\boldsymbol{b}_{i}} = = (\mathbf{I}_{n_{i}p} \otimes \mathbf{Z}_{i}^{\top}) \left[ \left\{ (\mathbf{C}_{i}^{\top} \otimes \mathbf{X}_{i}^{\top}) \, \mathbf{\Omega}_{i} \otimes \mathbf{I}_{n_{i}} \right\} \mathbf{\Omega}_{i} \left( \dot{\mathbf{M}}_{i} - 2\mathbf{D}_{\rho_{i}} \mathbf{A}_{i} \right) + \phi \left( \mathbf{I}_{n_{i}} \otimes \mathbf{X}_{i}^{\top} \, \boldsymbol{\tau}_{i} \otimes \mathbf{I}_{n_{i}} \right) \mathbf{\Omega}_{i} \mathbf{M}_{i} + \phi \left( \mathbf{I}_{n_{i}p} \otimes \mathbf{M}_{i} \mathbf{\Omega}_{i}^{\top} \right) \left( \mathbf{T}_{i} \otimes \mathbf{\Pi}_{n_{i}} \left( \mathbf{X}_{i} \right) \, \mathbf{1}_{n_{i}} \right) \right] \mathbf{Z}_{i} + \phi \, \mathbf{\Lambda}_{p} \left( \mathbf{Z}_{i}^{\top} \, \mathbf{T}_{i} \right) \mathbf{X}_{i}^{\top} \, \mathbf{M}_{i} \, \mathbf{Z}_{i} \, .$$

 $\lambda_i \mathbf{f}_i^{\mathsf{T}}$ : Table B.2 presents expressions for d  $\mathbf{f}_i$  and d  $\lambda_i$ . Thus, we have

d vec  $(\lambda_i \mathbf{f}_i^{\mathsf{T}}) = (\mathbf{f}_i \otimes \mathbf{N}_q) (\mathbf{I}_q \otimes \mathbf{b}_i) d \mathbf{b}_i + \phi (\mathbf{T}_i \mathbf{Z}_i \otimes \lambda_i) d \mathbf{b}_i$  and

$$d \operatorname{vec} \left\{ \left( \mathbf{I}_{q} \otimes \boldsymbol{b}_{i}^{\mathsf{T}} \right) \left( \boldsymbol{f}_{i}^{\mathsf{T}} \otimes \mathbf{N}_{q} \right) + \phi \left( \mathbf{Z}_{i}^{\mathsf{T}} \mathbf{T}_{i} \otimes \boldsymbol{\lambda}_{i}^{\mathsf{T}} \right) \right\}$$

$$= \left( \mathbf{f}_{i} \otimes \mathbf{N}_{q} \otimes \mathbf{I}_{q} \right) d \operatorname{vec} \left( \mathbf{I}_{q} \otimes \boldsymbol{b}_{i}^{\mathsf{T}} \right) + \left( \mathbf{I}_{n_{i}q^{3}} \otimes \boldsymbol{b}_{i}^{\mathsf{T}} \right) d \operatorname{vec} \left( \mathbf{f}_{i}^{\mathsf{T}} \otimes \mathbf{N}_{q} \right) + \phi \left\{ \mathbf{I}_{n_{i}} \otimes \boldsymbol{\Pi}_{q} \left( \boldsymbol{\lambda}_{i}^{\mathsf{T}} \right) \right\} d \operatorname{vec} \left( \mathbf{Z}_{i}^{\mathsf{T}} \mathbf{T}_{i} \right)$$

$$+ \phi \mathbf{\Lambda}_{q^{2}} \left( \mathbf{Z}_{i}^{\mathsf{T}} \mathbf{T}_{i} \right) d \boldsymbol{\lambda}_{i}$$

$$= \left( \mathbf{f}_{i} \otimes \mathbf{N}_{q} \otimes \mathbf{I}_{q} \right) \mathbf{\Lambda}_{q} \left( \mathbf{I}_{q} \right) d \boldsymbol{b}_{i} + \left( \mathbf{I}_{n_{i}q^{3}} \otimes \boldsymbol{b}_{i}^{\mathsf{T}} \right) \left( \mathbf{I}_{n_{i}} \otimes \boldsymbol{\Pi}_{1} \left( \mathbf{N}_{q} \right) \right) d \mathbf{f}_{i}$$

$$+ \phi \left( \mathbf{I}_{n_{i}} \otimes \boldsymbol{\lambda}_{i} \otimes \mathbf{I}_{q} \right) \left( \mathbf{I}_{n_{i}} \otimes \mathbf{Z}_{i}^{\mathsf{T}} \right) \mathbf{\Omega}_{i} \mathbf{M}_{i} \mathbf{Z}_{i} d \boldsymbol{b}_{i} + \phi \mathbf{\Lambda}_{q^{2}} \left( \mathbf{Z}_{i}^{\mathsf{T}} \mathbf{T}_{i} \right) \mathbf{N}_{q} \left( \mathbf{I}_{q} \otimes \boldsymbol{b}_{i} \right) d \boldsymbol{b}_{i}$$

$$= \left( \mathbf{f}_{i} \otimes \mathbf{N}_{q} \otimes \mathbf{I}_{q} \right) \mathbf{\Lambda}_{q} \left( \mathbf{I}_{q} \right) d \boldsymbol{b}_{i} + \phi \left( \mathbf{I}_{n_{i}q^{3}} \otimes \boldsymbol{b}_{i}^{\mathsf{T}} \right) \left( \mathbf{I}_{n_{i}} \otimes \operatorname{vec} \left( \mathbf{N}_{q} \right) \right) \mathbf{T}_{i} \mathbf{Z}_{i} d \boldsymbol{b}_{i}$$

$$+ \phi \left( \mathbf{I}_{n_{i}} \otimes \boldsymbol{\lambda}_{i} \otimes \mathbf{Z}_{i}^{\mathsf{T}} \right) \mathbf{\Omega}_{i} \mathbf{M}_{i} \mathbf{Z}_{i} d \boldsymbol{b}_{i} + \phi \mathbf{\Lambda}_{q^{2}} \left( \mathbf{Z}_{i}^{\mathsf{T}} \mathbf{T}_{i} \right) \mathbf{N}_{q} \left( \mathbf{I}_{q} \otimes \boldsymbol{b}_{i} \right) d \boldsymbol{b}_{i} ,$$

$$\Longrightarrow (\lambda_{i} \mathbf{f}_{i}^{\top})_{b_{i}b_{i}} = = (\mathbf{f}_{i} \otimes \mathbf{N}_{q} \otimes \mathbf{I}_{q}) \mathbf{\Lambda}_{q} (\mathbf{I}_{q}) + \phi (\mathbf{I}_{n_{i}q^{3}} \otimes \mathbf{b}_{i}^{\top}) (\mathbf{I}_{n_{i}} \otimes \operatorname{vec} (\mathbf{N}_{q})) \mathbf{T}_{i} \mathbf{Z}_{i} + \phi (\mathbf{I}_{n_{i}} \otimes \lambda_{i} \otimes \mathbf{Z}_{i}^{\top}) \mathbf{\Omega}_{i} \mathbf{M}_{i} \mathbf{Z}_{i} + \phi \mathbf{\Lambda}_{q^{2}} (\mathbf{Z}_{i}^{\top} \mathbf{T}_{i}) \mathbf{N}_{q} (\mathbf{I}_{q} \otimes \mathbf{b}_{i}).$$

 $\dot{\mathbf{f}}_i^{\mathsf{T}} \mathbf{C}_i$ : In Table B.2 expressions for d  $\dot{\mathbf{f}}_i$  and d vec ( $\mathbf{C}_i$ ) are displayed. Thus, we have

$$d \operatorname{vec} \left( \dot{\mathbf{f}}_{i}^{\top} \mathbf{C}_{i} \right) = d \operatorname{vec} \left\{ \left( \mathbf{C}_{i}^{\top} \dot{\mathbf{f}}_{i} \right)^{\top} \right\} = d \operatorname{vec} \left( \mathbf{C}_{i}^{\top} \dot{\mathbf{f}}_{i} \right) = d \left( \mathbf{C}_{i}^{\top} \dot{\mathbf{f}}_{i} \right)$$
$$= \mathbf{C}_{i}^{\top} d \dot{\mathbf{f}}_{i} + \left( \dot{\mathbf{f}}_{i}^{\top} \otimes \mathbf{I}_{n_{i}} \right) d \operatorname{vec} \left( \mathbf{C}_{i}^{\top} \right) = \mathbf{C}_{i}^{\top} \mathbf{T}_{i} \mathbf{Z}_{i} d \mathbf{b}_{i} + \left( \dot{\mathbf{f}}_{i}^{\top} \otimes \mathbf{I}_{n_{i}} \right) \mathbf{K}_{n} d \operatorname{vec} \left( \mathbf{C}_{i} \right)$$
$$= \mathbf{C}_{i}^{\top} \mathbf{T}_{i} \mathbf{Z}_{i} d \mathbf{b}_{i} + \left( \dot{\mathbf{f}}_{i}^{\top} \otimes \mathbf{I}_{n_{i}} \right) \left( \mathbf{1}_{n_{i}} \otimes \mathbf{T}_{i} \mathbf{Z}_{i} \right) d \mathbf{b}_{i} = \mathbf{C}_{i}^{\top} \mathbf{T}_{i} \mathbf{Z}_{i} d \mathbf{b}_{i} + \phi v_{i} \mathbf{T}_{i} \mathbf{Z}_{i} d \mathbf{b}_{i} \text{ and }$$

$$d \operatorname{vec} \left( \mathbf{Z}_{i}^{\top} \mathbf{T}_{i} \mathbf{C}_{i} + \phi v_{i} \mathbf{Z}_{i}^{\top} \mathbf{T}_{i} \right)$$

$$= \left( \mathbf{I}_{n_{i}} \otimes \mathbf{Z}_{i}^{\top} \mathbf{T}_{i} \right) d \operatorname{vec} \left( \mathbf{C}_{i} \right) + \left( \mathbf{C}_{i}^{\top} \otimes \mathbf{I}_{q} \right) d \operatorname{vec} \left( \mathbf{Z}_{i}^{\top} \mathbf{T}_{i} \right) + \phi \operatorname{vec} \left( \mathbf{Z}_{i}^{\top} \mathbf{T}_{i} \right) \mathbf{\tau}_{i}^{\top} \mathbf{Z}_{i} d \mathbf{b}_{i}$$

$$+ \phi v_{i} d \operatorname{vec} \left( \mathbf{Z}_{i}^{\top} \mathbf{T}_{i} \right)$$

$$= \phi \left( \mathbf{T}_{i} \mathbf{Z}_{i} \otimes \mathbf{Z}_{i}^{\top} \mathbf{\tau}_{i} \right) d \mathbf{b}_{i} + \left( \mathbf{C}_{i}^{\top} \otimes \mathbf{Z}_{i}^{\top} \right) \mathbf{\Omega}_{i} \mathbf{M}_{i} \mathbf{Z}_{i} d \mathbf{b}_{i} + \phi \operatorname{vec} \left( \mathbf{Z}_{i}^{\top} \mathbf{T}_{i} \right) \mathbf{\tau}_{i}^{\top} \mathbf{Z}_{i} d \mathbf{b}_{i}$$

