

**Joint modeling of longitudinal markers and
survival data: an approach via dynamic
hierarchical models**

Pamela Massiel Chiroque Solano



Universidade Federal do Rio de Janeiro

2018



Joint modeling of longitudinal markers and survival data: an approach via dynamic hierarchical models

Pamela Massiel Chiroque Solano

Tese de doutorado submetida ao Programa de Pós-graduação em Estatística do Instituto de Matemática - Departamento de Métodos Estatísticos - UFRJ, como parte dos requisitos necessários à obtenção do título de Doutor em Estatística.

Orientador: Helio S. Migon

Rio de Janeiro, RJ - Brasil

Dezembro de 2018

Modelagem conjunta de marcadores longitudinais e de sobrevivência: uma abordagem via modelos hierárquicos dinâmicos

Pamela Massiel Chiroque Solano

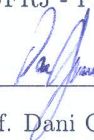
Tese de doutorado submetida ao Corpo Docente do Instituto de Matemática - Departamento de Métodos Estatísticos da Universidade Federal do Rio de Janeiro - UFRJ, como parte dos requisitos necessários do grau de Doutor em Estatística.

Aprovada por:



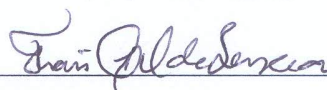
Prof. Helio S. Migon.

Ph.D. - IM - UFRJ - Presidente, Orientador.



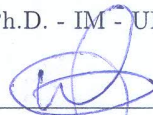
Prof. Dani Gamerman.

Ph.D. - IM - UFRJ.



Profª. Thais Cristina Oliveira da Fonseca.

Ph.D. - IM - UFRJ.



Prof. Fábio Nogueira Demarqui

Ph.D. - ICEX - UFMG.



Prof. Leonardo Soares Bastos

Ph.D. - Fiocruz.

Rio de Janeiro, RJ - Brasil

Dezembro de 2018

Ficha Catalográfica

C541j Chiroque Solano, Pamela Massiel
 Joint modeling of longitudinal markers and
 survival data: an approach via dynamic hierarchical
 models / Pamela Massiel Chiroque Solano. -- Rio de
 Janeiro, 2018.
 144 f.

 Orientador: Helio dos Santos Migon.
 Tese (doutorado) - Universidade Federal do Rio
 de Janeiro, Instituto de Matemática, Programa de Pós
 Graduação em Estatística, 2018.

 1. Bayesian survival analysis. 2. Joint survival
 analysis. 3. Longitudinal data. 4. Quantile
 Regression. 5. Regime Change. I. dos Santos Migon,
 Helio, orient. II. Título.

Resumo

Modelos para dados longitudinais e tempo até a ocorrência de um evento de interesse são propostos no contexto de modelagem conjunta. Três estruturas de regressão diferentes para os dados longitudinais são apresentadas com evolução baseada em modelo linear dinâmico hierárquico generalizado. Essas estruturas de regressão são padrão, quantílica e regressão com mudança de regime markoviana. Em termos da função de risco base de sobrevivência algumas formas paramétricas e semi-paramétricas são usadas e validadas com resultados simulados. Algumas relações de dependência são exploradas para a função semi-paramétrica. Em todas as propostas, a ligação entre os dois sub-modelos é uma escolha flexível que é discutida. O procedimento de inferência é desenvolvido sob a abordagem Bayesiana. Métricas de comparação são adaptadas e usadas para validação de modelo além de métricas padrão. Quatro conjuntos de dados são usados para ilustrar as metodologias propostas.

Abstract

Models for longitudinal and time-to-event data are proposed in the context of joint modeling. Three alternative regression models for the longitudinal data are presented with time evolution based on Generalized Hierarchical Dynamic Linear Model. These regression models are standard, quantile and Markov switching regression. For the survival baseline hazard function some parametric and semi-parametric forms are used and validated through simulated results. Some dependence relations are explored for the semi-parametric function. In all proposals, the link between the two sub-models is a flexible choice which is discussed. Inference procedure is developed under the Bayesian approach. Comparison metrics are adapted and used for model validation alongside standard metrics. Four data sets are used to illustrate the proposed methodologies.

Contents

1	Introduction	1
1.1	Brief historical review	3
1.2	Research setting	5
1.2.1	The bridge term	8
2	Joint Modeling of Time-to-event and Longitudinal Measurements Data: A Dynamic Generalized Hierarchical Approach	12
2.1	Introduction	12
2.1.1	The Dataset Structure	15
2.1.2	Notation and Definitions	16
2.2	Joint Model Buildup	17
2.2.1	Modeling the time-to-event data	17
2.2.2	The Multivariate Longitudinal Model	18
2.3	Inference	24
2.3.1	Posterior Inference	25
2.4	Simulation Study	26
2.5	Real Data Analysis	28
2.5.1	Schizophrenia Dataset	29
2.5.2	AIDS Dataset	34
2.6	Discussion and Future Directions	38
3	A flexible joint model for time-to-event and a sequence of longitudinal biomarker data	40
3.1	Introduction	41
3.2	Model	42
3.2.1	Semi-parametric survival model formulation	43

3.2.2	Dynamic modeling of the longitudinal markers	46
3.3	Inference	49
3.3.1	Prior	49
3.3.2	Likelihood	49
3.4	Simulation study	50
3.5	Liver cirrhosis longitudinal data	57
3.6	Discussion	61
4	Quantile joint model for the longitudinal marker and event time dataset	63
4.1	Introduction	63
4.1.1	Preliminaries	66
4.2	Quantile regression	66
4.3	Bayesian linear hierarchical quantile joint model	68
4.3.1	Inference	71
4.3.2	Prediction with the dynamic hierarchical quantile joint model . . .	74
4.4	Application	75
4.4.1	Schizophrenia dataset	76
4.4.2	Liver data analysis	82
4.5	Conclusions and future work	88
5	Joint model with Markov switching regimes to longitudinal marker and event time dataset	90
5.1	Introduction	90
5.2	Joint model with Markov-switching regimes	91
5.2.1	Mathematical definition	92
5.2.2	Two states: A particular case	95
5.3	Inference	96
5.3.1	Prior distribution	98
5.3.2	Posterior distribution	99
5.4	Application	101
5.4.1	Children with episodes of diarrhea	101
5.5	Conclusion	109
6	Conclusions and future work	110

A Appendix	115
A.1 The general model in detail: the longitudinal component	115
A.2 Supplementary material for chapter 2	117
A.2.1 Revisiting Some Special Cases in the Literature	117
A.2.2 Dynamic Hierarchical Joint Model for the Schizophrenia Dataset . .	119
A.2.3 Decomposition of the Comparison Criteria	120
A.2.4 Posterior histogram for the parameters for the Schizophrenia Dataset model	121
A.2.5 Posterior histogram for the parameters for the Aids Dataset model	122
A.2.6 Simulated data generator for chapter 2	123
A.3 Supplementary material for chapter 3	127
A.3.1 Bayesian Inference	127
A.3.2 Simulated data generator function for chapter 3	128
A.3.3 Posterior histogram for the parameters for the Liver Dataset. . . .	134
A.4 Jags code for Chapter 4	135
A.5 Jags code for Chapter 5	136

List of Figures

1.1	Graphical representation of common evolution (above) and random effect for the intercept (below) dynamic models	9
2.1	(a) Observed trajectories of PANSS score for the 150 patients and average trajectory (blue curve). (b) Kaplan-Meier curve for the follow-up time and 95% pointwise confidence band.	30
2.2	Survival probabilities calculated with the posterior mean and 95% credible interval of: solid line our proposal (LD); dashed line a survival regression without the longitudinal process; the dotted lines model from Henderson et al. (2000) (GD).	33
2.3	Survival probabilities computed with the posterior mean of the parameters. Solid line: our proposal for all individuals, dashed line: patients under treatment placebo, dashed-dotted line: patients under treatment risperidone, dotted line: patients under treatment haloperidol, solid line: all patients, and stepped lines: Kaplan-Meier curves.	35
2.4	(a) Histogram of the observed CD4 cell counts for 467 patients. (b) Kaplan-Meier curves for the follow-up time and 95% pointwise confidence bands.	36
3.1	Comparison of Gamma Process vs Log Normal : Baseline hazard functions using Simulation with Gompertz baseline hazard $\exp\{t\}$	54
3.2	Results based on 1000 simulated data sets. Bias is computed with the mean as the measure of location and multiply by 100 with sizes 50 and 100, K= 3,5,7. The fully dynamic joint model Gamma Process (GP) and Log Normal Process (NP) were compared.	55

3.3	Results based on 1000 simulated data sets. The mean square error (MSE) is computed with the mean as the measure of location with sizes 50 and 100, $K=3,5,7$. The fully dynamic joint model Gamma Process (GP) and Log Normal Process (NP) were compared.	56
3.4	<i>Baseline hazard function by interval k: estimated using the package <code>joiner</code> based on Henderson et al. (2000) SRE approach (M3), Rizopoulos and Ghosh (2011) (dashed and dotted, respectively, M2) and our proposal (hard line with credible intervals, M1)</i>	59
3.5	<i>Longitudinal data: for each j interview time, the level μ_j with credibility interval (blue hard line, M1) and estimated linear trajectories by Rizopoulos and Ghosh (2011) (red dotted, M2) and Henderson et al. (2000) (green dash-dotted, M3)</i>	60
4.1	PANSS score for a sample of 150 patients with three quantile regression curves: The curves were obtained for the posterior mean of $\mu_{j1}, j = 1, \dots, 6$ which represent the mean level of the trajectories of the ν^{th} 0.1, 0.5 and 0.9 quantiles.	78
4.2	Density of the posterior marginal distributions of β_{21} , the treatment effect in the longitudinal sub-model for each quantile.	79
4.3	Boxplot of the posterior marginal distributions of the Weibull r shape parameter for each quantile.	81
4.4	<i>(a) Boxplots of Prothrombin index for the follow-up times, the points represents the median values. (b) Empirical survival probability (Kaplan-Meier) for Prednisone ($arm = 1$) versus Placebo ($arm = 0$).</i>	83
4.5	Prothrombin measurements for a sample of 488 patient with three quantile regression curves: The curves were obtained for $\mu_{j1}, j = 1, \dots, 17$ that represent the mean level of the trajectories of the ν^{th} quantiles.	85
4.6	Boxplot of the posterior marginal distributions of the Weibull r shape parameter for each quantile.	86
4.7	Densities of the posterior marginal distributions of β_{21} for each quantile.	87
5.1	Markov Switching model for two states is represents through directed acyclic graphs (DAG).	96

5.2	<i>Empirical survival probability (Kaplan-Meier) for the childhood with episodes of diarrhea by treatments vitamim A (Vit A) and placebo (pla).</i> . . .	103
5.3	Boxplot of pediatric diarrhea target enteral volume (ml/kg/d) for 860 patients in 27 interview.	104
5.4	Baseline hazard function by interval k : estimated using the Rizopoulos model Rizopoulos and Ghosh (2011) under SRE approach (PE_{Rizo}), (dashed) and our proposal (solid line with credible intervals, PE_1)	106
5.5	Posterior density of the effect bridge estimated using the Rizopoulos model Rizopoulos and Ghosh (2011), (dashed line, PE_{Rizo}) and our proposal (solid line, PE_1) with their credible intervals.	107
5.6	Posterior density of the parameters estimated using our proposal (PE_1). Where P_{11} and P_{22} are the transition probability. The densities for μ_1 and μ_2 represent the level trajectory.	108
A.1	<i>The α_k effect by interval k: estimated using the package Philipson et al. (2017) based on Henderson et al. (2000) SRE approach (M3), Rizopoulos and Ghosh (2011) (dashed and dotted, respectively, M2) and our proposal (hard line with credible intervals, M1)</i>	135

List of Tables

2.1	Summaries from 500 simulated data sets under the proposed model (P: parameter, True: true value, Bias: computed with the mean as the measure of location, rMSE: root of the mean-squared error, and SD: average of the posterior standard deviations).	28
2.2	Comparison of Bayesian model selection methods for the locally dynamic model (LD) and globally dynamic model (GD)	32
2.3	Comparison of Bayesian model selection methods for the fully Bayesian model (FullBayes) and posterior mean of the latent term model (PostMean).	32
2.4	Bayesian model selection criteria for the dynamic longitudinal sub-model: Poisson response (Poisson) and normal longitudinal sub-model (Normal). Metrics divided by 10.	37
3.1	Comparison of posterior mean. They are computed with the mean based on 1000 posterior means calculated for each simulated dataset for all scenarios. The fully dynamic joint model Gamma Process (GP) and Log Normal Process (NP) were compared.	57
3.2	Comparison of Bayesian model selection methods for the fully dynamic joint model considering Gamma Process (GP) compared to the standard joint model (SRE) using the package in Rizopoulos (2016).(Metrics divided by 100).	59
4.1	Comparison of Bayesian model selection methods for the quantile joint models considering two choices of the bridge term ξ_i	80
4.2	Summary of the estimated regression coefficients based on the posterior distributions related the survival sub-model. 1-dimensional bridge term.	87

4.3	Comparison of Bayesian model selection methods for the quantile joint models considering two choices of the bridge term ξ_i	88
5.1	Comparison of Bayesian model selection methods.	105

Glossary

Symbols and nomenclatures

\propto	Proportionality
\sim	Has distribution
\in	Is an element of
$[\cdot]$	density of \cdot
N	Total of individuals followed-up
i^{th}	An individual
C_i	Censoring time
T_i^*	True survival time
T_i	Survival time random variable $\min(T_i^*, C_i)$
t_i	Observed value of T_i
j	A visit time
J_i	Number of follow-up measurements (visits) of individual i
J	Number of visits when they are the same for all individuals
s_{ij}	Scheduled dates of individual i at j visit
s_j	Visit scheduled dates at time j when they are the same for all individuals
d_i	Indicator of right censoring for the i^{th} individual
r	Shape parameter of the Weibull distribution
$h(t)$	Hazard function at time t
$h_0(t)$	Baseline hazard function at time t
\mathbf{Z}_i	Full matrix covariates associated with the time-to-event sub-model
\mathbf{X}_{1i}	Fixed covariates, a block of \mathbf{Z}
$\boldsymbol{\xi}_i$	Bridge term: Latent component which links two sub-models, a block of \mathbf{Z}_i
$\boldsymbol{\gamma}_i$	Effects of \mathbf{Z}_i
$\boldsymbol{\beta}_1$	Effect of \mathbf{X}_{1i} in the survival sub-model
$\boldsymbol{\alpha}$	Effect of $\boldsymbol{\xi}_i$ in the survival sub-model
$\boldsymbol{\gamma}$	Mean of $\boldsymbol{\gamma}_i$ in the survival sub-model

ζ_i	Hierarchical error of γ_i (frailty term)
Σ_ζ	Variance of ζ_i
q_1	Number of columns fixed regressors \mathbf{X}_{1i} in the survival sub-model
q^*	Number of columns of latent covariates ξ_i
δ_{ij}	Time between consecutive visits for individual i at visit j
δ_j	Time between consecutive visits at visit j for all individuals
y_{ij}	Longitudinal marker
\mathbf{y}_j	N -dimensional longitudinal markers vector at visit j
Λ	Latent $N \times J$ -dimensional matrix related to the longitudinal markers
λ_{ij}	Latent value for each y_{ij}
$\lambda_{\cdot j}$	The j^{th} column of Λ
$\lambda_{i \cdot}$	The i^{th} row of Λ
p	Polynomial order
\mathbf{F}_i	Design matrix related to the longitudinal sub-model
\mathbf{X}_{2i}	Fixed covariates, a block of \mathbf{F}_i
q_2	Number of columns fixed regressors \mathbf{X}_{2i} related to the longitudinal marker
ϵ_j	Perturbation term of \mathbf{y}_j at the j^{th} visit
Σ	Variance-covariance matrix of the observation equation
θ_j	Common mean trajectory
μ_j	Polynomial effects
β_2	Effect of \mathbf{X}_{2i}
\mathbf{G}_{ij}	Transition matrix
\mathbf{w}_j	Perturbation term
θ_0	Mean trajectory at the initial time
\mathcal{D}_j	Information accumulated until time j
\mathcal{D}_0	Initial information
\mathbf{m}_0	Initial mean prior
\mathbf{C}_0	Initial variance prior

L_p	The transition matrix for the p -order polynomial
e_p	Design vector relative to the polynomial model of order p
Σ	Variance-covariance matrix $N \times N$
W	Covariance matrix related to the polynomial and covariates effects
v_i	Random effect
η_{ij}	the natural parameter in the exponential family
ψ	dispersion parameter in the exponential family
V	Disturbance term v_i variance
$g(\cdot)$	known monotonic link function at least twice-differentiable
$a(\cdot), b(\cdot)$	Exponential family identifying functions
Θ	the complete unknown parameters vector
θ	J -dimensional vector of latent vector terms θ_j
θ_{-j}	$J - 1$ -dimensional vector θ without the j^{th} term
$LPML_{Surv Long}$	LPML comparison metric of the joint model
$\Delta LPML_{Surv}$	LPML contribution of the survival sub-model
$DIC_{Surv Long}$	DIC comparison metric of the joint model
ΔDIC_{Surv}	DIC contribution of the survival sub-model

Operators and functions

$ A $	Determinant of A matrix
A'	Transpose of A matrix (or vector)
$F^{-1}(\cdot)$	inverse of the function $F(\cdot)$
$E(T)$	Expectation of the random variable T
$Var(T)$	Variance of the random variable T
$min(a, b)$	Minimum between a and b
$I(d(x))$	Indicator function of the set d : (1 if $x \in d$, 0 otherwise)
$\log(\cdot)$	Logarithmic function of base e (ln)
$\exp(\cdot)$	Exponential function
$diag(a)$	Square diagonal matrix with a on the main diagonal

Probability distributions

$\mathcal{N}(m, C)$	Normal distribution with mean m and covariance matrix C
$IG(a; b)$	Inverse Gamma distribution parametrized with expected value $\frac{b}{a-1}$
$IW(\mathbf{R}; g)$	Inverse-Wishart with scale matrix \mathbf{R} and g degrees of freedom
$Exp(c)$	Exponential distribution with parameter c
$\mathcal{W}(r, \exp\{\mathbf{X}'_{1i}\boldsymbol{\beta}_1 + \boldsymbol{\xi}'_i\boldsymbol{\alpha}\})$	Weibull distribution with parameters r and $h(t)$

Abbreviations

GHDLM	Generalized Hierarchical Dynamic Linear Model
AIDS	Acquired Immune Deficiency Syndrome
CD4	Cluster of differentiation 4
SRE	Shared random effects
EM	Expectation maximization
GD	Global dynamic
LD	Local dynamic
DIC	Deviance information criterion
LPML	Logarithm of the pseudomarginal likelihood
pD	Effective number of parameters
HIV	Human immunodeficiency virus
ddI	Didanosine
ddC	Zalcitabine
AZT	Zidovudine therapy
DLM	Dinamic linear model
PANSS	Positive and Negative Syndrome Scale
JAGS	Just Another Gibbs Sampler
MAR	Missing at random
MC	Monte Carlo
MCMC	Markov chain Monte Carlo
SD	Standard deviation
WinBUGS	Windows-based Bayesian inference using Gibbs sampling

Chapter 1

Introduction

All scientific innovation which aims to be used on the real world must be justified by data driven results. Medical research is not an exception and new technological instruments allow for immediately available and reliable data, which leads to requiring more consideration from Statisticians. This allows for data based decisions in terms of diagnosis and treatment.

However, the use of an instrument to collect data implies that there might be some disparity between the measured value and the desired information. This could be the consequence of measuring error, information withholding by the patient or human error, for example. Such disparity should be taken into account in the decision making process, and can be done via statistical modeling.

A way to statistically take this measuring disparity into account is to take the same measure repeatedly over time. If these measures are related to the desired information, commonly called true value, in a particular form then the repeated measures can aggregate knowledge towards the true value. This form can be a parametric distribution where the true value is the distribution expected value. One way to think this is to say that the measured values are the true value plus a random noise.

The very fact that the measures are taken over time means that the desired true value may change between one measuring session and the other. This can be understood as a dynamic, which may be represented through a temporal dependence. In order to represent this dynamic, a large variety of time series models is available for use. This

process is usually called longitudinal measurements modeling and there are numerous ways to represent this dependence and not all of them are time series models. Nonetheless the nature of the data is directly relatable to such models.

Sometimes this longitudinal information influences the risk of occurrence of an event for that patient. Some examples are the count of CD4 cells, which may be measured with error, influences the risk of death of patients infected with HIV. Another example is the PANSS score, an index which is calculated and therefore may misrepresent the desired information, influences the risk of crisis of schizophrenic patients. Traditionally the risk of occurrence of an event is modeled with a class of models called survival models, which allow the inclusion of independent variables that are related to the event occurrence risk. Then, we can use the learned information from the longitudinal process as a variable to explain this risk.

Adding the longitudinal filtered information into the occurrence risk characterizes a class of models called joint models, which is the focus of this thesis. There are many ways for this information to be added and this configures a choice for the researcher.

The main contribution provided is the dynamic treatment of the longitudinal measurements through a time series. A direct consequence of this is that future predictions of the longitudinal information are easily obtained. Hence occurrence times prediction become also available, which is desired in the context of survival models. Other available models also provide longitudinal prediction, but only through extrapolation.

Other contributions include some options based on already existing ideas adapted to the proposed longitudinal modeling with evolution over time. One example is using quantile regression to model the longitudinal measurement. One extension which cannot be found in other works incorporates shifting levels for the longitudinal trajectory with Markov switching regimes.

Next a brief historical review of joint models and the fundamental groundwork for the thesis is presented.

1.1 Brief historical review

This thesis works with settings where the sample units, also called individuals, are followed-up over time and typically different types of outcomes are collected. These measurements we denominate as longitudinal outcomes, i.e., the same outcome measured repeatedly over time for the same individual, and event time outcomes, that is, the time until a specific event of interest occurs. The main assumption of the data is that the longitudinal outcomes contain information that influences the event time or at least the risk that the event will happen as a function of time.

Originally the longitudinal measurements were used as raw data input into relative risk survival models for the event times. However, some works in the early 1990's noted that modeling the longitudinal measurements is preferable (Ibrahim et al. 2004) Section 7.1.1. Since then, separate analysis of the longitudinal and time-to-event data were common, but (Ibrahim et al. 2010) argue that this may lead to inefficient or biased results.

With this in mind, DeGruttola and Tu (1994), Faucett and Thomas (1996), Wulfsohn and Tsiatis (1997), Ibrahim et al. (2004) and Wang and Taylor (2001) combine all information and represent both structures through a single model called joint model, where DeGruttola and Tu (1994) and Wulfsohn and Tsiatis (1997) develop a frequentist solution and the others develop a Bayesian solution. Throughout this thesis, we refer to each structure as sub-model, that is, the joint model combines a longitudinal and a survival sub-model.

In the above references models in the class called Shared Random Effects (SRE) are developed to which the proposals in this thesis belong as well. Other solutions are well explained in Klein et al. (2014).

Most of the SRE proposals are generalized by Henderson et al. (2000). The Bayesian approach introduced by Guo and Carlin (2004) is described below. Let y_{ij} represent the endogenous longitudinal outcomes for individual i taken at interview j (possibly missing), where $i = 1, \dots, N$ and $j = 1, \dots, J_i$. All N individuals are followed over a time interval

$[0, \tau]$. Denote the interview times as s_{ij} . Each individual i has a (possibly censored) survival time t_i to a certain endpoint. The joint model is given by

$$\begin{aligned} h_i(t) &= h_0(t) \exp(\mathbf{x}'_{1i}(t)\boldsymbol{\beta}_1 + W_{1i}(t)), \quad t > 0 \\ y_{ij} &= \mu_{2i}(s_{ij}) + W_{2i}(s_{ij}) + \epsilon_{ij}, \end{aligned} \tag{1.1}$$

where the first equation represents the hazard function for the survival sub-model and assuming that $t_i \sim \text{Weibull}(r, \mu_{1i}(t))$, where $\log(\mu_{1i}(t)) = \mathbf{x}'_{1i}(t)\boldsymbol{\beta}_1 + W_{1i}(t)$ and $r > 0$. The connection between these two equations lies with the fact that $W_{1i}(t)$ and $W_{2i}(s_{ij})$ are correlated. They will be detailed below. Note that this is similar to a *Cox Proportional Hazards Model* (Cox 1972) where the baseline hazard function assumes a parametric Weibull form $h_0(t) = rt^{r-1}$, but any baseline function can be used, and also adding the process $W_{1i}(t)$. The vectors $\mathbf{x}_{1i}(t)$ and $\boldsymbol{\beta}_1$ represent (possibly time-dependent) explanatory variables and their corresponding regression coefficients. The individual specific intercept of $W_{1i}(t)$ is often called a *frailty*.

For the longitudinal sub-model represented by the second line of (1.1), $\mu_{2i}(\mathbf{s}) = \mathbf{x}'_{2i}\boldsymbol{\beta}_2$ is the mean response, $W_{2i}(\mathbf{s}) = \mathbf{d}'_{2i}\mathbf{U}_i$ incorporates subject-specific random effects, and $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2)$ is a sequence of mutually independent measurement errors. The $W_{2i}(\mathbf{s})$ can be viewed as the true individual-level longitudinal trajectory after they have been adjusted for the overall mean trajectory and other fixed effects. The form of $W_{2i}(\mathbf{s})$ is similar to $W_{1i}(t)$, including subject-specific covariate effects and an intercept. The vectors $\mathbf{x}_{2i}(\mathbf{s})$ and $\boldsymbol{\beta}_2$ represent possibly time-varying explanatory variables and their corresponding regression coefficients, respectively. They may or may not have elements in common with \mathbf{x}_{1i} and $\boldsymbol{\beta}_1$ in the survival sub-model. The \mathbf{U}_i are vectors of random effects corresponding to the explanatory variables $\mathbf{d}_{2i}(\mathbf{s})$ (which may be a subset of $\mathbf{x}_{2i}(\mathbf{s})$) and are typically modeled as iid $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ random variables.

Association between the longitudinal and survival processes can arise in two ways. One is through explanatory variables present in both equations. The other is through stochastic dependence between W_{1i} and W_{2i} . This can be achieved by jointly modeling the two processes via a latent zero-mean bivariate Gaussian process on $(W_{1i}, W_{2i})'$, which is independent across different subjects.

Another view deserves special attention and is described in Rizopoulos (2012, Chapter 4), where instead of correlated latent processes in the two sub-models, some true unobserved trajectory is included into the hazard function as a latent covariate. Then his model is described by:

$$\begin{aligned} h_i(t) &= h_0(t) \exp(\mathbf{x}'_{1i}(t)\boldsymbol{\beta}_1 + \alpha m_i(t)), \quad t > 0, \\ y_i(t) &= m_i(t) + \epsilon_i(t), \quad m_i(t) = \mathbf{x}'_{2i}\boldsymbol{\beta}_2 + \mathbf{z}'_i(t)\mathbf{b}_i, \end{aligned} \tag{1.2}$$

where $\{\mathcal{M}_i(t) = m_i(u), 0 \leq u < t\}$ denotes the history of the true unobserved longitudinal process up to time point t , the α term quantifies the effects of the underlying longitudinal outcome to the risk for the event. The term $m_i(t)$ is a mixture of fixed and random components $\mathbf{x}'_{2i}\boldsymbol{\beta}_2 + \mathbf{z}'_i(t)\mathbf{b}_i$, where $\mathbf{z}'_i(t)\mathbf{b}_i$ is some function of t , say a polynomial. Note that every mixed effects model can be easily recovered by a hierarchical structure.

An important feature of the models presented by Rizopoulos (2012, Chapter 5) and Rizopoulos and Ghosh (2011), is the addition of other ways that $m_i(t)$ is inserted into the hazard function, such as $\frac{\partial m_i(t)}{\partial t}$. This sets a precedent in the sense that productions in joint modeling should consider this flexibility as they are proposed. This thesis is not indifferent to this relevant discussion (Section 1.2.1) to which the term bridge is used to make reference.

Both the models from Henderson et al. (2000) and Rizopoulos (2012) provide desirable features. The former provides a rigorous structure for the longitudinal sub-model while the latter quantifies the impact of the longitudinal sub-model on the hazard function. This aspect of this class of models is the motivation for this thesis, which is to create a hazard-quantifiable rigorous structure for the longitudinal sub-model.

1.2 Research setting

This section presents a general model framework which can be classified as a Shared Random Effects model. The data described in Section 1.1 is modeled assuming that the censoring of event times is non-informative and that data from different subjects is generated by independent processes.

One of the main characteristics of longitudinal data widely known in the literature is

time dependence, as discussed. Another is that the longitudinal trajectories have some common feature, which is usually considered as the mean response. On the other hand the mean response is not the only feature that can be analyzed. One could be interested, for example, in a particular quantile of the distribution of the longitudinal outcomes, especially when they are not symmetrical.

The time dependence can be modeled by a Dynamic Linear Model (DLM) framework as described by West and Harrison (1997), albeit with few longitudinal measurements for each individual. This class of models can locally approximate complex evolutions, as well as allow for the prediction of future observations. Extensions of DLM are available to deal with non Gaussian and non linear responses, which are useful since some longitudinal outcomes can be counts, proportions or dichotomies for instance.

The hierarchy is modeled by a common value added to a random effect which captures individual-specific features. An important advantage of this is the sharing of common information between individuals. Fortunately, hierarchy has already been combined into DLMS in Gamerman and Migon (1993). Adding the generalization from West et al. (1985), we obtain a Generalized Hierarchical Dynamic Linear Model (GHDLMS), which can be used to model the longitudinal outcomes. Remember that it is known that all joint mixed random effects models can be written in a hierarchical form. This means that the proposal is comparable to benchmark models, but with a dynamic structure added.

Thus we note that all proposals in this thesis is the pertinent combination of already established models in the literature, both in the longitudinal and survival viewpoints. The novelty in this work is to combine these models in a joint model.

The general model structure is given by

Observation equation:

$$y_{ij} \sim Ef(\eta_{ij}, \psi), \quad E(y_{ij} | \eta_{ij}, \psi) = g^{-1}(\lambda_{ij}),$$

Structural equation:

$$\lambda_{ij} = \mathbf{F}'_i \boldsymbol{\theta}_{ij} + v_i, \quad v_i \sim \mathcal{N}(0, V) \quad (1.3)$$

System equation:

$$\boldsymbol{\theta}_{ij} = \mathbf{G}_j \boldsymbol{\theta}_{i,j-1} + \mathbf{w}_j, \quad \mathbf{w}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{W})$$

Initial information:

$$\boldsymbol{\theta}_0 | \mathcal{D}_0 \sim \mathcal{N}(\mathbf{m}_0, \mathbf{C}_0),$$

where the longitudinal observations are described in the Exponential family (Ef). As is commonly presented, $[y_{ij} | \eta_{ij}, \psi] = b(y_{ij}, \psi) \exp(\psi(y_{ij}\eta_{ij} - a(\eta_{ij})))$, with mean $E(y_{ij} | \eta_{ij}, \psi) = g^{-1}(\lambda_{ij})$, where g^{-1} is assumed to be some twice-differentiable function of λ_{ij} , where η_{ij} and ψ denote the natural and dispersion parameters in the exponential family, respectively.

The linear predictor, which is the $g(\cdot)$ transformation of the mean trajectory of the observations, is represented by the term λ_{ij} . We call this the general model since we can use the versatility of Dynamic Linear Models construction, through the choice of \mathbf{F}_i , \mathbf{G}_j , $\boldsymbol{\theta}_{ij}$ and \mathbf{W} , as well as the hierarchical structure through v_i .

A clear example is the ability to recover the linear growth version of the popular mixed model shown in (1.2). By choosing $\mathbf{F} = (1, 0, 0)'$, $\boldsymbol{\theta}_{ij} = (\theta_{0ij}, \theta_{1j}, b_i)'$, the first line of the three-dimensional matrix \mathbf{G} is $(1, \delta_j, \delta_j)$, where δ_j is the timespan between the $j - 1^{th}$ and the j^{th} visit, the second line is $(0, 1, 0)$ and the third line is $(0, 0, 1)$. Finally, \mathbf{W} has a 0 for the third variance. That way, the random effect for the intercept is v_i and for the slope it is b_i . The exact derivation can be found in the Appendix A.1.

Another advantage of this general structure is the possibility of inclusion of covariates in \mathbf{F} and their respective time evolving effects per individual in $\boldsymbol{\theta}_{ij}$. If eventually some effects do not evolve in time, all that is needed is to set their respective variance in the matrix \mathbf{W} to $\mathbf{0}$.

For simplicity, all development in this thesis is on a particular case of (1.3) where $\theta_{ij} \equiv \theta_j, \forall i = 1, \dots, N$, that is, θ_j is a common term which evolves over time. Although this is a simplification, it still allows for great complexity, as is seen throughout the thesis.

This represents a model with random effects on the intercept term, that is a random fluctuation on the mean per individual. Figure 1.1 shows a model without the random effect v_i and one with this effect.

A component that still needs discussion in order for the model to be fully specified is presented next.

1.2.1 The bridge term

Bearing in mind the original motivation of the joint model, the bridge term is understood as the filtered information from the longitudinal data relevant to the hazard function. This information is sometimes simple to understand, such as CD4 cells true count measured with error by an instrument, but not always so, such as schizophrenia symptoms severity measured by a questionnaire that results in the PANSS score, which is a constructed measure. When the information is simple, then it is easy to define the bridge term by using, say, the mean trajectory g -transform (i.e. the linear predictor λ_{ij}).

When it is not straightforward then this bridge term should be discussed in order to ensure that it carries the desired information. For this reason, the proposed class of models does not specify a bridge term, therefore the joint model is completely specified only after the choice of this term. To make this choice, note that whatever term in the longitudinal sub-model which is not the y_{ij} themselves represents some filtered information. For instance, choosing λ_{ij} or θ_j , which are different components of the distribution of y_{ij} , represent distinct information to be bridged over to the hazard function. It is, therefore, the researcher's job to realize what is most appropriate from the logical point of view. The bridge term choice based on what makes most sense to the problem should matter more than model comparison metrics. Note that this could also mean a quantile of the distribution of y_{ij} .

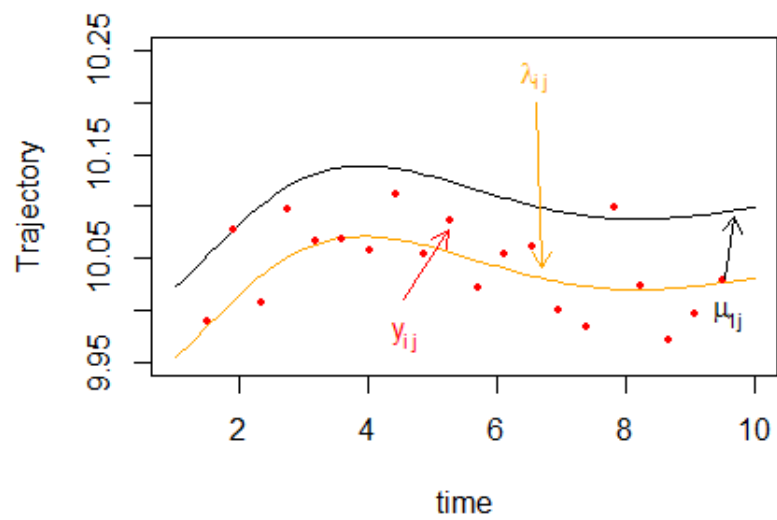
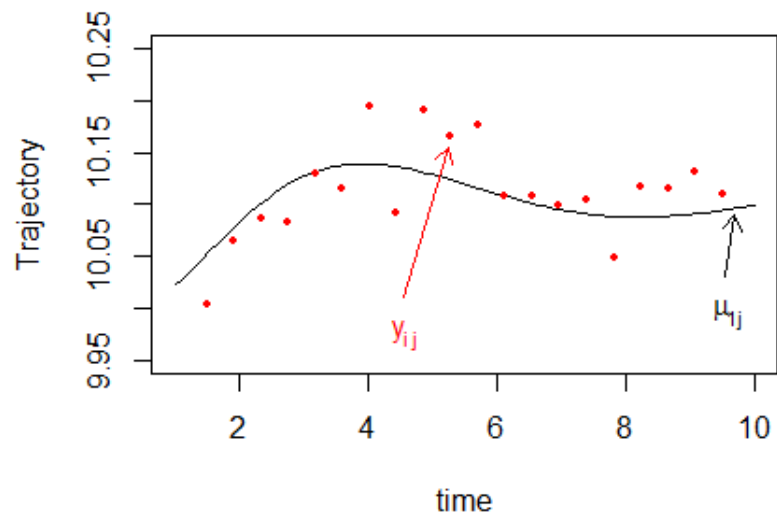


Figure 1.1: Graphical representation of common evolution (above) and random effect for the intercept (below) dynamic models

Note that the model deals with missing data due to patients who miss their interview or some other reason, which can be informative or not. Missing data also includes censorship, however this thesis does not include modeling for informative censorship. Missing data can also affect the choice of the bridge term, since the uncertainty raised by this situation is bridged to the hazard function differently depending on this choice. In other words, if one bridge choice is punishing with respect to missing data, then another bridge might be needed.

Although Rizopoulos and Ghosh (2011) present different bridge options, they do not discuss the implications of each option. To show the importance of this choice we draw attention to the hazard function of the standard model:

$$\begin{aligned} h_i(t) &= h_0(t) \exp\{\mathbf{Z}'\boldsymbol{\beta}_1 + \alpha(a_{0i} + a_{1i}t)\} \\ &= h_0(t) \exp\{\mathbf{Z}'\boldsymbol{\beta}_1\} \exp\{b_{0i} + b_{1i}t\}. \end{aligned} \tag{1.4}$$

where a_{0i} and a_{1i} are features from the longitudinal sub-model. Note that the observed times t_i are included in the likelihood function both in $h_0(t)$ and in $b_{0i} + b_{1i}t$ and also $\exp\{b_{0i} + b_{1i}t\}$ forms a Gompertz baseline hazard function. This means that whatever form the true baseline hazard function has, $h_0(t)$ is sure to underestimate it to compensate for the Gompertz exponential growth.

Therefore we consider that this discussion is not currently present in the literature and propose that it should be.

This discussion is started in all chapters of this thesis, which are summarized below. Chapters 2 and 3 provide the first versions of the proposal by changing the longitudinal sub-model then exploring on the baseline hazard function.

Further development is in chapter 4 where there is no parametric form to model y_{ij} and λ_{ij} represents the distribution's quantile instead of expected value.

Finally, the quality of life index will be considered as a longitudinal outcome variable. Although there are methods that analyze this variable in a survival context, most of

them do not consider joint models. Those that do, in their turn, do not take into account what is considered a core feature of the quality of life, which is the change of health states. With that in mind, the model in (1.3) can easily incorporate level states that change at any given time. This is done in Chapter 5.

All inference procedure described in the thesis is based on the Bayesian paradigm. The sampling of the posterior distribution is obtained via MCMC methods.

The theory in this introduction is meant as a general version and each chapter of this thesis focuses on its own subject. This dissertation has been organized in four reports format, which means that Chapters 2 through 5 are meant to be standalone texts. As a consequence, the base theory is presented in each of the chapters, therefore some discussions are repeated in them. In every chapter of this thesis we use comparison metrics based on the ideas from Zhang et al. (2017) to validate the model. These metrics are detailed in appendix Section A.2.3, and is the same for all chapters.

Some real data examples are used throughout the thesis. Each chapter briefly describes the data sets used in them, but the following references provide deeper descriptions as well as the context in which they were collected. The Schizophrenia data set used in Chapters 2 and 5 is detailed by Diggle et al. (1998). AIDS data set used in Chapter 2 is detailed by DeGruttola and Tu (1994). The Liver cirrhosis data set used in Chapters 3 and 4 in Andersen et al. (1993) and the diarrhea study with Brazilian children in the state of Bahia used in Chapter 5 is detailed in Barreto et al. (1994).

Chapter 2

Joint Modeling of Time-to-event and Longitudinal Measurements Data: A Dynamic Generalized Hierarchical Approach

This chapter introduces a broad class of dynamic generalized hierarchical models for jointly analyzing the behavior of a sequence of longitudinal measurements and time-to-event data. The distribution of the time-to-event is conditional on the filtered longitudinal measurements and follows a dynamic generalized hierarchical structure. This class includes and extends a number of specific models proposed in recent years. Bayesian inference is implemented via Markov chain Monte Carlo methods. The proposed methodology is applied to the analysis of clinical trial datasets on schizophrenia and AIDS.

2.1 Introduction

Over the past decade joint models for sequences of longitudinal measurements and time-to-event data have become a valuable tool in chronic diseases studies. Many clinical trials consist of following up a group of individuals over time to examine the relationship between one or more explanatory variables (longitudinal measurement) and the risk of developing a disease (time-to-event) (Ibrahim et al. 2010).

In practice, what is being modeled are situations in which individuals are observed and monitored on a regular or intermittent schedule until the event's occurrence. For example, in HIV infected patients clinical trials a natural biomarker for the evolution of the disease is the CD4 lymphocyte count. This biomarker is measured intermittently

and its relationship with time to seroconversion or death is of interest.

Wen and Menggang (2014) mention three classes of joint models which can be grouped into two main families which are widely used in practice. One is the shared random effects (*SRE*) class of models, which includes characteristics of the longitudinal marker as predictors in the model for the time-to-event data. This family mainly has three alternative approaches. The first models the time-to-event data given the longitudinal markers, the second models the longitudinal markers given the time-to-event, and the third models both components given a common term. A less well-known approach is the joint latent model family, which involves assuming that the population is divided into homogeneous groups of individuals with respect to both the marker trajectory and the event risk, as in Proust-Lima et al. (2014). This class of models will not be pursued here. Instead, in this thesis, a multivariate model is proposed, but due to its complexity most of the work is focused on a hierarchical univariate model, which fits in the *SRE* class.

Both families of joint models consider that the observations from all individuals are not independent and identically distributed, indicating that each individual shares a common average, being disturbed by some noise per individual usually known as a random effect.

Under the parametric and semi-parametric approaches, a wide class of models was proposed by Henderson et al. (2000), whose inference was developed via the EM algorithm. Alternatively, Guo and Carlin (2004) provide a Bayesian model in the class of *SRE* models. The hierarchical model developed by Guo and Carlin (2004) contains intercept and slope coefficients and the linking of sub-models is done through latent parameters common to both components describing a global hierarchical regression.

Almost all recent developments in joint models assume a longitudinal Gaussian model for the marker and a right-censored survival model for the time-to-event component (Brown et al. 2005; Chi and Ibrahim 2006; Ibrahim et al. 2010; Guo and Carlin 2004). However, in most studies on chronic diseases, the variable for prediction does not always follow a Gaussian distribution. The repeated measurements, say single or multiple progression markers of a disease, are continuous or discrete random quantities. On the other hand,

the time-to-event data can be of a different nature: single or multiple events, recurrent events, competing risks, etc. Nevertheless, our contribution covers the case of a single marker of disease progression and a single event time. Our effort concerns the joint distribution of all components of the model. These include the time-to-event component as well as the longitudinal component and all the uncertainty presented. In short, this means that the event time is described conditional on parameters and on the longitudinal component, which in turn has its own parameters.

When this type of data is modeled, the reason for the probabilistic description of the longitudinal markers is that they are assumed to be measured with error. Our purpose is to decompose this marker measurement to obtain its filtered unobservable true trajectory. Following that, some adequately chosen (preferably smooth) terms of this sub-model are used as inputs in the survival model. These terms will be referred to as link components throughout the text. It will be shown that some models in the literature can be written as the proposed model when a convenient link is used. Note that if the marker measurement has no uncertainty, the survival data can be analyzed as simply as if it had known covariates, and any standard model can be used, for instance the extended Cox model Rizopoulos (2012, Chapter 3). Thus, given the link, the time-to-event model is completely defined.

The longitudinal process is described given a temporal trend. In this context, this trend is usually modeled globally, that is, through fixed parameters of a polynomial (Chi and Ibrahim 2006) or spline (Brown et al. 2005) component. Instead, we propose a dynamic polynomial trend, i.e., the parameters evolve over time, which determines a local temporal trend. An important advantage of this trend description is that the proposed model allows the prediction of survival probabilities based on dynamic prediction, i.e., dynamic risk predictions per individual in the sense of Rizopoulos (2011).

We explore the incorporation of the temporal trend into the sub-model in two ways. The first is the Gaussian multivariate version, whose general construction includes regular follow-up times for each individual. Although it is theoretically feasible, its computational cost is high. In the second approach, we come closer to models more

used in practice by making simplifying hypotheses that yield a hierarchical dynamical model (Gamerman and Migon 1993). Furthermore, the hierarchical version can be easily extended to response variables that belong to the exponential family (Paez and Gamerman 2011), which is not usually seen in the context of the *SRE* class. In short, the model we are proposing is an extension of the Guo and Carlin (2004) model.

In order to compare the proposed model with others in the recent literature, we apply the comparison criterion introduced by Zhang et al. (2017), useful to measure the contribution of the longitudinal component.

The applications presented in this article came from the studies of Henderson et al. (2000) and Guo and Carlin (2004). The former authors investigated the relationship between the patient's mental state measured through a score from a weekly follow-up and the times of events for reasons related to the mental state score. The AIDS dataset was analyzed in Guo and Carlin (2004) and Rizopoulos (2011) and both work with transformed data. We use the dataset to illustrate the proposed method and work with the data in the original scale, which is an advantage as will be shown.

We can extend the idea of joint approach to other types of models, such as frailty, cure rate and multistate models. The thesis is divided into four sections. Section 2.2 presents the structure of the model. Section 2.3 discusses the inference procedure. In Section 2.4 of the thesis appendix a small simulation study is presented. In Section 2.5, two datasets are analyzed using the proposed methodology. Finally, Section 2.6 concludes the thesis with a discussion and remarks on future works.

2.1.1 The Dataset Structure

The main data to be modeled are time-to-event observations, which is right-censored. In order to explain and predict these times, suppose that covariates are available for each individual. One of these covariates is observed and monitored on a regular schedule until the occurrence of the event and is assumed to be measured with error. As a way to deal with this, a longitudinal component is added to the model.

This component, in turn, also has available covariates which are used to better model its measurement error. Any given covariate can be used in the time-to-event part, the longitudinal part, or both.

An example, as is seen later in Section 2.5.1, could be a patient with Schizophrenia who is followed up until he experiences a crisis. In each interview, this patient fills up a form which will be used to constitute a score (called PANSS¹ score) which quantifies his syndrome in a scale. In this example, the PANSS score is the longitudinal measurement and the crisis is the event.

2.1.2 Notation and Definitions

Suppose that we have a set of N individuals followed-up and measured with respect to one marker over a time interval $(0, \tau]$, where τ is the maximum event time. Those measurements may be collected at different time points for each individual and provide information for the measurement until the occurrence of the event of interest.

Specifically, the observed time of the event of interest, called time-to-event, for the i^{th} individual is defined by $T_i = \min(T_i^*, C_i)$, where C_i denotes the censoring time and T_i^* denotes the true event time.

Formally, the sequence of scheduled dates, s_{ij} , must satisfy the relationship $0 \leq s_{i1} < s_{i2} < \dots < s_{iJ_i} \leq t_i$, where t_i is the observed value of T_i and J_i is the number of follow-up measurements for the i^{th} individual, $j = 1, \dots, J_i$.

Let d_i represent the indicator of right censoring for the i^{th} individual, that is $d_i = 1$ if individual i experiences the event at time $T_i^* \leq C_i$ and $d_i = 0$ if it is censored.

Finally, we denote density functions by $[\cdot | \cdot]$, which represents a conditional density function.

¹Positive And Negative Syndrome Scale

2.2 Joint Model Buildup

The model involves two components. The first determines how the covariates influence the time-to-event component and the second is the model for the marker measurement. This component will be modeled with different levels of complexity.

2.2.1 Modeling the time-to-event data

The model for the event time is done as in Guo and Carlin (2004), but rewritten in hierarchical form, as follows:

$$\begin{aligned}
 T_i | r, \boldsymbol{\xi}_i &\sim \mathcal{W}(r, \exp\{\mathbf{X}'_{1i}\boldsymbol{\beta}_1 + \boldsymbol{\xi}'_i\boldsymbol{\alpha}\}), \\
 \log(h_i(t)) &= \mathbf{Z}'_i\boldsymbol{\gamma}_i, \quad \mathbf{Z}'_i = (\mathbf{X}'_{1i}, \boldsymbol{\xi}'_i), \\
 \boldsymbol{\gamma}_i &= \boldsymbol{\gamma} + \boldsymbol{\zeta}_i, \quad \boldsymbol{\zeta}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\zeta), \quad i = 1, \dots, N.
 \end{aligned} \tag{2.1}$$

Note that this model is equivalent to modeling the hazard function for each individual h_i as a relative risk function, that is, $h_i(t) = h_0(t) \exp(\mathbf{Z}'_i\boldsymbol{\gamma}_i)$, where $h_0(t)$ is some baseline hazard function and $\mathbf{Z}'_i\boldsymbol{\gamma}_i$ is the same as $\mathbf{Z}'_i\boldsymbol{\gamma}_i$, only without the first term. The time-to-event of the i^{th} individual follows a parametric Weibull hierarchical regression, denoted by $\mathcal{W}(\cdot, \cdot)$, whose relative hazard is described by a multiplicative model, with $r > 0$ as the shape parameter and $h_0(t) = rt^{r-1}$ as the Weibull baseline hazard function.

The regressor \mathbf{Z}_i is a $(q_1 + q^*)$ -dimensional vector. This term includes two blocks. The first is composed of q_1 regressors, denoted as \mathbf{X}_{1i} , assumed to be known without measurement error. We include the index 1 to the covariates associated with the time-to-event sub-model. The second is an arbitrary q^* dimensional latent term $\boldsymbol{\xi}_i$, which can be understood as a placeholder related to a longitudinal marker taken over time. This placeholder is chosen by the researcher and different choices, such as longitudinal mean or median, imply different models. We discuss this in more details in Section 2.2.1.

The regression coefficient $\boldsymbol{\gamma}'_i = (\boldsymbol{\beta}'_{1i}, \boldsymbol{\alpha}')$ is a $(q_1 + q^*)$ -dimensional vector capturing the effect of \mathbf{X}_{1i} and also $\boldsymbol{\xi}_i$. This term is decomposed as the sum of a common mean $\boldsymbol{\gamma}$, with a individual-specific latent component, $\boldsymbol{\zeta}_i$, which is often called the frailty term, is normally distributed with zero mean and variance $\boldsymbol{\Sigma}_\zeta$.

Consider that the measurements for each individual may occur at different follow-up times, so irregularly spaced interview dates should be considered. For this, let $\delta_{ij} = s_{ij} - s_{i(j-1)}, j = 1, \dots, J_i$ be defined as the time between consecutive visits. The situation where all J_i are potentially different for each individual will be called the non-regular case. In contrast, the regular case is defined as the case where $J_i = J, \forall i$, although the times s_{ij} need not be the same.

The Latent Term ξ_i : The Link Component

We address now the flexibility of the proposed model through the choice of the ξ_i term. This term carries the information of one measurement taken over time, called the longitudinal measurements. A common strategy is to model this measurement, as it is assumed to be taken with error, and use the random effect or the average trajectory in ξ_i , which is generally the linear predictor of the longitudinal model. For non-Gaussian longitudinal outcomes a transformation of the average must be used in ξ_i . However, the proposed method has a wider range of possibilities, and the researcher may choose the terms which make the most sense to their study. In this chapter the average trajectory will be used and in one of the applications an alternative is presented.

Our proposal is focused on the modeling this measurement. For this we assume a dynamic linear structure. Thus, the model is supported by the dynamic linear model (DLM) theory in Migon et al. (2005). This allows for the forecasting of yet unattained values in a natural and probabilistic fashion, according to DLM theory. These predictions can be used in turn to predict future failure probabilities, which is comparable to Rizopoulos (2011) and Guo and Carlin (2004).

2.2.2 The Multivariate Longitudinal Model

This model assumes that the measurement markers are Gaussian and independent across individuals. A multivariate polynomial dynamic linear Gaussian model is proposed stacking the individuals' measurements at the visit j in a vector assuming the regular case.

Let \mathbf{y}_j be an N -dimensional vector, $\mathbf{y}_j = (y_{1j} \dots, y_{Nj})'$ representing the N individuals evaluated at visit times $j = 1, \dots, J$. The vector \mathbf{y}_j potentially contains missing

information, so we call $\mathbf{\Lambda} = (\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_J)$ the true unobserved $N \times J$ -dimensional matrix, so that each λ_{ij} is the expected value of each y_{ij} . We represent the N -dimensional j^{th} column of $\mathbf{\Lambda}$ as $\boldsymbol{\lambda}_j$.

An intuitive choice to use as $\boldsymbol{\xi}_i$ in (2.1), i.e. the term which connects equations (2.1) and (2.2), is $\boldsymbol{\lambda}_i$, that is, the J -dimensional i^{th} row of $\mathbf{\Lambda}$, which represents the average trajectory per individual at all observed times. Consequently, $q^* = J$ in (2.1). Note that this implies that all individuals contribute with all J times to the hazard function, which may not be compatible with the reality of the study, since some individuals have their event before the end of the study. This means that the interview times after their event times have no observed value. By treating these interviews as missing data, Bayesian inference can easily incorporate their uncertainty into the hazard function, and their predictive distribution does not have a strong influence on the final result.

A normal multivariate DLM with a polynomial component of order p and q_2 covariate effects associated with the j^{th} interview for all individuals jointly is given by

$$\begin{aligned} \mathbf{y}_j \mid \boldsymbol{\lambda}_j &= \boldsymbol{\lambda}_j + \boldsymbol{\epsilon}_j, \quad \boldsymbol{\epsilon}_j \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \\ \boldsymbol{\lambda}_j &= \mathbf{F}'\boldsymbol{\theta}_j, \\ \boldsymbol{\theta}_j &= \mathbf{G}_j\boldsymbol{\theta}_{j-1} + \mathbf{w}_j, \quad \mathbf{w}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{W}), \\ \boldsymbol{\theta}_0 \mid \mathcal{D}_0 &\sim \mathcal{N}(\mathbf{m}_0, \mathbf{C}_0), \quad j, \dots, J, \end{aligned} \tag{2.2}$$

where $\boldsymbol{\theta}'_j = (\boldsymbol{\mu}'_j, \boldsymbol{\beta}'_{2j})$ is a $(p + q_2)N$ -dimensional column vector and consists of the effect of the elements of \mathbf{F} for all individuals; the term $\boldsymbol{\mu}_j$ is a pN -dimensional vector describing the polynomial effect, and the term $\boldsymbol{\beta}_{2j}$ is a q_2N -dimensional vector describing the covariates effects. These terms are allowed to evolve smoothly over time, according to the evolution equation.

The $\boldsymbol{\theta}_j$ evolution equation defines a multivariate dynamic linear model with quadruple $\{\mathbf{F}, \mathbf{G}_j, \boldsymbol{\Sigma}, \mathbf{W}\}$ structure based on West and Harrison (1997).

The component \mathbf{F} is a $N(p + q_2) \times N$ design matrix and contains components representing hierarchical, temporal and fixed effects. More specifically, $\mathbf{F}' = \text{blockdiag}\{\mathbf{F}'_i\}$, where

$\mathbf{F}'_i = (\mathbf{e}'_p, \mathbf{X}'_{2i})$ and the term $\mathbf{e}_p = (1, 0, \dots, 0)$, as a p -dimensional vector. The term \mathbf{X}_{2i} is a q_2 -dimensional vector.

The $(p + q_2)N \times (p + q_2)N$ -dimensional transition matrix \mathbf{G}_j is obtained based on δ_{ij} , which describes the time span between two consecutive appointment schedules for individual i . At $j = 1$ the appointment scheduled is $s_{i0} = 0$, and $\delta_{i1} = s_{i1}$. Specifically, $\mathbf{G}_j = \text{blockdiag}\{\text{diag}(\mathbf{L}_p(\delta_{ij}), \mathbf{I}_{q_2 \times q_2})\}$, where $\mathbf{L}_p(\delta_{ij})$ is a $p \times p$ matrix in which the diagonal elements are all equal to 1 and those above the diagonal are δ_{ij} . This block diagonal transition matrix is defined for each individual. For instance, with $p = 2$, for individual i at interview j , the transition matrix is an upper triangular 2×2 matrix $\mathbf{L}_2(\delta_{ij})$ with elements $a_{11} = a_{22} = 1; a_{12} = \delta_{ij}; a_{21} = 0$.

The terms \mathbf{L}_p and \mathbf{e}_p ensure that a polynomial model of order p is recovered. The term $\mathbf{I}_{q_2 \times q_2}$ indicates that the covariate coefficients follow a random walk evolution but can also be fixed throughout time. Many alternative evolution's structure can be considered as can be seen in Section 2.5.

The covariance matrix $\mathbf{\Sigma}$ in the multivariate proposal leads us to estimate the very dense $N \times N$ matrix with $N(N + 1)/2$ free unknown parameters. To work around this difficulty, we will require that a structure be imposed on the elements of $\mathbf{\Sigma}$ and some restrictions must be assumed. This point will be taken up later.

The last term \mathbf{W} defines the $(p + q_2) \times (p + q_2)$ covariance matrix with elements on the diagonal associated with the variance related to polynomial and covariates effects. Thus $\mathbf{W} = \text{blockdiag}\{\mathbf{W}_p, \mathbf{0}_{q_2 \times q_2}\}$, specifically $\mathbf{W}_p = \text{diag}(W_p)$ is a p -dimensional matrix in which each element is associated with the polynomial effects and the term $\mathbf{0}_{q_2 \times q_2}$ is a $q_2 \times q_2$ matrix for the covariates effects and defines the coefficients as fixed, i.e., we set the last q_2 variances of \mathbf{W} to zero. Then β_{2j} are constant and the coefficients are parameterized as a fixed linear function recovering the longitudinal part as is widely used in literature.

At time $j = 0$, define the initial prior information $\boldsymbol{\theta}_0 \mid \mathcal{D}_0 \sim \mathcal{N}(\mathbf{m}_0, \mathbf{C}_0)$, for

known \mathbf{m}_0 and \mathbf{C}_0 . The term \mathcal{D}_j contains all the information collected until time j , $\mathcal{D}_j = \{\mathcal{D}_{j-1} \cup \mathbf{y}_j\}$ including all individuals, for $j = 1, \dots, J$. The initial information is summarized in \mathcal{D}_0 , containing all hyperparameters associated with the prior distributions.

Note that the term $\mathbf{F}'\boldsymbol{\theta}_j$ represents the mean trajectory of all individuals at time j . Furthermore, given $\boldsymbol{\xi}_i$, the failure time t_i in (2.1) is conditionally independent of y_{ij} . This means that changes in the modeling of y_{ij} influence the time-to-event model in a filtered manner through the appropriate i^{th} block of $\mathbf{F}'\boldsymbol{\theta}_j$ when $\boldsymbol{\xi}_i = \boldsymbol{\lambda}_i$.

Returning to the dense $\boldsymbol{\Sigma}$ matrix with many different and non-null elements, we need to take care of its structure. To do that, we assume the same follow-up times for all individuals s_j with J visits, which is theoretically restrictive, but is not unusual to find in practice. This allows us to establish a single common representation of the transition matrix \mathbf{G}_j at visit j for all individuals, which leads to a common $\boldsymbol{\theta}_j$ term for all individuals. A direct consequence is the sharing of information of all individuals for the estimation of $\boldsymbol{\theta}_j$, where the expected value of each individual y_{ij} can be obtained simply with a random effect. This is ultimately equivalent to $\boldsymbol{\Sigma}$ being a diagonal matrix.

Common Evolution $\boldsymbol{\theta}_j$

The change in the modeling structure of the component related to the single marker forces the term $\boldsymbol{\theta}_j$ to be common for every individual.

Note that now in equation (2.2), the term \mathbf{G}_j is a $(p + q_2) \times (p + q_2)$ transition matrix and is the same for all individuals; the $\boldsymbol{\theta}_j$ component is a $(p + q_2)$ -dimensional vector capturing the effect of the elements of \mathbf{F}_i ; the perturbation \mathbf{w}_j is a $(p + q_2)$ -dimensional vector following a zero mean multivariate normal distribution with diagonal covariance matrix \mathbf{W} ; the other terms coincide with the previous Section 2.2.2. Furthermore, note that what once had dimension $N(p + q_2)$ now has dimension $(p + q_2)$.

One can see that the joint model is simple, since it is constructed as a relative risk Weibull regression model of the time-to-event data, conditional to a latent longitudinal dynamic linear Gaussian model. A powerful feature of this model is that the *SRE* class of models

presented in Ibrahim et al. (2010), Henderson et al. (2000), Guo and Carlin (2004) and Rizopoulos (2012, Chapter 4), can all be written as special cases of it. In Section A.2.1 of the thesis appendix we discuss how these cases are recovered using our proposal.

Considering the regular case, we present an extension of the model in case the longitudinal marker is not normal.

Hierarchical Dynamic Generalized Model

A hierarchical structure is commonly used when modeling this type of longitudinal variable, adding a random effect per individual to λ_{ij} , say v_i . This term ensures that the mean trajectory is different between all individuals beyond the covariates in the F_i . Also, this makes the proposed model comparable with standard models in the literature. Usually, models assume that the mean trajectory g -transform (i.e. the linear predictor λ_{ij}) is some continuous function of time which has been observed at a finite set of points, thus it is usually modeled as a Gaussian process (Henderson et al. 2000). We do not go that far and instead do not make any assumptions beyond what is observed other than they are Gaussian at those specific time points.

Notice that because of the change of dimension of \mathbf{G}_j and consequently $\boldsymbol{\theta}_j$, the unidimensional λ_{ij} term is now modeled individually. Also, it becomes more practical only to use \mathbf{F}_i multiplied by the common term $\boldsymbol{\theta}_j$ to model λ_{ij} , instead of using all of them in a single equation through \mathbf{F} multiplied by a huge $N(p + q_2)$ -dimensional $\boldsymbol{\theta}_j$. Theoretical discussions on this can be seen in Ibrahim et al. (2010) and Rizopoulos (2012, Chapter 2). This addition is applied in the context of the model described below.

Suppose that the density functions of the outcomes y_{ij} are conditionally independent and come from the exponential family of distributions, denoted as *Ef*. The model for the observations $y_{i1}, y_{i2}, \dots, y_{iJ}$ for the i^{th} individual, at times $s_{i1}, s_{i2}, \dots, s_{iJ}$ is defined as the hierarchical generalized dynamic linear model to model the longitudinal component and

it is given by

$$\begin{aligned}
y_{ij} &\sim Ef(\eta_{ij}, \psi), & g(E(y_{ij} | \eta_{ij}, \psi)) &= \lambda_{ij}, \\
\lambda_{ij} &= \mathbf{F}'_i \boldsymbol{\theta}_j + v_i, & v_i &\sim \mathcal{N}(0, V), \\
\boldsymbol{\theta}_j &= \mathbf{G}_j \boldsymbol{\theta}_{j-1} + \mathbf{w}_j, & \mathbf{w}_j &\sim \mathcal{N}(0, \mathbf{W}), \\
\boldsymbol{\theta}_0 | \mathcal{D}_0 &\sim \mathcal{N}(\mathbf{m}_0, \mathbf{C}_0).
\end{aligned} \tag{2.3}$$

Note that equation (2.3) differs from (2.2) in that the former is a univariate hierarchical form while the latter is a multivariate Gaussian representation. To put it more formally, $[y_{ij} | \eta_{ij}, \psi] = b(y_{ij}, \psi) \exp(\psi(y_{ij}\eta_{ij} - a(\eta_{ij})))$, with mean $E(y_{ij} | \eta_{ij}, \psi) = g^{-1}(\lambda_{ij})$, which is assumed to be some twice-differentiable function of λ_{ij} , where η_{ij} and ψ denote the natural and dispersion parameters in the exponential family, respectively. The terms $a(\cdot)$, $b(\cdot)$ are known functions specifying the member of the exponential family. Different choices for these functions include the Binomial, Poisson, Gamma, and Normal distributions among others (van den Hout and Muniz-Terrera 2016).

We complete the specification of the first level by defining $g(\cdot)$ as a known monotonic link function. The latent term λ_{ij} is a time-varying component for each individual, modeled via the dynamic linear model with regression components associated with the effects $\boldsymbol{\theta}_j$, which is allowed to evolve smoothly over time. The other terms coincide with the previous longitudinal Gaussian model (2.2). Here, the reader can easily verify that model (2.3) is a special case of equation (2.2) under the regular case, when the Gaussian family is chosen.

In many applications, the individuals present highly nonlinear longitudinal trajectories, so it is advisable to consider flexible time functions to describe them. Often this is done via high-order local polynomials or splines. It is well known, that a dynamic polynomial model is equivalent to polynomial splines (Wahba 1978; Biller and Fahrmeir 1997).

The link component between the longitudinal and survival processes, $\boldsymbol{\xi}_i$, can be achieved in many alternative ways beyond $\boldsymbol{\lambda}_i$. Some authors, like Ibrahim et al. (2010) and Rizopoulos (2012, Chapter 5), link these components through the average longitudinal trajectory, while others link them through a latent stochastic dependence, such as a *latent* zero-mean bivariate Gaussian process as in Henderson et al. (2000). Many alternative are

present in Rizopoulos and Ghosh (2011). The link component between the two sub-models is very flexible in this model. Exploring this flexibility, though potentially powerful, could add much more complexity to the model description and will not be discussed further in this work.