$$+ \phi v_{i} \left( \mathbf{I}_{n_{i}} \otimes \mathbf{Z}_{i}^{\top} \right) \mathbf{\Omega}_{i} \mathbf{M}_{i} \mathbf{Z}_{i} d \mathbf{b}_{i}$$

$$= \left( \mathbf{I}_{n_{i}} \otimes \mathbf{Z}_{i}^{\top} \right) \left\{ \phi \left( \mathbf{T}_{i} \otimes \mathbf{\tau}_{i} \right) + \left( \mathbf{C}_{i}^{\top} \otimes \mathbf{I}_{n_{i}} \right) \mathbf{M}_{i} + \phi \operatorname{vec} \left( \mathbf{T}_{i} \right) \mathbf{\tau}_{i}^{\top} + \phi \mathbf{\Omega}_{i} \mathbf{M}_{i} \right\} \mathbf{Z}_{i} d \mathbf{b}_{i},$$

$$\implies \left(\dot{\mathbf{f}}_{i}^{\mathsf{T}} \mathbf{C}_{i}\right)_{\boldsymbol{b}_{i} \boldsymbol{b}_{i}} = \left(\mathbf{I}_{n_{i}} \otimes \mathbf{Z}_{i}^{\mathsf{T}}\right) \left\{\phi\left(\mathbf{T}_{i} \otimes \boldsymbol{\tau}_{i}\right) + \left(\mathbf{C}_{i}^{\mathsf{T}} \otimes \mathbf{I}_{n_{i}}\right) \mathbf{M}_{i} + \phi \operatorname{vec}\left(\mathbf{T}_{i}\right) \boldsymbol{\tau}_{i}^{\mathsf{T}} + \phi \boldsymbol{\Omega}_{i} \mathbf{M}_{i}\right\} \mathbf{Z}_{i}.$$

 $\mathbf{f}_i$ : From Table B.2 we have that  $d \mathbf{f}_i = \phi \mathbf{T}_i \mathbf{Z}_i d \mathbf{b}_i$  and  $d \mathbf{Z}_i^\top \mathbf{T}_i = (\mathbf{I}_{n_i} \otimes \mathbf{Z}_i^\top) \mathbf{\Omega}_i \mathbf{M}_i \mathbf{Z}_i d \mathbf{b}_i$ . Thus,

$$(\mathbf{f}_i)_{\boldsymbol{b}_i \boldsymbol{b}_i} = \phi \left( \mathbf{I}_{n_i} \otimes \mathbf{Z}_i^{\mathsf{T}} \right) \mathbf{\Omega}_i \mathbf{M}_i \mathbf{Z}_i.$$

 $\dot{\mathbf{f}}_i$ : Note that  $d \dot{\mathbf{f}}_i = \phi^{-1} d \mathbf{f}_i$ . The procedure is as before and we have

$$(\dot{\mathbf{f}}_i)_{\boldsymbol{b}_i \boldsymbol{b}_i} = (\mathbf{I}_{n_i} \otimes \mathbf{Z}_i^{\mathsf{T}}) \boldsymbol{\Omega}_i \mathbf{M}_i \mathbf{Z}_i.$$

#### 130 APPENDIX B

# Appendix C

# Within-subject correlations

Let us consider the GLMM as follows

$$Y_{ij} | \boldsymbol{b}_i \stackrel{\text{ind}}{\sim} \text{EF}(\mu_{ij}; \phi),$$
  

$$g(\mu_{ij}) = \boldsymbol{x}_{ij}^{\mathsf{T}} \boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\mathsf{T}} \boldsymbol{b}_i,$$
  

$$\boldsymbol{b}_i \stackrel{\text{iid}}{\sim} N_q(\boldsymbol{0}; \boldsymbol{\Xi}) \quad \text{for} \quad j = 1, 2, \dots, n_i; i = 1, 2, \dots, N_i$$

Thus, for 1 = 1, ..., N,  $1 \le r, s \le n_i$  and  $r \ne s$ , we have that

$$\operatorname{Cov}(y_{ir}, y_{is}) = \operatorname{Cov} \{ \operatorname{E}(Y_{ir} | \boldsymbol{b}_i\}, \operatorname{E}(Y_{is} | \boldsymbol{b}_i)) + \operatorname{E} \{ \operatorname{Cov}(Y_{ir}, Y_{is} | \boldsymbol{b}_i) \} \\ \operatorname{Cov} \{ \operatorname{g}^{-1}(\boldsymbol{x}_{ir}^{\top} \boldsymbol{\beta} + \boldsymbol{z}_{ir}^{\top} \boldsymbol{b}_i), \operatorname{g}^{-1}(\boldsymbol{x}_{is}^{\top} \boldsymbol{\beta} + \boldsymbol{z}_{is}^{\top} \boldsymbol{b}_i) \}.$$

Also,

$$Var (Y_{ir}) = E \{Var (Y_{ir} | \boldsymbol{b}_i)\} + Var \{E (Y_{ir} | \boldsymbol{b}_i)\}$$
$$= E \{\phi^{-1} V (\mu_{ir})\} + Var (\mu_{ir}),$$

where  $V(\mu_{ir})$  is the variance function and  $\mu_{ir} = g^{-1} (\mathbf{x}_{ir}^{T} \boldsymbol{\beta} + \mathbf{z}_{ir}^{T} \boldsymbol{b}_{i})$ .

### Logarithmic link

A special case occurs when  $g(\cdot) = \log(\cdot)$ . In this case, we obtain the model

$$Y_{ij} | \boldsymbol{b}_i \stackrel{\text{ind}}{\sim} \text{EF}(\mu_{ij}; \boldsymbol{\phi}),$$
  

$$\log(\mu_{ij}) = \boldsymbol{x}_{ij}^{\mathsf{T}} \boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\mathsf{T}} \boldsymbol{b}_i,$$
  

$$\boldsymbol{b}_i \stackrel{\text{iid}}{\sim} N_q(\boldsymbol{0}; \boldsymbol{\Xi}) \quad \text{for} \quad j = 1, 2, \dots, n_i; i = 1, 2, \dots, N$$

and

$$Cov (y_{ir}, y_{is}) = Cov \left\{ exp \left( \boldsymbol{x}_{ir}^{\top} \boldsymbol{\beta} + \boldsymbol{z}_{ir}^{\top} \boldsymbol{b}_{i} \right), exp \left( \boldsymbol{x}_{is}^{\top} \boldsymbol{\beta} + \boldsymbol{z}_{is}^{\top} \boldsymbol{b}_{i} \right) \right\}$$

$$= exp \left\{ \left( \boldsymbol{x}_{ir}^{\top} + \boldsymbol{x}_{is}^{\top} \right) \boldsymbol{\beta} \right\} Cov \left\{ exp \left( \boldsymbol{z}_{ir}^{\top} \boldsymbol{b}_{i} \right), exp \left( \boldsymbol{z}_{is}^{\top} \boldsymbol{\beta} \right) \right\}$$

$$= exp \left\{ \left( \boldsymbol{x}_{ir}^{\top} + \boldsymbol{x}_{is}^{\top} \right) \boldsymbol{\beta} \right\} \left( E \left[ exp \left\{ \left( \boldsymbol{z}_{ir}^{\top} + \boldsymbol{z}_{is}^{\top} \right) \boldsymbol{b}_{i} \right\} \right] - E \left\{ exp \left( \boldsymbol{z}_{ir}^{\top} \boldsymbol{b}_{i} \right) \right\} E \left\{ exp \left( \boldsymbol{z}_{is}^{\top} \boldsymbol{b}_{i} \right) \right\} \right)$$

$$= exp \left\{ \left( \boldsymbol{x}_{ir}^{\top} + \boldsymbol{x}_{is}^{\top} \right) \boldsymbol{\beta} \right\} \left[ exp \left\{ \frac{1}{2} \left( \boldsymbol{z}_{ir} + \boldsymbol{z}_{is} \right)^{\top} \Xi \left( \boldsymbol{z}_{ir} + \boldsymbol{z}_{is} \right) \right\} - exp \left( \frac{1}{2} \boldsymbol{z}_{ir}^{\top} \Xi \boldsymbol{z}_{ir} \right) exp \left( \frac{1}{2} \boldsymbol{z}_{is}^{\top} \Xi \boldsymbol{z}_{is} \right) \right]$$

$$= exp \left\{ \left( \boldsymbol{x}_{ir}^{\top} + \boldsymbol{x}_{is}^{\top} \right) \boldsymbol{\beta} \right\} exp \left\{ \frac{1}{2} \left( \boldsymbol{z}_{ir}^{\top} \Xi \boldsymbol{z}_{ir} + \boldsymbol{z}_{is}^{\top} \Xi \boldsymbol{z}_{is} \right) \right\} \left\{ exp \left( \boldsymbol{z}_{ir}^{\top} \Xi \boldsymbol{z}_{is} \right) - 1 \right\}.$$
(C.1)

Note that in this procedure was used the moment generating function for the random vector  $\boldsymbol{b}_i \sim N_q(\boldsymbol{0}; \boldsymbol{\Xi})$ , which is given by

$$\mathbf{M}_{\boldsymbol{b}_i}(\boldsymbol{t}) = \exp\left(\frac{1}{2}\boldsymbol{t}^\top \Xi \, \boldsymbol{t}\right), \quad \boldsymbol{t} \in \mathbb{R}^q \quad \text{and} \quad i = 1, \dots, N.$$
(C.2)

#### Gamma - normal GLMM

Let us consider the GLMM as follows

$$Y_{ij} | \boldsymbol{b}_i \stackrel{\text{ind}}{\sim} G(\mu_{ij}; \phi),$$
  

$$\log(\mu_{ij}) = \boldsymbol{x}_{ij}^{\mathsf{T}} \boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\mathsf{T}} \boldsymbol{b}_i,$$
  

$$\boldsymbol{b}_i \stackrel{\text{iid}}{\sim} N_q(\boldsymbol{0}; \boldsymbol{\Xi}) \quad \text{for} \quad j = 1, 2, \dots, n_i; i = 1, 2, \dots, N.$$