Joint Distribution and Likelihood Function

Let us denote the available information as $\{y_{ij}, t_i, d_i; j = 1, \dots, J; i = 1, \dots, N\}$. Then the joint density can be written as $[y_{ij}, t_i, d_i | \Theta] = [t_i, d_i | \lambda_i, \Theta][y_{ij} | \lambda_i, \Theta]$ and the joint likelihood of the joint dynamic hierarchical model (2.3) is

$$L(\Theta | \mathbf{y}, \mathbf{t}, \mathbf{d}) = \prod_{i=1}^N [t_i, d_i | \xi_i, \gamma_i, r][\gamma_i | \gamma, \Sigma_\zeta] \\ \times \prod_{j=1}^J [y_{ij} | \lambda_{ij}, \psi][\lambda_{ij} | \theta_j, V][\theta_j | \theta_{j-1}, \mathbf{W}][\theta_0].$$

We denote the complete unknown parameters vector by $\Theta = \{\psi, \Lambda, \theta, V, \mathbf{W}, \Sigma_\zeta, \gamma, r\}$.

2.3 Inference

In order to complete the requirements for Bayesian analysis, prior distributions must be defined. In this context, we will be restricted to standard proper prior distributions, with hyperparameter values chosen so that the priors have minimal impact relative to the data.

The term ψ in (2.3) is an uncertainty related to the longitudinal measurement and in the Gaussian model this term is denoted by σ^2 . We use the inverse Gamma prior distribution $IG(a_g, b_g)$, where a_g, b_g are known hyperparameters for the variance components $\{\sigma^2, V\}$. To \mathbf{W} and Σ_ζ matrix, we use inverse-Wishart, $IW(\mathbf{R}, g)$ scale matrix \mathbf{R} and g degrees of freedom.

To reduce notation of the posterior distribution, we define the vector $\theta' = (\theta'_1, \dots, \theta'_J)$ which represents the latent terms that drive the system evolution. As already defined in equation (2.3), the first p components of each θ_j is time dependent while the last q_2 components are time-invariant and follow a $\mathcal{N}(\mathbf{m}_\beta, \mathbf{C}_\beta)$. For the parameters $\gamma \sim \mathcal{N}(\mathbf{m}_\gamma, \mathbf{C}_\gamma)$ related to the regression in the survival sub-model, we use a multivariate normal distribution with mean \mathbf{m}_γ and variance \mathbf{C}_γ . Considering the comment about

estimating the shape parameter r presented by Guo and Carlin (2004), we used $r \sim \text{Exp}(c)$, where $c = 0.1$ in the applications. This prior distribution is slowly decreasing on the positive real line. In our examples, the other hyperparameters were chosen to represent fairly vague prior information, in particular the hyperparameters of inverse Gamma priors are given by $a_g = b_g = 0.01$ which has infinite mean and variance. For the multivariate Normal distribution the mean vector as $\mathbf{0}$ and covariance matrix $\text{diag}(100)$ and for the inverse-Wishart the hyperparameters are \mathbf{R} as diagonal matrix with 100 in every entry of the diagonal and adequate dimension for the problem, i.e. p and $g = 10$.

2.3.1 Posterior Inference

Since the $\{\mathbf{F}_i, \mathbf{G}_j, V, \mathbf{W}\}$ quadruple depends on unknown parameters, a sequential analysis of $\boldsymbol{\theta}$ would not produce an adequate inference (not completely specified). Instead, all unknown components of the model must be treated as parameters and an adequate inference procedure is to find their posterior distribution.

Assuming independence in the prior distributions, the posterior distribution is given by

$$\begin{aligned}
[\psi, \boldsymbol{\theta}, V, \mathbf{W}, \boldsymbol{\gamma}, \boldsymbol{\Sigma}_\zeta, r \mid \mathcal{D}_J] &\propto \prod_{i,j=1}^{N,J} b(y_{ij}, \psi) \exp(\psi(y_{ij}\eta_{ij} + a(\eta_{ij}))) \\
&\propto [\lambda_{ij} \mid \boldsymbol{\theta}_j, V][\boldsymbol{\theta}_j \mid \boldsymbol{\theta}_{j-1}, \mathbf{W}][\boldsymbol{\theta}_{j+1} \mid \boldsymbol{\theta}_j, \mathbf{W}] \times (rt_i^{r-1} \exp(\mathbf{Z}'_i \boldsymbol{\gamma}_i))^{d_i} \exp\left(-\int_0^{t_i} h_i(u) du\right) \\
&\quad \times [\boldsymbol{\gamma}_i \mid \boldsymbol{\gamma}, \boldsymbol{\Sigma}_\zeta][\psi][V][\mathbf{W}][\boldsymbol{\Sigma}_\zeta][\boldsymbol{\gamma}][r][\boldsymbol{\theta}_0],
\end{aligned} \tag{2.4}$$

where $b(\cdot, \cdot)$ and $a(\cdot)$ characterize the distribution in the exponential family and ψ is the respective dispersion parameter. Since the joint posterior distribution in (2.4) does not have a known closed form, we propose the use of MCMC methods to obtain samples from it. The main concern is to sample from the full conditional density of $\boldsymbol{\Theta}$, which also does not have a closed form.

Here we present the inference scheme to obtain posterior distribution samples for the unknown quantities in the model based on our joint model defined in (2.1) and (2.3). Formally, we used equations (2.1) and (2.3) to find the full conditional density.

Implementation of the Gibbs sampler is straightforward. Let $[\boldsymbol{\theta}_j \mid D_J, \cdot]$ denote the

distribution of $\boldsymbol{\theta}_j$ given the remaining parameters, for example. Then at each iteration of the Gibbs sampler with Metropolis steps, we proceed as follows:

- Sample the conditional density of $\boldsymbol{\theta}_j$ from conditional density $\boldsymbol{\theta}_j \mid \boldsymbol{\theta}_{-j}, \mathcal{D}_J, \cdot \sim \mathcal{N}((NV^{-1} + |\mathbf{W}|^{-1})^{-1}(|\mathbf{W}|^{-1}\mathbf{G}_j\boldsymbol{\theta}_{j-1} + V^{-1}\sum_{i=1}^N \lambda_{ij}), (NV^{-1} + |\mathbf{W}|^{-1})^{-1})$;
- Sample \mathbf{W} from parametric model $\mathbf{W} \mid \mathcal{D}_J, \cdot \sim IW(\mathbf{R} + \hat{\mathbf{H}}, g + p + J)$, where $\mathbf{h}_j = \boldsymbol{\theta}_j - \mathbf{G}_j\boldsymbol{\theta}_{j-1}$, $\hat{\mathbf{H}} = \sum_{j=1}^J \mathbf{h}_j\mathbf{h}'_j$;
- Sample ψ from parametric model $[\psi \mid, \mathcal{D}_J, \cdot] = IG(N+J+a_g, \sum_{i,j}^{N,J} (y_{ij} - \lambda_{ij})^2/2 + b_g)$;
- Sample V from parametric model $[V \mid, \mathcal{D}_J, \cdot] = IG(N+J+a_g, \sum_{i,j}^{N,J} (\lambda_{ij} - \mathbf{F}'_i\boldsymbol{\theta}_j)^2/2 + b_g)$;
- sample $\boldsymbol{\Sigma}_\zeta$ from parametric model $\boldsymbol{\Sigma}_\zeta \mid \mathcal{D}_J, \cdot \sim IW(\mathbf{R} + \sum_{i=1}^N \boldsymbol{\gamma}'_i\boldsymbol{\gamma}_i, N + g + q_1 + q^*)$,
- Sample $\boldsymbol{\gamma}$ from the kernel of conditional density $\log[\boldsymbol{\gamma} \mid \mathcal{D}_J, \cdot] \propto \sum_{i=1}^N (d_i\mathbf{Z}'_i\boldsymbol{\gamma} - \exp(\mathbf{Z}'_i\boldsymbol{\gamma})t_i) - \frac{1}{2}(\boldsymbol{\gamma}'\mathbf{C}\boldsymbol{\gamma} - 2\boldsymbol{\gamma}\mathbf{C}\mathbf{m}_\boldsymbol{\gamma})$;
- Sample $r \mid, \mathcal{D}_J, \cdot$ from $\log[r \mid, \mathcal{D}_J, \cdot] \propto \sum_{i=1}^N d_i \log(rt_i^{r-1}) - cr + \sum_{i=1}^N -rt_i^{r-1} \exp(\mathbf{Z}'_i\boldsymbol{\gamma}_i)t_i$.

In our blocked Metropolis sampling step, all proposed steps are Cauchy or (exponential) with mean equal to the current value and variances tuned to get acceptance rates to get around a 20-50%.

2.4 Simulation Study

To assess the performance of the proposed dynamic hierarchical joint modeling approach, we conducted a small simulation study to investigate parameter estimation for our joint model. We used only three scheduled longitudinal measurements for each individual ($J = 3$). The measurement points used in the simulation are considered at times $s_1 = 1$, $s_2 = 2$ and $s_3 = 3$. We generated scenarios with 500 replications each with $N = 250$ subjects. The true longitudinal trajectory was simulated as $\lambda_{ij} \sim \mathcal{N}(\mathbf{F}'_i\boldsymbol{\theta}_j, V)$ where the initial distribution of $\boldsymbol{\theta}_0$ was $\boldsymbol{\theta}_0 \sim \mathcal{N}(\mathbf{m}_0, \mathbf{C}_0)$, with mean and covariance matrix given by $\mathbf{m}'_0 = (0.5, 0.001)$ and $\mathbf{C}_0 = 10\mathbf{I}_2$. The variance of the random term v_i in (2.3) was chosen as $V = 2$ and the vector $\mathbf{F}'_i = (1, 0)$ is the i^{th} row of the design matrix,

for $i = 1, \dots, N$ and $j = 1, 2, 3$. To describe the evolution equation $\boldsymbol{\theta}_j$ we used the matrix $\mathbf{G}_j = \mathbf{L}_2(1) = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, for $j = 1, 2, 3$. The covariance matrix of the evolution equation was chosen as a diagonal matrix with diagonal inputs $W_1 = 0.05$ and $W_2 = 0.01$ corresponding to the variance of the evolution level and the variance of the evolution increasing term, respectively. Finally, the observed longitudinal data was simulated from a Gaussian distribution with 0 mean and variance $\sigma^2 = 0.05$. Note that $\boldsymbol{\theta}_j = \boldsymbol{\mu}_j$, i.e. there are no covariates in this particular sub-model, so that $\boldsymbol{\beta}_{2j}$ has no components. Given $\boldsymbol{\xi}'_i = (\lambda_{i1}, \dots, \lambda_{iJ})$, the term $\mathbf{Z}'_i = (1, \boldsymbol{\xi}'_i)$ is built. We consider no covariates for the time-to-event sub-model and set the intercept $\beta_1 = 1$. Finally, the main interest in the inference is the vector $\boldsymbol{\alpha}$, chosen as $\boldsymbol{\alpha}' = (0.1, -0.15, -0.05)$, for this vector represents the impact of the longitudinal trajectory on the hazard function. Using these effects, the survival function $S(t) = \exp\left(-\int_0^t h(u)du\right)$ is computed for each time point t . By applying the inversion method, the survival times are generated from equation (2.1) with parameter $r = 1$ associated with the Weibull baseline hazard function $h_0(t) = rt^{r-1}$. Since it is not desirable to have event times much later than the last longitudinal visit, we used the exponential distribution with mean 2 for the right censoring times, so that we expect about 17.4% of the censored times to be above 3.5. Note that we set $\boldsymbol{\Sigma}_\zeta = \mathbf{0}$, which means that $\boldsymbol{\gamma}'_i = (\beta_1, \boldsymbol{\alpha}')$, for $i = 1, \dots, N$. Code used for data generation can be found in Section A.2.6.

The computational implementation is developed in the R language (R Development Core Team 2018) running on an Intel® i5 2300 Windows PC with 4GB RAM memory. The prior distributions considered were described in Section 2.3. We fit our proposed model for each simulated data set.

Table 2.1 shows simulation results based on 500 simulated data sets. The overall results of the scenarios show that the bias of the estimated values of the variances of the longitudinal sub-model are slightly negative in contrast to the variance of the Gaussian perturbation of the equation for y_{ij} in (2.3), σ^2 which presents positive bias. Observe that the standard deviations of the variance terms W_1, W_2 can be considered high. This clearly demonstrates that this and other aspects of the estimation need more research. Although our simulation study was brief, it shows that these true parameter values are (approximately) recovered

Table 2.1: Summaries from 500 simulated data sets under the proposed model (P: parameter, True: true value, Bias: computed with the mean as the measure of location, rMSE: root of the mean-squared error, and SD: average of the posterior standard deviations).

P	True	Bias	rMSE	sd
α_1	0.1000	0.0617	0.0712	0.0356
α_2	-0.1500	0.0380	0.0533	0.0374
α_3	-0.0500	0.0219	0.0416	0.0354
β_{11}	1.0000	0.0525	0.0829	0.0642
r	1.0000	0.0258	0.0453	0.0372
V	2.0000	-0.0122	0.0230	0.0195
W_1	0.0500	-0.0495	0.2303	0.2249
W_2	0.0100	-0.0244	0.1027	0.0997
σ^2	0.0500	-0.0136	0.2630	0.2249

by fitting our model. Other statistics such as relative variation show unfavorable results, which leads us to believe deeper simulation studies may be required.

2.5 Real Data Analysis

The aim of this Section is to show the performance of our proposal compared with competing models. It is also desirable to validate the modeling of the longitudinal component through a dynamic structure in the joint model. In order to do this, we adapted the comparison criteria of Zhang et al. (2017), which is detailed in appendix Section A.2.3.

The dataset analyses are conducted using the R software (R Development Core Team 2018). The prior distribution specification is based on Section 2.3. The parameters are estimated by using MCMC sampling. We ran the chain for 500,000 iterations and discarded the first 100,000 iterations to avoid any transient effect. Thinning of 100 was used to eliminate autocorrelation, providing a final sample of 4000 iterations from which to draw posterior distributions. Convergence was verified with the diagnostics in the R package coda (Plummer et al. 2006). Some of these convergence diagnosis are simple

graphical ways of summarizing the data.

Computational time (calculations were done with an Intel ®i5 2300 Windows PC with 4GB memory) for the Schizophrenia dataset was approximately 4.2 minutes and for the CD4 count dataset, around 9 minutes for each model. Model selection is performed using the comparison criteria mentioned above.

2.5.1 Schizophrenia Dataset

The commonly analyzed Schizophrenia dataset is used to exemplify our proposed method. Henderson et al. (2000) presented a thorough analysis of the use of the shared random effect in joint models using this dataset.

This dataset consists of $N = 150$ patients who were assessed using the Positive and Negative Syndrome Scale (PANSS). This is a medical scale used for measuring symptom severity of patients with schizophrenic conditions. It is decomposed in three scales: positive scale (7-49), negative scale (7-49) and general psychopathology scale (16-112) https://en.wikipedia.org/wiki/Positive_and_Negative_Syndrome_Scale. Lower PANSS scores indicate weaker symptoms and higher scores imply more severe conditions.

Assuming the regular situation as described in Section 2.2.1, the score is denoted by y_{ij} and describes the mental state assessment for the i^{th} individual at scheduled times $s_j = \{0, 1, 2, 4, 6, 8\}$. The assessments are weekly.

Chronically ill mental health patients were randomly allocated on one of the following three treatments: placebo (1), risperidone (2) and haloperidol (3). All groups had forcibly 50 patients in them. Other available information identifies the events due to reasons that were thought to be related to their mental state, i.e. the event times are collected based on the patient's mental state. Eight weeks without news of the patient's mental state configure a non-informative censorship.

Since the event time can also be non-informative, in cases where the event was not related to the individual's mental state, a right censoring indicator was created, receiving

the value 0 when the event is non-informative, and 1 otherwise.

Informative event times occurred for 40% of the individuals; 50% of the remaining ones dropped out for unidentified reasons. The censoring indicator was set to 0 in these cases.

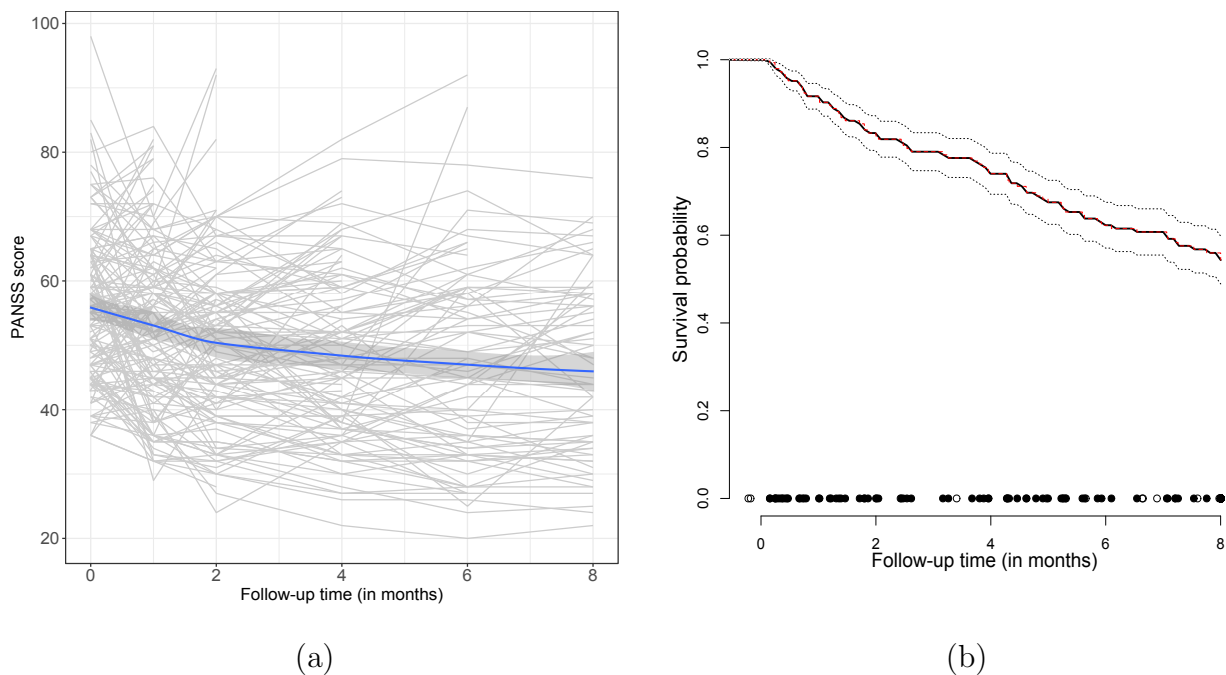


Figure 2.1: (a) Observed trajectories of PANSS score for the 150 patients and average trajectory (blue curve). (b) Kaplan-Meier curve for the follow-up time and 95% pointwise confidence band.

Figure 2.1(a) shows the observed longitudinal measures plotted against time for the 150 patients included in the analysis. The heterogeneity of the patients' PANSS score is apparent in this plot. The smooth curve represents the average trajectory by treatment with its confidence interval. This indicates that a random effect on the average score might be adequate. Figure 2.1(b) summarizes the survival probability based on the nonparametric Kaplan-Meier approach (Kaplan and Meier 1958). The filled dots on the horizontal axis represent uncensored data.

The fitting procedure can be seen in detail in appendix Section A.2.2. Our analysis begins by fitting the best model developed by Henderson et al. (2000) in a Bayesian setup

to explore some aspects of our proposal. In particular, we wish to assess the added value of incorporating a dynamic structure by comparing the contribution of the inclusion of the latent component to explain the hazard function when such a structure is not present. An exploration of the link component between the longitudinal and survival models can also be performed by using components other than λ_{ij} , which shows the versatility of the model. We do not present this here, as our purpose is only to introduce the model in a simple and intuitive way.

In order to reach this objective, we first set the evolution's variance $\mathbf{W} \equiv \mathbf{0}$, which corresponds to assuming time invariance for the vector $\boldsymbol{\theta}_j$ in equation (2.3). It is equivalent to the model of Henderson et al. (2000) and will be denoted the *global dynamic* (GD) model as a benchmark. In opposition, the proposed model is denoted *local dynamic* (LD).

The following Bayesian comparison criteria are used: deviance information criterion (*DIC*) and logarithm of the pseudomarginal likelihood (*LPML*), as described in Zhang et al. (2017), as well as their components pD and $LPML_{Surv|Long}$, which represents the logarithm of the pseudomarginal likelihood restricted to the survival sub-model. Two comparison metrics are derived based on Zhang's idea of contribution differential, called $\Delta LPML_{Surv}$ and ΔDIC_{Surv} . They measure the contribution to the respective metrics of adding the longitudinal measurements sub-model to the joint distribution. For further details, see Section A.2.3 of the thesis appendix.

Both models are compared using the aforementioned criteria. According to Table 2.2, clearly the LD model attains the largest values *LPML* and smallest *DIC*. Thus, the *local dynamic* model is preferred by these methods. Note that the $LPML_{Surv|Long}$ component is very similar for both models, while the other metrics have bigger differences. This is not unexpected, as $LPML_{Surv|Long}$ measures the *LPML* metric for the survival sub-model only, which is the same in both cases.

Also in Table 2.2, the value of including the temporal structure can be verified by the $\Delta LPML_{Surv}$ and ΔDIC_{Surv} values. These comparison criteria are consistently higher for our proposal.

Table 2.2: Comparison of Bayesian model selection methods for the locally dynamic model (LD) and globally dynamic model (GD)

	DIC	pD	$LPML$	$LPML_{Surv Long}$	$\Delta LPML_{Surv}$	ΔDIC_{Surv}	ΔpD_{Surv}
LD	336.446	331.560	-2.924	-0.478	-0.021	336.164	331.550
GD	509.368	505.015	-5.057	-0.585	-0.128	509.085	505.004

Table 2.3: Comparison of Bayesian model selection methods for the fully Bayesian model (FullBayes) and posterior mean of the latent term model (PostMean).

	DIC	pD	$LPML$	ΔDIC_{Surv}	ΔpD_{Surv}
FullBayes	318.005	314.243	-2.090	317.722	314.232
PostMean	497.968	492.859	-2.938	497.685	492.849

Now we discuss the issue of choosing the link between the two components. This point was also discussed in Henderson et al. (2000). The treatment covariate was included both in the dynamic and survival components. To explore the possibilities, two particular models related to our proposal were fitted; the fully Bayesian (*FullBayes*) model links all the information present in the posterior distribution of $\xi_i = \lambda_i$, the latent component of interest. The performance of the *FullBayes* model was compared to a model that links the components through $E(\lambda_{ij} | \mathcal{D}_J)$, that is the posterior mean of the latent term. This model is called *PostMean*.

The results of Table 2.3 heavily favor the (*FullBayes*) model, indicating that linking the components by merely substituting the latent term with an estimate might mean ignoring a great deal of information. This information difference could indicate that the posterior distribution of the latent term might be very skewed.

Also note that the full information of the posterior of $\mathbf{\Lambda}$ carries the hierarchical structure of the model, evidencing the borrowed strength among patients present in the common mean score of PANSS.

In order to single out the value of the dynamic modeling of the longitudinal component, we plotted a graphic with a comparison of the hazard functions. The average hazard functions were plotted. This average is evaluated at every iteration of the MCMC

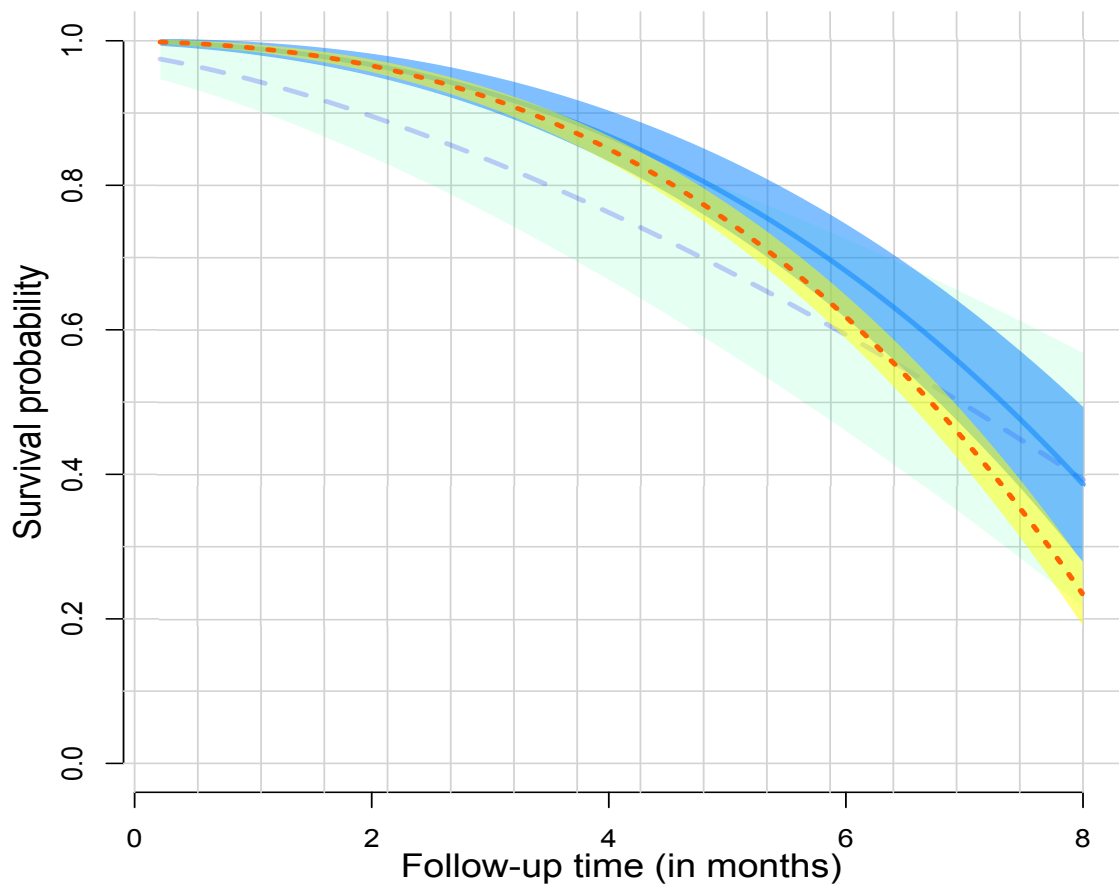


Figure 2.2: Survival probabilities calculated with the posterior mean and 95% credible interval of: solid line our proposal (LD); dashed line a survival regression without the longitudinal process; the dotted lines model from Henderson et al. (2000) (GD).

algorithm. Figure 2.2 shows the posterior curves with their credible intervals, where the difference between the three survival functions can be seen. It is noteworthy that until the third follow-up time, all the functions are roughly the same. Between times 3 and 5.5, while the solid and dashed lines decrease at approximately the same rate, the dotted line decreases more rapidly. After time 5.5, the red (dotted) survival function decreases so fast that it drops below even the violet (dashed) one.

One could interpret this difference by noting that since the violet (dashed) line represents the unfiltered covariate and it evolves over time at a rate comparable to the green (solid) line, then they must have something in common that the red (dotted) one does not. The only difference between the red (dotted) line and the green (solid) one is the dynamic structure of the covariate filter. A logical conclusion is that the green (solid) line successfully captures a structure present in the actual data which makes a difference in the long-term evolution of the survival function. This happens as a consequence of the progressive filtering of the dynamic model.

Another important result involves the comparison of the treatment, which shows the posterior mean of the survival probability curve conditional on the PANSS score (joint model) compared to the survival curve without the longitudinal process. These probabilities show evidence of the difference between the survival probability for patients treated with risperidone and haloperidol at four months when we consider the joint model. See Figure 2.5.1 for further details. One can see that treatment haloperidol provides lesser risk than risperidone, but both are outperformed by the placebo. The next analysis exemplifies the use of our proposal for count data.

2.5.2 AIDS Dataset

The number of CD4 cells per cubic millimeter of blood is widely used as a biomarker for progression of the AIDS syndrome.

In practice, based on a high degree of skewness toward high CD4 counts (see the Figure 2.4(a)), some authors suggest the use of y^a , i.e., some power transformation of the longitudinal measurements. For instance, in the joint model specified for the shared effect

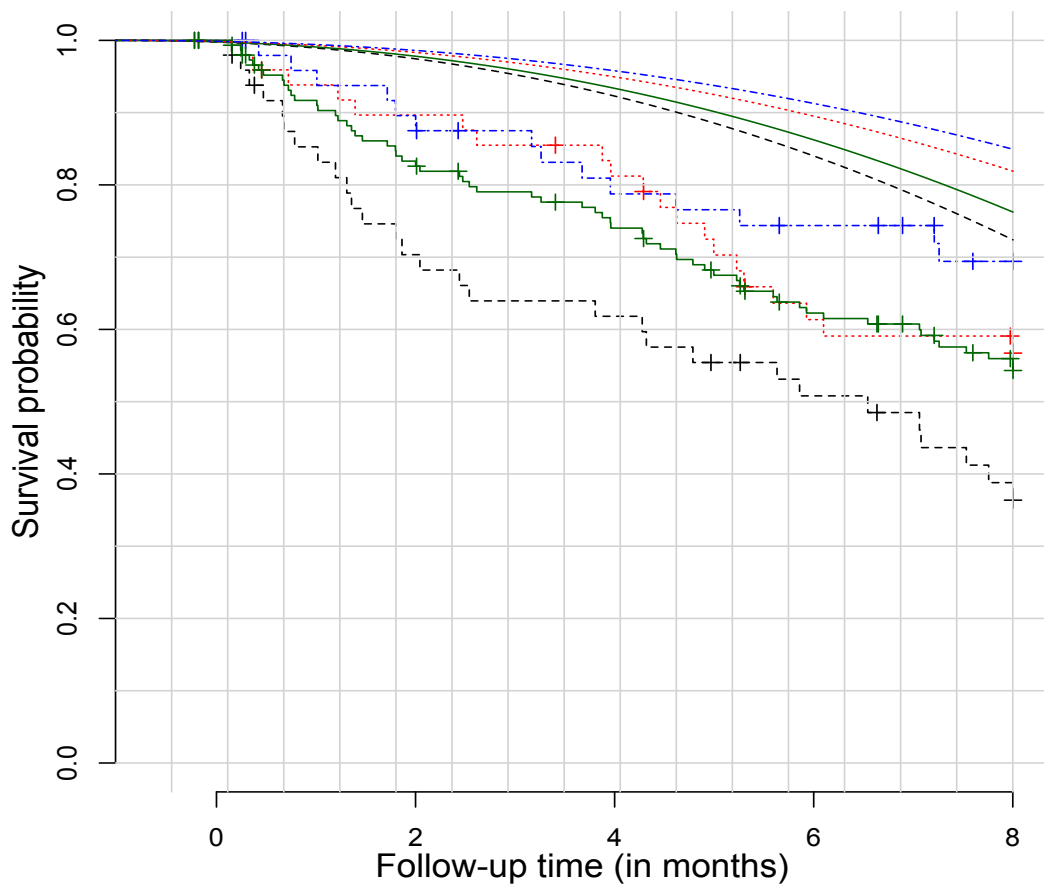


Figure 2.3: Survival probabilities computed with the posterior mean of the parameters. Solid line: our proposal for all individuals, dashed line: patients under treatment placebo, dashed-dotted line: patients under treatment risperidone, dotted line: patients under treatment haloperidol, solid line: all patients, and stepped lines: Kaplan-Meier curves.

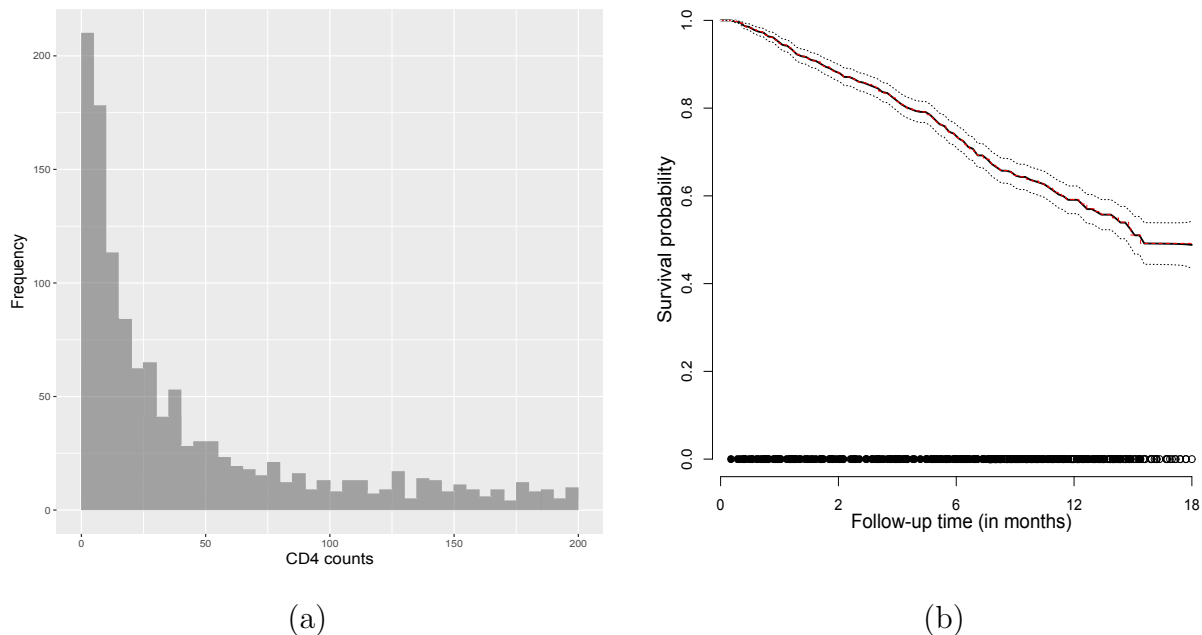


Figure 2.4: (a) Histogram of the observed CD4 cell counts for 467 patients. (b) Kaplan-Meier curves for the follow-up time and 95% pointwise confidence bands.

approach, Guo and Carlin (2004) fitted the joint model using $a = 1/2$, while Rizopoulos (2011) adjusted the CD4 cell counts while working with $a = 1/4$. We could consider an asymmetrical distribution for this, but for now, we focus on the exponential family of distributions.

Deciding which transformation captures the most information about the process of the dataset is a challenge. Our objective is to show that good results can be obtained by analyzing the data in their original scale.

Figure 2.4(a) shows the skewness toward the right in the CD4 cell counts distribution. From the frequency histogram, it can be seen that the lower the counts, the higher the frequency.

For the AIDS dataset, 467 HIV-infected patients were enrolled and randomly assigned to two antiretroviral drugs: didanosine (ddI) and zalcitabine (ddC). The objective was to compare their efficacy and safety. CD4 cell counts were recorded at five time points

$s_j, j = 1, \dots, 5$, representing the times: 0 (baseline), 2, 6, 12 and 18 months. We combined a longitudinal model for the treatment with a model for the survival experience of the patients who had failed (died) or were intolerant of the zidovudine (AZT) therapy. Figure 2.4(b) summarizes the survival probability based on the nonparametric Kaplan-Meier approach, where the filled dots on the horizontal axis represent uncensored times.

We denote the CD4 count for the i^{th} individual at time j as y_{ij} . There are four explanatory variables which were included in the analysis: drug ($ddI = 1, ddC = 0$); gender ($male = 1, female = -1$); previous infection at study entry ($prev = 1$ when AIDS was diagnosed and -1 otherwise); and stratum (AZT $failure = 1, AZT intolerance = -1$).

To show the performance of our proposal, we fit the CD4 count as a joint generalized polynomial DLM for the biomarkers. Two scenerios under regular situations are examined: The CD4 cells as $y_{ij}^{1/2}$ and y_{ij} are fitted by the polynomial model assuming Gaussian and Poisson responses, respectively.

The former is equivalent to the model developed in Guo and Carlin (2004), and the latter is the proposed model where the data are kept in the original scale. For this, the variance \mathbf{W} was set to 0 at all times $j = 1, \dots, J$. For the survival component, model (2.1) was used in both cases. After defining the prior distributions (see Section 2.3), we estimate the parameters by using MCMC sampling.

Table 2.4: Bayesian model selection criteria for the dynamic longitudinal sub-model: Poisson response (Poisson) and normal longitudinal sub-model (Normal). Metrics divided by 10.

	DIC	pD	$LPML$	$LPML_{Surv Long}$	$\Delta LPML_{Surv}$	ΔDIC_{Surv}	ΔpD_{Surv}
Poisson	11.214	1.379	-9.093	-2.022	-0.032	10.105	1.370
Normal	21.718	12.265	-6.807	-2.024	-0.034	20.609	12.256

A similar conclusion is obtained when comparing the Poisson and Gaussian models when

the temporal structure is described in the local sense using $\xi_i = \lambda_{iJ}$ as a link component between sub-models. Table 2.4 indicates that the largest values of $LPML$ and the smallest values of DIC are obtained when the Poisson distribution is assumed for the longitudinal measurements in the original scale. Again, as in the last application, the difference in $LPML_{Surv|Long}$ is very small between both models, since the survival sub-model is the same. More interesting are the Δ metrics. $\Delta LPML_{Surv}$ has a slight preference towards the Poisson model, yet the values are very small, indicating that adding the longitudinal sub-model contributes little to the forecasting in this particular dataset. On the other hand, the ΔDIC_{Surv} value is very large, compared with the absolute DIC metrics of the model. This means that the longitudinal component greatly improves the model fit. Not only that, the Poisson model is strongly preferred over the Normal one.

2.6 Discussion and Future Directions

The main novelty in this thesis is related to the modeling of the longitudinal sub-model, using dynamic and different link components between sub-models. An important point to discuss is the use of the forecast function as a link component between sub-models for prediction in longitudinal data, whenever the hazard function for a individual who has not yet failed is relevant in future times.

Thus, we proposed a new class of joint parametric models for longitudinal and survival data. Although our proposed models are a natural extension of Guo and Carlin (2004), our proposed methods are different from theirs in several aspects.

First, the models proposed by Henderson et al. (2000) are within the static joint model framework. Second, the proposed dynamic models allow transmitting the information using the latent true structure equation and knowing the effect at every measured point in time. Furthermore, this term makes it possible to filter the uncertainty of the measures of the longitudinal part, which makes it more suitable for the survival sub-model.

Third, dynamic models provide great flexibility in modeling the longitudinal measurement because with our longitudinal model, there is no restriction on polynomial's order, as shown by our examples. Fourth, we formally carried out Bayesian model comparisons

via *DIC* and *LPML* decomposition. Fifth, the proposal shows good results when the analysis has non Gaussian outcomes. The applications show the advantage of including local dynamics over global dynamics. A direct future work would be to test the forecasting capabilities of the proposed model.

About the implementation, it is important to emphasize that the algorithms when the response variable is Gaussian was developed in R, while the analysis of count data was implemented in Jags (Su and Yajima 2015) due to easy implementation and fast convergence.

An important analysis that has not been done for this article is robustness verification. All results from our simulations are from estimating models which were the same as the generating ones. In order to verify robustness of the proposal, we need to simulate data from other models and see how well our proposal fits these data, as well as predict them.

We expect that data generated from the models in Rizopoulos (2012, Chapter4) and Guo and Carlin (2004) would be well estimated with our model, since these models are special cases of our proposal, as can be seen in Section A.2.1 of the thesis appendix. However, this actual verification has not been performed for this article.

In this context, recently it has become impossible to ignore the power of analytical approximation of posterior distributions. One cannot ignore that MCMC methods give approximate results to integrals, but numerical INLA in Rue et al. (2009) has been producing impressive results with insignificant computing times. Our intention, then, is to not miss the opportunity in our approach.

The model presented here can be seen as a generalization of what is available in the literature so far. Much has been developed to extend particular cases of the available models. This means that many of these can be adapted for use with the proposal here. We intend to do so, hoping that it can be done with less effort since it is just an adaptation.

Chapter 3

A flexible joint model for time-to-event and a sequence of longitudinal biomarker data

In this chapter we introduce a class of full dynamic hierarchical joint models that relates the simultaneous behaviour of a sequence of longitudinal measurements and the time-to-event for each individual. To allow for greater flexibility, the key components of the model are treated dynamically. For the individual-specific longitudinal evolutions we use a polynomial generalized hierarchical approach while a semi-parametric approach is used for the event time, both of them dynamically structured. This chapter presents a theory which is well established in a survival model context and inserts it when modeling the event time jointly with a stochastic longitudinal component. The advantage of doing this comes in the form of great versatility in structuring the baseline hazard function. The estimation is done under the Bayesian paradigm using MCMC methods.

To compare models, metrics are developed, with a similar logic seen in Zhang et al. (2017), based on the decomposition of the *DIC* and *LPML* allowing to determine the relevance and contribution of the longitudinal trajectory to the model; the performance between the proposed model and a standard model is compared.

The methodological novelties are illustrated through an extensive simulation study and an application on real liver cirrhosis data. The results show the gain of the use of our proposed model.

3.1 Introduction

In many medical trials where the time span before an important event is of interest, the individuals go to repeated interviews or appointments until such an event happens. Such studies are commonly known as follow-up studies. In these interviews, measurements are taken, which are assumed to be related to the studied event, and are commonly referred to as longitudinal measurements.

Many statistical models have been developed for this type of data and are known as joint models. This constitutes a very active area of statistical research that has received a lot of attention in the recent years. In this context the seminal papers of Faucett and Thomas (1996) and Wulfsohn and Tsiatis (1997) introduce what could be nowadays called the standard joint models to treat longitudinal measurements with a survival dataset. Numerous papers have appeared proposing several extensions focused mainly on the mixed model for the longitudinal data and relative risk model for the time-to-event.

Although the standard joint model, i.e. the model proposed in Ibrahim et al. (2010), is an extremely powerful tool for fitting and predicting there are some gaps in the literature. For example, modeling the temporal evolution of the data via DLM could be an advantage. In this context, under a parametric approach, chapter 2 develops a dynamic model to treat the longitudinal markers. Now we take another look at the time-to-event sub-model by proposing a semi-parametric approach. This not only provides more flexibility to the baseline hazard function, but it also allows for some or all the covariate effects in the hazard function to vary over different time intervals.

In particular, a piecewise exponential distribution is used, which may have a temporal dependence structure. The hazard function may be influenced by the longitudinal markers differently at each interval, and this effect might also have a temporal dependence. This proposal provides medically relevant summary measures, which will in fact help accuracy for the decision making process. A consequence of this choice of baseline hazard function is the natural challenge of the choice of that structure. Two forms were taken in this work and are compared with benchmarks in the literature, namely the Markovian process

(Gamerman 1991) and the Gamma process presented in Nieto-Barajas (2002).

Even though the example in this chapter is based on health research, the model is proposed in a very general context being usable in many areas of research, such as sociology, civil engineering, petroleum engineering, industry, economics, social dynamics, education, epidemiological studies among others. For instance, some illustrative examples can be found in Rizopoulos (2012, Chapter 1).

This chapter is organized as follows: the proposed model is introduced in Section 3.2 and the inference procedure is presented in Section 3.3. The new methodology is illustrated with a simulation Section 3.4 and the analysis of real data involving patients diagnosed with cirrhosis in Section 3.5. Finally, in Section 3.6 some conclusion and discussion about the proposed model are drawn.

3.2 Model

The proposed joint model is defined by two sub-models which are linked by some components present in both of them. This link term will be called bridge component. To better understand the model, we first present the data required by it and notation used throughout the chapter.

Suppose a set of N individuals are followed-up and measured with respect to a single marker over the study time length interval $[0, \tau^*)$, where τ^* is the maximum dropout time. Those measurements may be collected at different time points for each individual until the occurrence of the event of interest.

The survival data usually, called time-to-event, consists of the time until the occurrence of an event. It is denoted by $T_i, i = 1, \dots, N$, for the i^{th} individual and is defined by $T_i = \min(T_i^*, C_i)$, where C_i denotes the censoring time and T_i^* denotes the true event time.

Let d_i represent the realization of the random variable D_i which is an indicator of right censoring for the i^{th} individual, that is $d_i = 1$, if individual i experiences the event $T_i^* \leq C_i$ and $d_i = 0$, if it is censored. The time-to-event for each individual is assumed

to be influenced by some regression variables one of which, called bridge, is related to a longitudinal measurement of that individual.

The longitudinal data consist of observed repeated measurements, denoted by y_{ij} , for each individual i in the study. Let its respective random variable be denoted by Y_{ij} . Each individual is followed up in J determined interview dates denoted $s_{ij}, j = 1, \dots, J$. We denote the vector $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iJ})$. Missing-at-random data may occur and the respective y_{ij} for the missing interview s_{ij} are treated as an unknown quantity and can be inserted into the parameter vector of the model.

It is important to note that the sequence of scheduled dates, s_{ij} , must satisfy the relationship $0 \leq s_{i1} < s_{i2} < \dots < s_{iJ} \leq t_i$, where t_i is the observed value of T_i .

For simplicity, all data considered in this work are assumed to be regular. That means that all individuals have the same number of interviews. Thus the situation where the number of visits are potentially different for each individual will be called the non-regular case. The extension to the non-regular case can be developed in a future work through Gaussian processes (Henderson et al. 2000).

Next the two sub-models will be defined. First, the time-to-event model, then the model for the longitudinal response.

3.2.1 Semi-parametric survival model formulation

The survival data is modeled using a flexible piecewise exponential class of distributions (PE) (Ibrahim et al. 2004, Chapter 3). It is reasonable to assume a time structure for the hazard function values, since any true hazard function is probably a smooth function of time, thus two intervals in time would probably not have very different values of the function, which can be intuitively modeled by a time dependence. The procedure to understand the piecewise exponential class consists of the segmentation of the time axis in K disjoint intervals determined by the partition $\boldsymbol{\tau} = \{0 = \tau_0, \tau_1, \dots, \tau_K < \infty\}$ in \mathcal{R}^+ , where $\tau_K > T_{max} = \max\{t_1, \dots, t_n\}$, that is, τ_K is larger than the largest observed time. To denote in which interval the event times fall, we define the variable $I_k = (\tau_{k-1}, \tau_k]$.

In Rizopoulos and Ghosh (2011) an argument is made that the knots positions $\tau_k, k = 1, \dots, K$, will affect the smoothness of the marginal survival function $S(t)$. In this work, once K has been chosen, we divide the time axis in K equally long intervals.

The choice of the number of intervals K is crucial and has been widely discussed in the recent literature (Harrel 2001; Rizopoulos and Ghosh 2011). We take advantage of the inclusion of the longitudinal sub-model in order to deal with this issue. Assuming that our objective is to know the effects of J interviews, we consider J as a lower bound for K . On the other hand it is undesirable that K grows too large, as this increases the number of parameters to estimate. This will be further discussed later.

One way to model the hazard function $h(t)$ is to represent it by a log-linear function of a linear regression whose intercept may be interpreted as the log-baseline hazard function $\log(h_{0k}), k = 1, \dots, K$. This regression's covariates may be decomposed into a set of q_1 known covariates \mathbf{X}_{1i} and one element of the q^* -dimensional latent term $\boldsymbol{\xi}_i = (\xi_{i1}, \dots, \xi_{iK})'$ which carries the filtered information from the longitudinal sub-model. Define the effects $\boldsymbol{\beta}_1$ relative to \mathbf{X}_{1i} and α_k to ξ_{ik} .

Formally, we assume N observations, each potentially right censored, and fixed and known partitions $\boldsymbol{\tau}$. Then, with a fully specified hazard function, the random variable time-to-event has a PE distribution in $\cup_k I_k$ given by the model:

$$\begin{aligned} T_i \mid \boldsymbol{\xi}_i &\sim PE[\mathbf{h}_i(\cdot \mid \boldsymbol{\xi}_i), \boldsymbol{\tau}], \mathbf{h}_i = (h_{i1}, \dots, h_{ik}, \dots, h_{iK}); \\ \log h_{ik}(t \mid \boldsymbol{\xi}_i) &= \mathbf{Z}'_{ik} \boldsymbol{\gamma}_k; \quad k = 1, \dots, K, \end{aligned} \tag{3.1}$$

where \mathbf{Z}_{ik} and $\boldsymbol{\gamma}_k$ are $(1 + q_1 + 1)$ -dimensional vectors. Vector $\mathbf{Z}_{ik} = (1, \mathbf{X}'_{1i}, \xi_{ik})'$ unites all regressors, and vector $\boldsymbol{\gamma}_k = (\log(h_{0k}), \boldsymbol{\beta}'_1, \alpha_k)'$ includes all the effects of \mathbf{Z}_{ik} . The representation of the hazard function is rewritten as:

$$\log h_{ik}(t \mid \boldsymbol{\xi}_i) = \log(h_{0k}(t)) + \mathbf{X}'_{1i} \boldsymbol{\beta}_1 + \xi_{ik} \alpha_k; \quad k = 1, \dots, K. \tag{3.2}$$

Note that the covariates are included as it is assumed that the hazard function is not the same for every individual, and these differences are explained by $\mathbf{Z}_{ik} \boldsymbol{\gamma}_k$. Additionally, it is possible to include a frailty term to expand on this notion, although this has not been

done in this chapter.

In order to complete the sub-model specification. We consider the dependence of the γ_k over time. The dynamic part is given by

$$\begin{aligned}\gamma_k &= \mathcal{G}(\Delta)\gamma_{k-1} + \omega_k, \quad \omega_k \sim \mathcal{N}(0, \mathbf{\Omega}), \\ \gamma_0 | \mathcal{D}_0 &\sim \mathcal{N}(\mathcal{M}_0, \mathcal{C}_0).\end{aligned}\tag{3.3}$$

$\mathcal{G}(\Delta)$ is the system evolution of the interval I_k and $\Delta = \tau_k - \tau_{k-1}$ defines the constant length of the intervals. One can see that the first equation in (3.1) and the first equation in (3.3) when put together is a non-Gaussian non-linear dynamic model as described by West et al. (1985). Such a survival analysis has been done in Gamerman (1991). In the remainder of this chapter, only the baseline hazard function and the effects α_k are modeled as being time-variant. For that purpose the variances relative to β_1 in $\mathbf{\Omega}$ must be set to 0 and its respective blocks of $\mathcal{G}(\Delta)$ to an identity block. We allow α_k to evolve over time, since we expect that the filtered information of the longitudinal measurements should correspond to its respective interval. In other words, we expect that the k^{th} longitudinal measurement is only influential at the k^{th} interval.

Alternatively, $h(t)$ can be written as $h_{ik}(t | \xi_i) = h_{0k}(t) \exp\{\mathbf{X}'_{1i}\beta_1 + \xi_{ik}\alpha_k\}$, so that $h_{0k}(t)$ can be modeled in its original scale. Since this function is constant by parts, it is only necessary to describe the joint distribution of the values they assume at every interval. Although the simplest model would consider them to be independent, we can describe more complex structures.

Consider a Markovian structure for $h_{0k}|h_{0(k-1)}$ as a Gamma Process, based on Nieto-Barajas (2002). In short the Gamma process starts with $h_{01} \sim G(a_1, b_1)$ and define $u_k | h_{0k} \sim Po(c_k h_{0k})$ and $h_{0(k+1)} | u_k \sim G(a_{k+1} + u_k, b_{k+1} + c_k)$ where different choices for a_k , b_k and c_k yield different process structures. For further details, read Nieto-Barajas (2002). Note that two dynamic structures for the baseline hazard function will be considered in this chapter, but any dynamic structure is easy to adapt in this proposal.

Since the model is conditional to the latent term $\xi_i, i = 1, \dots, N$, we assume a probabilistic description for this term. We further assume that this sub-model draws

information from longitudinal measurements taken with error (Rizopoulos 2012, Chapter 1), which happen in J different times. Although $\boldsymbol{\xi}_i$ can be chosen to be any component of the longitudinal sub-model, one very direct choice is the J filtered values of those longitudinal measurements, represented by $\boldsymbol{\xi}_i = (\lambda_{i1}, \dots, \lambda_{iJ})$, defined in the next Section. In other words, $\xi_{ik} = \lambda_{ik}$ for the sake of example.

An important detail in this model is the choice of the number of intervals K . Since in this work it is otherwise of no interest, we force that $K \geq J$. If $K = J$ and $\boldsymbol{\xi}_i$ has J components, then we can “plug” each component of the longitudinal sub-model in the hazard function at each interval simply as a covariate. This is done throughout the rest of this chapter. However, if $K > J$, then this is also possible, except that there will be intervals whose respective component of the longitudinal sub-model have had no measurements. This is worked around by a dependence structure, so that in those intervals, the latent covariate borrows information from the ones adjacent to it. This dependence structure is a dynamic model, as will be described below.

3.2.2 Dynamic modeling of the longitudinal markers

The longitudinal component is modeled as a dynamic hierarchical linear model. Essentially, this component models a covariate that is assumed to be measured over time with some error. To allow for flexibility, a generalized dynamic hierarchical linear model is assumed, as described by Gamerman and Migon (1993). In mathematical terms, suppose that the density function of the outcomes $y_{i1}, y_{i2}, \dots, y_{iJ}$ for the i^{th} individual, at times $s_{i1}, s_{i2}, \dots, s_{iJ}$ are exchangeable and come from the exponential family of distributions *Ef*, i.e., $[y_{ij} \mid \eta_{ij}, \psi] = b(y_{ij}, \psi) \exp\{\psi(y_{ij}\eta_{ij} - a(\eta_{ij}))\}$, with mean $E[y_{ij} \mid \eta_{ij}, \psi] = g^{-1}(\lambda_{ij})$, which is assumed to be some twice differentiable function of λ_{ij} , where η_{ij} and ψ denote the natural and dispersion parameters in the exponential family, respectively. We can assume that y_{ij} is the covariate measured with error and $g^{-1}(\lambda_{ij})$ is its respective true unobserved value. The terms $a(\cdot)$ and $b(\cdot)$ are known functions specifying the member of the exponential family. Thus the hierarchical generalized dynamic linear model for the

longitudinal component is given by

$$\begin{aligned}
y_{ij} &\sim Ef[\eta_{ij}, \phi], \quad i, \dots, N; \\
g(E[y_{ij}|\eta_{ij}, \phi]) &= \lambda_{ij}, \quad j, \dots, J; \\
\lambda_{ij} &= \mathbf{F}'_i \boldsymbol{\theta}_j + v_i, v_i \sim \mathcal{N}(0, V); \\
\boldsymbol{\theta}_j &= \mathbf{G}(\delta_j) \boldsymbol{\theta}_{j-1} + \mathbf{w}_j, \mathbf{w}_j \sim \mathcal{N}(0, \mathbf{W}); \\
\boldsymbol{\theta}_0 | \mathcal{D}_0 &\sim \mathcal{N}(\mathbf{m}_0, \mathbf{C}_0).
\end{aligned} \tag{3.4}$$

The latent term λ_{ij} is a time varying component for each individual at interview j , modeled via the dynamic linear model, where the regression components associated to the effects $\boldsymbol{\theta}'_j = (\boldsymbol{\mu}'_j, \boldsymbol{\beta}'_{2j})$ is a $(p + q_2)$ dimensional vector and consists of the effects of the elements of \mathbf{F}_i for all individuals; the term $\boldsymbol{\mu}_j$, is a p -dimensional vector describing the polynomial effect and the term $\boldsymbol{\beta}_{2j}$, is a q_2 -dimensional vector describing the covariates effects. These terms are allowed to evolve smoothly over time according to the evolution equation, characterized by the matrix $\mathbf{G}(\delta_j)$, which is also denoted by \mathbf{G}_j for simplicity. \mathbf{G}_j and \mathbf{w}_j are explained in better detail later in this Section.

The quadruple $(\mathbf{F}_i, \mathbf{G}_j, V, \mathbf{W})$ identifies the dynamic model, as described in West and Harrison (1997). Here, v_i represents the random effects whose variance is V . \mathbf{F}_i and \mathbf{G}_j are chosen to represent a second order polynomial and the covariates from the model so that $\boldsymbol{\theta}_j$ includes their respective effects.

The vector $\mathbf{F}'_i = (\mathbf{e}'_p, \mathbf{X}'_{2i})$ is composed by the block $\mathbf{e}'_p = (1, 0, \dots, 0)$, a p -dimensional vector and the block \mathbf{X}_{2i} , a q_2 -dimensional vector. In summary, for the i^{th} individual, \mathbf{F}_i is the i^{th} row of the design matrix with $(p + q_2)$ terms and contains components which are multiplied by the hierarchical, temporal and fixed effects.

The transition matrix \mathbf{G}_j whose dimension is $(p + q_2) \times (p + q_2)$ is obtained based on $\delta_j = s_j - s_{j-1}$, which describes the time span between two consecutive appointment schedules, assumed to be the same for all individuals. At $j = 1$, δ_1 is set to s_1 as $s_0 = 0$. Specifically $\mathbf{G}_j = \text{blockdiag}\{\text{diag}(\mathbf{L}_p(\delta_j), \mathbf{I}_{q_2 \times q_2})\}$, where $\mathbf{L}_p(\delta_j)$ relates to the superposition of dynamic linear models, except that the time increments is δ_j instead of

1. In the special case of $p = 2$, which is the one used throughout this chapter,

$$\mathbf{L}_2(\delta_j) = \begin{pmatrix} 1 & \delta_j \\ 0 & 1 \end{pmatrix}.$$

The terms \mathbf{L}_p and \mathbf{e}_p ensure that a polynomial linear growth model is recovered. The term $\mathbf{I}_{q_2 \times q_2}$ of \mathbf{G}_j indicates that the covariate coefficients follow a random walk evolution, but it can be fixed in time too by setting the respective block in \mathbf{W} as a matrix of zeros.

The last term \mathbf{W} defines the $(p + q_2)$ -dimensional covariance matrix whose elements in the diagonal are associated with the variance related to the polynomial and covariates effects of $\boldsymbol{\theta}$.

At time $j = 0$, define the initial prior information $\boldsymbol{\theta}_0 \mid \mathcal{D}_0 \sim \mathcal{N}(\mathbf{m}_0, \mathbf{C}_0)$, for known \mathbf{m}_0 and \mathbf{C}_0 . The initial information is summarized in \mathcal{D}_0 containing all hyper parameters associated with the prior distributions.

Note that the joint model is fully defined by equations (3.1) and (3.4) given some baseline hazard function. Also, note that the joint model is obtained by $[T_i, D_i \mid \boldsymbol{\xi}_i][\mathbf{Y}_i \mid \boldsymbol{\Lambda}]$, while observing that T_i , D_i and \mathbf{Y}_i are independent given the choice of $\boldsymbol{\xi}_i$. The bridge term $\boldsymbol{\xi}_i$ can be any term from the longitudinal sub-model and its choice may be relevant to the analysis, but it must make sense from a physical viewpoint. As an illustration, the choice for this chapter is to use $\boldsymbol{\xi}_i$ in (3.2) as $\boldsymbol{\lambda}_{i\cdot}$, that is, the i^{th} row of the $N \times J$ -dimensional matrix $\boldsymbol{\Lambda}$, composed of the λ_{ij} terms. Then the vector $\boldsymbol{\lambda}_{i\cdot}$ represents the average trajectory g -transform (i.e. the linear predictor λ_{ij}) per individual.

Note that this implies that all individuals contribute with all J times to the hazard function, which may not be compatible with the reality of the study, since some individuals have their event before the end of the study. This means that the interview times after their event times have no observed value. By treating these interviews as missing data, Bayesian inference can easily incorporate their uncertainty into the hazard function, and their predictive distribution does not have a strong influence on the final result. In the next Section, the inference procedure will be described.

3.3 Inference

Since the analysis is done under the Bayesian paradigm, the inference is done over the posterior distribution of the unknown quantities in the model. The model construction described above provide the groundwork of the likelihood function $L_{\mathbf{t}, \mathbf{d}, \mathbf{y}}(\cdot)$ of the parameters, which will be described in detail in Section 3.3.2. The posterior distribution is obtained by

$$\pi(\cdot | \mathbf{t}, \mathbf{d}, \mathbf{y}) \propto L_{\mathbf{t}, \mathbf{d}, \mathbf{y}}(\cdot) \pi(\cdot),$$

where $\pi(\cdot)$ represents the prior distribution of the parameters. To complete the Bayesian inference specification, prior distributions must be set for all unknown quantities in the model.

3.3.1 Prior

To specify the prior distribution, consider the parameter vector $\Theta = (\gamma_k, \alpha, \psi, \Lambda, \theta, \beta_2, V, W_1, W_2)$, where $\theta = (\theta'_1, \dots, \theta'_J)$. We assume independent and non informative marginal prior distributions for the longitudinal effects $\gamma_k \sim \mathcal{N}(\mathbf{0}; 1000\mathbf{I}_J)$, $k = 2, \dots, q_1 + 1$. For all dispersion parameters, $\psi, V, W_1, W_2 \sim IG(0.01, 0.01)$. The prior, then, is $\pi(\Theta) = \pi(\gamma_k)\pi(\psi)\pi(V)\pi(W_1)\pi(W_2)$.

3.3.2 Likelihood

The likelihood function is obtained through the product $[t_i, d_i | \xi_i][\mathbf{y}_i | \xi_i]$, when viewed as a function of the parameters. The first factor comes from the survival model and depends on the individual survival functions $S_i(t)$ and the individual hazard functions $h_i(t)$. But

$$\begin{aligned} S_{ik}(t | \xi_i, \cdot) &= \exp \left\{ - \int_0^t h_{ik}(s) ds \right\}, k = 1, \dots, K; \\ h_{ik}(t) &= h_{0k}(t) \exp \{ \mathbf{Z}'_i \gamma_k + \xi_{ik} \alpha_k \}, \quad \forall t \in I_k, \end{aligned} \quad (3.5)$$

where $\log h_{ik}(t) = \log h_{0k}(t) + \mathbf{X}'_{1i} \beta_1 + \xi_{ik} \alpha_k$. The likelihood contribution associated to the survival component for the i^{th} individual is given by: $[t_i, d_i | \xi_i] = \prod_{k=1}^K \{ h_{ik}(t)^{d_{ik}} S_{ik}(t) \}$, where the term $d_{ik} = d_i I(\tau_{k-1} < t \leq \tau_k)$ is the right censored indicator in interval I_k for

individual i .

The likelihood contribution from longitudinal measurements is:

$[\mathbf{y}_i | \boldsymbol{\xi}_i] = \prod_{j=1}^J b(y_{ij}, \psi) \exp\{\psi(y_{ij}\eta_{ij} - a(\eta_{ij}))\}$, where η_{ij} and ψ denote the natural and dispersion parameters in the exponential family, respectively. The terms $a(\cdot)$, $b(\cdot)$ are known functions specifying the member of the exponential family.

Finally the likelihood function is evaluated by $L_{\mathbf{t}, \mathbf{d}, \mathbf{y}}(\boldsymbol{\Theta}) = \prod_i [t_i, d_i | \boldsymbol{\xi}_i][\mathbf{y}_i | \boldsymbol{\xi}_i]$. Details about its derivation can be found in Section A.3.1. We find the posterior distribution from $\pi(\boldsymbol{\Theta} | \mathbf{t}, \mathbf{d}, \mathbf{y}) \propto L_{\mathbf{t}, \mathbf{d}, \mathbf{y}}(\boldsymbol{\Theta})\pi(\boldsymbol{\Theta})$.

The resulting posterior distribution is not known in closed form. Therefore, MCMC methods are employed to obtain samples from the distribution. An important advantage of the proposed model and the MCMC inference method is the automatic solution to two integrals which, traditionally, bring trouble to the analysis. First, the hazard function is easy to integrate, since it is constant in intervals, which means that it is analytically integrable. The second one is related to the bridge term. In other works, the bridge effect is a random effect and must be integrated out for the inference procedure, which is usually the main computation bottleneck. In our model, even if a random effect is used for the bridge term, the MCMC method provides a simple way to integrate it out.

3.4 Simulation study

In this Section, our effort is to present a simulation study to illustrate the performance of the proposed model using a sample generated from a relative risk parametric survival model and dynamic non-Gaussian sub-model for the longitudinal measures. The chosen baseline hazard function was the Gompertz function, that is, $h_0(t) = e^{a_g + b_g t}$, where $a_g = 0$ and $b_g = 1$ are the scale and shape parameters of the Gompertz distribution, respectively.

Beyond varying the baseline hazard function, we varied the total of longitudinal measurements J and the sample size N . The different values for J were 3, 5 and 7, and for two different sample sizes: 50 and 100. Note that, in every case, $K = J$. Model fit

was done with two proposals of baseline hazard functions: the log Normal process and the Gamma process. Each scenario was repeated 1000 times. This means that, in total, $2 * 3 * 1000 = 6000$ datasets were simulated, and for each, 2 models were fitted, meaning that in total, $2 * 2 * 3 * 1000 = 12000$ results were analyzed. The simulation procedure to obtain $\{\mathbf{y}, \mathbf{t}, \mathbf{d}\}$ is detailed below.

Hierarchical dynamic joint using $\xi_{ij} = \lambda_{ij}$ as a bridge function.