In this case the variance matrix is given by  $V(\mu_{ir}) = \mu_{ir}^2$  and therefore, for  $1 \le i \le N$  and  $1 \le r \le n_i$  we have that

$$\begin{aligned} \operatorname{Var} \left( \mathbf{Y}_{ir} \right) &= \phi^{-1} \operatorname{E} \left( \mu_{ir}^{2} \right) + \operatorname{Var} \left( \mu_{ir} \right) \\ &= \phi^{-1} \operatorname{E} \left( \mu_{ir}^{2} \right) + \operatorname{E} \left( \mu_{ir}^{2} \right) - \left\{ \operatorname{E} \left( \mu_{ir} \right) \right\}^{2} \\ &= \left( \phi^{-1} + 1 \right) \operatorname{E} \left( \mu_{ir}^{2} \right) - \left\{ \operatorname{E} \left( \mu_{ir} \right) \right\}^{2} \\ &= \left( \phi^{-1} + 1 \right) \exp \left( 2 \, \boldsymbol{x}_{ir}^{\mathsf{T}} \boldsymbol{\beta} \right) \operatorname{E} \left\{ \exp \left( 2 \, \boldsymbol{z}_{ir}^{\mathsf{T}} \boldsymbol{b}_{i} \right) \right\} - \exp \left( 2 \, \boldsymbol{x}_{ir}^{\mathsf{T}} \boldsymbol{\beta} \right) \left[ \operatorname{E} \left\{ \exp \left( \boldsymbol{z}_{ir}^{\mathsf{T}} \, \boldsymbol{b}_{i} \right) \right\} \right]^{2} \\ &= \exp \left( 2 \, \boldsymbol{x}_{ir}^{\mathsf{T}} \boldsymbol{\beta} \right) \left( \left( \phi^{-1} + 1 \right) \operatorname{E} \left\{ \exp \left( 2 \, \boldsymbol{z}_{ir}^{\mathsf{T}} \, \boldsymbol{b}_{i} \right) \right\} - \left[ \operatorname{E} \left\{ \exp \left( \boldsymbol{z}_{ir}^{\mathsf{T}} \, \boldsymbol{b}_{i} \right) \right\} \right]^{2} \right) \\ &= \exp \left( 2 \, \boldsymbol{x}_{ir}^{\mathsf{T}} \boldsymbol{\beta} \right) \left\{ \left( \phi^{-1} + 1 \right) \exp \left( 2 \, \boldsymbol{z}_{ir}^{\mathsf{T}} \, \boldsymbol{\Xi}_{ir} \right) - \exp \left( \boldsymbol{z}_{ir}^{\mathsf{T}} \, \boldsymbol{\Xi}_{ir} \right) \right\} . \end{aligned}$$
(C.3)

Note that in the procedure above was used the expression (C.2).

Thus, from expressions (C.1) and (C.3), we can conclude that

$$\rho_{rs}^{(i)} = \operatorname{Corr}\left(Y_{ir}, Y_{is}\right) = \frac{\operatorname{Cov}\left(Y_{ir}, Y_{is}\right)}{\left\{\operatorname{Var}\left(Y_{ir}\right)\right\}^{1/2} \left\{\operatorname{Var}\left(Y_{is}\right)\right\}^{1/2}}$$

$$= \frac{\exp\left\{\left(x_{ir}^{\mathsf{T}} + x_{is}^{\mathsf{T}}\right)\boldsymbol{\beta}\right\} \exp\left\{\frac{1}{2}\left(z_{ir}^{\mathsf{T}} \Xi z_{ir} + z_{is}^{\mathsf{T}} \Xi z_{is}\right)\right\} \left\{\exp\left(z_{ir}^{\mathsf{T}} \Xi z_{is}\right) - 1\right\}}{\left\{\operatorname{Var}\left(Y_{ir}\right)\right\}^{1/2} \left\{\operatorname{Var}\left(Y_{is}\right)\right\}^{1/2}} \\ = \frac{\exp\left\{\frac{1}{2}\left(z_{ir}^{\mathsf{T}} \Xi z_{ir} + z_{is}^{\mathsf{T}} \Xi z_{is}\right)\right\} \left\{\exp\left(z_{ir}^{\mathsf{T}} \Xi z_{is}\right) - 1\right\}}{\left\{\left(\phi^{-1} + 1\right)\exp\left(2z_{ir}^{\mathsf{T}} \Xi z_{ir}\right) - \exp\left(z_{ir}^{\mathsf{T}} \Xi z_{is}\right)\right\}^{1/2} \left\{\left(\phi^{-1} + 1\right)\exp\left(2z_{is}^{\mathsf{T}} \Xi z_{is}\right) - \exp\left(z_{is}^{\mathsf{T}} \Xi z_{is}\right)\right\}^{1/2}} \\ = \frac{\exp\left(z_{ir}^{\mathsf{T}} \Xi z_{is}\right) - 1}{\left\{\left(\phi^{-1} + 1\right)\exp\left(z_{ir}^{\mathsf{T}} \Xi z_{is}\right) - 1\right\}^{1/2} \left\{\left(\phi^{-1} + 1\right)\exp\left(z_{is}^{\mathsf{T}} \Xi z_{is}\right) - 1\right\}^{1/2}}.$$

#### Gamma-normal random intercept GLMM

The expression above may be simplified by considering a gamma-normal random intercept GLMM with logarithmic link. That is, if we consider the model

$$Y_{ij} | b_i \stackrel{\text{ind}}{\sim} G(\mu_{ij}; \phi),$$

$$\log(\mu_{ij}) = \mathbf{x}_{ij}^{\mathsf{T}} \boldsymbol{\beta} + b_i,$$

$$b_i \stackrel{\text{iid}}{\sim} N(0; \xi) \quad \text{for} \quad j = 1, 2, \dots, n_i; i = 1, 2, \dots, N,$$
(C.4)

then the correlation among two observations within *ith* subject, induced by the assumption of the random effects distribution, is given by

$$\rho_{rs}^{(i)} = \rho = \operatorname{Corr}\left(Y_{ir}, Y_{is}\right) = \frac{\exp\left(\xi\right) - 1}{\left(\phi^{-1} + 1\right)\exp\left(\xi\right) - 1},\tag{C.5}$$

for  $1 \le i \le N$ ,  $1 \le r$ ,  $s \le n_i$  and  $r \ne s$ .

Several conclusions may be obtained from (C.5). Note that

- Under the model (C.4) the correlation structure within *ith* subject is a uniform correlation. That is, the  $n_i \times n_i$  correlation matrix for the *ith* subject assumes the form
  - $\begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix}.$
- Under the model (C.4), the structure of the correlation matrix in each subject is the same and only depends on a parameter denoted by *ρ* and exposed in (C.5). In general, this is not true.
- It is well known that the correlation coefficient is a measure of two random

variables whose values are between -1 and 1. However, from (C.5), we can see that  $0 < \rho < 1$  under the model (C.4). Also, when  $\xi \to \infty$  we have that  $\rho \to \frac{\phi}{\phi+1}$ .

#### Poisson - normal GLMM

Let us consider the GLMM as follows

$$Y_{ij} | \boldsymbol{b}_i \stackrel{\text{ind}}{\sim} \operatorname{Po}(\mu_{ij}),$$
  

$$\log(\mu_{ij}) = \boldsymbol{x}_{ij}^{\mathsf{T}} \boldsymbol{\beta} + \boldsymbol{z}_{ij}^{\mathsf{T}} \boldsymbol{b}_i,$$
  

$$\boldsymbol{b}_i \stackrel{\text{iid}}{\sim} \operatorname{N}_q(\boldsymbol{0}; \boldsymbol{\Xi}) \quad \text{for} \quad j = 1, 2, \dots, n_i; i = 1, 2, \dots, N.$$

In this case the variance matrix is given by  $V(\mu_{ir}) = \mu_{ir}$  and  $\phi = 1$ . Therefore for  $1 \le i \le N$  and  $1 \le r \le n_i$  we have that

$$\begin{aligned} \operatorname{Var}\left(\mathbf{Y}_{ir}\right) &= \operatorname{E}\left\{\operatorname{V}\left(\mu_{ir}\right)\right\} + \operatorname{Var}\left(\mu_{ir}\right) = \operatorname{E}\left\{\mu_{ir}\right\} + \operatorname{Var}\left(\mu_{ir}\right) \\ &= \exp\left(\mathbf{x}_{ir}^{\mathsf{T}}\boldsymbol{\beta}\right) \operatorname{E}\left\{\exp\left(\mathbf{z}_{ir}^{\mathsf{T}}\boldsymbol{b}_{i}\right)\right\} + \exp\left(2\,\mathbf{x}_{ir}^{\mathsf{T}}\boldsymbol{\beta}\right) \left[\operatorname{E}\left\{\exp\left(2\,\mathbf{z}_{ir}^{\mathsf{T}}\,\boldsymbol{b}_{i}\right)\right\} - \operatorname{E}\left\{\exp\left(z_{ir}^{\mathsf{T}}\,\boldsymbol{b}_{i}\right)\right\}^{2}\right] \\ &= \exp\left(2\,\mathbf{x}_{ir}^{\mathsf{T}}\,\boldsymbol{\beta}\right) \left[\exp\left(-\mathbf{x}_{ir}^{\mathsf{T}}\,\boldsymbol{\beta}\right) \operatorname{E}\left\{\exp\left(\mathbf{z}_{ir}^{\mathsf{T}}\,\boldsymbol{b}_{i}\right)\right\} + \operatorname{E}\left\{\exp\left(2\,\mathbf{z}_{ir}^{\mathsf{T}}\,\boldsymbol{b}_{i}\right)\right\} - \operatorname{E}\left\{\exp\left(z_{ir}^{\mathsf{T}}\,\boldsymbol{b}_{i}\right)\right\}^{2}\right] \\ &= \exp\left(2\,\mathbf{x}_{ir}^{\mathsf{T}}\,\boldsymbol{\beta}\right) \left\{\exp\left(-\mathbf{x}_{ir}^{\mathsf{T}}\,\boldsymbol{\beta}\right) \exp\left(\frac{1}{2}\mathbf{z}_{ir}^{\mathsf{T}}\,\mathbf{\Xi}\,\mathbf{z}_{ir}\right) + \exp\left(2\mathbf{z}_{ir}^{\mathsf{T}}\,\mathbf{\Xi}\,\mathbf{z}_{ir}\right) - \exp\left(\mathbf{z}_{ij}^{\mathsf{T}}\,\mathbf{\Xi}\,\mathbf{z}_{ir}\right)\right\} \\ &= \exp\left(2\,\mathbf{x}_{ir}^{\mathsf{T}}\,\boldsymbol{\beta}\right) \exp\left(\mathbf{z}_{ir}^{\mathsf{T}}\,\mathbf{\Xi}\,\mathbf{z}_{ir}\right) \left\{\exp\left(-\mathbf{x}_{ir}^{\mathsf{T}}\,\boldsymbol{\beta}\right) \exp\left(-\frac{1}{2}\mathbf{z}_{ir}^{\mathsf{T}}\,\mathbf{\Xi}\,\mathbf{z}_{ir}\right) + \exp\left(\mathbf{z}_{ir}^{\mathsf{T}}\,\mathbf{\Xi}\,\mathbf{z}_{ir}\right) - 1\right\}. \end{aligned}$$

$$(C.6)$$

Note that in the procedure above was used the expression (C.2).

Thus, from expressions (C.1) and (C.6), we can conclude that

$$\begin{split} \rho_{rs}^{(i)} &= \frac{\operatorname{Cov}\left(Y_{ir}, Y_{is}\right)}{\left\{\operatorname{Var}\left(Y_{ir}\right)\right\}^{1/2} \left\{\operatorname{Var}\left(Y_{is}\right)\right\}^{1/2}} \\ &= \frac{\exp\left\{\left(x_{ir}^{\top} + x_{is}^{\top}\right)\beta\right\} \exp\left\{\frac{1}{2}\left(z_{ir}^{\top} \Xi z_{ir} + z_{is}^{\top} \Xi z_{is}\right)\right\} \left\{\exp\left(z_{ir}^{\top} \Xi z_{is}\right) - 1\right\}}{\left\{\operatorname{Var}\left(Y_{ir}\right)\right\}^{1/2} \left\{\operatorname{Var}\left(Y_{is}\right)\right\}^{1/2}} \\ &= \frac{\exp\left(z_{ir}^{\top} \Xi z_{is}\right) - 1}{\left\{\exp\left(z_{ir}^{\top} \Xi z_{ir}\right) - 1\right\}^{1/2} \left\{\exp\left(z_{is}^{\top} \Xi z_{is}\right) - 1\right\}^{1/2} \left\{\exp\left(-x_{is}^{\top} \beta\right) k_{is} + 1\right\}^{1/2}}, \end{split}$$

where

$$k_{ir} = \left\{ \exp\left(\frac{3}{2} z_{ir}^{\mathsf{T}} \Xi z_{ir}\right) - \exp\left(\frac{1}{2} z_{ir}^{\mathsf{T}} \Xi z_{ir}\right) \right\}^{-1} \text{ and } k_{is} = \left\{ \exp\left(\frac{3}{2} z_{is}^{\mathsf{T}} \Xi z_{is}\right) - \exp\left(\frac{1}{2} z_{is}^{\mathsf{T}} \Xi z_{is}\right) \right\}^{-1}.$$

#### Poisson - normal random intercept GLMM

The expression above may be simplified by considering a Poisson-normal random intercept GLMM with logarithmic link. That is, if we consider the model

$$Y_{ij} | b_i \stackrel{\text{ind}}{\sim} \operatorname{Po}(\mu_{ij}; \phi),$$

$$\log(\mu_{ij}) = \mathbf{x}_{ij}^{\mathsf{T}} \boldsymbol{\beta} + b_i,$$

$$b_i \stackrel{\text{iid}}{\sim} \operatorname{N}(0; \xi) \quad \text{for} \quad j = 1, 2, \dots, n_i; i = 1, 2, \dots, N,$$
(C.7)

the correlation among two observations within *ith* subject, induced by the assumption of the random effects distribution, is given by

$$\rho_{rs}^{(i)} = \operatorname{Corr}\left(Y_{ir}, Y_{is}\right) = \frac{1}{\sqrt{\exp\left(-\boldsymbol{x}_{ir}^{\top}\boldsymbol{\beta}\right) k + 1} \sqrt{\exp\left(-\boldsymbol{x}_{is}^{\top}\boldsymbol{\beta}\right) k + 1}}, \quad (C.8)$$

for  $1 \le i \le N$ ,  $1 \le r$ ,  $s \le n_i$  and  $r \ne s$  and  $k = \left\{ \exp\left(\frac{3}{2}\xi\right) - \exp\left(\frac{1}{2}\xi\right) \right\}^{-1}$ . It is important to highlight the following facts:

- Under the model (C.7) we can see that  $0 < \rho_{rs}^{(i)} < 1$ , for  $1 \le i \le N$  and  $1 \le r, s \le n_i$ .
- Unlike what happens in model (C.4), from equation (C.8), the correlation between two observations of the same subject depends on the associated covariate values. Thus, the correlation structure for the *ith* subject, in general, is unstructured.

#### Binomial negative - normal random intercept GLMM (v fixed)

Let us consider the GLMM as follows

$$Y_{ij} | b_i \stackrel{\text{ind}}{\sim} BN(\mu_{ij}; \nu),$$
  

$$\log(\mu_{ij}) = \mathbf{x}_{ij}^{\mathsf{T}} \boldsymbol{\beta} + b_i,$$
  

$$b_i \stackrel{\text{iid}}{\sim} N_q(\mathbf{0}; \xi) \quad \text{for} \quad j = 1, 2, \dots, n_i; i = 1, 2, \dots, N$$

It is possible to prove that that variance function is  $V(\mu_{ij}) = \mu_{ij}(1 + \nu^{-1}\mu_{ij})$  for  $1 \le j \le n_i$  and  $1 \le i \le N$ . Thus, following a similar procedure to that previously performed, it is possible to prove that the correlation between two observations of the *ith* subject, induced by the assumption of the random effects distribution, is given by

$$\rho_{rs}^{(i)} = \operatorname{Corr}\left(Y_{ir}, Y_{is}\right) = \frac{1}{\sqrt{\exp\left(2\xi\right)\nu^{-1} + \exp\left(\boldsymbol{x}_{ir}^{\top}\boldsymbol{\beta}\right)k + 1}\sqrt{\exp\left(2\xi\right)\nu^{-1} + \exp\left(\boldsymbol{x}_{is}^{\top}\boldsymbol{\beta}\right)k + 1}},$$

where  $k = \left\{ \exp\left(\frac{3}{2}\xi\right) - \exp\left(\frac{1}{2}\xi\right) \right\}^{-1}$ ,  $1 \le r, s \le n_i$ , i = 1, ..., N and  $r \ne s$ . Again,  $0 < \rho_{rs}^{(i)} < 1$  and the correlation structure for the *ith* subject is unstructured,

Again,  $0 < \rho_{rs}^{(i)} < 1$  and the correlation structure for the *ith* subject is unstructured, since  $\rho_{rs}^{(i)}$  depends on the associated covariate values to the observations  $Y_{ir}$  and  $Y_{is}$ .