Input: N, J , mean 2 for the right censoring, $\mathbf{F}, \mathbf{G}, V, \mathbf{W}, a, b, \mathbf{m}_0, \mathbf{C}_0$

- Initial information: $(\boldsymbol{\theta}_0) \leftarrow \mathcal{N}(\mathbf{m}_0, \mathbf{C}_0), \quad j = 1;$
- System equation: $(\boldsymbol{\theta}_j) \leftarrow \mathcal{N}(\mathbf{G}_j \boldsymbol{\theta}_{j-1}, \mathbf{W})$ to $j = 2, \dots, J;$
- Structure equation: $(\lambda_{ij}) \leftarrow \mathcal{N}(\mathbf{F}'_i \boldsymbol{\theta}_j, V);$
- Observation equation: $y_{ij} \leftarrow Po(\exp(\lambda_{ij}));$
- The true times t_i^* using the inverse probability method
 $(t_i^*) \leftarrow S(t) = \int_0^{t_i^*} h_0(u \mid a, b) \exp(\sum_{j=1}^J \alpha_j \xi_{ij}) du;$
- Obtain the censored time (c_i) using exponential distribution;
- Calculate the observed time $(t_i, d_i) (t_i; d_i) \leftarrow \min(t_i^*, c_i); I(t_i^* \leq c_i);$
- Return $\mathbf{y}, \mathbf{t}, \mathbf{d}$

The algorithm to elaborate our inference procedure is developed in R with an Intel ®i5 2300 Windows PC with 4GB memory. The prior distributions considered were described in Section 3.3. The mean computational time was 16.545 minutes per adjusted sample. We ran the chain for 500,000 iterations and discarded the first 250,000 iterations as a burn-in period. Thinning of 50 was used to eliminate autocorrelation, providing a final sample of 5000 iterations from which to draw samples from the posterior distributions. Details about the data generation procedure and full conditional distributions can be seen in Section A.3.1 of the thesis appendix

For results verification, note that the purpose of this simulation study is to verify

the model's ability to reconstruct the true baseline hazard function. Since it is not straightforward how to measure the quality of estimation for each individual, we limit the comparison to the model parameters and baseline hazard function. Furthermore, since a semi-parametric hazard function is an output of the model, it is desirable to understand the model's sensitivity to the discretization of the time axis through τ .

Estimation by MCMC of the baseline hazard functions can be seen in Figure 3.1. The black solid line represents the function that generated the data (i.e. e^t), the lighter boxplots represent the log Normal process and the darker boxplots represent for the Gamma process. The boxplots are built with the average posterior mean resulting from all 1000 simulations.

Although the log Normal process performs poorly when $K = 3$, it captures the hazard function well when $K = 5$, especially when $n = 100$. For $K = 7$ the estimation seems to accommodate the true baseline function with exponential growth by increasing the posterior variance. Also note that the narrower the last interval, the higher the posterior variance of h_{0K} . This is due to the fact that we are approximating a function with very fast growth by a constant function in that interval.

For the Gamma process, however, all graphics show that it can come very close to the true function, although it only outperforms the log Normal process when $K = 3$. Therefore, an argument can be made in benefit of both propositions. For $K = 7$ the posterior variance in the last interval does not cover most of the function. This is due to the fact that in the Gamma distribution mean and variance are linearly connected. Also the Markovian dependence seems to be too strong to allow the last interval's estimation to be higher.

The results of the proposed model are shown through the mean square error (MSE) in Figure 3.3 and absolute bias in Figure 3.2. All the biases are small both for the Gamma process and the log Normal process. For all parameters it is no higher than 1.2% and there is a clear advantage for the datasets with $n = 100$. For the α_6 and α_7 parameters there is higher bias since most of the individuals have failed by the last intervals so there is few data for estimation. This is more apparent for the Gamma process. The parameters with highest bias are W_1 and W_2 . However, if we look at the true values of

the variances of the longitudinal process, we see that the signal-noise ratio is small (0.2), therefore the information for W_1 and W_2 is not abundant.

Similar results are obtained when we analyze the MSE. In summary all parameters show that the model was able to estimate the values very well because the bias and the MSE are small.

Posterior statistics based on 1000 datasets are shown in Table 3.1. This presents the mean posterior estimates for all scenarios. The global results show that these true parameter values are mostly recovered by fitting our model. As in chapter 2, the relative variation shows bad results, which may call for more simulated studies.

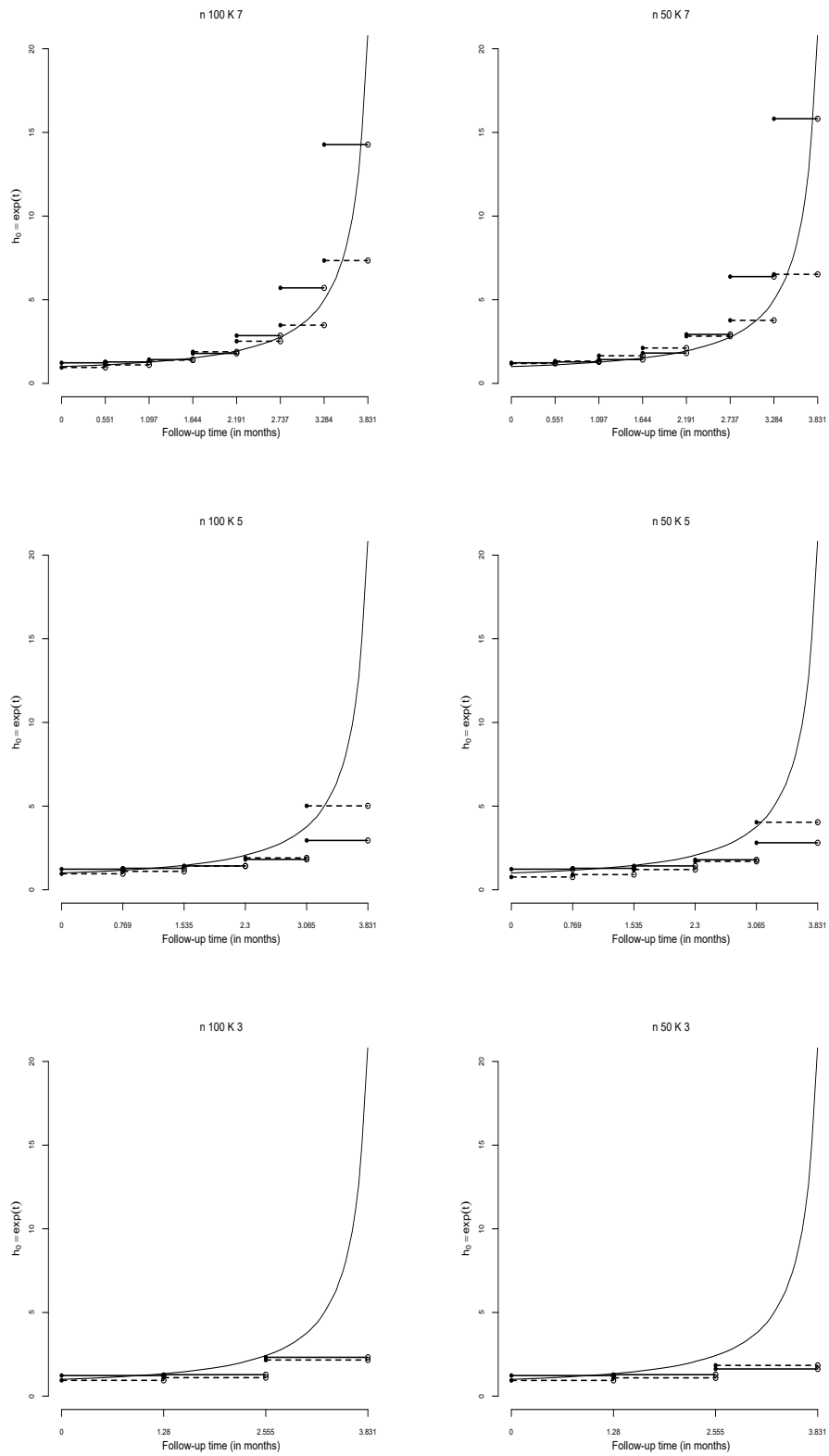
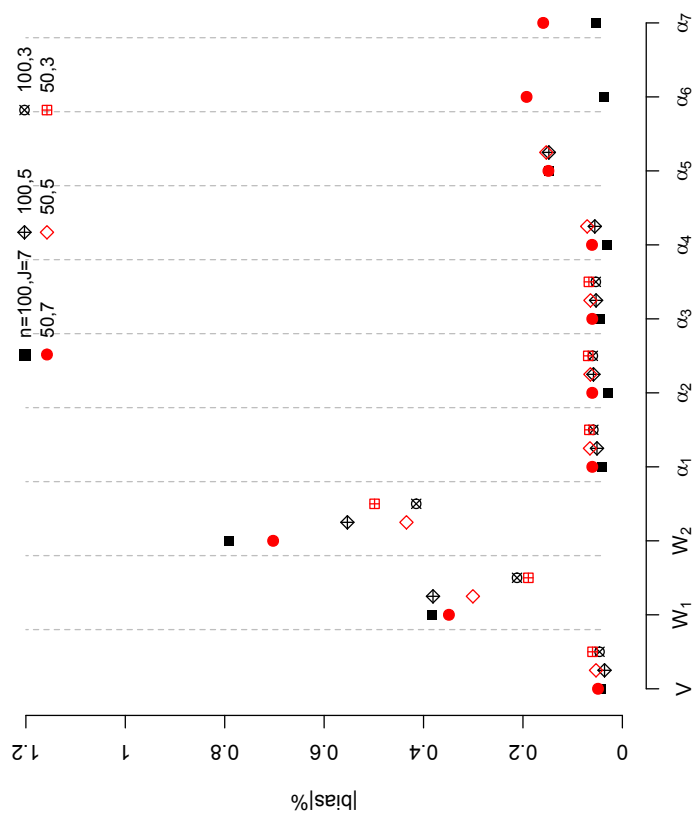


Figure 3.1: Comparison of Gamma Process vs Log Normal : Baseline hazard functions using Simulation with Gompertz baseline hazard $\exp\{t\}$

Figure 3.2: Results based on 1000 simulated data sets. Bias is computed with the mean as the measure of location and multiply by 100 with sizes 50 and 100, $K=3,5,7$. The fully dynamic joint model Gamma Process (GP) and Log Normal Process (NP) were compared.

GP



NP

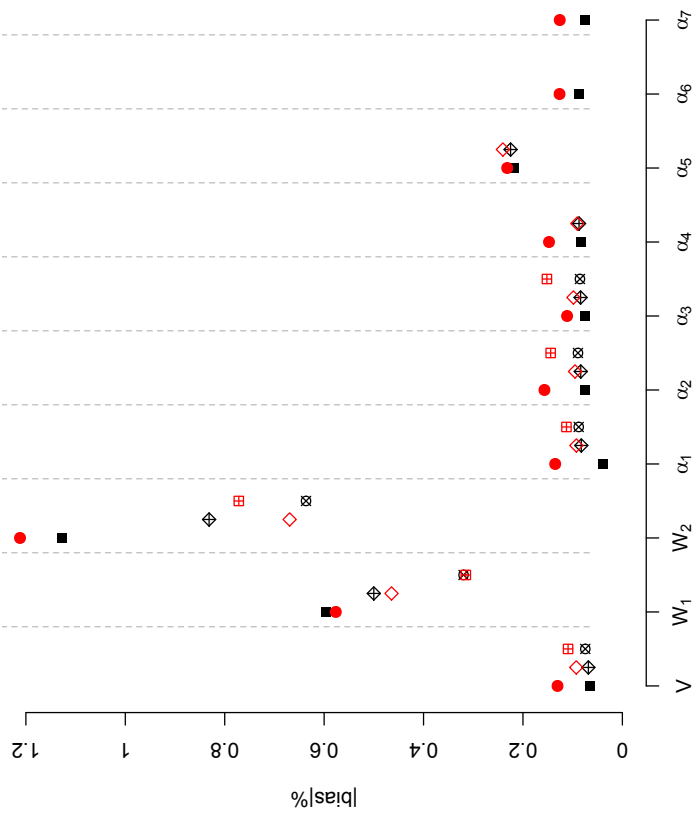


Figure 3.3: Results based on 1000 simulated data sets. The mean square error (MSE) is computed with the mean as the measure of location with sizes 50 and 100, $K=3,5,7$. The fully dynamic joint model Gamma Process (GP) and Log Normal Process (NP) were compared.

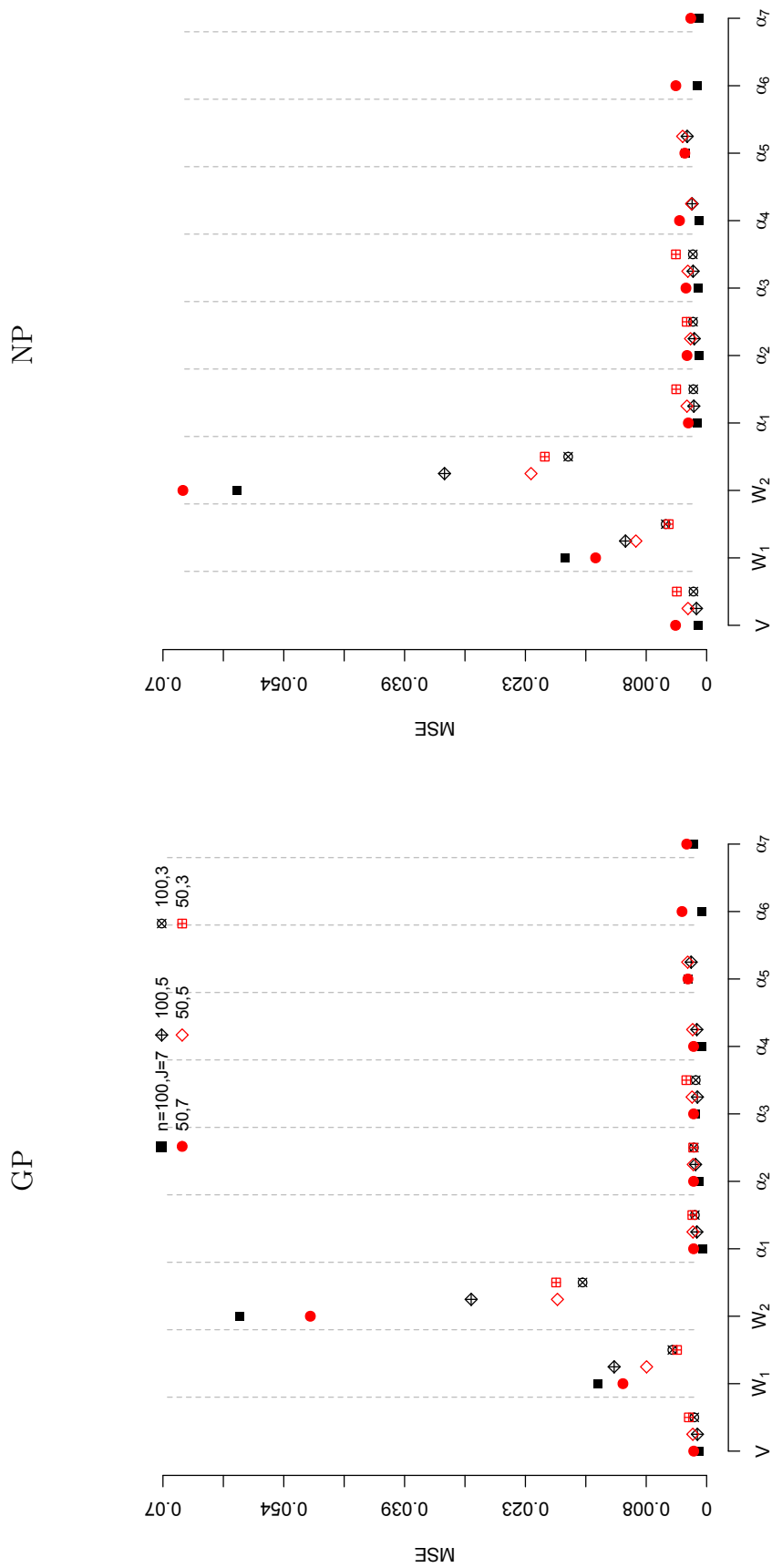


Table 3.1: Comparison of posterior mean. They are computed with the mean based on 1000 posterior means calculated for each simulated dataset for all scenarios. The fully dynamic joint model Gamma Process (GP) and Log Normal Process (NP) were compared.

P	GP						NP						
	K=7		5		3		7		5		3		
	True $N = 100$	50	100	50	100	50	100	50	100	50	100	50	
V	0.10	0.109	0.098	0.098	0.089	0.102	0.088	0.110	0.109	0.097	0.093	0.102	0.090
W_1	0.02	0.018	0.024	0.021	0.012	0.019	0.037	0.040	0.041	0.045	0.008	0.040	0.050
W_2	0.01	0.011	0.006	0.057	0.064	0.063	0.054	0.021	0.028	0.013	0.049	0.011	0.023
α_1	0.10	0.099	0.089	0.094	0.087	0.090	0.091	0.095	0.065	0.089	0.059	0.088	0.077
α_2	-0.20	-0.199	-0.179	-0.191	-0.180	-0.172	-0.179	-0.190	-0.185	-0.189	-0.180	-0.187	-0.179
α_3	0.20	0.197	0.199	0.201	0.187	0.195	0.191	0.194	0.197	0.202	0.193	0.201	0.190
α_4	-0.10	-0.106	-0.089	-0.099	-0.097	-	-	-0.099	-0.089	-0.098	-0.087	-	-
α_5	0.30	0.299	0.293	0.304	0.088	-	-	0.296	0.102	0.303	0.311	-	-
α_6	0.20	0.210	-0.199	-	-	-	-	0.204	0.156	-	-	-	-
α_7	-0.10	-0.110	0.103	-	-	-	-	-0.097	-0.065	-	-	-	-

To further show the advantage of modeling the baseline hazard function with a Markovian structure on top of the longitudinal measurements, we compare our proposition with other models in the literature by fitting them with benchmark real datasets.

Next, we present results obtained from fitting the proposed model on real data. The dataset is taken from Andersen et al. (1993). It concerns the measurement of liver function in cirrhosis patients treated either with standard or novel therapy.

The dataset is included in two packages, namely JMbays in Rizopoulos (2016) and joineR in Philipson et al. (2017). It can be accessed with the command `data(liver)`.

3.5 Liver cirrhosis longitudinal data

Here we illustrate a fully Bayesian joint model proposed using Gamma process for the baseline hazard function and Gaussian response for the longitudinal sub-model. This dataset gives the longitudinal observations of Prothrombin index, a measure of

liver function, for 488 patients from a controlled trial into Prednisone treatment of liver cirrhosis. The time-to-event observations consist of time of death and associated censoring indicators are also recorded along with a single baseline covariate - the allocated treatment arm in this instance. The data are originally analyzed in Andersen et al. (1993)(p. 19) and were also analyzed in Henderson and Diggle (2002). The variable Prothrombin index is measured in % and their follow-up times are measured in years with a total of $J = 17$ interviews. The patient treatment indicator is coded as 0 = placebo; 1 = Prednisone. The patient event time to death is measured in years and the censoring indicator is coded as 1 = died (59.8% of the cases) and 0 = censored.

We did some preliminary tests to define the minimum MCMC configuration to achieve convergence. In this case, we settled with 100000 iterations after a burn-in period of 50000 iterations. The effective iterations were thinned by storing every 10th iteration in order to decrease autocorrelation in the sample. The main results are shown below. Samples of the posterior were drawn using the software Jags (Su and Yajima 2015).

In Figure 3.5, the level, that is the average trend of the longitudinal measurement, can be seen, as well as the 3 quartiles of the observations at each follow-up time. It can be seen that the trajectory is not always linear, as is usually assumed, both when looking at the unfiltered data and the modeled filtered trajectory. The proposed trajectory, on the other hand, seems to follow the distribution of the observed measurements correctly.

Figure 3.4, the baseline hazard functions are compared. The independent Gamma function from Rizopoulos and Ghosh (2011) and the Gompertz function from Henderson et al. (2000) are compared with the Gamma process as was proposed. The independent Gamma function seems to overestimate the hazard function done with both other methods in the early follow-up times, while the Gompertz function seems underestimate it in the late follow-up times. Although different, the independent Gamma function is a special case of what was proposed, so an argument can be made that the more general form is more appropriate.

Table 3.2 compares the proposed model using the Gamma process represented by Fd

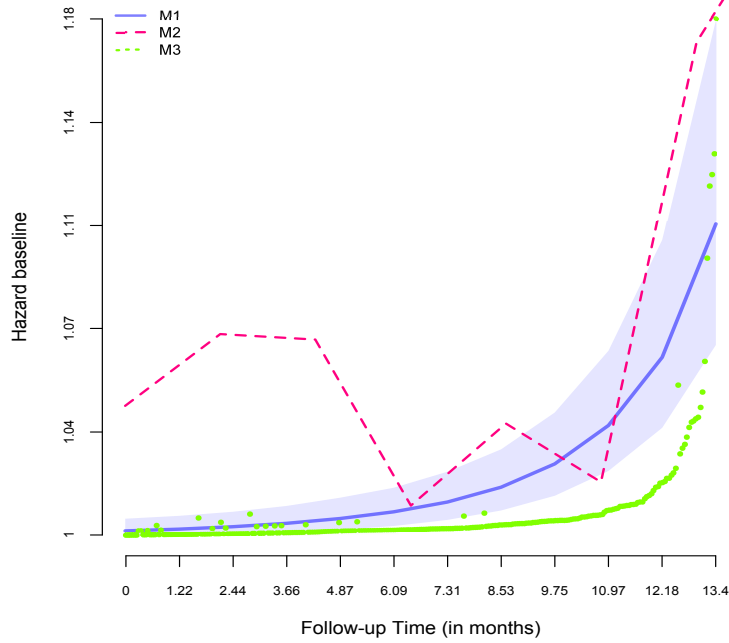


Figure 3.4: *Baseline hazard function by interval k : estimated using the package `joiner` based on Henderson et al. (2000) SRE approach (M3), Rizopoulos and Ghosh (2011) (dashed and dotted, respectively, M2) and our proposal (hard line with credible intervals, M1)*

Table 3.2: Comparison of Bayesian model selection methods for the fully dynamic joint model considering Gamma Process (GP) compared to the standard joint model (SRE) using the package in Rizopoulos (2016). (Metrics divided by 100).

	DIC	pD	$LPML$	$\Delta LPML_{Surv}$	ΔDIC_{Surv}	ΔpD_{Surv}
Fd (GP)	346.333	330.791	-166.552	-24.304	44.674	44.614
SRE	346.981	9.756	-175.249	-	-	-

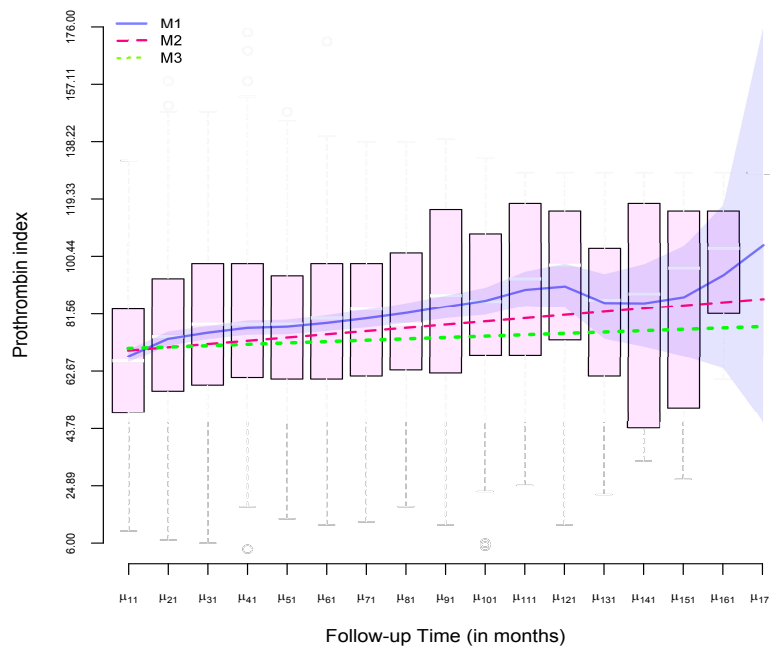


Figure 3.5: *Longitudinal data: for each j interview time, the level μ_j with credibility interval (blue hard line, M1) and estimated linear trajectories by Rizopoulos and Ghosh (2011) (red dotted, M2) and Henderson et al. (2000) (green dash-dotted, M3)*

(GP) with a benchmark model represented by SRE. It can be seen that the full dynamic model is preferred, both in a fitting sense and a forecasting sense. Note that the column relative to the pD metric prefers the benchmark model. This column penalizes the inclusion of parameters, so this means that the inclusion of parameters in the proposed model is compensated by the improvement in the fit.

The three last columns of the Table are contributions to the metrics from the survival sub-model. They cannot be compared since the package in Rizopoulos (2016) does not provide them as an output. However we note that they are much smaller than the full metric, which leads us to believe that the longitudinal sub-model is the main contributor to the metrics.

3.6 Discussion

We relaxed parametric distributional assumptions for the survival process and modeling the relative hazards using a semi-parameter model for the joint modeling of longitudinal measurements and time-to-event data where both sub-models evolve with a Markovian structure, and demonstrated its use through a real data example and a simulation study. The main strength of this framework is that it effectively handles the dependence structure between hazard increments. In addition, it was shown in the simulation study that the two possibilities for h_0 are good options.

The type of data to be applied on our model already has a solid literature background. However, for some reason, the longitudinal process has not been modeled dynamically. This is surprising as it is a very natural framework for time evolving information. Moreover, it seems more flexible to use a semi parametric baseline hazard function such as the piecewise exponential. As much as this hazard function is already used for this type of data in Rizopoulos (2012), the Markovian dependence is new. A very important aspect to note is that all parts of the proposed model are not new by themselves but they had not been put together as in this text.

An interesting exercise could be done about the choice of ξ_i . While we chose the mean trajectory g -transform (i.e. the linear predictor λ_{ij}), many other choices are possible

and potentially relevant. The global trend μ_1 is also an interesting choice as well as the random effect v_i . Although we believe that much could be learned from looking into this, this was not the main objective of this chapter. However, the flexibility is very promising.

Chapter 4

Quantile joint model for the longitudinal marker and event time dataset

In this chapter, the problem of using the information of a longitudinal measurement in a survival framework is considered linking the sub-models which is known as joint model. This type of model has been gaining attention in the statistical literature, but little has been published modeling the longitudinal measurement with a dynamic evolution, despite it being a natural structure for this time varying data. More than this, some data are not adequately modeled by its mean trajectory, as the longitudinal measurement's distribution might be skewed or heavily tailed. Quantile regression (QR) is used as a way to reduce vulnerability to these issues, while maintaining the dynamic structure. This particular approach is new in the area as far as we know. We show one advantage of this procedure by developing the prediction of survival probabilities, which can be achieved due to the natural forecasting feature of dynamic models. We exemplify the model with two examples using real data. We also show an additional advantage of the model by exploring two different links between the sub-models and discuss the results.

4.1 Introduction

In the last fifteen years there has been focus on clinical trial research with longitudinal data where the outcomes are measured at multiple times for each individual until an event of interest occurs. Many applications in survival analysis deals with time dependent responses measured with error in a framework usually called time-to-event joint models (JM). This explains the large amount of contributions to the longitudinal and time-to-event joint models developed in the recent literature. A thorough review of

the area is provided by Ibrahim et al. (2010).

Within the traditional JM framework, a linear random model for the longitudinal expected values, which we call moment based model, is used to describe the longitudinal outcome. However, in many clinical studies it is of interest to make inference or prediction on median, lower, or higher quantiles of the longitudinal outcome's distribution. In other words we are interested in the conditional quantile, rather than the conditional mean. In contrast to modeling the regression mean, original QR models offer a flexible framework that relaxes some distributional assumptions, and provides a way to study covariate effects on various quantiles of the outcome. Since its introduction by Koenker and Bassett (1978) QR modeling can be understood as a distribution free approach, no matter what the original distribution of the data is, while Kotz et al. (2012) was able to represent the density of a quantile regression by assuming that the response follows an asymmetric Laplace Distribution (ALD). This representation is central for the Bayesian approach introduced by Yu and Moyeed (2001), where the inference is based on a sampling algorithm exploring a location-scale mixture representation of the ALD.

The flexibility of the quantile regression method in Yu et al. (2003) has allowed this new thinking to be applicable in many situations, consolidating it as a convenient tool with very attractive properties, including less vulnerability to heavy tails and skewness. These advantages are not ignored and are brought to the JM context. Thus we set aside the modeling of the conditional mean of the response distribution and bring forward a tool for researchers interested in predicting distribution quantiles, particularly when the interest is in the distribution's tails. Nevertheless, quantile regression has been extensively studied for cross sectional data while less developed for longitudinal data. Some researchers considered marginal quantile regression models for analyzing longitudinal data; a simple and complete review is given by Leng and Zhang (2014). Quantile regression has already been proposed with a fully Bayesian approach in JMs by Wen and Menggang (2014). Farcomeni and Viviani (2015) incorporated a linear quantile mixed model into a JM for longitudinal and survival processes, for which they used a Monte Carlo Expectation Maximization algorithm for parameter estimation.

We propose a dynamic quantile regression method in state space form to accommodate the temporal dependence of the longitudinal data. Furthermore, hierarchical characteristics have been included in the longitudinal data analysis. We are specifically interested in estimating the risk of the event for those individuals who have not yet experienced the event. To this end, the dynamic framework offers a natural way of making such personalized predictions of future probabilities. This dynamic framework has already been explored in a moments based model context in Chapter 2. To complete the model specification, we describe the survival sub-model as a well established relative risk model with Weibull baseline hazard function.

Recently a dynamic version quantile of QR was introduced by Gonçalves et al. (2018). This approach takes into account all facilities inherent to the well known class of dynamic linear model (DLM), including forecasting. Two alternative forms of inference were implemented in their work. The first uses MCMC and the other applies linear Bayes. We extend the Gonçalves et al. (2018) work by including a hierarchical component. In order to predict event probabilities for event-free patients, we explore a key feature of these dynamic predictions which is that the predictions can be dynamically updated as soon as an additional longitudinal measurement from the target individual become available, providing instantaneous risk assessment.

Our approach differs from standard models in a few ways. We consider a full Bayesian dynamic hierarchical QR joint model for statistical inference inspired in Yang et al. (2018) and extending Gonçalves et al. (2018). Also, by taking advantage of the posterior distribution of model parameters and subject-specific latent effects, we develop a dynamic prediction procedure for future event probability based on the proposed QR and dynamically update that probability. Lastly, we provide a discussion on the choice of component(s) that links the two sub-models, survival and longitudinal, which we call the bridge component. These characteristics make our proposed method a flexible alternative.

After this introductory Section the rest of this chapter is structured as follows. Section 4.2 reviews the standard quantile regression. Section 4.3 contains our proposal for longitudinal and time-to-event data. In Section 4.3.1 and Section 4.3.2, we give details about the

inference and the predictions for our model. Sections 4.4.1 and 4.4.2 provide some illustration of our proposal with real data. Some conclusions and future work are presented in Section 4.5.

4.1.1 Preliminaries

Suppose that we have a set of N individuals followed-up and measured with respect to one marker over a time interval $(0, \tau]$, where τ is the maximum dropout time. Those measurements may be collected at different time points for each individual and provide information for the measurement until the occurrence of the event of interest.

Specifically, the observed time of the event of interest, called time-to-event, for the i^{th} individual is defined by $T_i = \min(T_i^*, C_i)$, where C_i denotes the censoring time and T_i^* denotes the true event time.

Let d_i represent the indicator of right censoring for the i^{th} individual, that is $d_i = 1$ if individual i experiences the event at time $T_i^* \leq C_i$ and $d_i = 0$ if it is censored.

Formally, the sequence of scheduled dates, s_{ij} , must satisfy the relationship $0 \leq s_{i1} < s_{i2} < \dots < s_{iJ_i} \leq t_i$, where t_i is the observed value of T_i and J_i is the number of follow-up measurements for the i^{th} individual, $j = 1, \dots, J_i$. The situation where all J_i are potentially different for each individual will be called the non-regular case. In contrast, the regular case is defined as the case where $J_i = J, \forall i$, although the times s_{ij} need not be the same.

To ease the description of the inference algorithm we denote a marginal distribution function by $[\cdot]$ and a conditional one by $[\cdot | \cdot]$.

4.2 Quantile regression

In this work a dynamic hierarchical QR model is developed to analyze longitudinal markers and a parametric model for the time-to-event data as relative risk sub-model. This demonstrates the flexibility which can be achieved through the component ξ_i , which

bridges the longitudinal and survival sub-models. This component is referred to as bridge.

The usual assumption in the standard joint models is that the time-to-event is conditional to the mean of the longitudinal marker response (the mean trajectory). In this thesis, we deal with cases where another feature of the longitudinal measurements' distribution is more appropriate to the study, namely a particular quantile. The interest in this type of modeling varies per application, so some examples will be given in the appropriate Sections. Additionally, many examples and justifications can also be seen in Yu et al. (2003). In this context, it is necessary to explore the use of a new way to make predictions based on quantiles of the longitudinal process. To this end, quantile regression provides a flexible, distribution-free way to study covariate effects at different quantiles of the longitudinal outcomes and it is robust not only with respect to deviation from normality, but also to outlying observations (Yang et al. 2018). We would like to note that the goal of this work is to present a methodology to model the data using a quantile which is most adequate according to the data analyst or by the nature of the application. It is not our interest, in this chapter, to determine the most adequate quantile for a given application nor provide a method to determine it.