# Appendix D

## **Implementation** in R software

## To use this function it is necessary to load the following packages. ## **require** (matrixcalc) require (gamlss) ## DESCRIPTION ## ## ## ## Returns potentially influential subjects and observations of a fitted ## ## GLMM through gamlss() function jointly with the re() function of the ## gamlss package. This function is built for several distributions ## ## in the exponential family and several link functions. Among them we ## ## ## have ## ## ## Distribution ## Link functions ## "inverse", "log" and "identity". Gamma(GA) ## ## "log"," sqrt" and "identity". Poisson (PO) ## ## ## Normal (NO) "identity" and "log". ## "logit". Bernoulli ## ## ## Negative binomial(NBI) "log" (fixed nu). ## Inverse Gaussian(IG) "1/mu^2" and "log". ## ## ## ## Currently, this function allows to obtain conformal normal curvature ## ## measures for random intercept GLMMs and random slope ## and intercept ## ## GLMMs. ## ## ## Arguments ## ## ## ## ## model a fitted GLMM object through the gamlss() and re() functions. ## For example, for a random intercept GLMM we have ## ## ## model <- gamlss (response ~ cov1 + cov2 + cov3 + ##

```
##
          re(random=~1|factor), data= ..., family= ...).
                                                                 ##
##
                                                                 ##
   data
          data set used by fitting the model.
##
                                                                 ##
##
                                                                 ##
          use "subjects" to discriminate potentially influential sub-
##
   type
                                                                 ##
          jects and "observations" to highlight potentially influential ##
##
          observations. If this argument is omitted, the function
##
                                                                 ##
          provides two plots for highlighting potentially influential
##
                                                                 ##
          subjects and observations.
##
                                                                 ##
##
                                                                 ##
##
                                                                 ##
local. influence .GMMk-function (model, data, type=c("subjects", "observations")){
 a<-model
 dat<-data
 diagi<-function(a){
   if(length(a)==1)
   {
     da < -as . matrix(a)
   }
   else \{ da < -diag(a) \}
   return (da)
   }
##This function allows to extract the estimated variance-covariance matrix##
         for the fitted model by means of the gamlss() function.
##
                                                                 ##
getSigmab <- function(obj)</pre>
 {
   vc <- VarCorr(obj)
   suppressWarnings(storage.mode(vc) <- "numeric")</pre>
   if(ncol(vc)>2){
     variances<-diag(vc[, "Variance"][1:length(vc[, "Variance"])-1])</pre>
     covariance<-prod(vc[, "StdDev"][1:length(vc[, "StdDev"])-1])*</pre>
     vc[,"Corr"][!is.na(vc[,"Corr"])]
     variances [1,2] <- covariance
     variances [2, 1] <- variances [1, 2]
     covariancematrix<-variances
   } else {
     covariancematrix<-diagi(vc[, "Variance"][1:length(vc[, "Variance"])-1])</pre>
   }
     return (covariancematrix)
 }
*****
 kr<-function(a,b){kronecker(a,b)}
```

```
Ko<-function (q, r)
   if(q==1){
    Kqr<-diag(r)
   else if (r = = 1)
    Kqr < -diag(q)
   ł
   else {
    Kqr<-commutation. matrix(q, r)
   return(Kqr)}#commutationmatrix
 G < -function(q) 
   if (q==1){
    G \leftarrow diag(1)
   }
   else {
    G-duplication.matrix(q)
   return(G) ### duplication matrix
 N<-function (q){Ko(q,q)+diag(q^2)}###Nmatrix
##
       Lambda and Pi functions from definition in section 3.4.
                                                       ##
H \leftarrow function(A,q) \{ kr(diag(ncol(A)), Ko(q, nrow(A))) \% * \% kr(vec(A), diag(q)) \} # Lambda
Gm \leftarrow function(B,m) \{ kr(Ko(ncol(B),m), diag(nrow(B))) \% \% kr(diag(m), vec(B)) \} \# Pi
**********
##
            Extracting some components of the fitted model.
                                                       ##
Y<-as.matrix(a$y)#Response variable
 del<-as.vector(which(is.na(coef(a))))
 X<-model.matrix(a)[,-del]##Design matrix
 beta<-coef(a)[!is.na(coef(a))]##Fixed effects
 p<-length(beta)##Number of fized effects
 d-getSigmab(getSmo(a))##Variance-covariance matrix of random-effects
 q<-ncol(d)##Number of randome-effects
 NT<-nrow(Y)##total number of observations
 piij<-fitted(a, "mu")</pre>
*****
      Extracting the predicted random effects from the fitted model
##
                                                       ##
##
              and building the design matrix Zi.
                                                       ##
if (ncol(ranef(getSmo(a))) > 1)
 {
```

```
slopeglmm<-names(ranef(getSmo(a)))[2]</pre>
   interceptglmm<-names(a$mu.coefSmo[[1]]$coefficients$random)
   zetas<-cbind(rep(1,NT), dat[, colnames(dat)==slopeglmm])</pre>
 } else {
   interceptglmm<-names(a$mu.coefSmo[[1]]$coefficients$random)
   zetas < -as.matrix(rep(1,NT))
        }
 cluster <- rle (as.vector(dat[, colnames(dat)==interceptglmm]))$lengths
 k<-length(cluster)
 u<-matrix(unlist(ranef(getSmo(a))), ncol=nrow(d))
 u \leq -lapply(1:k, function(i)\{u[i, ]\})
##
      Function to obtain separately the observations for each subject.
                                                                     ##
pos<-c(0,cumsum(cluster))</pre>
 divisor <- function (A) {
   if (ncol(A) !=1)
   {
    \operatorname{spt} - \operatorname{function}(i, A) \{B - A[(\operatorname{pos}[i]+1): \operatorname{pos}[i+1]]\}
    B<-as.matrix(B)
    if(ncol(B) = = 1) \{B < -t(B)\}
    else \{B < -B\}
    return(as.matrix(B))}
    lis<-lapply(1:k, spt,A=A)}
   else {
    \operatorname{spt}-function (i, A) {B-A[(pos[i]+1): pos[i+1],]
    return(B)}
    lis<-lapply(1:k, spt,A=A)
   }
   return(lis)}
##
       Function to distinguish the hessian matrix and the matrix of
                                                                     ##
##
               Hessians as described in section 2.4.1.1.
                                                                     ##
jointhessian <- function (A,d){
   kk < -nrow(A)/(d)
   spt<-function(i,A){</pre>
      B < -A[((i-1)*d+1):(i*d),]
      return(B)}
   a<-lapply(1:kk,spt,A=A)
   initial <- a [[1]]
   if(length(a)!=1)
   {
   for(i in 1:(kk-1))
   { b<-cbind(initial,a[[i+1]])</pre>
```

```
initial<-b
  }} else {
   b<-initial
  ł
  return(b)
  }
Separating some components of the fitted model for each subject.
##
                                                  ##
Xi < -divisor(X)
 Yi<-divisor(Y)
 Zi<-divisor(zetas)
 zetas<-Zi
 pii<-divisor(as.matrix(piij))</pre>
Function to obtain the omega matrix defined in the definition 3 of
##
                                                  ##
##
                   section 3.4.
                                                  ##
Om <-function(A)
  if (ncol(A) !=1)
    {
   T \le as.matrix(A[,1])
    for(j in 1:(ncol(A)-1)){
     acu<-direct.sum(T, as.matrix(A[, j+1]))
     T<-acu
    }
  }
  else {T<-as.matrix(A)}
   return(T)
 }
 omegas<-lapply(lapply(cluster, diag),Om)</pre>
Required components for computing the conformal normal curvature for
##
                                                  ##
##
   different distributions in the exponential family and different link
                                                  ##
   functions.
                                                  ##
##
##
  Poisson distribution with logarithmic, identity and square root link
                                                  ##
##
  function.
                                                  ##
if (a family [1] = = "PO")
  phi<-1
  Vi<-lapply(divisor(as.matrix(piij)), diagi)
  l<-Y*log(piij)-piij-lgamma(Y+1)</pre>
  if (a$mu. link=="log"){
   Wi<-Vi
```

```
Ai<-Vi
     Aip<-Ai
     Aip2<-Ai
      roi<-lapply(lapply(cluster, function(a){rep(0,a^2)}),</pre>
                            function (a) { matrix (a, nrow=sqrt(length(a))) })
     roip<-roi
     roipp<-roi
    }
    if(a$mu.link=="identity"){
     Wi<-lapply(divisor(as.matrix(piij^(-1))), diagi)
      Ai<-lapply(cluster, diag)
     Aip<-lapply(lapply(cluster, function(a){rep(0,a^2)}),
                            function(a){matrix(a, nrow=sqrt(length(a)))})
      Aip2<-Aip
      roi <- lapply (divisor (as. matrix (-piij^(-2))), diagi)
      roip<-lapply(divisor(as.matrix(2*piij^(-3))),diagi)</pre>
      roipp<-lapply(divisor(as.matrix(-6*piij^(-4))), diagi)</pre>
    }
    if(a$mu.link=="sqrt"){
     Wi<-lapply(4*cluster, diag)
     Ai<-lapply(divisor(as.matrix(2*sqrt(piij))), diagi)
      Aip<-lapply(2*cluster, diag)</pre>
      Aip2<-lapply(lapply(cluster, function(a){rep(0,a^2)}),
                             function (a) { matrix (a, nrow=sqrt(length(a))) } )
      roi<-lapply(divisor(as.matrix(-2*piij^(-1))), diagi)</pre>
      roip < -lapply(divisor(as.matrix(4*piij^(-3/2))), diagi)
      roipp<-lapply(divisor(as.matrix(-12*piij^(-2))), diagi)</pre>
    }
  }
Bernoulli distribution with canonical link.
##
                                                                           ##
if (a family [1] = = "BI")
    {
    phi<-1
    Vi<-lapply(as.vector(divisor(as.matrix(piij*(1-piij)))), diagi)
    l \leftarrow choose(1,Y) + Y \times log(piij/(1-piij)) + log(1-piij)
    if(a$mu.link=="logit")
    {
   Wi<-Vi
    Ai<-Vi
   Aip<-lapply(divisor(as.matrix(piij*(1-piij)*(1-2*piij))), diagi)
   Aip2<-lapply(divisor(as.matrix((6*piij<sup>2</sup>-6*piij+1)*piij*(1-piij))), diagi)
    roi<-lapply(lapply(cluster, function(a){rep(0,a^2)}),
                            function(a){matrix(a, nrow=sqrt(length(a)))})
```

```
roip<-roi
   roipp<-roi
   }
 }
##
      Negative binomial distribution (with fixed) nu and logarithmic
                                                                 ##
##
      link function.
                                                                 ##
if (a family [1] = = "NBI")
 {
   phi<-1
   nu<-exp(coef(a, "sigma"))</pre>
   Vi<-lapply(divisor(as.matrix(piij*({nu^{-1}}*piij+1))), diagi)
   Wi<-lapply(divisor(as.matrix(piij/(1+nu^{-1}*piij))), diagi)
   Ai<-lapply(pii, diagi)
   Aip<-Ai
   Aip2<-Ai
   roi < -lapply(divisor(as.matrix(-nu^{-1}*piij)^{2})), diagi)
   roip<-lapply(divisor(as.matrix(nu^{-1}*piij*
                     (nu^{-1} * piij -1)/(1+nu^{-1} * piij)^{3})), diagi)
   roipp<-lapply(divisor(as.matrix(-(nu<sup>{-1}</sup>*piij*
         ((nu^{-1}*piij)^{2}-4*nu^{-1}*piij+1))/(1+nu^{-1}*piij)^{4})), diagi)
   l<-phi*Y*log((piij)/(piij+nu))-nu*log(piij+nu)+(lgamma(nu+Y))-</pre>
                     lgamma(Y+1) - lgamma(nu) + nu * log(nu)
 }
##Gamma distribution with logarithmic, inverse and identity link functions.##
if (a family [1] = = "GA")
 {
   phi < -1/(coef(a, "sigma"))^2
   l < -phi*((-Y/piij)-log(piij))+(phi-1)*log(Y)+phi*log(phi)-lgamma(phi)
   lp < -as.vector(Y) * piij^{(-1)} + log((phi * as.vector(Y)) / (piij)) + 1 - digamma(phi)
   Vi<-lapply(divisor(as.matrix(piij^2)), diagi)
   cpp<-lapply(divisor(as.matrix(rep((1/phi)-trigamma(phi),NT)))),
             function(a){sum(a)})
   if(a$mu.link=="log")
   {
     Wi<-lapply(lapply(cluster, function(a){rep(1,a)}), diagi)
     Ai<-lapply(pii, diagi)
     Aip<-Ai
     Aip2<-Ai
     roi <- lapply (Ai, function (a) {- solve (a) })
     roip<-lapply(Vi, function(a){solve(a)})</pre>
```

```
roipp<-lapply(divisor(as.matrix(-2*piij^(-3))), diagi)</pre>
    }
    if(a$mu.link=="inverse")
    {
     Wi<-Vi
    Ai<-lapply(divisor(as.matrix(-piij^2)), diagi)
     Aip<-lapply(divisor(as.matrix(2*piij^3)), diagi)
     Aip2<-lapply(divisor(as.matrix(-6*piij^4)), diagi)</pre>
     roi<-lapply(lapply(cluster, function(a){rep(0,a^2)}),</pre>
                           function(a){matrix(a, nrow=sqrt(length(a)))})
     roip<-roi
     roipp<-roi
    }
    if(a$mu.link=="identity")
    {
     Wi<-lapply(divisor(as.matrix(piij^(-2))),diagi)
     Ai<-lapply(cluster, diag)
     Aip<-lapply(lapply(cluster, function(a){rep(0,a^2)}),
                           function(a){matrix(a, nrow=sqrt(length(a)))})
     Aip2<-Aip
     roi<-lapply(divisor(as.matrix(-2*piij^(-3))), diagi)</pre>
     roip<-lapply(divisor(as.matrix(6*piij^(-4))), diagi)</pre>
     roip<-roip<-lapply(divisor(as.matrix(-24*piij^(-5))), diagi)</pre>
    }
  }
##
        Normal distribution with identity and log link functions.
                                                                         ##
if (a family [1] = = "NO")
    phi < -1/(coef(a, "sigma"))^2
    l<-phi*(piij*Y-0.5*piij^2)-0.5*(log(2*pi*phi^(-1))+phi*Y^2)</pre>
   lp < -(piij * Y - 0.5 * piij^2) - 0.5 * (-phi^{(-1)} + Y^2)
   cpp<-lapply(divisor(as.matrix(rep(-1/(2*phi^2),NT))), function(a){sum(a)})
    Vi<-lapply(cluster, diag)
   if (a$mu. link=="identity"){
   Wi<-Vi
   Ai<-Vi
   Aip<-lapply(lapply(cluster, function(a){rep(0,a^2)}),
                       function(a){matrix(a, nrow=sqrt(length(a)))})
   Aip2<-Aip
    roi<-Aip
    roip<-Aip
    roipp<-Aip
```

```
}
  if (a mu. link == "log"){
     Wi<-lapply(divisor(as.matrix(piij^2)), diagi)
     Ai<-lapply(divisor(as.matrix(piij)), diagi)
     Aip<-Ai
     Aip2<-Ai
     roi<-Ai
     roip<-Ai
     roipp<-Ai
    }
 }
Inverse Gaussian distribution with "1/mu^2" and log link functions.
##
                                                                 ##
if (a family [1] = = "IG")
 phi < -1/(coef(a, "sigma"))^2
 1 < -phi*(-Y/(2*piij^{2})+1/piij) - 0.5*(log((2*pi*Y^{3})/phi)+phi/Y)
 lp < -Y/(2 * piij^{2}) + 1/piij - 0.5 * ((-1/phi) + 1/Y)
 cpp<-lapply(divisor(as.matrix(rep(-(0.5)*phi^{-2},NT))), function(a){sum(a)})
 Vi<-lapply(divisor(as.matrix(piij^3)), diagi)
  if(a\$mu.link == "1/mu^2"){
    Wi<-lapply(divisor(as.matrix((0.25)*piij^3)), diagi)
    Ai < -lapply(divisor(as.matrix(-(0.5)*piij^3)), diagi)
    Aip<-lapply(divisor(as.matrix((3/4)*piij^5)),diagi)
    Aip2<-lapply(divisor(as.matrix(-(15/8)*piij^7)), diagi)
    roi<-lapply(lapply(cluster, function(a){rep(0,a^2)}),</pre>
                       function (a) { matrix (a, nrow = sqrt(length(a))) }
    roip<-roi
    roipp<-roi
      }
  if (a$mu. link=="log"){
    Wi<-lapply(divisor(as.matrix(1/piij)),diagi)
    Ai<-lapply(divisor(as.matrix(piij)), diagi)
    Aip<-Ai
    Aip2<-Ai
    roi<-lapply(divisor(as.matrix(-2*piij^{-2})), diagi)</pre>
    roip<-lapply(divisor(as.matrix(4*piij^{-2})),diagi)</pre>
    roipp<-lapply(divisor(as.matrix(-8*piij^{(-2)})), diagi)
     }
   }
##
  Let's begin with the computation of the conformal normal curvature.
                                                                 ##
```

```
**********
##
            M matrices defined in section 3.6.
                                              ##
mi < -function(a, b, c, d, e, f) \{ diagi(a-b) \% \% c - (d)^{(1/2)} \% (solve(e))^{(1/2)} \% \% \}
 Mi<-mapply(mi,Yi,pii,roi,Wi,Vi,Ai,SIMPLIFY=FALSE)
 Mip<-mapply(mi,Yi,pii,roip,Wi,Vi,Aip,SIMPLIFY=FALSE)</pre>
 Mipp<-mapply(mi,Yi,pii,roipp,Wi,Vi,Aip2,SIMPLIFY=FALSE)
Hessian matrix (or expected Hessian matrix), with respect to bi, of
##
                                              ##
   the joint "log-likelihood" function of yi and bi. (Appendix B.2).
##
                                              ##
hesconj <-function(a,b) \{-phi * t(a) \% * \% b \% * \% a - solve(d)\}
 Hesconi<-mapply(hesconi, zetas, Wi, SIMPLIFY=FALSE)</pre>
R1 function given in the definition 5 in section 3.5.2.
##
                                              ##
Laplaced - function (a, b, c) { as . matrix (a) - kr(diag(nrow(as.matrix(a)))),
 t(vec(diag(q))))%%kr(b, diag(q))%%(kr(diag(ncol(as.matrix(a)))),
 vec(solve(c))))}
##
     Delta matrix under first perturbation scheme. Section 3.7.1.
                                              ##
function (a, b, c, d, e) { t(a) %*% ((b)^{(1/2)}) %*% (solve (c)^{(1/2)} %*% (d-e) }
Funcion1<-mapply(funcion1, Xi, Wi, Vi, Yi, pii, SIMPLIFY=FALSE)
hessiana1<-function (a, b, c, d, e, f) { kronecker (t(a), t(b))%*%c%*%(d-2*e%*%f)%*%b}
H1<-mapply(hessiana1,Xi,zetas,omegas,Mip,roi,Ai,SIMPLIFY=FALSE)
E1<-mapply(Laplaced, Funcion1, H1, Hesconj, SIMPLIFY=FALSE)
d < -as . matrix(d)
funcion2<-function(a){kr(a,a)}
```

```
Function2<-lapply(u,function2)</pre>
```

```
H2<-lapply(rep(1,k), function(a){kr(N(q), diag(q))%*%H(diag(q),q)*a})
E2<-mapply(Laplaced, Funcion2, H2, Hesconj, SIMPLIFY=FALSE)
```

```
scobeta<-lapply(E1, function(a){phi*a})</pre>
```

```
if (a$family[1]=="NBI" || a$family[1]=="PO" || a$family[1]=="BI")
      Fullyscore <- mapply (function (a,b,c) {rbind (a,b)}, scobeta, scov)
    }
  if (a family [1] = "GA" || a family [1] = "NO" || a family [1] = "IG")
    {
        lip<-divisor(as.matrix(lp))</pre>
        Function10<-lapply(lip,function(a){sum(a)})</pre>
       H10<-mapply(function(a,b){t(a)%*%b%*%a},Zi,Mi,SIMPLIFY=FALSE)
        E10<-mapply(Laplaced, Funcion10, H10, Hesconj, SIMPLIFY=FALSE)
        Scoredispersao<-E10
        Fullyscore <-mapply(function(a,b,c){rbind(a,b,c)}, scobeta, scov,
        Scoredispersao)
    }
Delta matrix under second perturbation scheme. Section 3.7.2.
##
                                                                                                                                                      ##
li <- divisor (as. matrix(1))
    bi \leftarrow function(c,d) \{ diag(c) + rep(1,c) \% (d) \}
    Bi<-mapply(bi,cluster,li,SIMPLIFY=FALSE)</pre>
    funcion3<-function(a,b,c,d,e,f)</pre>
                \{t(a)\%\%(b^{(1/2)})\%\%solve(c)^{(1/2)}\%\%diagi(d-e)\%\%f
    Funcion3<-mapply(funcion3, Xi, Wi, Vi, Yi, pii, Bi, SIMPLIFY=FALSE)
    ti <-function(a, b, c, d, e) \{ diagi(a-b) \% * \% solve(c)^{(1/2)} \% * \% d^{(1/2)} \}
    Ti<-mapply(ti,Yi,pii,Vi,Wi,SIMPLIFY=FALSE)
    hessiana3art < -function (a,b,c,d,e,f,g,h,i,j,k){
        phi*kr(diag(a*p), t(b))\%*\%(phi^{-1}*kr(t(kr(c,d)))\%*\%e, diag(a))\%*\%
        e^{*}(f-2*g^{*}(h)+kr(kr(diag(a),k),diag(a)))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(a),k),diag(a))^{*}(h)+kr(kr(diag(
        phi*(H(t(b)\%*\%j,p)\%*\%t(d)\%*\%i\%*\%b) +
        phi*kr(diag(a*p),t(b)%*%i%*%t(e))%*%kr(j,Gm(d,a)%*%rep(1,a))%*%b}
    H3art<-lapply(mapply(hessiana3art, cluster, Zi,
    Bi, Xi, omegas, Mip, roi, Ai, Mi, Ti, Funcion1, SIMPLIFY=FALSE), jointhessian, d=p*q)
    E3art<-mapply(Laplaced, Funcion3, H3art, Hesconj, SIMPLIFY=FALSE)
    Function4art<-lapply(li, function(a){matrix(a, ncol=1)})
    hessiana4art <- function (a, b, c, d) { phi * kr (diag(a), t(b))%*%c%*%d%*%b }
    H4art<-mapply(hessiana4art, cluster, Zi, omegas, Mi, SIMPLIFY=FALSE)
    E4art <-- mapply (Laplaced, Funcion4art, H4art, Hesconj, SIMPLIFY=FALSE)
```

```
scorebetaomega<-mapply(function(a,b,c){phi*a-phi*b%*%t(c)},</pre>
```

```
funcion5<-function(a,b){as.matrix(kr(a,a))%*%t(b)}
Funcion5<-mapply(funcion5,u,li,SIMPLIFY=FALSE)
hossions5arts_function(a,b,c,d,a,f,g,b)[kr(kr(a,diag(g)))]</pre>
```