A comment on the interpretation of the bridge component is in order. The quantile of the distribution of the longitudinal measurements y_{ij} can be used as the bridge value. An usual interpretation is to say that the chosen quantile is the true underlying longitudinal marker of the measurement y_{ij} (Yang et al. 2018). This particular interpretation makes a further assumption on the sampling process, which we do not verify, therefore we do not use in this work. If, for example, the chosen quantile is 0.9, then saying that the quantile is the true underlying longitudinal marker means that the observed values y_{ij} are typically *much* lower than the true values. This is true if y_{ij} consistently under-represents the true values. In order to state that, this repeated pattern must be verified, or at least assumed, which can only be done with an appropriate expert's opinion, which is not done in the sample applications of Sections 4.4.1 and 4.4.2.

For the joint QR modeling, we focus on the longitudinal sub-model as an extension of the Bayesian method of Gonçalves et al. (2018) based on the location-scale representation of

the asymmetric Laplace distribution to a dynamic hierarchical context. Our development is motivated by two real datasets. The first is a mental study of schizophrenic disease where the primary clinical interest is to utilize longitudinal scores and other covariates to predict the risk of developing schizophrenic disease, see Henderson and Diggle (2002) and Henderson et al. (2000). The second is related to the liver cirrhosis longitudinal data previously analyzed by Andersen et al. (1993).

This chapter proceeds as follows. In 4.3, we give details of the dynamic hierarchical QR and statistical methods used for inference and dynamic predictions. In 4.4.1 and 4.4.2, we present two real case studies to motivate the proposed method where we show how our proposal provides good performance to model this kind of data. We end this chapter with a conclusion and discussion about future work in Section 4.5.

4.3 Bayesian linear hierarchical quantile joint model

The standard linear quantile models use random effects and assume that the conditional quantile of the outcome is a linear function of covariates. The ν^{th} quantile of a random variable Y is defined as $\mathcal{Q}_Y(\nu) = F_Y^{-1}(\nu) = \inf\{y : F_Y(y) \geq \nu\}$ for $\nu \in [0, 1]$. The term $F_Y(y)$ represents the cumulative distribution function of the random variable Y . The model proposed in this thesis does not use random effects for covariates.

The Bayesian QR was introduced by Yu and Moyeed (2001) using the asymmetric Laplace distribution (ALD). The inference is based on a sampling algorithm exploring a location-scale mixture representation of the ALD (Kotz et al. 2012). Under this parameterization, the random error is represented as $\epsilon_{ij} = \kappa_{1\nu}\epsilon_{ij} + \kappa_{2\nu}\sqrt{\sigma\epsilon_{ij}}\varsigma_{ij}$ with $\varsigma_{ij} \sim \mathcal{N}(0, 1)$, $\epsilon_{ij} \sim \exp(1/\sigma)$, $\kappa_{1\nu} = \frac{1-2\nu}{\nu(1-\nu)}$ and $\kappa_{2\nu}^2 = \frac{2}{\nu(1-\nu)}$. This re-parameterization leads to the longitudinal sub-model of our proposal to be written as

$$\begin{aligned}
 y_{ij} &= \lambda_{ij} + \kappa_{1\nu}\epsilon_{ij} + \kappa_{2\nu}\sqrt{\sigma\epsilon_{ij}}\varsigma_{ij}, \\
 \lambda_{ij} &= \mathbf{F}'_i\boldsymbol{\theta}_j + v_i, v_i \sim \mathcal{N}(0, V) \\
 \boldsymbol{\theta}_j &= \mathbf{G}_j\boldsymbol{\theta}_{j-1} + \mathbf{w}_j, \quad \mathbf{w}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{W}) \\
 \boldsymbol{\theta}_0 &| \mathcal{D}_0 \sim \mathcal{N}(\mathbf{m}_0, \mathbf{C}_0),
 \end{aligned} \tag{4.1}$$

where the ν^{th} quantile λ_{ij} is \mathbf{F}'_i multiplied by the common term $\boldsymbol{\theta}_j$ plus a random effect

per individual denoted by v_i . The term v_i is included since it differentiates the quantiles of the individuals beyond the covariates in \mathbf{F}_i . Also, it allows us to compare our model with others which include a comparable term. The term V represents the variance of v_i assuming the Gaussian distribution. The $(p + q_2)$ -dimensional vector $\mathbf{F}'_i = (\mathbf{e}'_p, \mathbf{X}'_{2i})$ where $\mathbf{e}_p = (1, 0, \dots, 0)$ is a p -dimensional vector and \mathbf{X}_{2i} is a q_2 -dimensional covariates vector.

The common evolution $\boldsymbol{\theta}_j = (\boldsymbol{\mu}'_j, \boldsymbol{\beta}'_{2j})$ is a $(p + q_2)$ -dimensional vector capturing the effects of the elements of \mathbf{F}_i . The term $\boldsymbol{\mu}_j$ is a p -dimensional vector describing the polynomial effect, and the term $\boldsymbol{\beta}_{2j}$ is a q_2 -dimensional vector describing the covariates effects. These terms are allowed to evolve smoothly over time, according to the evolution equation. The transition matrix \mathbf{G}_j is $(p + q_2) \times (p + q_2)$ -dimensional. The construction of this term depends on the times between consecutive visits $\delta_{ij} = s_{ij} - s_{i(j-1)}, j = 1, \dots, J, \forall i$. We work with the situation where all measurements for each individual may occur at equal follow-up times, the regular case, which means $J_i = J, \forall i$. The linear predictor λ_{ij} represents the ν^{th} quantile of the distribution of y_{ij} , and μ_{1j} , the first term of $\boldsymbol{\theta}_j$, represents the mean level of the trajectories of the ν^{th} quantile. The perturbation \mathbf{w}_j is a $(p + q_2)$ -dimensional vector following a zero mean multivariate normal distribution with diagonal covariance matrix \mathbf{W} . Finally, the initial terms \mathbf{m}_0 and \mathbf{C}_0 are known prior information.

It is worth to point out that the goal of the dynamic hierarchical QR analysis is to describe the evolution of a quantile of the longitudinal measurements' distribution. The uncertainty component v_i personalizes each individual from the information sharing structure $\mathbf{F}'_i \boldsymbol{\theta}_j$. The resulting information is carried by a component $\boldsymbol{\xi}_i$ which can be defined as, say, $\boldsymbol{\xi}_i = (\lambda_{i1}, \dots, \lambda_{iJ})'$, where $\lambda_{ij} = \mathcal{Q}_{y_{ij}}(\nu)$ are the quantiles of y_{ij} . This choice is discussed in the end of this Section. In the log hazard function of the time-to-event sub-model, described below, the effect of the $\boldsymbol{\xi}_i$ is quantified by a parameter vector $\boldsymbol{\alpha}$. A positive α_j , for a given j , indicates that the hazard rate will be $\exp(\alpha_j)$ times higher with a unit increase of the ν^{th} conditional quantile at that interview.

To build the joint model we use the relative risk Weibull regression model for the time-to-event data. This sub-model was rewritten from Guo and Carlin (2004) considering its

hierarchical form, as follows:

$$\begin{aligned}
T_i|r, h_i &\sim \mathcal{W}(r, \exp\{\mathbf{X}'_{1i}\boldsymbol{\beta}_1 + \boldsymbol{\xi}'_i\boldsymbol{\alpha}\}), i = 1, \dots, N \\
\log(h_i(t)) &= \log(h_0(t)) + \mathbf{X}'_{1i}\boldsymbol{\beta}_{1i} + \boldsymbol{\xi}'_i\boldsymbol{\alpha} \\
\boldsymbol{\gamma}_i &= \boldsymbol{\gamma} + \boldsymbol{\zeta}_i, \quad \boldsymbol{\zeta}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\zeta), \boldsymbol{\gamma}'_i = (\boldsymbol{\beta}'_{1i}, \boldsymbol{\alpha}')
\end{aligned} \tag{4.2}$$

It is cleaner to write the second line of (4.2) with a single predictor, say $\mathbf{Z}'_i = (\mathbf{X}'_{1i}, \boldsymbol{\xi}'_i)$, so that the second line can be written as $h_i(t) = h_0(t) \exp(\mathbf{Z}'_i\boldsymbol{\gamma}_i)$, where $h_0(t)$ is some baseline hazard function. For this work, we have chosen it as the well known Weibull function. It is defined for a shape parameter $r > 0$ and $h_0(t) = rt^{r-1}$ is the baseline hazard function.

The regressor \mathbf{Z}_i is a $(q_1 + q^*)$ -dimensional vector. This term includes two blocks. The first is composed of q_1 regressors, denoted as \mathbf{X}_{1i} , assumed to be known without measurement error. We include the index 1 to the covariates as they are associated with the time-to-event sub-model. The second is an arbitrary q^* dimensional latent term $\boldsymbol{\xi}_i$, discussed below.

The regression coefficient $\boldsymbol{\gamma}_i$ is decomposed as the sum of a common mean $\boldsymbol{\gamma}$, with a individual-specific latent component, $\boldsymbol{\zeta}_i$ which is normally distributed with zero mean and variance $\boldsymbol{\Sigma}_\zeta$. $\boldsymbol{\zeta}_i$ is commonly called the frailty term.

Combining the quantile regression with the time-to-event sub-models characterizes the joint model. This combination takes place through the term $\boldsymbol{\xi}_i$, which has many possible choices. To properly make this choice, we interpret what the model does with the information. We assume that the longitudinal measurements y_{ij} include some, at some level, information relevant to the hazard function. Whatever this information, we would like to filter it into the hazard function through $\boldsymbol{\xi}_i$. The key point of a stochastic description of the y_{ij} is to use some characteristic of its distribution, which we assume carries the desired information. This chapter gives the possibility to make this characteristic be a feature of the chosen quantile of the distribution. This feature can be the evolving trajectory μ_{1j} , the cumulated trajectory μ_{1J} or the quantiles themselves λ_{ij} or even the cumulated quantile λ_{iJ} . The model in (4.1) uses $\boldsymbol{\xi}_i = (\lambda_{i1}, \dots, \lambda_{iJ})'$ and the examples from Sections 4.4.2 and 4.4.1 compare this choice with $\boldsymbol{\xi}_i = \frac{1}{J} \sum_j \lambda_{ij}$. Note

that, in this last case, $\boldsymbol{\alpha}$ is unidimensional.

Note that $\boldsymbol{\xi}_i = (\lambda_{i1}, \dots, \lambda_{iJ})'$ implies that all individuals contribute with all J times to the hazard function, which may not be compatible with the reality of the study, since some individuals have their event before the end of the study. This means that the interview times after their event times have no observed value. By treating these interviews as missing data, Bayesian inference can easily incorporate their uncertainty into the hazard function, and their predictive distribution does not have a strong influence on the final result.

Once the model is completely defined, we proceed with the following Section which details the inference procedure.

4.3.1 Inference

The location–scale mixture representation of the ALD described in Section 4.3 is employed to ease the implementation of the full Bayesian inference approach via Markov Chain Monte Carlo (MCMC). The complete likelihood function is given by the equation below.

$$\begin{aligned}
L(\boldsymbol{\Theta} \mid \mathcal{D}_J) &= \prod_{i=1}^N (h_0(t) \exp(\mathbf{X}'_{1i} \boldsymbol{\beta}'_{1i} + \boldsymbol{\xi}'_i \boldsymbol{\alpha}))^{d_i} \\
&\times \exp\left(-\int_0^t h_0(u) \exp(\mathbf{X}'_{1i} \boldsymbol{\beta}'_{1i} + \boldsymbol{\xi}'_i \boldsymbol{\alpha}) du\right) [\boldsymbol{\gamma}'_i \mid \boldsymbol{\gamma}, \boldsymbol{\Sigma}_\zeta] \\
&\times [\boldsymbol{\theta}_0] \prod_j^J \mathcal{N}(y_{ij} \mid \lambda_{ij} + \kappa_{1\nu} \varepsilon_j, \kappa_{2\nu}^2 \sigma \varepsilon_{ij}) \\
&\times \mathcal{N}(\lambda_{ij} \mid \boldsymbol{\theta}, V) \mathcal{N}(\boldsymbol{\theta}_j \mid \boldsymbol{\theta}_{j-1}, \mathbf{W}),
\end{aligned}$$

where the collection of unknown parameters related with our proposed model is $\boldsymbol{\Theta} = \{\sigma, \boldsymbol{\Lambda}, V, \mathbf{W}, \boldsymbol{\theta}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\alpha}, r, \boldsymbol{\Sigma}_\zeta\}$, where $\boldsymbol{\theta} = (\boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_J)$. The term $\mathcal{D}_J = (\mathbf{y}, \mathbf{t}, \mathbf{d})$ contains all the information collected until time J . The posterior distribution of the model parameters is given by $[\boldsymbol{\Theta} \mid \mathcal{D}_J] \propto L(\boldsymbol{\Theta} \mid \mathcal{D}_J) [\boldsymbol{\Theta}]$. We consider the case where the variance block of \mathbf{W} relative to the covariates effects is given by $\mathbf{W} \boldsymbol{\beta}_{2j} \equiv \mathbf{0}$. Consequently, without an evolution structure for the covariates effects of the longitudinal

component, our model only evolves the polynomial terms of $\boldsymbol{\theta}_j$, i.e. $\boldsymbol{\mu}_j$.

A fully Bayesian inferential approach via MCMC for the joint model to estimate all parameters simultaneously is proposed. We assume that the disturbance terms ϵ_{ij} , v_i , and \boldsymbol{w}_j are mutually independent of each other. To complete the Bayesian inference specification scheme, we next need to specify prior distributions for all of the unknown parameters in Θ . We used weakly informative prior distributions for the parameters.

We assume normal distribution priors for $\boldsymbol{\beta}_1$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}_2$, with $\boldsymbol{\beta}_1 \sim \mathcal{N}(\mathbb{M}_1, \mathbb{V}_1)$, $\boldsymbol{\beta}_2 \sim \mathcal{N}(\mathbb{M}_2, \mathbb{V}_2)$, $\boldsymbol{\alpha} \sim \mathcal{N}(\mathbb{M}_3, \mathbb{V}_3)$. An inverse gamma (IG) prior is specified for σ , i.e., $\sigma \sim \text{IG}(a_0, b_0)$. For the disturbance variance V , we use the inverse gamma as prior $V \sim \text{IG}(a_1, b_1)$. The prior to the evolution effects variance matrices, \boldsymbol{W} , are set as an Inverse-Wishart (IW) distribution, $\boldsymbol{W} \sim \text{IW}_p(\boldsymbol{R}_0, g_0)$. For the parameter r of the baseline hazard function we assume the exponential distribution, $r \sim \text{Exp}(b_2)$. For simplicity, in the application exercises, no frailty term was used, i.e., $\boldsymbol{\Sigma}_\zeta \equiv 0$

In Section 4.4.1 and Section 4.4.2 (the real data analysis), the prespecified hyperparameters are given by $\mathbb{M}_1 = \mathbb{M}_2 = \mathbb{M}_3 = \mathbf{0}$, $\mathbb{V}_1 = 1000I_{q_1}$, $\mathbb{V}_2 = 1000I_{q_2}$, $\mathbb{V}_3 = 10000I_{q^*}$, $a_0 = 0.001$, $b_0 = 0.001$, $g_0 = 2$, $R_0 = 0.001I_2$, $a_1 = 0.001$, $b_1 = 0.001$ and $b_2 = 0.1$. We assumed that all parameters are independent a priori, which means that matrices such as \boldsymbol{W} can be assumed to be diagonal.

After specifying the models for the observed data denoted by $\mathcal{D}_J = \{(y_{i1}, \dots, y_{iJ}, t_i, d_i), i = 1, \dots, N\}$ and the prior distributions for the unknown model parameters, we can make statistical inference for the parameters based on their posterior distributions under the Bayesian framework. The posterior distribution can be seen in equation (4.3).

$$\begin{aligned}
& [\Theta \mid \mathbf{y}, \mathbf{t}, \mathbf{d}] \\
& \propto \prod_{i=1}^N [t_i, d_i \mid \gamma_i, r][\gamma'_i \mid \gamma, \Sigma_\zeta] \\
& \quad \times \prod_j^J [y_{ij} \mid \lambda_{ij}, \sigma][\lambda_{ij} \mid \theta_j, V][\theta_j \mid \theta_{j-1}, \mathbf{W}][\theta_0] \\
& \quad \times [\sigma][V][\mathbf{W}][\beta_2][r][\alpha][\beta_1].
\end{aligned} \tag{4.3}$$

Due to the complexity of the joint model, the posterior distribution of $[\Theta \mid \mathcal{D}_J]$ in (4.3) is not known in closed form, thus we obtain samples from it through MCMC methods. An advantage of the hierarchical representation of our proposed model is that they are easily implemented using the package Jags (Su and Yajima 2015) of the R free software (R Development Core Team 2018). Furthermore, one feature of this package is that it is not necessary to explicitly specify the full conditional distributions. The program code is available in the Appendix Section A.4.

The advantages of using a full Bayesian approach include that the uncertainty of the parameter estimates is fully captured in the posterior distribution and no asymptotic theory is needed to derive the standard error. Another advantage of our proposal is related to estimating the risk of different filtered values of the longitudinal marker for those individuals who are still vulnerable to the event. To this end, the proposed model framework offers a natural way of making such personalized dynamic predictions of future survival probabilities, based on the ideas of Rizopoulos (2011) and Taylor et al. (2013).

A key feature of this dynamic prediction framework is that the predicted longitudinal markers can be dynamically updated as additional longitudinal measurements from the target individual become available, providing on-line risk assessment.

To make dynamic predictions in the proposed model we base ourselves on Gonçalves et al. (2018). The next Section details the prediction procedure further.

4.3.2 Prediction with the dynamic hierarchical quantile joint model

A by-product of the parameter estimation is the ability to obtain personalized dynamic predictions. Our goal is to predict survival probabilities. There are two cases to be considered: a new subject is susceptible to the event. After observing the data available of all individuals, the event time for the new individual is to be predicted on the basis of its covariates and on the information gathered so far; a subject who is still in the study but the event has not yet occurred for them. During the observation process, we may be interested in predicting the (remaining) event time of an individual for whom the event has not yet occurred given longitudinal markers. Only the first case will be developed, with the second being directly derived from the first.

After fitting a sample of size N to our proposed model, we are interested in predicting survival probabilities for a new specific individual, at a specific time point s during follow-up. We would like to utilize all available information we have at hand to produce predictions of survival probabilities. The new subject i' must provide a set of longitudinal measurements $\mathbf{y}_{i'} = \{y_{i'j}; j = 1, \dots, J^* < J\}$, meaning that the new subject must have at least one interview, but not all J interviews.

Since most individuals fail before time J and the ones who do fail after that time do so shortly after, then it is not recommended that the survival probability be forecast much beyond time J , as this would mean an extrapolation on the observed data. A new individual with J^* interviews is assumed that they have not yet experienced the event by that time. Then the goal of the forecast is to predict the survival function at time $s, J^* < s \leq J$. This function is completely determined by the hazard function. Since all non-individual specific components are estimated from the N individuals, all that is left is to forecast the remaining $J - J^*$ values of the quantiles for the new individual.

In order to obtain predictions of the quantiles at times (interviews) $j' = J^* + 1, \dots, J$, we use the distribution associated with $\lambda_{i'j'} \mid \Theta$, which is identical to $[\lambda_{i'j'} \mid \theta_{j'}, V, \mathcal{D}_{J^*}] = [\lambda_{i'j'} \mid \theta'_{j'}]$. The notation for the accumulated data \mathcal{D}_{J^*} and the variance V have been

suppressed for visualization reasons. Since the available data only goes as far as J^* , then we need $[\lambda_{i'j'} \mid \boldsymbol{\theta}_{J^*}]$. The distribution can be obtained by

$$\begin{aligned}
[\lambda_{i'j'} \mid \boldsymbol{\theta}_{j'-1}] &= \int [\lambda_{i'j'} \mid \boldsymbol{\theta}_{j'}, \boldsymbol{\theta}_{j'-1}] [\boldsymbol{\theta}_{j'} \mid \boldsymbol{\theta}_{j'-1}] d\boldsymbol{\theta}_{j'} \\
[\lambda_{i'j'} \mid \boldsymbol{\theta}_{j'-2}] &= \int [\lambda_{i'j'} \mid \boldsymbol{\theta}_{j'-1}, \boldsymbol{\theta}_{j'-2}] [\boldsymbol{\theta}_{j'-1} \mid \boldsymbol{\theta}_{j'-2}] d\boldsymbol{\theta}_{j'-1} \\
&\vdots \\
[\lambda_{i'j'} \mid \boldsymbol{\theta}_{J^*}] &= \int [\lambda_{i'j'} \mid \boldsymbol{\theta}_{J^*+1}, \boldsymbol{\theta}_{J^*}] [\boldsymbol{\theta}_{J^*+1} \mid \boldsymbol{\theta}_{J^*}] d\boldsymbol{\theta}_{J^*+1}.
\end{aligned} \tag{4.4}$$

We can obtain $[\lambda_{i'j'} \mid \boldsymbol{\theta}_{j'}, \boldsymbol{\theta}_{j'-1}]$ in the first line of (4.4) since the model is Markovian, which means that $[\lambda_{i'j'} \mid \boldsymbol{\theta}_{j'}, \boldsymbol{\theta}_{j'-1}] = [\lambda_{i'j'} \mid \boldsymbol{\theta}_{j'}]$. The same applies to the other lines. Also, we already have $[\boldsymbol{\theta}_{j'} \mid \boldsymbol{\theta}_{j'-1}]$ in the first line of (4.4) and the equivalents of the other lines from the evolution equation.

Next, we use the $[\lambda_{i'j'} \mid \boldsymbol{\theta}_{j'}, \boldsymbol{\theta}_{j'-1}]$ distribution as the ‘‘bridge’’ term to construct \mathbf{Z}_i and the relative risk function $h_i(t)$, in the survival component. The probability of surviving at time $s > J^*$ given survival until J^* , that is, $[T \geq s \mid T > J^*, \mathcal{D}_{J^*}]$, can be derived from

$$\begin{aligned}
&[T \geq s \mid T > J^*, \mathcal{D}_{J^*}] \\
&= \int \int [T \geq s \mid T > J^*, \boldsymbol{\Theta}, \boldsymbol{\lambda}, \mathcal{D}_{J^*}] [\boldsymbol{\lambda} \mid \boldsymbol{\Theta}, \mathcal{D}_{J^*}] [\boldsymbol{\Theta} \mid \mathcal{D}_{J^*}] d\boldsymbol{\Theta} d\boldsymbol{\lambda} \\
&= \int \int \frac{S(s \mid \boldsymbol{\Theta}, \boldsymbol{\lambda})}{S(J^* \mid \boldsymbol{\Theta}, \boldsymbol{\lambda})} [\boldsymbol{\lambda} \mid \boldsymbol{\Theta}, \mathcal{D}_{J^*}] [\boldsymbol{\Theta} \mid \mathcal{D}_{J^*}] d\boldsymbol{\Theta} d\boldsymbol{\lambda},
\end{aligned}$$

where the term $S(J^* \mid \boldsymbol{\Theta})$ is the survival function at J^* time conditional to the MCMC outputs with B iterations, and $\boldsymbol{\lambda} = (\lambda_{i'1}, \dots, \lambda_{i'J})'$. We use the MCMC outputs to obtain the posterior distribution of the predicted survival function. Formally, we sample $\boldsymbol{\Theta}$ from $[\boldsymbol{\Theta} \mid \mathcal{D}_{J^*}]$, and obtain a numerical Monte Carlo approximation of $[T \geq s \mid T > J^*, \mathcal{D}_{J^*}]$ and it is given by $[T \geq s \mid T > J^*, \mathcal{D}_{J^*}] \approx \frac{1}{B} \sum_{b=1:B} \frac{S(s \mid \boldsymbol{\Theta}_b)}{S(J^* \mid \boldsymbol{\Theta}_b)}$.

The second case, where a subject in the study has not yet experienced the event, is easily adapted from these results.

4.4 Application

Both applications are implemented using the package Jags (Su and Yajima 2015) of the R free software (R Development Core Team 2018). We obtained a Markov chain sample with

500000 iterations and discarded the first 250000 iterations as the Burn-in period. In order to reduce sample autocorrelations we thin each chain by keeping every 5th simulated draw.

Visual methods, such as trace plots and correlation diagnostics were used for convergence assessment. The chains convergence were confirmed using standard tools within the coda package in R (Plummer et al. 2006).

In both applications, we compare two different choices for ξ_i , as discussed in the end of Section 4.3. One choice, referred to as J -dimensional bridge is $\xi_i = (\lambda_{i1}, \dots, \lambda_{iJ})'$. For the other choice, referred to as 1-dimensional bridge, we have $\xi_i = \frac{1}{J} \sum_j \lambda_{ij}$, added for comparison with other methods as they insert only one value to quantify the effect of the longitudinal measurement.

4.4.1 Schizophrenia dataset

This dataset (Henderson and Diggle 2002) contains information regarding the treatment of patients suffering from mental health issues. It is known that the PANSS (Positive and Negative Syndrome Scale) score is a useful tool that influences the risk that a patient will have a schizophrenic crisis. A patient that has a high score for a long period of time requires a treatment much more intense than one who maintains a low score. If the modeling of the crisis hazard is done through the expected value of the score, then patients with extremely high scores will pull all the patients' average up, which may lead to unnecessary treatment.

If, however, the modeling is done through a well chosen quantile of the distribution, then these discrepancies will be taken into account as we expect that quantile based analysis will be less impacted by extreme values. Also, as a customizing feature of the model, the biomedical specialist will be able to determine the quantile which most makes sense in terms of the analysis.

The dataset involves $N = 150$ patients who were assessed using the PANSS score. This is a medical scale used for measuring symptom severity of patients with schizophrenic conditions. A low PANSS score indicates that a patient has weak symptoms and a high

score implies a severe condition.

Assuming the regular case, the scale is denoted by y_{ij} and describes the mental state assessment for the i^{th} individual at scheduled times $s_j = \{0, 1, 2, 4, 6, 8\}$. The assessments are weekly timed.

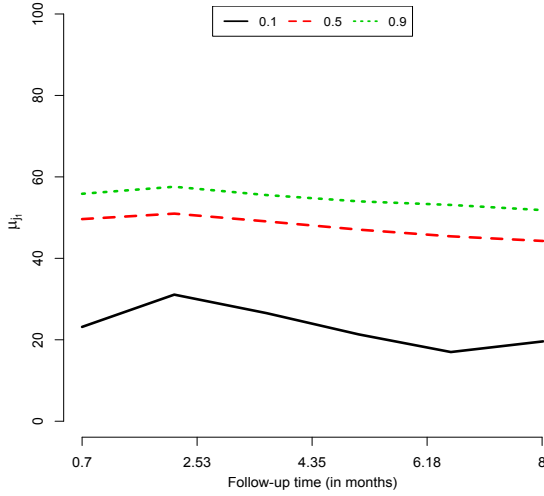
Chronically ill mental health patients were randomly allocated on one of the following three treatments: placebo (1), risperidone (2) and haloperidol (3). All groups had forcibly 50 patients in them. Other available information identifies the events due to reasons that were thought to be related to their mental state, i.e. the event times are collected based on the patient's mental state. Eight weeks without news of the patient's mental state configure a non-informative censorship.

Since the event time can also be non-informative, in cases where the event was not related to the individual's mental state, a right censoring indicator was created, receiving the value 0 when the event is non-informative, and 1 otherwise.

To show the performance of the model for different choices of ν , levels of quantiles are shown: 0.1, 0.5 and 0.9. Figure 4.1 shows how the choice of quantile can greatly influence the conclusions. The graphics show the mean levels of the trajectories of each of the quantiles for each choice of ξ_i . An interesting thing to note is that in both graphics, the trajectories of quantiles 0.5 and 0.9 are very similar when compared with quantile 0.1. This is evidence that the distribution of y_{ij} is not symmetric and that the choice of the quantile ν in the analysis is a relevant task. The lack of symmetry can indicate that using the mean of the distribution of y_{ij} instead of a quantile might be as damaging as a poor choice of ν .

It is also noticeable that both graphics are somewhat similar. In order to analyze that, note that the trajectories in the graphics are not only influenced by the longitudinal measurements, but also by the event times, since the inference is done simultaneously on all parameters of the joint model and they compose most of the bridge terms ξ_i , although the 1-dimensional bridge less so. Therefore the similarity between both graphics is not

(a) 6-dimensional bridge



(b) 1-dimensional bridge

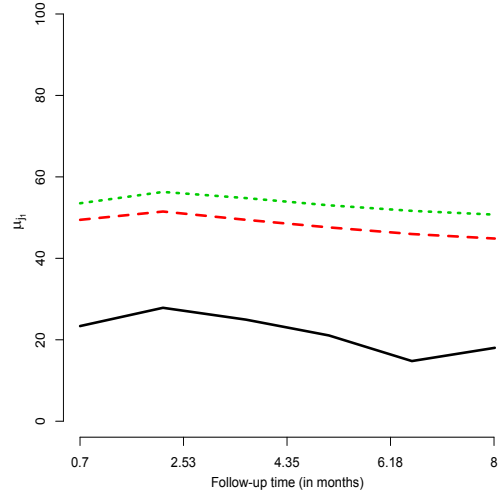


Figure 4.1: PANSS score for a sample of 150 patients with three quantile regression curves: The curves were obtained for the posterior mean of $\mu_{j1}, j = 1, \dots, 6$ which represent the mean level of the trajectories of the ν^{th} 0.1, 0.5 and 0.9 quantiles.

a surprise and is even an expected result, as it shows a consistent estimation of these mean trajectories, which is evidence that they are in fact representing features of the longitudinal measurements and are not being heavily influenced by the observed event times.

Figure 4.2 provides the posterior marginal distributions for the effect of the treatment on the PANSS score, which is expected to be negative, as the treatment is expected to reduce the score. The only quantile showing such behavior is 0.1, which may show that it can provide more information to the analysis than the other 2 quantiles, which show light positive effect. This again shows the importance of the choice of ν . Both choices of ξ_i show similar patterns with this respect. Most of all, it is clear that quantile regression provides a powerful and versatile tool.

Table 4.1 compares the results between these models. Columns 2 through 5 describe the models with a 6-dimensional bridge, where $\xi_i = (\lambda_{i1}, \dots, \lambda_{i6})'$, and columns 6 through 9 describe the models with a 1-dimensional bridge, where $\xi_i = J^{-1} \sum_j \lambda_{ij}$.

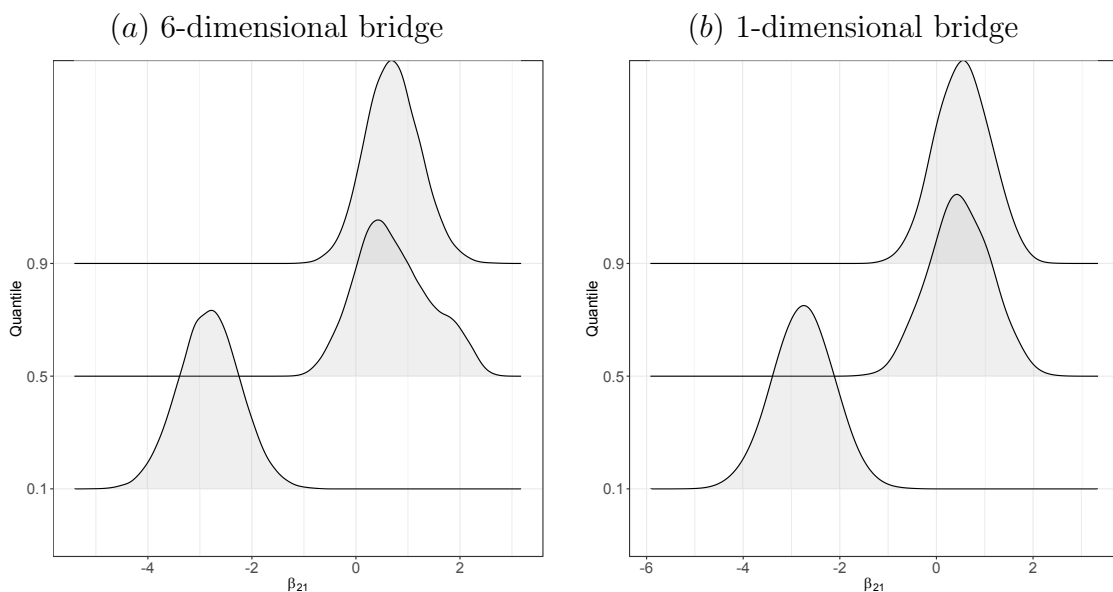


Figure 4.2: Density of the posterior marginal distributions of β_{21} , the treatment effect in the longitudinal sub-model for each quantile.