E3art, E1, E4art, SIMPLIFY=FALSE)

```
H5art<-lapply(mapply(hessiana5art,Funcion4art,cluster,u,Ti,Zi,Funcion2,
omegas,Mi,SIMPLIFY=FALSE),jointhessian,d=q^3)
```

E5art<-mapply(Laplaced, Funcion5, H5art, Hesconj, SIMPLIFY=FALSE)

```
scorecovomega<-mapply(function(a,b,c){(1/2)*t(G(q))%*%
kr(solve(d),solve(d))%*%(a-b%*%t(c))},E5art,E2,E4art,SIMPLIFY=FALSE)</pre>
```

```
taui<-mapply(function(a,b,c,d){a^{1/2}%*%solve(b)^{1/2}%*%(c-d)},Wi,Vi,
Yi,pii,SIMPLIFY=FALSE)
```

```
{
```

```
Funcion14<-mapply(function(a,b){t(a)%*%b},lip,Bi,SIMPLIFY=FALSE)
hessiana14art<-function(a,b,c,d,e,f,g,h){phi*kr(diag(a),t(b))%*%(
    kr(c,d)+phi^{-1}*kr(t(e),diag(a))%*%f%*%g+vec(c)%*%t(d)+
    kr(t(rep(1,a))%*%h,f%*%g))%*%b}</pre>
```

```
H14art<-lapply(mapply(hessiana14art, cluster, Zi, Ti, taui, Bi, omegas, Mi, lip, SIMPLIFY=FALSE), jointhessian, d=q)
E14art<-mapply(Laplaced, Funcion14, H14art, Hesconj, SIMPLIFY=FALSE)
```

```
scorephiomega<-mapply(function(a,b,c){a-b%*%t(c)},E14art,E10,E4art,
SIMPLIFY=FALSE)
```

```
fullyscoreomega<-mapply(function(a,b,c){rbind(a,b,c)},scorebetaomega,
scorecovomega,scorephiomega,SIMPLIFY=FALSE)
```

```
scoreobs<-function(a){
  scores<-a[[1]]
  for(i in 1:(k-1))
  { A<-cbind(scores,a[[i+1]])
   scores<-A
  }
  return(A)}
scorepsiomega<-scoreobs(fullyscoreomega)</pre>
```

```
}
if (a family [1] = "NBI" || a family [1] = "PO" || a family [1] = "BI")
   {
       fullyscoreomega < -mapply(function(a,b){rbind(a,b)}, scorebetaomega,
                                               scorecovomega, SIMPLIFY=FALSE)
       scoreobs<-function(a){</pre>
          scores<-a[[1]]</pre>
          for(i in 1:(k-1))
          { A<-cbind(scores, a[[i+1]])
          scores<-A
          }
          return (A)}
       scorepsiomega<-scoreobs(fullyscoreomega)</pre>
   }
Approximated Hessian matrix for the marginal log-likelihood (2.8).
##
                                                                                                                              ##
       Section 2.2.2.
                                                                                                                              ##
##
Function6<-mapply(function(a,b){t(a)%*%b%*%a},Xi,Mi,SIMPLIFY=FALSE)
hessiana6<-function (a,b,c,d,e,f,g,h) { kr(kr(t(a),t(a))) * b, t(c) } * b, t(c) * 
                             (d-3*(e\%*\%f+g\%*\%h))\%*\%c
H6<-lapply (mapply (hessiana6, Xi, omegas, Zi, Mipp, Ai, roip, Aip, roi, SIMPLIFY=FALSE),
                    jointhessian ,d=p*q)
E6<-mapply(Laplaced, Funcion6, H6, Hesconj, SIMPLIFY=FALSE)
Function7<-lapply(Function1, function(a)(a%*%t(a)))</pre>
hessiana7art < function (a,b,c,d,e,f,g,h,i) { kr(N(p),t(a))%*(kr(kr(t(b),t(b))**c),
diag(d))%*%e%*%(f-2*g%*%h)+kr(diag(p^2), i%*%b)%*%H(diag(p), p)%*%t(b)%*%i)%*%a}
H7art<-lapply(mapply(hessiana7art,Zi,Xi,taui,cluster,omegas,Mip,roi,Ai,Mi,
SIMPLIFY=FALSE), jointhessian, d=p*q)
E7art <- mapply (Laplaced, Function7, H7art, Hesconi, SIMPLIFY=FALSE)
##Approximated Hessian matrix in beta for the marginal log-likelihood (2.8).#
hessianabeta <- function (a, b, c) { (phi<sup>2</sup>) *a-(phi<sup>2</sup>) *b%*%t (b)+phi*c }
Hessianabeta <- mapply (hessianabeta, E7art, E1, E6, SIMPLIFY=FALSE)
Function8<–lapply (Function2, function(a) \{a\%*\%t(a)\})
```

 $\begin{aligned} &\text{hessiana8art} < -\text{function}(a,b) \{ kr(N(q^2), diag(q)) \% (kr(N(q) \% kr(diag(q), a)), \\ &\text{vec}(t(kr(diag(q), a)) \% \% (q)) + kr(H(t(b), q^2) \% \% (q), diag(q)) \% \% H(diag(q), q)) \} \end{aligned}$ 

H8art<-lapply(mapply(hessiana8art,u,Funcion2,SIMPLIFY=FALSE),jointhessian,d=q^3) E8art<-mapply(Laplaced,Funcion8,H8art,Hesconj,SIMPLIFY=FALSE) PP<-kr(diag(q),Gm(solve(d),q))+kr(H(solve(d),q),diag(q))

```
(1/2) * t(G(q)) % * (diag(q^2) - kr(t(a), diag(q^2)) * PP + (1/2) * solve(kr(d,d)) * (b-a% * t(a))) * solve(kr(d,d)) * G(q)
```

```
}
```

```
{
```

Funcion11<-lapply(Funcion10, function(a){a^2})</pre>

hessianallart<-function(a,b,c,d){2\*t(a)%\*%(b\*c+d%\*%t(d))%\*%a} H11art<-mapply(hessianallart,Zi,Funcion10,Mi,taui,SIMPLIFY=FALSE) E11art<-mapply(Laplaced,Funcion11,H11art,Hesconj,SIMPLIFY=FALSE) Hessianadispersion<-mapply(function(a,b,c){a+b-c^2},cpp,E11art,E10, SIMPLIFY=FALSE)

```
}
```

{

```
\begin{aligned} &\text{hessiana9art} < - \textit{function}(a, b, c, d, e, f, g, h, i, j) \{ kr(kr(\textit{diag}(q^2), t(a)), t(b)) \% \% \\ &(kr(c, d) \% \% (e-2*f \% \% g) \% \% b+kr(N(q), vec(h)) \% \% kr(\textit{diag}(q), i)) + \\ &kr(kr(N(q), t(a) \% \% j), \textit{diag}(q)) \% \% H(\textit{diag}(q), q) + H(kr(\textit{diag}(q), t(i)) \% \% N(q), p) \% \% \\ &t(a) \% \% h \% \% b  \} \end{aligned}
```

```
H9art<-lapply(mapply(hessiana9art,Xi,Zi,Funcion2,omegas,Mip,roi,Ai,Mi,u,taui,
SIMPLIFY=FALSE),jointhessian,d=p*q)
```

```
Funcion12<-mapply(function(a,b){a%*%b}, Funcion1, Funcion10, SIMPLIFY=FALSE)</pre>
```

```
hessiana12art<-function (a, b, c, d, e, f, g, h, i, j) { t (kr (a, b))%*% 
(c*d%*%(e-2*f%*%g)+vec (h)%*%t (i)+N(j)%*%kr (h, i))%*%b}
```

H12art<-mapply(hessiana12art,Xi,Zi,Funcion10,omegas,Mip,roi,Ai,Mi,taui, cluster,SIMPLIFY=FALSE)

```
H13art<-mapply(hessiana13art,Funcion10,u,taui,Zi,Funcion2,Mi,
SIMPLIFY=FALSE)
```

Hessianacovphi<-mapply(hessianacovphi, E13art, E2, E10, SIMPLIFY=FALSE)

{

}

}

hessianmodelbcd<-function (a,b,c){rbind(cbind(a,b),cbind(t(b),c))} Hessianmodelbcd<-Reduce('+',mapply(hessianmodelbcd,Hessianabeta,Covfixhess,

```
Hessianacov , SIMPLIFY=FALSE))
```

InvHessianmodelbcd<-solve(Hessianmodelbcd)

```
if (a family[1] = "GA" || a family[1] = "NO" || a family[1] = "IG")
   {
      hessianmodelbcd < -function (a,b,c,d,e,f) { rbind (cbind (a,b,c), cbind (t(b),d,e),
                           cbind(t(c), t(e), f))}
      Hessianmodelbcd<-Reduce('+', mapply(hessianmodelbcd, Hessianabeta, Covfixhess,
                            Hessianabetaphi, Hessianacov, Hessianacovphi,
                            Hessianadispersion , SIMPLIFY=FALSE))
      InvHessianmodelbcd<-solve(Hessianmodelbcd)
   }
Conformal normal curvatures for each subject and observation.
##
                                                                                                                          ##
conformal < -function(a,b) \{ abs(diag((t(a)) * b) * abs(diag((t(a)) * b) * abs(diag((t(a)) * b) * abs(diag((t(a)) * b) * b) * abs(diag((t(a)) * b) * b) * abs(diag((t(a)) * abs(diag
                              %*%(t(a)%*%b%*%a))))))))
   ILTOTALconf<-conformal (Fullyscore, InvHessianmodelbcd)
   cuttoflogconf<-2*mean(ILTOTALconf)
confobs<-conformal (scorepsiomega, InvHessianmodelbcd)
cuttofconformal<-mean(confobs)+4*sd(confobs)
##
                           Index plots to detect influential points.
                                                                                                                          ##
if(!(missing(type))){
if (type=="subjects"){
   par(mfrow=c(1,1))
par(mar = c(5.1, 5.1, 4.1, 2.1))
plot(1:k, ILTOTALconf, ylab=expression(B[i](psi)), xlab="Subjects",
vlim = c(min(ILTOTALconf) - 0.000001, max(cuttoflogconf, max(ILTOTALconf)) + 0.01),
           cex = .60, cex.lab = 1, cex.axis = 0.8, cex.main = 1.2, main = "", pch = 19)
   #abline(cuttoflogconf, 0, lty = 2)
identify(1:k, as.vector(ILTOTALconf), n=7, cex=0.7)
}
if (type=="observations"){
ss<-unlist(lapply(1:k, function(i){1:cluster[i]}))</pre>
labelsprint= paste("(", as.vector(dat[,colnames(dat)==interceptglmm]), ",",
                                ss, ")", sep="")
par(mfrow=c(1,1))
par(mar = c(5.1, 5.1, 4.1, 2.1))
plot(as.vector(rep(c(1:k), cluster)), as.vector(confobs), type = "p", lty = 1,
lwd = 1,ylab=expression(B[ij](psi)),xlab="Subjects",
vlim=c(min(confobs)-0.00000001, max(cuttofconformal, max(confobs))+0.0001),
```

```
cex = .60, cex . lab = 1, cex . axis = 0.8, cex . main = 1.2, main = "", pch = 18)
  #abline(cuttofconformal,0,lty=2)
identify(as.vector(rep(c(1:k), cluster)), labels=labelsprint,
             as.vector(confobs), n=7, cex = 0.7)
}
}else{
  par(mfrow=c(1,2))
  par(mar = c(5.1, 5.1, 4.1, 2.1))
  plot(1:k,ILTOTALconf,ylab=expression(B[i](psi)),xlab="Subjects",
  ylim=c(min(ILTOTALconf) - 0.000001, max(cuttoflogconf, max(ILTOTALconf)) + 0.01),
       cex = .60, cex.lab = 1, cex.axis = 0.8, cex.main = 1.2, main = "", pch = 19)
  identify (1:k, as.vector(ILTOTALconf), n=7, cex=0.7)
  ss<-unlist(lapply(1:k, function(i){1:cluster[i]}))</pre>
  labelsprint= paste("(", as.vector(dat[,colnames(dat)==interceptglmm]), ",",
        ss, ")", sep="")
  plot(as.vector(rep(c(1:k), cluster)), as.vector(confobs), type = "p", lty = 1,
 lwd = 1,ylab=expression(B[ij](psi)),xlab="Subjects",ylim=c(min(confobs)-
        0.00000001, max(cuttofconformal, max(confobs))+0.0001),
       cex = .60, cex.lab = 1, cex.axis = 0.8, cex.main = 1.2, main = "", pch = 18)
  identify(as.vector(rep(c(1:k), cluster)), labels=labelsprint,
  as. vector (confobs), n=7, cex = 0.7)
}
  }
```

#### 154 APPENDIX D

# Bibliography

- Abadir and Magnus (2005) Karim M. Abadir and Jan R. Magnus. *Matrix Algebra*. Cambridge University Press. From page 11, 12, 13, 21, 22, 27, 112, 117, 122
- **Aitkin (1999)** Murray Aitkin. A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics*, 55(1):117–128. From page 1
- Breslow and Clayton (1993) N. E. Breslow and D. G. Clayton. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association*, 88 (421):9–25. From page 1, 7, 8, 14, 18, 19, 26, 30, 31
- **Broström and Holmberg (2011)** Göran Broström and Henrik Holmberg. Generalized linear models with clustered data: Fixed and random effects models. *Computational Statistics & Data Analysis*, 55(12):3123–3134. From page 83
- Buuren and Fredriks (2001) Stef van Buuren and Miranda Fredriks. Worm plot: a simple diagnostic device for modelling growth reference curves. *Statistics in Medicine*, 20(8): 1259–1277. From page 65
- **Cook (1986)** R. Dennis Cook. Assessment of local influence. *Journal of the Royal Statistical Society. Series B (Methodological)*, 48(2):133–169. From page 2, 3, 5, 9, 10, 18, 19, 20, 45
- **Cook (1977)** R Dennis Cook. Detection of influential observation in linear regression. *Technometrics*, 19(1):15–18. From page 2, 86
- **Cook and Weisberg (1982)** R Dennis Cook and Sanford Weisberg. *Residuals and Influence in Regression*. New York: Chapman and Hall. From page 2
- **De Jong and Heller (2008)** Piet De Jong and Gillian Z Heller. *Generalized Linear Models for Insurance Data*. Cambridge University Press. From page 89
- **Demidenko (2013)** Eugene Demidenko. *Mixed Models: Theory and Applications with R*. John Wiley, second ed. From page 1, 3, 5, 7, 11, 14, 15, 16, 26, 29