The following Bayesian comparison criteria are used: deviance information criterion (*DIC*) and logarithm of the pseudomarginal likelihood (*LPML*), as described in Zhang et al. (2017), as well as their components pD . It is noteworthy that the *DIC* and *LPML* are best used when the response follows a parametric distribution, which is not the case in quantile regression. The choice has been made for consistency of the thesis. Also a comparison metrics is derived based on Zhang’s idea of contribution differential, called $\Delta LPML_{Surv}$. They measure the contribution metric of adding the longitudinal measurements sub-model to the joint distribution. $\Delta LPML_{Surv} = LPML - LPML_{Surv}$, where $LPML_{Surv}$ is the logarithm of the pseudomarginal likelihood for a survival model which uses the longitudinal measurements as covariates of the relative risk function. For further details, see the supplementary material.

The models with a 6-dimensional bridge are much closer in terms of the selection criteria *LPML* and *DIC*. However, the indices for the models with a 1-dimensional bridge are very heterogeneous with great differences.

For each bridge choice, a different $LPML_{Surv}$ is used for reference, since they come from different models, that is, for the 6-dimensional bridge, all 6 longitudinal measurements are included as separate covariates in the relative risk function, while in the 1-dimensional

Table 4.1: Comparison of Bayesian model selection methods for the quantile joint models considering two choices of the bridge term ξ_i .

Q	6-dimensional bridge				1-dimensional bridge			
	$\Delta LPML_{Surv}^a$	$LPML$	DIC	pD	$\Delta LPML_{Surv}$	$LPML$	DIC	pD
10%	84.744	-32.527	57.721	0.998	-27.562	-32.176	58.059	1.223
50%	85.174	-32.698	60.028	2.661	-4.158	-8.770	9491.134	9469.403
90%	85.226	-32.641	59.330	1.960	2.814	-1.799	5754.909	5744.786

^aAbbreviations: $\Delta LPML_{Surv}$, Contribution metric of adding the longitudinal measurements sub-model to the joint distribution; $LPML$, Logarithm of the Pseudomarginal Likelihood; DIC , Deviance Information Criterion; pD , DIC component.

model, the average of the 6 values are used as a single covariate. The values of $LPML_{Surv}$ are approximately -117 and -4.6 respective to the 6- and 1-dimensional bridges. These values allow us to expect the 1-dimensional bridge models to have better $LPML$ metrics.

A note on $\Delta LPML_{Surv}$ is in order, since it can indicate the importance of both the longitudinal sub-model and the bridge choice. The models with a 6-dimensional bridge have great contribution to model prediction, that is, all of them are above 84, which means that modeling the longitudinal component has great positive impact on predictions. However, for the 1-dimensional bridge, only quantile 0.9 seems to be giving a positive impact on prediction. For the other quantiles, more research is needed to analyze why they provide negative contributions.

The DIC clearly favors the 6-dimensional model. The 0.1 quantile is the preferred model for this metric for both bridge choices. Note that this quantile in the 1-dimensional bridge is almost as preferred as the 6-dimensional bridge. This might show that $\nu = 0.1$ is ideal, especially if we join the results of Table 4.1 with the results of Figure 4.2.

Another important parameter in the model is the shape parameter r in the Weibull survival function. Its estimation shows interesting differences between the models and can be seen in Figure 4.3. The estimation of r is different between 1- and 6-dimensional bridges. 1-dimensional bridge leads to very similar estimation of the parameter for all

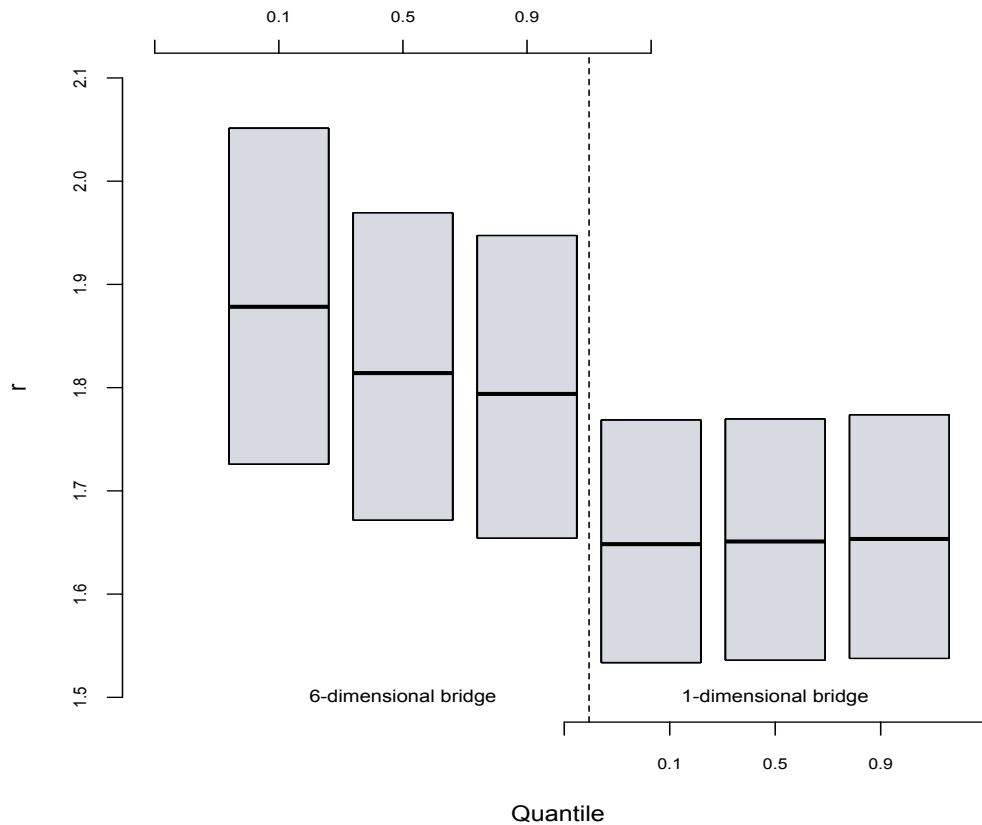


Figure 4.3: Boxplot of the posterior marginal distributions of the Weibull r shape parameter for each quantile.

quantiles, which seem to be slightly lower than the ones estimated by the 6-dimensional bridge models.

We interpret this by noting that changing covariates in a survival (sub-)model can lead to changes in the baseline hazard function, here represented by changes in the estimation of r , since each individual's hazard function is the product of the baseline function and the exponential of the linear predictor. This happens in the 6-dimensional bridge, but not in the 1-dimensional, which is consistent with the fact that the longitudinal information is condensed into a single value. This might indicate that the fitting power of the linear predictors of the 1-dimensional bridge models might be somewhat similar. This raises a question of how much different the 1-dimensional bridge QR analysis is compared to a similar model using the expected values of y_{ij} . Even more so, the choice of the ξ_i bridge component is relevant.

There is a strong point for the 1-dimensional bridge, since it summarizes the longitudinal information well, meaning that two studies with different values of J have comparable bridge terms if they both have a 1-dimensional bridge. In summary, we do not recommend any particular bridge in this study over the other.

The results shown in this Section show the potential power to be explored with our proposal. Possible extensions of this application are discussed in the conclusion of this chapter. Next, we use the survival dataset collected in randomized clinical trials with the objective of comparing Prednisone versus Placebo in Liver Cirrhosis Patients.

4.4.2 Liver data analysis

This dataset is taken from Andersen et al. (1993)(p. 19) and was analyzed in Henderson and Diggle (2002). It consists of the longitudinal observations of a Prothrombin index, a measure of liver function, for 488 patients from a controlled trial of liver cirrhosis. The time-to-event observations consist of time of death and associated (uninformative) censoring indicator, which happen due to cirrhosis related impossibility to remain in the study (40.2% of the cases), are also recorded along with a single baseline covariate - the allocated treatment in this instance. The variable Prothrombin index is measured in % and their follow-up times are measured in years with a total of $J = 17$ interviews. The treatment indicator is coded as 0 = placebo; 1 = Prednisone. Our work is motivated by the inclusion of various levels of the Prothrombin index in the hazard function, through different choices of the quantile ν (see Section 4.3). A small exploratory analysis is done on the data before model estimation to get familiarized with the dataset.

Figure 4.4(a) shows the boxplots of observed Prothrombin indexes plotted against time for the 488 patients included in the analysis compared by treatment Prednisone $arm = 1$ versus Placebo $arm = 0$.

The overdispersion and heterogeneity of the longitudinal markers is apparent in this plot. The dots represent the medians inside each box. The heavy tails and asymmetry of the markers distributions suggest that the analysis of their quantiles might be adequate

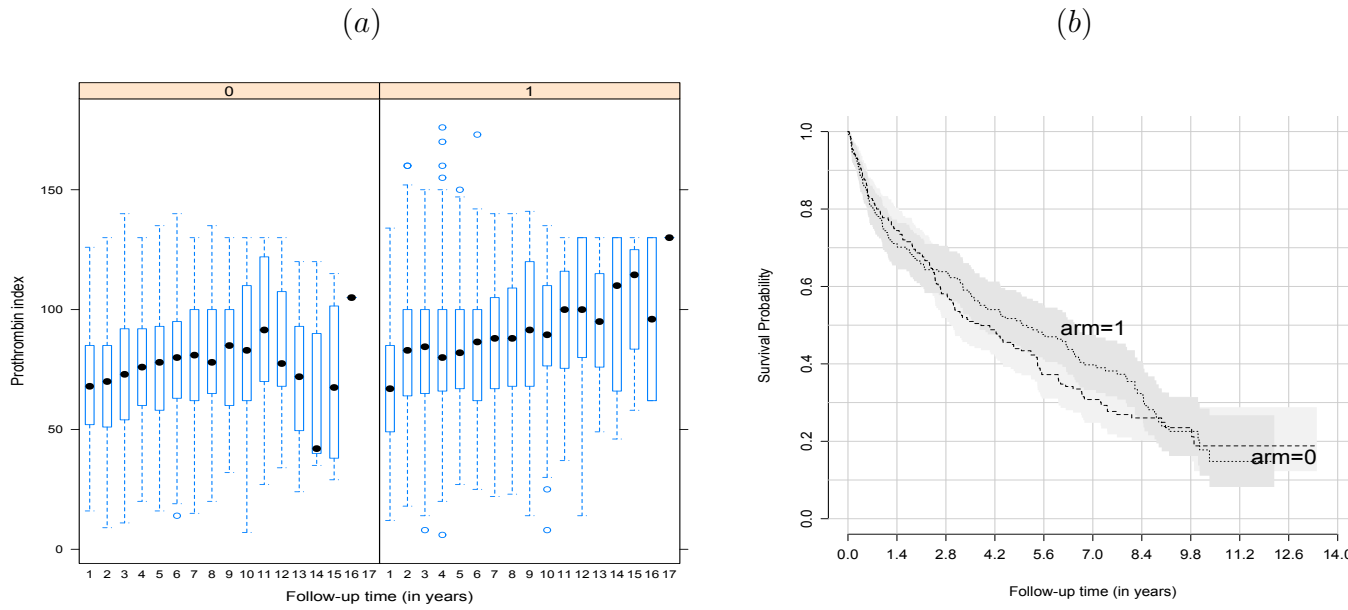


Figure 4.4: (a) Boxplots of Prothrombin index for the follow-up times, the points represents the median values. (b) Empirical survival probability (Kaplan-Meier) for Prednisone ($arm = 1$) versus Placebo ($arm = 0$).

instead of the usual moment regression. Figure 4.4 (b) summarizes the survival's probability and their respective confidence intervals based on the non parametric Kaplan-Meier approach (Kaplan and Meier 1958) by treatment. It can be seen that patients on treatment Prednisone appear to have a higher survival rate than those on Placebo. This occurs between percentiles 25% and 80% of the empirical survival distribution. The median event time for the patients on Placebo appears to be 6.3 years versus about 8 years for the ones on Prednisone. This indicates that the latter might be significantly important to prevent patients from dying.

We use the proposed dynamic quantile regression joint model to combine this information, especially since the Prothrombin index is so heterogeneous.

In the proposed model, all parameters are functions of quantile ν . Thus, by choosing different quantiles, one can conduct a comprehensive analysis of the relationship between the Prothrombin and event time. The research aimed at the Liver dataset is made with the goal of assessing the treatment effects. It is known that the Prothrombin index shows significant interaction with the treatment in such a manner that high Prothrombin index

is associated with a beneficial effect of the Prednisone treatment (Andersen et al. 1993). Since it is an index, the Prothrombin can be modeled as a longitudinal measurement. We can take different strategies to utilize the flexibility of our proposal. For example, conducting a study over the entire conditional distribution of the longitudinal outcome, we can fit the dynamic quantile regression joint model through a set of pre-selected quantiles, collect and compare the resulting parameter estimations.

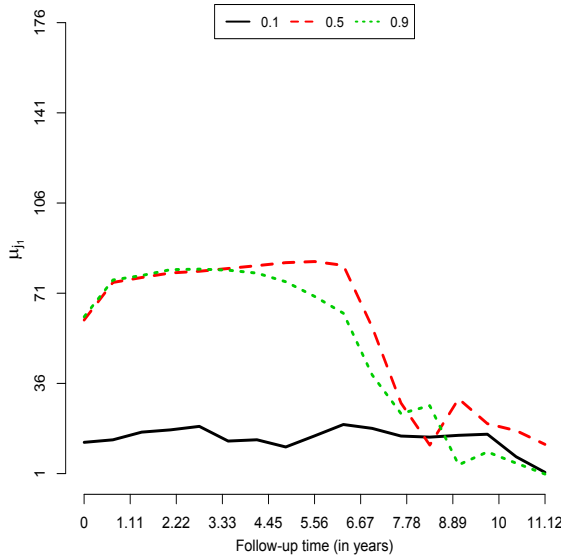
Less varying values of the Prothrombin index indicate a relatively stable effect on the hazard function. No matter how it varies, we expect the relative effect α in the hazard function to be negative, since greater index is expected to yield smaller risk.

Different quantiles were chosen for the analysis: $\nu \in \{0.1, 0.5, 0.9\}$. The median was chosen as a mean equivalent and the other two to represent the distribution extremes. Additionally, as in Section 4.4.1, the two same bridge components were considered. All Markov Chains have converged as verified by the methods discussed in Section 4.3.1.

The use of quantile regression is supported by Figure 4.5 which represents the mean levels of the trajectories for the ν^{th} quantiles for the model where the bridge is 1-dimensional. These suggest that the associated conditional distributions are skewed to the left, since the fitted median curve is nearly equal to the 0.9 quantile curve, being visibly different only from year 8 onwards. The choice of ν in an analysis should represent which quantile is pertinent to the analysis. If, say, patients with high level of the Prothrombin index with respect to the population are relevant, then a high value of ν is used. Figure 4.5 shows that the bridge choice also influences the longitudinal modeling. In both graphics the last points show lower quantiles above higher quantiles, which is an indication of lack of information. This is usual in this type of data since over the course of time, there are less and less observations due to the survival process.

Figure 4.6 shows the estimation of the shape parameter $r > 0$ of the Weibull distribution, used to model the baseline hazard function. It can be seen that, for the 1-dimensional bridge, the parameter marginal posterior distribution is concentrated around slightly above 0.9, which indicates a decreasing hazard function, and is stable over all quantiles.

(a) 17-dimensional bridge



(b) 1-dimensional bridge

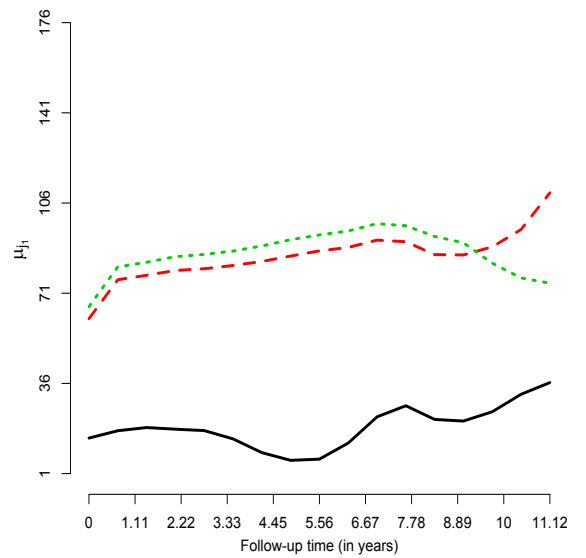


Figure 4.5: Prothrombin measurements for a sample of 488 patient with three quantile regression curves: The curves were obtained for $\mu_{j_1}, j = 1, \dots, 17$ that represent the mean level of the trajectories of the ν^{th} quantiles.

A different scenario is observed for the 17-dimensional bridge. They are all larger than 0.9 for all quantiles. Also, two clusters can be identified. For quantile 0.1, the distribution has a median slightly larger than 1, while for quantiles 0.5 and 0.9 the distribution is also roughly the same, but around almost 0.95.

Figure 4.7 shows the effects of Prednisone on the Prothrombin index. We can see that its effect is large and very different depending on the bridge choice. Clearly the effect is smaller for quantile 0.1. Also, we can see that its posterior for quantile 0.9 in the 17-dimensional bridge might not be uni-modal or that it is not very informative.

Table 4.2 contains the estimated regression coefficients based on the posterior distributions for the 1-dimensional bridge models. That which is more easily interpretable are the exponentiated coefficients, provided by the column $\exp(coef)$, which denote the multiplicative change in the risk due to each covariate. We observe that the Prednisone treatment effect β_{11} in Table 4.2 reduces the risk for the event by about 6%, consistent over all quantiles. It coincides with the conclusion of Andersen et al. (1993). On the

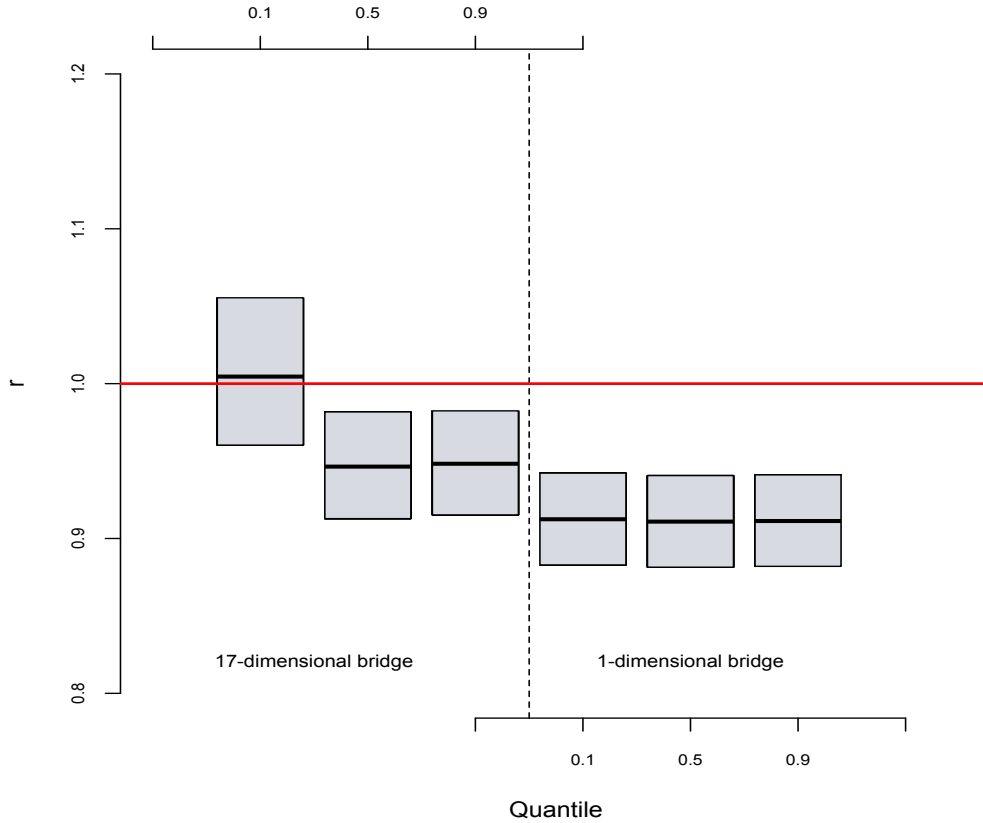


Figure 4.6: Boxplot of the posterior marginal distributions of the Weibull r shape parameter for each quantile.

other hand, one unit decrease in the ξ_i (bridge) is associated with a 17.5% increase in the risk for the 0.1 quantile model. Observe that the coefficient posterior distribution includes the value 0 in its 95% credible interval, which could mean that the effect is not statistically significant. Further research may include an attempt to model the fact that both high and low levels of Prothrombin index are associated with high risk. For this, a 2-dimensional longitudinal model may be necessary to model a quadratic level.

Finally, to conclude the analysis of this dataset, the models are compared with the measures explained in Section 4.4.1. The results show preference towards a 1-dimensional bridge, with better DIC and $LPML$ metrics. Quantile 0.1 shows preference in both cases when compared with the other quantiles, with better $LPML$ in the 17-dimensional bridge column and with better DIC in the 1-dimensional bridge column. Clearly there is no preferred bridge choice, as the metrics are usually very close. For quantile 0.9 the same discussion from Section 4.4.1 about the bridge choice applies. This is still a work

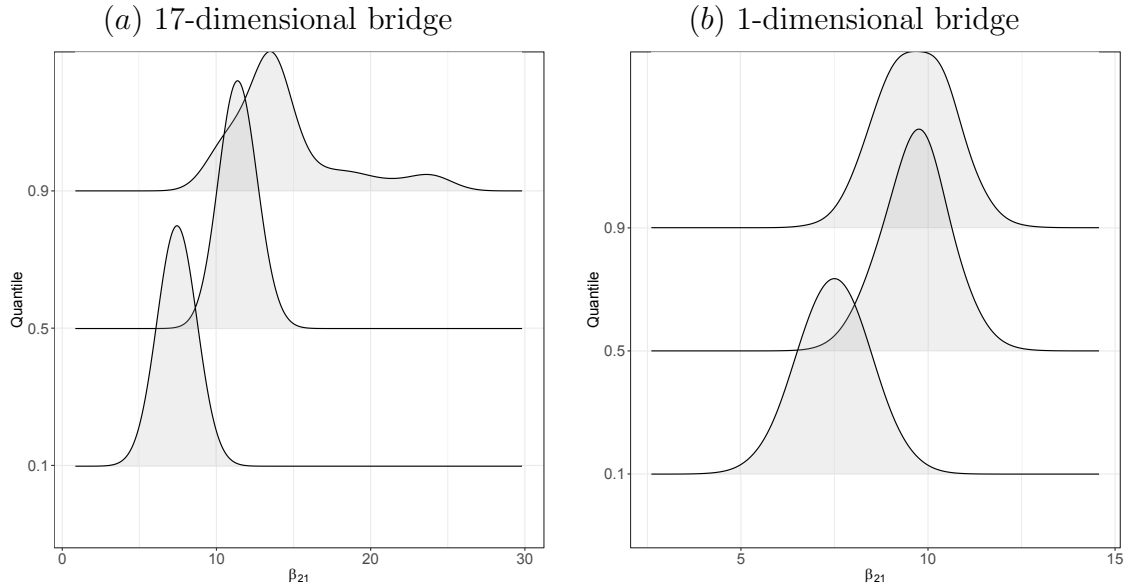


Figure 4.7: Densities of the posterior marginal distributions of β_{21} for each quantile.

Table 4.2: Summary of the estimated regression coefficients based on the posterior distributions related the survival sub-model. 1-dimensional bridge term.

Q^a	P	median	exp(coef)	exp(-coef)	mean	sd	2.5%	25%	75%	97.5%
10%	α	-0.161	0.851	1.175	-0.057	0.545	-0.931	-0.390	0.143	1.292
	β_{11}	-0.064	0.938	1.066	-0.063	0.117	-0.293	-0.141	0.015	0.164
50%	α	-0.003	0.997	1.003	-0.006	0.016	-0.056	-0.007	0.002	0.010
	β_{11}	-0.074	0.929	1.077	-0.076	0.117	-0.310	-0.153	0.004	0.149
90%	α	-0.007	0.993	1.007	-0.045	0.073	-0.200	-0.109	0.001	0.065
	β_{11}	-0.075	0.928	1.078	-0.076	0.118	-0.304	-0.157	0.005	0.155

^aAbbreviations: Q, quantile; P, parameter; sd, standard deviation; 2.5%, 25%, 75%, 97.5% quantiles of distribution.

Table 4.3: Comparison of Bayesian model selection methods for the quantile joint models considering two choices of the bridge term ξ_i .

Q	6-dimensional bridge				1-dimensional bridge			
	$\Delta LPML_{Surv}^a$	LPML	DIC	pD	$\Delta LPML_{Surv}$	LPML	DIC	pD
10%	95.816	-331.414	296.446	6.635	-316.489	-330.631	291.613	1.163
50%	96.076	-356.726	319.891	27.467	-312.047	-326.215	25206.869	24923.989
90%	96.330	-379.396	463.602	168.904	30.367	16.213	84526.242	84488.225

^aAbbreviations: $\Delta LPML_{Surv}$, Contribution metric of adding the longitudinal measurements sub-model to the joint distribution; *LPML*, Logarithm of the Pseudomarginal Likelihood; DIC, Deviance Information Criterion; pD, DIC component.

in progress, as it is still necessary to compare the model with other models without a dynamic component, as well as some modeling nuances, such as frailty, which can be added.

4.5 Conclusions and future work

A joint model for survival and longitudinal data was presented using quantile regression for the longitudinal sub-model. A Weibull baseline hazard function was used although any function can be used as well. Two choices of bridge terms were tested in two real datasets which have already been analyzed in other works. Despite the lack of a clear conclusion about which quantile or bridge function to use, it is clear that the proposal provides flexibility and the possibility of including information into the joint model from a new perspective.

Table 4.3 illustrates that using quantiles away from 50% provide better fit for the dataset, giving evidence that quantile regression based models have useful versatility in the analysis. However, when the study does not provide an interpretative indication of which quantile should be used to draw appropriate conclusions, then tables like 4.3 could help determine which quantile to use based on goodness of fit. Thus, a possible extension is the estimation of the quantile ν . If regarded as an unknown parameter, we

can estimate it jointly with other parameters in the model to infer on the quantile that provides the best fit for that particular dataset. This can be useful when a particular quantile cannot be provided by the researcher, so it becomes more attractive to use the quantile which best fits the data. This idea is new in our context but not in the marginal QR approach (Chokethaworn et al. 2017).

Another important feature of the model, which also provides versatility, is the choice of the bridge term. Its choice has also not been fully probed, and that could be an interesting follow-up to the analysis. It could be also of value to verify how much the time-to-event data is influencing the longitudinal trajectories. There is also knowledge to be obtained from comparing the results obtained in this chapter with a joint model using moment regression, particularly a model whose only difference is with respect to the mean/median.

Chapter 5

Joint model with Markov switching regimes to longitudinal marker and event time dataset

This chapter discusses joint models for a longitudinal biomarker and a failure time outcome in the setting of including a latent term employing a Markov switching regression model. This inclusion directly affects the mean longitudinal trajectory which allows a bidirectional regime change. This characteristic provides more flexibility to the proposal compared to the quality adjusted to lifetime model. The inference procedure is done under the Bayesian paradigm and MCMC methods are used to obtain samples of the posterior distribution. The proposed methodology is applied to the analysis of a clinical trial dataset. We analyze the case study of children with episodes of diarrhea.

5.1 Introduction

It is a known fact that quality of life impacts the health of patients with chronic diseases. A widely used novel technique is known as palliative care and has the goal to improve the quality of life of patients undergoing treatment. The method for measuring quality of life is subjective - and prone to arbitrariness - and currently varies per disease usually through filling forms. In Fiteni et al. (2015) a procedure is described for pancreatic cancer, for instance.

Measurements of quality of life have been treated in the medical context in a lifetime analysis framework since the 1980's. The widely used methodology called Quality Adjusted Lifetime (QAL) is introduced by Gelber et al. (1989). The health state induced by the quality of life is Markovian according to Chen and Sen (2001). Ghosh and

Mukhopadhyay (2007) use a Bayesian approach for a one-directional state evolution, that is, the states can only change in one direction or not move at all. Silva et al. (2009) develops a frequentist approach for bidirectional states.

The problem is essentially characterized by the presence of two components. One is associated with the quality of life and the other with the event time. It is intuitive that these components can be treated as done in Chi and Ibrahim (2006), that is, treated jointly. Additionally, the quality of life can also be non-progressive, that is, bidirectional, and the state levels are not necessarily known.

The method proposed in this work follows this joint model and attempts to incorporate these desired properties. We model quality of life and event time simultaneously using the joint modeling of a longitudinal marker and time-to-event data. This well accepted approach in the literature of survival models is clearly explained in Rizopoulos (2012). In this context, we consider a multi-state model which uses unknown states at different times. These unknown states, when used correctly, may be considered the true longitudinal trajectory in a joint model.

Furthermore, the proposed model can be extended to be used in any situation where there is a change in regime for a given response, so it is not restricted to quality of life modeling. This is done in the example application. The next Section reviews a Markov-switching regression model and introduces it in the joint model context. Section 5.3 presents the Bayesian inference procedure for the model. In Section 5.4 we exemplify the model with data of children with diarrhea episodes and we conclude the work in Section 5.5 with a discussion.

5.2 Joint model with Markov-switching regimes

In this Section we discuss situations in which the occurrence of an event is influenced by one or more variables with potential bidirectional regime change. Markov switching models, see Hamilton (1989), have been successfully used for fitting economic and financial time series, for example, interest rates, exchange rates, consumption costs and so on. We use this structure so that the longitudinal marker process assumes sudden

discrete shifts in level due to sudden abnormal events. Note that we use the first-order Markov switching process, see Kim and Nelson (1999), applied to longitudinal markers data, where only the markers' first moment varies. This way, transitions are unrestricted with respect to direction, opposed to the QAL approach.

The proposed joint model with Markov-switching regimes is based on the introduction of a latent structure (states) into the mean longitudinal component trajectory. A survival sub-model is built conditional on the states in a multiplicative way with respect to the hazard function. This proposal extends the models in Chapters 2 and 3, where the longitudinal sub-model is replaced by a dynamic linear Markov-switching structure.

The inference procedure is done under the Bayesian paradigm and MCMC methods are used to obtain samples of the posterior distribution. Decomposable comparison criteria that evaluate fit and prediction of the model are used to determine the relevance of including the longitudinal biomarkers modeling based on Zhang et al. (2017).

5.2.1 Mathematical definition

We focus on the class of linear regime-switching models (MSW) assuming that the regimes occurring at all times $j = 1, \dots, J$ can not be observed. They are defined by an unobservable process, which we denote as S_j . The model is defined so that regimes are updated stochastically. The class of MSW models with \mathcal{K} regimes is based on the assumption that a regime is determined by the discrete ergodic first order Markov process, hence it is only necessary to consider the current and the previous states: $P(S_j = m | S_{j-1} = l, S_{j-2} = k, \dots) = P(S_j = m | S_{j-1} = l) = p_{lm}$ for $1 \leq l, m \leq \mathcal{K}$, where p_{lm} is a state transition probability. The transition means that the process moves from state l at the time $j - 1$ to the state m at time j in the Markov chain. Some restrictions for the probabilities are $0 \leq p_{lm} \leq 1, \sum_{m=1}^{\mathcal{K}} p_{lm} = 1$. For details on MSW models see, e.g., Hamilton (1989) and Hamilton (1994).

For the survival sub-model, let $T_i = \min(T_i^*, C_i)$ be the random variable relative to the observed event time for the i^{th} subject, where T_i^* is the true underlying event time, C_i is

the censoring time and $i = 1, \dots, N$ where N is the number of individuals. The observed value of T_i is represented by t_i . Let d_i be the event censored indicator for the i^{th} subject (1 if the event is observed, and 0 otherwise), and D_i its respective random variable. Let Y_{ij} be the continuous longitudinal outcome for individual i measured at time j . Its realizations y_{ij} are measured at times a_{ij} , for times of visit $j = 1, \dots, J_i$. Assume, for simplicity, that the follow-up times can be different, but the number of visits must be the same, that is $J_i = J, \forall i = 1, \dots, N$, which is called the regular case. A first direct consequence of this is that the measurements times a_{iJ_i} will be comparable along all individuals, which means we can denote all visits simply as $a_j, j = 1, \dots, J$.

A common assumption for the longitudinal measurements Y_{ij} is that they are a function of time, but it is observed only for a finite set of times, which we call j . Since no further information of these measurements is available, then we assume that the observed values are fixed values per individual added to random values, which means that we make no assumptions about the longitudinal process outside the observed times. We consider a hierarchical model for the fixed value per individual. If we assume that this fixed value is constant over a period of time, and also that the range of possible values is not only finite, but also known, then a Markov-switching regimes model is not unrealistic.