- **Diggle** *et al.* **(1994)** Peter J Diggle, Kung-Yee Liang and Scott L. Zeger. *Analysis of Longitudinal Data*. Oxford University Press. From page 61
- Espinheira et al. (2008) Patrícia L Espinheira, Silvia LP Ferrari and Francisco Cribari-Neto. Influence diagnostics in beta regression. *Computational Statistics and Data Analysis*, 52(9):4417–4431. From page 20
- **Fong** *et al.* **(2010)** Youyi Fong, Håvard Rue and Jon Wakefield. Bayesian inference for generalized linear mixed models. *Biostatistics*, 11(3):397–412. From page 1
- Harville (1997) David A. Harville. *Matrix Algebra From a Statician's Perspective*. Springer. From page 12, 21, 22
- Hubert and Vandervieren (2008) Mia Hubert and Ellen Vandervieren. An adjusted boxplot for skewed distributions. *Computational statistics & data analysis*, 52(12):5186– 5201. From page 61, 77
- Jiang (2007) Jiming Jiang. *Linear and Generalized Linear Mixed Models and their Applications*. Springer Science & Business Media. From page 1
- Lee and Xu (2004) Sik-Yum Lee and Liang Xu. Influence analyses of nonlinear mixedeffects models. *Computational Statistics & Data Analysis*, 45(2):321–341. From page 19
- Lee *et al.* (2006) Sik-Yum Lee, Bin Lu and Xin-Yuan Song. Assessing local influence for nonlinear structural equation models with ignorable missing data. *Computational statistics & Data Analysis*, 50(5):1356–1377. From page 19, 20, 74, 101
- Lesaffre and Verbeke (1998) Emmanuel Lesaffre and Geert Verbeke. Local influence in linear mixed models. *Biometrics*, 54(2):570–582. From page 2, 18, 20, 36, 49, 70, 84, 96, 103
- Liang and Zeger (1986) Kung-Yee Liang and Scott L Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22. From page 78
- Magnus and Neudecker (1999) Jan R. Magnus and Heinz Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley, second ed. From page 11, 12, 13, 27, 103, 112, 115, 122
- McCullagh and Nelder (1989) P. McCullagh and J.A. Nelder. *Generalized Linear Models, Second Edition*. Chapman & Hall, second ed. From page 1
- McCulloch (1997) Charles E McCulloch. Maximum likelihood algorithms for generalized linear mixed models. *Journal of the American Statistical Association*, 92(437): 162–170. From page 1

- McCulloch and Neuhaus (2005) Charles E McCulloch and John M Neuhaus. Generalized linear mixed models. *Encyclopedia of Biostatistics*, 3. From page 1, 9
- McCulloch and Neuhaus (2011) Charles E McCulloch and John M Neuhaus. Prediction of random effects in linear and generalized linear models under model misspecification. *Biometrics*, 67(1):270–279. From page 66, 81, 94
- McCulloch and Searle (2000) Charles E. McCulloch and Shayle R. Searle. *Generalized, Linear and Mixed Models*. Wiley, New York. From page 1, 3, 6, 7, 8, 22, 25, 26, 29, 31
- **McGilchrist (1994)** CA McGilchrist. Estimation in generalized mixed models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(1):61–69. From page 1, 8
- Molenberghs and Kenward (2007) Geert Molenberghs and Michael Kenward. *Missing Data in Clinical Studies*. John Wiley & Sons. From page 61
- Myers *et al.* (2002) Raymond H Myers, Douglas C Montgomery and G Geoffrey Vining. *Generalized Linear Models: with Applications in Engineering and the Sciences*. John Wiley & Sons. From page 76, 78, 85
- Nelder et al. (2006) John A Nelder, Yudi Pawitan and Hyong Joo Lee. Generalized Linear Models with Random Effects: Unified Analysis via H-likelihood. Chapman and Hall/CRC. From page 1, 8, 19, 25, 48
- **Ouwens** *et al.* **(2001)** Mario JNM Ouwens, Frans ES Tan and Martijn PF Berger. Local influence to detect influential data structures for generalized linear mixed models. *Biometrics*, 57(4):1166–1172. From page 2, 17, 18
- Paula (2013a) Gilberto A Paula. On diagnostics in double generalized linear models. Computational Statistics & Data Analysis, 68:44–51. From page 20, 106
- **Paula (2013b)** Gilberto Alvarenga Paula. *Modelos de Regressão: com Apoio Computacional*. IME-USP São Paulo. From page 86, 88, 101
- Pawitan (2001) Yudi Pawitan. In all likelihood: Statistical Modelling and Inference using Likelihood. Oxford University Press. From page 19, 31
- **Pinheiro and Bates (2000)** José C. Pinheiro and Douglas M. Bates. *Mixed-effects Models in S and S-PLUS*. Springer, New York, NY [u.a.]. From page 14
- Pinho et al. (2015) Luis Gustavo B Pinho, Juvêncio S Nobre and Julio M Singer. Cook's distance for generalized linear mixed models. *Computational Statistics & Data Analysis*, 82:126–136. From page 2, 90, 91, 96, 97, 101
- **Poon and Poon (1999)** W-Y Poon and Yat Sun Poon. Conformal normal curvature and assessment of local influence. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(1):51–61. From page 5, 10, 19
- **R Core Team (2018)** R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2018. URL https://www.R-project.org/. From page 4, 22, 26, 28, 30, 34, 64, 68, 83, 88, 104
- **Rabe-Hesketh** *et al.* **(2002)** Sophia Rabe-Hesketh, Anders Skrondal and Andrew Pickles. Reliable estimation of generalized linear mixed models using adaptive quadrature. *The Stata Journal*, 2(1):1–21. From page 1
- **Rakhmawati** *et al.* **(2017)** Trias Wahyuni Rakhmawati, Geert Molenberghs, Geert Verbeke and Christel Faes. Local influence diagnostics for generalized linear mixed models with overdispersion. *Journal of Applied Statistics*, 44(4):620–641. From page 1, 2, 18, 21
- **Rigby and Stasinopoulos (2005)** Robert A Rigby and D Mikis Stasinopoulos. Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3):507–554. From page 30, 31, 105
- Stasinopoulos and Rigby (2007) D Mikis Stasinopoulos and Robert A Rigby. Generalized additive models for location scale and shape (gamlss) in r. *Journal of Statistical Software*, 23(7):1–46. From page 89
- Stasinopoulos *et al.* (2017) Mikis D Stasinopoulos, Robert A Rigby, Gillian Z Heller, Vlasios Voudouris and Fernanda De Bastiani. *Flexible Regression and Smoothing: using GAMLSS in R.* Chapman and Hall/CRC. From page 26, 29, 30, 64, 65, 66, 68, 79, 105
- **Stokes** *et al.* **(1995)** Maura E Stokes, Charles S Davis and Gary G Koch. Categorical Data Analysis Using the SAS System. SAS Institute. *Inc., Cary, NC,* pages 34–35. From page 76, 78
- **Stroup (2012)** Walter W Stroup. *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications.* CRC press. From page 1
- Tan et al. (2001) Frans ES Tan, Mario JN Ouwens and Martijn PF Berger. Detection of influential observations in longitudinal mixed effects regression models. *Journal of* the Royal Statistical Society: Series D (The Statistician), 50(3):271–284. From page 2
- Tang et al. (2006) Nian-Sheng Tang, Bo-Cheng Wei and Wen-Zhuan Zhang. Influence diagnostics in nonlinear reproductive dispersion mixed models. *Statistics*, 40(3): 227–246. From page 2, 7, 18, 19, 105

- Tapia et al. (2019a) Alejandra Tapia, Victor Leiva, Maria del Pilar Diaz and Viviana Giampaoli. Sensitivity analysis of longitudinal count responses: a local influence approach and application to medical data. *Journal of Applied Statistics*, 46(6):1021– 1042. From page 3, 19, 25
- Tapia et al. (2019b) Alejandra Tapia, Victor Leiva, Maria del Pilar Diaz and Viviana Giampaoli. Influence diagnostics in mixed effects logistic regression models. TEST, pages 1–23. From page 3, 19, 25
- Tuerlinckx et al. (2006) Francis Tuerlinckx, Frank Rijmen, Geert Verbeke and Paul De Boeck. Statistical inference in generalized linear mixed models: a review. *The British Journal of Mathematical and Statistical Psychology*, 59 Pt 2:225–255. From page 1, 8
- **Venables and Ripley (2002)** William N Venables and Brian D Ripley. *Modern Applied Statistics with S.* Springer Science, fourth ed. From page 31
- **Vonesh and Chinchilli (1996)** Edward Vonesh and Vernon M Chinchilli. *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. CRC press. From page 1, 14
- Wu (2009) Lang Wu. *Mixed Effects Models for Complex Data*. Chapman and Hall/CRC. From page 1
- Xiang et al. (2002) Liming Xiang, Siu-Keung Tse and Andy H Lee. Influence diagnostics for generalized linear mixed models: applications to clustered data. *Computational Statistics & Data Analysis*, 40(4):759–774. From page 2
- Xiang et al. (2003) Liming Xiang, Andy H Lee and Siu-Keung Tse. Assessing local cluster influence in generalized linear mixed models. *Journal of Applied Statistics*, 30 (4):349–359. From page 2
- Xu et al. (2006) Liang Xu, Sik-Yum Lee and Wai-Yin Poon. Deletion measures for generalized linear mixed effects models. *Computational Statistics & Data Analysis*, 51 (2):1131–1146. From page 2
- **Zeger and Karim (1991)** Scott L Zeger and M Rezaul Karim. Generalized linear models with random effects; a Gibbs sampling approach. *Journal of the American Statistical Association*, 86(413):79–86. From page 8
- **Zhang and Lin (2008)** Daowen Zhang and Xihong Lin. Variance component testing in generalized linear mixed models for longitudinal/clustered data and other related topics. In *Random effect and latent variable model selection*, pages 19–36. Springer. From page 68, 83, 95

- Zhao et al. (2006) Yihua Zhao, John Staudenmayer, Brent A Coull and Matthew P Wand.
  General design Bayesian generalized linear mixed models. *Statistical Science*, pages 35–51. From page 8
- **Zhu and Lee (2003)** Hong-Tu Zhu and Sik-Yum Lee. Local influence for generalized linear mixed models. *Canadian Journal of Statistics*, 31(3):293–309. From page 2, 3, 19, 39
- **Zhu and Lee (2001)** Hong-Tu Zhu and Sik-Yum Lee. Local influence for incomplete data models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63 (1):111–126. From page 2, 19