Let $\mathcal{Y}_i(t) = \{y_{ij} : j = 1, \dots, J; j \leq t\}$ be the set of observed longitudinal measurements until time t . Define the \mathcal{K} -dimensional vector $M = \{\mu_1, \dots, \mu_{\mathcal{K}}\}$, where $\mu_1 > \dots > \mu_{\mathcal{K}}$. The ordering of the values $\mu_k, k = 1, \dots, \mathcal{K}$ is arbitrary but necessary for identifiability. Then let $\mathcal{M}_i(t) = \{\mu_{s_j} + v_i : j = 1, \dots, J; j \leq t; s_j \in \{1, \dots, \mathcal{K}\}\}$, where v_i is a random effect, be the filtered longitudinal process at the observed times per individual until time t . The states s_j defines the index of the level in M enabled at time j . The proposed idea is to make $\mathcal{Y}_i(t)$ equal to $\mathcal{M}_i(t)$ plus an appropriately sized Gaussian random vector. Then we denote $h_i(t | \boldsymbol{\xi}_i, \cdot)$ as the hazard function for the survival sub-model, where $\boldsymbol{\xi}_i$ is somehow related to $\mathcal{M}_i(J)$.

Unless otherwise noted, we will use $\boldsymbol{\xi}_i = \mathcal{M}_i(J)$, so the hazard function is $h_i(t | \mathcal{M}_i(J), \cdot)$. Note that this implies that all individuals contribute with all J times to the hazard function, which may not be compatible with the reality of the study, since some

individuals have their event before the end of the study. This means that the interview times after their event times have no observed value. By treating these interviews as missing data, Bayesian inference can easily incorporate their uncertainty into the hazard function, and their predictive distribution does not have a strong influence on the final result.

The term ξ_i is introduced to provide some versatility in the choice of the linking components and will be addressed again in the applications.

The Markov-switching regimes are the μ_{s_j} , modeled through a Markovian structure. Since the possibilities for their values are finite, then all we must do is define an appropriate transition matrix for the consecutive values. Note that the values are shared for all individuals through the hierarchy, although they are differentiated by the random effect v_i . The proposed Markov switching joint model can be written as a set of two sub-models

$$\begin{aligned} y_{ij} &= \lambda_{ij} + e_{ij}, & e_{ij} &\sim \mathcal{N}(0, \sigma^2) \\ \lambda_{ij} &= \mu_{s_j} + v_i, & v_i &\sim \mathcal{N}(0, V), \\ \log h_i(t) &= \mathbf{Z}'_i \boldsymbol{\gamma} + \log(\psi_i), \end{aligned} \tag{5.1}$$

where μ_{s_j} is described by a \mathcal{K} -state first-order Markov process introduced in Section 5.2.1, which models the longitudinal markers data. The second equation models a relative risk survival sub-model given an arbitrary baseline hazard function and will be detailed later in this Section.

The term λ_{ij} represents the level for subject i at time j . The random effect v_i is modeled as a Gaussian random variable with 0 mean and variance V whilst S_j is modeled according to a \mathcal{K} -state first-order Markov process, where $\mu_{s_j} \in M$. The initial probabilities for S_0 are given by the \mathcal{K} -dimensional vector \mathbf{q} . The $\mathcal{K} \times \mathcal{K}$ -dimensional transition matrix for this process is defined by the transition probabilities where every row must add up to 1 and the transition probability of the states μ_l to μ_m is 0 if $|l - m| \geq 2$. This 1-step only restriction is meant to represent a smooth state evolution. The transition probabilities are unknown model parameters.

To complete the model specification, the second sub-model in (5.1) must be defined.

By considering γ as composed by a four block vector $(\log(h_0(t)), \boldsymbol{\beta}', \boldsymbol{\alpha}', \log(\psi))'$ and also $\mathbf{Z}_i = (1, \mathbf{X}'_i, \boldsymbol{\lambda}'_i, 1)'$, we can rewrite the sub-model as $\log h_i(t \mid \boldsymbol{\lambda}_i) = \log h_0(t) + \mathbf{X}'_i \boldsymbol{\beta} + \boldsymbol{\lambda}'_i \boldsymbol{\alpha} + \log(\psi_i)$, where $\boldsymbol{\lambda}_i = (\lambda_{i1}, \dots, \lambda_{ij})'$, $h_0(t)$ is some baseline hazard function, \mathbf{X}_i is a covariates vector for subject i whose effects are represented by $\boldsymbol{\beta}$ and $\log(\psi_i) \sim \mathcal{N}(0, \Psi)$ is a frailty term.

The two sub-models are linked by the term $\boldsymbol{\lambda}_i$ and $\boldsymbol{\alpha}$ quantifies the effect of this term in the hazard function multiplicatively.

In this work, we use two parametric forms for the baseline hazard function $h_0(t)$, namely the Weibull distribution, whose mathematical form is rt^{r-1} for positive r , and the Gompertz distribution, which has a log representation by $\eta_0 + \eta_1 t$ for $\eta_0 \in \mathbb{R}$ and $\eta_1 > 0$.

We also use a semi-parametric form for the baseline hazard function. This form discretizes the interval $[0, \max(\mathbf{t}) + dt)$, where dt is an infinitesimal increment to ensure that the largest observed time is in the interval. The choice of number of intervals is not trivial and is justified in Chapter 3. For this model, we use $K = J$ intervals denoted by $I_k = [\tau_{k-1}, \tau_k)$, $k = 1, \dots, K$, where τ_k are the chosen equally spaced nodes for the discretization, restricted by $\tau_0 = 0$ and $\tau_K = \max(\mathbf{t}) + dt$. We denote $\delta = \tau_k - \tau_{k-1}$ the constant length of the intervals. For each I_k , $k = 1, \dots, K$, the function is defined by:

$$\begin{aligned} \log \mathbf{h}_{0k} &= \mathcal{G}(\delta) \log \mathbf{h}_{0(k-1)} + \boldsymbol{\omega}_k, \quad \boldsymbol{\omega}_k \sim \mathcal{N}(0, \boldsymbol{\Omega}), \\ \log \mathbf{h}_{00} \mid \mathcal{D}_0 &\sim \mathcal{N}(\mathbf{c}_0, \mathbf{C}_0), \end{aligned} \tag{5.2}$$

where $\mathcal{G}(\delta)$ is an evolution matrix for $\log \mathbf{h}_{0k}$. For versatility, one can define $\log \mathbf{h}_{0k}$ as a vector and achieve, say, polynomial growth models, but in our case, we use a 1-dimensional value for it, which means that $\mathcal{G}(\delta)$ is scalar. The matrix $\boldsymbol{\Omega}$ is some covariance matrix. In our 1-dimensional case, it is merely the variance of the evolution equation. Finally, the initial terms \mathbf{c}_0 and \mathbf{C}_0 are known prior information.

5.2.2 Two states: A particular case

As an example we implement our proposal assuming that the structural change model has two states, i.e., $\mu_{s_j} \in \{\mu_1, \mu_2\}$, where $\mu_1 > \mu_2$. In this case the true longitudinal marker

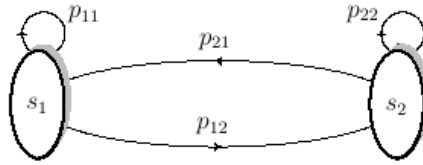


Figure 5.1: Markov Switching model for two states is represents through directed acyclic graphs (DAG).

λ_{ij} , follows a distribution conditional to a latent process $S_j \in \{1, 2\}$. Then

$$\begin{aligned}
 y_{ij} | \lambda_{ij} &\sim \mathcal{N}(\lambda_{ij}, \sigma^2) \\
 \lambda_{ij} | S_j &\sim \mathcal{N}(\mu_{s_j}, V) \\
 P(S_j = m | S_{j-1} = l) &= p_{lm}; \quad l, m \in \{1, 2\} \\
 P(S_0 = 1) &= 1 - P(S_0 = 2) = q_1,
 \end{aligned} \tag{5.3}$$

where q_1 is the first element of the initial probabilities vector \mathbf{q} , say, 0.5 if no other information is available. Note that $p_{11} + p_{12} = 1$ and $p_{22} + p_{21} = 1$, so for a binary process S_j , there are only two free parameters, say p_{11} and p_{22} . The DAG in Figure 5.1 shows the states dynamics of this particular case in an intuitive way.

The next Section we describe the inference procedure.

The usual challenge in the inference process is about the latent terms. Most of the parameters have a dependence with the S_j unobservable quantities. Many advanced techniques can be found in the literature, for instance Kim (1994) proposed an interesting implementation alternative. Kim and Nelson (1999) applied them in many practical situations. We use their idea for sampling from the full conditional of S_j .

5.3 Inference

The pseudo-likelihood function for the model with \mathcal{K} states is the product of two factors. To facilitate notation, consider $\mathbf{s} = s_0, s_1, s_2, \dots, s_J$, where s_0 is the initial state. To

include the transition probabilities, note that the matrix is

$$A = \begin{pmatrix} p_{11} & p_{12} & 0 & 0 & 0 & \dots & 0 \\ p_{21} & p_{22} & p_{23} & 0 & 0 & \dots & 0 \\ 0 & p_{32} & p_{33} & p_{34} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 0 & p_{\kappa-1,\kappa-2} & p_{\kappa-1,\kappa-1} & p_{\kappa-1,\kappa} \\ 0 & \dots & 0 & 0 & 0 & p_{\kappa,\kappa-1} & p_{\kappa,\kappa} \end{pmatrix}. \quad (5.4)$$

Since every row must sum to 1, we can concatenate the unknown parameters as $\mathbf{p} = p_{12}, p_{21}, p_{22}, p_{32}, p_{33}, \dots, p_{\kappa-1,\kappa-2}, p_{\kappa-1,\kappa-1}, p_{\kappa,\kappa-1}$. Then, the first factor of the pseudo-likelihood function L_i^l is given by

$$L_i^l(\mathbf{S}, \sigma^2, V, \mathbf{q}, \mathbf{p}, \boldsymbol{\lambda}_i, \boldsymbol{\mu} \mid \mathbf{y}_i) = \prod_{j=1}^J \mathcal{N}(y_{ij} \mid \lambda_{ij}, \sigma^2) \mathcal{N}(\lambda_{ij} \mid \mu_{s_j}, V) P(S_j = m), \quad (5.5)$$

where $P(S_j = m) = (\mathbf{q}' A^j)(m)$, with m being the state. Note that all values in $M =$ are unknown quantities of the model and must be estimated. This marginal probability can be derived in two steps. First, we find the conditional distribution $P(S_j = m \mid S_0 = l) = A^j(l, m)$, the position l, m from the j -step transition matrix (Gamerman and Lopes 2006). Then, we marginalize with respect to S_0 by $P(S_j = m) = \sum_l P(S_j = m \mid S_0 = l) P(S_0 = l)$. Since $P(S_0 = l) = q_l$, then this probability is as stated.

This factor of the pseudo-likelihood is not computable due to the data augmentation of \mathbf{S} , but this problem will be dealt with easily in the MCMC based inference, which is detailed later.

The second factor of the pseudo-likelihood function comes from the survival sub-model and depends on the baseline hazard function. For notation simplicity, call $R_i = (\mathbf{X}'_i \boldsymbol{\beta} + \boldsymbol{\lambda}'_i \boldsymbol{\alpha} + \log(\psi_i))$. Then the factor is, for the Weibull, Gompertz and Piecewise

exponential, respectively:

$$\begin{aligned}
L_i^s(r, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\lambda}_i, \Psi | t_i, d_i) &= (rt_i^{r-1})^{d_i} \exp(d_i R_i) \exp\left(-\int_0^{t_i} \exp(R_i) r u^{r-1} du\right) \\
L_i^s(\eta_0, \eta_1, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\lambda}_i, \Psi | t_i, d_i) &= (e^{\eta_0 + \eta_1 t_i})^{d_i} \exp(d_i R_i) \exp\left(-\int_0^{t_i} \exp(R_i) e^{\eta_0 + \eta_1 u} du\right) \\
L_i^s(\mathbf{h}_0, \boldsymbol{\Omega}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\lambda}_i, \Psi | t_i, d_i) &= [h_{00}] \prod_{k=1}^K (h_{0k})^{d_i} \exp(d_i R_i) \exp\left(-\int_0^{t_i} \exp(R_i) e^{h_{0k} u} du\right) \times \\
&\quad \times [h_{0k} | h_{0,k-1}].
\end{aligned} \tag{5.6}$$

We combine the parameters vector of the longitudinal sub-model into $\boldsymbol{\Theta}_l = (\mathbf{S}, \sigma^2, V, \mathbf{q}, \mathbf{p}, \boldsymbol{\lambda}_i, \boldsymbol{\mu})$ and of the survival sub-model into $\boldsymbol{\Theta}_s = (\text{base}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\Lambda}, \Psi)$, where base represents the parameters in the chosen baseline hazard function, so that base = $(\mathbf{h}_0, \boldsymbol{\Omega})$ for the semi-parametric baseline function, base = r for the Weibull baseline function and base = (η_0, η_1) for the Gompertz baseline function. The complete parameter vector is obtained by $\boldsymbol{\Theta} = \boldsymbol{\Theta}_l \cup \boldsymbol{\Theta}_s$. To complete the model specification the term $\boldsymbol{\xi}_i$ must be chosen as to represent the desired filtered information from the longitudinal sub-model to be included in the hazard function. The likelihood function of the joint model is obtained by:

$$L(\boldsymbol{\Theta} | \mathbf{t}, \mathbf{d}, \mathbf{y}) = \prod_{i=1}^N L_i^l(\boldsymbol{\Theta}_l | \mathbf{y}_i) L_i^s(\boldsymbol{\Theta}_s | t_i, d_i) \tag{5.7}$$

To complete the model specification for the Bayesian inference, we need to choose prior distributions for $\boldsymbol{\Theta}$.

5.3.1 Prior distribution

In order to complete the requisites for the Bayesian analysis, prior distributions must be defined. To begin with, the prior distribution for the initial probabilities $\mathbf{q} \sim \text{Dir}(\mathbf{u}_0)$, where \mathbf{u}_0 are known hyperparameters for $\mathbf{q} = (q_1, q_2, \dots, q_K)$. Inverse-Wishart and inverse Gamma distributions are assigned for the variance components. To $\sigma^2 \sim \text{IG}(a_{\sigma^2}, b_{\sigma^2})$, $V \sim \text{IG}(a_V, b_V)$, where a, b are known hyperparameters and to $\boldsymbol{\Omega} \sim \text{IW}(\mathbf{R}, g)$ we use inverse-Wishart, $\text{IW}(\mathbf{R}, g)$ with scale matrix \mathbf{R} and g degrees of freedom given hyperparameters. The prior distributions for components of A transition matrix $\mathbf{p}_l \sim \text{Dir}(\mathbf{u}_{l0})$, for $\mathbf{p}_l = (p_{l1}, p_{l2}, \dots, p_{lK})$ and $l = 1, \dots, K$. For the parameters related

to the mean trajectory $\boldsymbol{\mu} = (\mu_1, \dots, \mu_K)'$ we assume independent normal priors for the univariate terms of $\boldsymbol{\mu} \sim \mathcal{N}(\mathbf{m}_\mu, z_\mu \mathbf{I}_K)$. Technically speaking, to enforce that $\mu_2 > \mu_1$, we used in the application with 2 states a prior where $\mu_2 = \mu_1 + \zeta$ with $\zeta \sim \mathcal{N}(0, \sigma_\zeta^2)I(0, \infty)$, which has been inspired by Kim and Nelson (1999). Similar priors are used for the effects $\boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{m}_\alpha, z_\alpha \mathbf{I}_K)$ and for the effects $\boldsymbol{\beta}$ in survival sub-model we take $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{m}_\beta, z_\beta \mathbf{I}_x)$ and to complete the prior specification we assign the distribution exponential prior for the r parameter in the Weibull baseline hazard, it is $r \sim \text{Exp}(c)$. Finally for $\eta_0 \sim \mathcal{N}(m_{\eta_0}, C_{\eta_0})$ and for η_1 we use a truncated normal prior, it is $\eta_1 \sim \mathcal{N}_{(0, +\infty)}(m_{\eta_1}, C_{\eta_1})$.

In our applications the hyperparameters were chosen to represent fairly vague prior information.

After choice of the prior distribution for the unknown parameter and assuming independence prior we sample successively from the full conditional distributions.

5.3.2 Posterior distribution

All inference is done with respect to the posterior distribution $\pi(\boldsymbol{\Theta}|\mathbf{t}, \mathbf{d}, \mathbf{y}) \propto L(\boldsymbol{\Theta}|\mathbf{t}, \mathbf{d}, \mathbf{y})\pi(\boldsymbol{\Theta})$. It is not known in closed form for any of the choices of the baseline hazard function. Therefore, conditional on all information up to time J , posterior inference on a joint model with Markov Switching model can be done through a straightforward MCMC algorithm. Especially since they make the data augmentation scheme (Tanner and Wong 1987) of the states S_j uncomplicated.

MCMC methods are based on a sample of the posterior distribution, made possible with a Markov chain method. For this we need the full conditional distributions of all unknown quantities of the model.

For $V, \sigma^2, h_{0k}, r, \boldsymbol{\mu}, \eta_0, \eta_1, \boldsymbol{\beta}$ and Ω , the full conditional distributions are straightforward. We focus our comments on the states S_j , the transition probabilities p_{lm} of the transition matrix A in (5.4) and the link between the two sub-models λ_{ij} . To find the full conditionals, note that the prior distribution assumes independence between all parameters. This means that the dependence that matters for the full conditionals is in

the pseudo-likelihood function.

The state S_j appears in the likelihood function in the following way:

$$\prod_{i=1}^N \mathcal{N}(\lambda_{ij} \mid \mu_{s_j}, V) P(S_j = m). \quad (5.8)$$

Since these are the latent variables of the data augmentation, the prior for them is the model itself. This means that equation (5.8) is all that is needed for its sampling. Note that $P(S_j = m) = (\mathbf{q}'A^j)(m)$ provides \mathcal{K} states probabilities which are modified by $\mathcal{N}(\lambda_{ij} \mid \mu_{s_j}, V)$. Therefore we can obtain the full conditional probabilities from:

$$P(S_j = m \mid \cdot) \propto \prod_{i=1}^N \mathcal{N}(\lambda_{ij} \mid \mu_m, V) (\mathbf{q}'A^j)(m), \quad \mu_m \in M, \quad (5.9)$$

where $M = \{\mu_1, \dots, \mu_{\mathcal{K}}\}$. Note that the information from the data \mathbf{y} for these states come from $\boldsymbol{\lambda}_i, i = 1, \dots, N$.

The probabilities p_{lm} only appear in the likelihood in the terms $P(s_j = m)$. Since, in their full conditionals, the states $S_j, \forall j$ are known, then $(\mathbf{q}'A^j)(m) = (\mathbf{q}'A^j)(s_j)$ is a single term, which is a function of the probabilities in A . Then the joint full conditional of all probabilities in A is

$$\pi(\mathbf{p} \mid \cdot) \propto \prod_{j=1}^J (\mathbf{q}'A^j)(s_j) \pi(\mathbf{p}) \quad (5.10)$$

The above derivation is a direct result of the Markovian property of S_j . Other solution is provided by Hamilton's basic filter Hamilton and Susmel (1994). Some details of general MCMC algorithm for Markov switching stochastic volatility models are found in (So et al. 1998).

The mean trajectories λ_{ij} appear in the likelihood both in $L_i^l(\boldsymbol{\Theta}_l \mid \mathbf{y})$ and $L_i^s(\boldsymbol{\Theta}_s \mid \mathbf{t}, \mathbf{d})$. Like the states S_j , its prior is defined in the model itself. Then, their joint full conditional is

$$\pi(\lambda_{ij} \mid \cdot) \propto \mathcal{N}(\lambda_{ij} \mid \mu_{s_j}, V) \exp(d_i R_i) \exp\left(-\int_0^{t_i} \exp(R_i) h_0(u) du\right), \quad (5.11)$$

where $h_0(t)$ is the chosen baseline hazard function for the survival sub-model. The term R_i is as defined in (5.6).

5.4 Application

One dataset was used as example for the model. It consists of a diarrhea study with Brazilian children in the state of Bahia. We analyze a drop in feces volume which should indicate a health improvement.

For simplicity, in this application, we assume two states and do not incorporate the fragility term. The model was implemented with relatively vague priors for all model parameters. The prior for $q_1 \sim \text{Beta}(0.1, 0.1)$, $p_{11} \sim \text{Beta}(0.1, 0.1)$ and $p_{22} \sim \text{Beta}(0.1, 0.1)$, while the hyperparameters related to σ^2 and V are $a_{\sigma^2} = b_{\sigma^2} = a_V = b_V = 0.1$. For the mean trajectory $\boldsymbol{\mu}$, bridge effects $\boldsymbol{\alpha}$ and covariates effects $\boldsymbol{\beta}$, we assume independent normal priors with mean 0 and variance 100. Finally the hyperparameters for the priors relative to the baseline hazard function are defined depending on the chosen function. For $\boldsymbol{\Omega}$, when the semi-parametric baseline function is used, we chose $R_0 = 2$ and $g_0 = 0.1$. When the Weibull function is used, we chose $c = 0.01$. When the Gompertz baseline function is used, we chose $m_{\eta_0} = 0$, $C_{\eta_0} = 100$, $m_{\eta_1} = 0.1$, $C_{\eta_1} = 1$.

MCMC was run for 100,000 iterations, being the last 5,000 used as the posterior sample. Different starting values were tried as well as different MCMC burn-in lengths. In general the results were pretty much the same, with the chain converging, in practical terms, after 50,000 iterations. All computation was done in the free R software (R Development Core Team 2018) and MCMC computation was used with the Jags package (Su and Yajima 2015). In the Appendix Section A.5 we show some details related the algorithm. Convergence of Markov Chains was verified with the coda package (Plummer et al. 2006).

5.4.1 Children with episodes of diarrhea

This study is about episodes of diarrhea in the childhood. This dataset was collected by Federal University of Bahia, Bahia, in Serrinha, 170 km northwest of Salvador, capital of the state of Bahia, Brazil (between December/1990 and December/1991). It was originally analyzed in Barreto et al. (1994) and is available in Carvalho et al. (2012) (2012, p. 414). A new variable with event times was created on the original data in order for the proposed model to be used. The study was designed to determine whether there exists an effect of vitamin A supplementation on diarrhea and acute lower-respiratory-tract infections in young children in Brazil. We focus on the recovery of the diarrhea.

The childhood dataset includes 860 children in the age range of 6 – 48 months. The children were randomly assigned to treatment vitamin A (434 patients) or placebo (426) every 4 months for 1 year. They were followed up at home three times a week, and data about the occurrence and severity of diarrhea were collected. We understand the longitudinal marker by the target enteral volume (ml/kg/d) of children diagnosed with pediatric diarrhea according to the standard definition of 3 liquid or semi-liquid stools in 24 h. We understand the event of interest as the time until there is evidence of no diarrhea episode. These definitions were based on Barreto et al. (1994). Time-to-event information and censoring indicator are recorded along with a solitary baseline covariate - the allocated treatment arm in this instance.

In Figure 5.2 we can see the Kaplan-Meier empirical estimation of survival probability for both treatments. Both survival functions evolve separately until 40 weeks where they come closer at about 0.92 probability. After 60 weeks, the vitamin A treatment has a great survival decline, which may indicate that excessive treatment might be prejudicial to the patient. This could be a motivation to detect improvement sooner in order to avoid excessive treatment.

The objective of this joint model analysis is to take into account the effect of the dynamic trajectory of the mean severity of diarrhea on the time needed for the diarrhea to stop in children. Figure 5.3 shows the observed longitudinal measures plotted against time for the 860 patients included in the analysis. The heterogeneity of the patients’s immunological responses is apparent in this plot, both at every interview and between interviews. The most important feature of the plot for the model is the last 6 interviews, where there seems to be a shift of mean trajectory. This shift is meant to be modeled with the Markov Switch structure in order to answer the question whether the difference is relevant or not. Furthermore it is desired to know the magnitude of the difference and how much it impacts the hazard function.

The first 3 rows of Table 5.1 represent the proposed model with a 27-dimensional bridge, with baseline hazard function choices, respectively, Weibull, Gompertz and Piecewise exponential. The next 3 rows represent the same baseline hazard function, but with a 1-dimensional bridge choice $\xi_i = \frac{1}{27} \sum_j \lambda_{ij}$. This choice is easier to compare with other

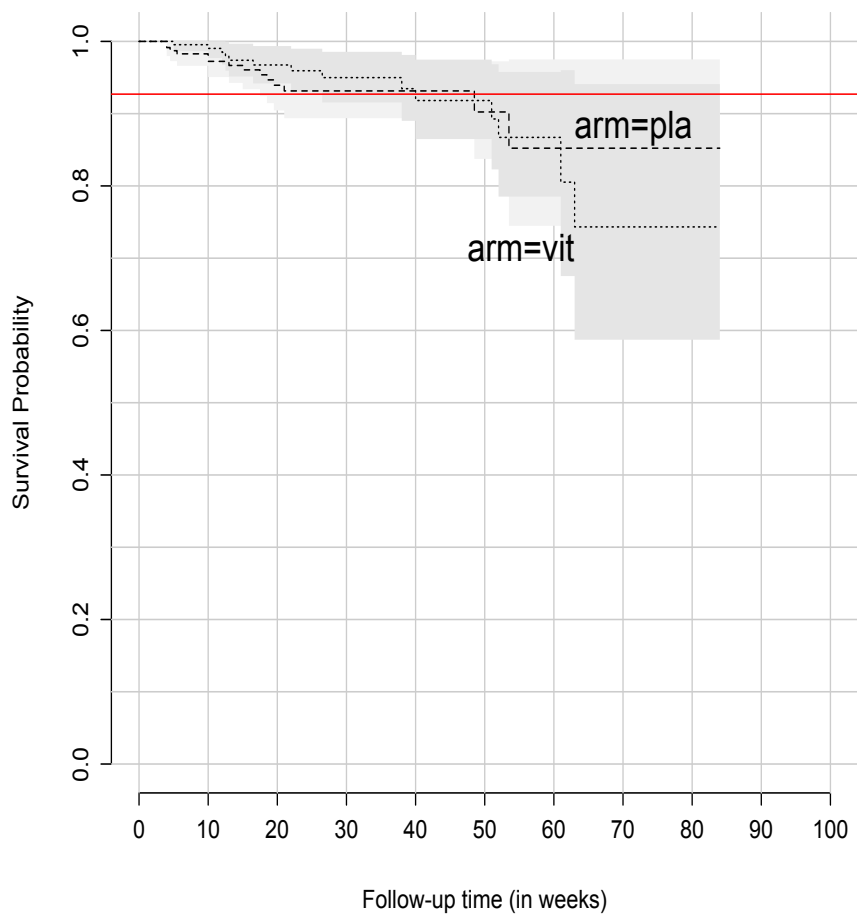


Figure 5.2: Empirical survival probability (Kaplan-Meier) for the childhood with episodes of diarrhea by treatments vitamin A (Vit A) and placebo (pla).

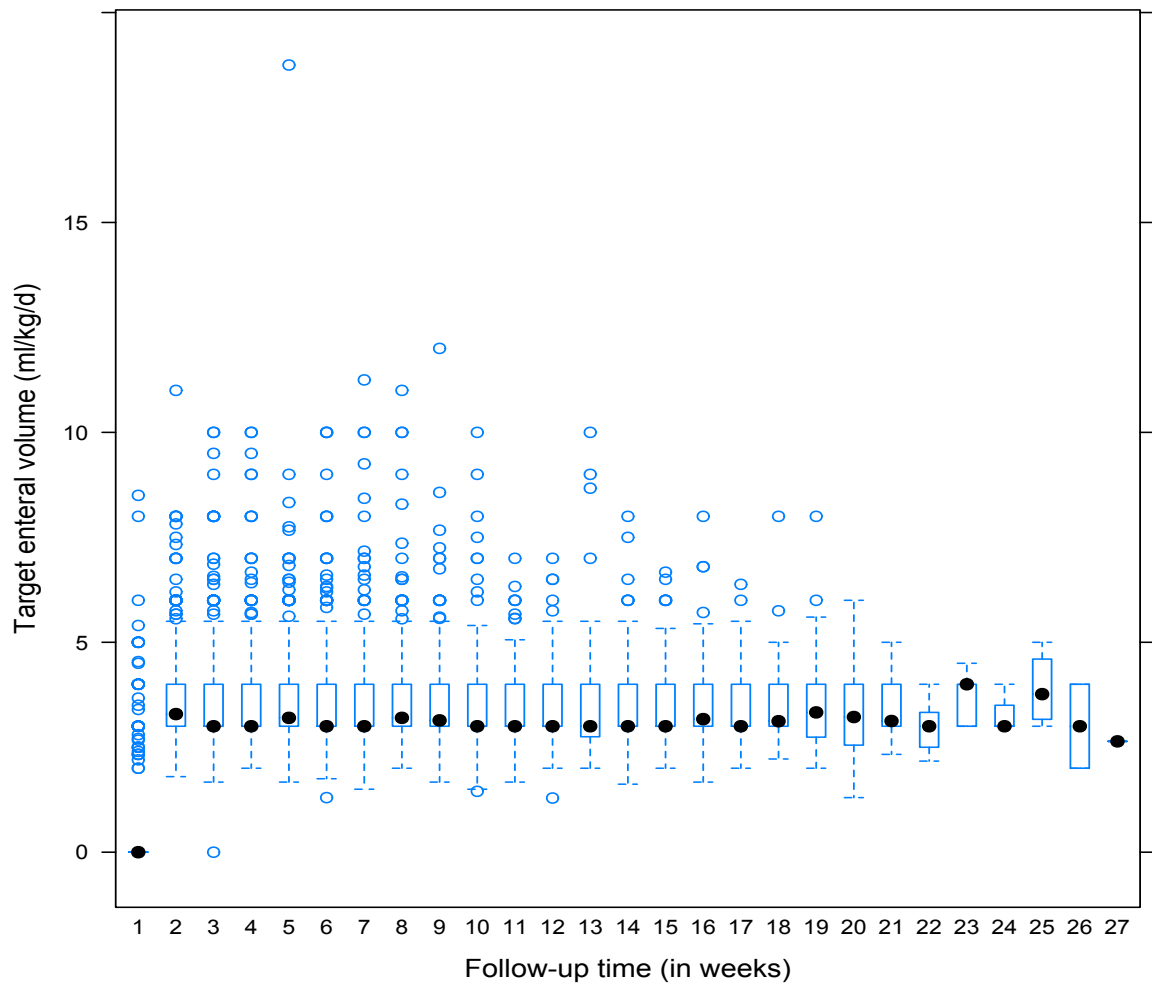


Figure 5.3: Boxplot of pediatric diarrhea target enteral volume (ml/kg/d) for 860 patients in 27 interview.

Md ^a	DIC	pD	LPML
<i>Weib</i> ₂₇	13080.358	845.019	-2976.659
<i>Gomp</i> ₂₇	13094.273	857.028	-2977.012
<i>PE</i> ₂₇	13096.534	860.002	-2977.017
<i>Weib</i> ₁	26464.021	49518.146	-12553.609
<i>Gomp</i> ₁	-42244.081	890.483	-2976.804
<i>PE</i> ₁	13205.538	841.095	-2976.922
<i>PE</i> _{Rizo}	19976.151	910.214	-10255.350

Table 5.1: Comparison of Bayesian model selection methods.

^aAbbreviations: *LPML*, Logarithm of the Pseudomarginal Likelihood; *DIC*, Deviance Information Criterion; *pD*, DIC component.

models in the literature, such as the one in (Rizopoulos and Ghosh 2011), which is represented in the last row of the table.

Note that the semi-parametric baseline hazard function is a generalization of the Gompertz function, since for the latter $h_0(t) = \exp(\eta_0 + \eta_1 t)$ and for the former, the equation (5.2) represents a DLM with linear growth, which means that $h_0(t)$ has log-linear growth, just as in the Gompertz function. The equivalence is achieved when the variance in (5.2) is zero.

The preferred comparison metric is *LPML*, since the main objective of the model is prediction of the event time, that is, the full recovery of the patient in this case. In this sense, the preferred models are *Weib*₂₇, *Gomp*₁ and *PE*₁. However, since the baseline hazard function of the *PE*_{Rizo} model is the same as *PE*₁, we select this model for the analysis. Note that this model has preferred metrics when compared to the benchmark *PE*_{Rizo} model. Both models estimate the same baseline hazard function, which can be seen in Figure 5.4.

To further compare the chosen model with the benchmark *PE*_{Rizo} model, the marginal posterior distribution of the effects of the bridge term from both models are shown in Figure 5.5.

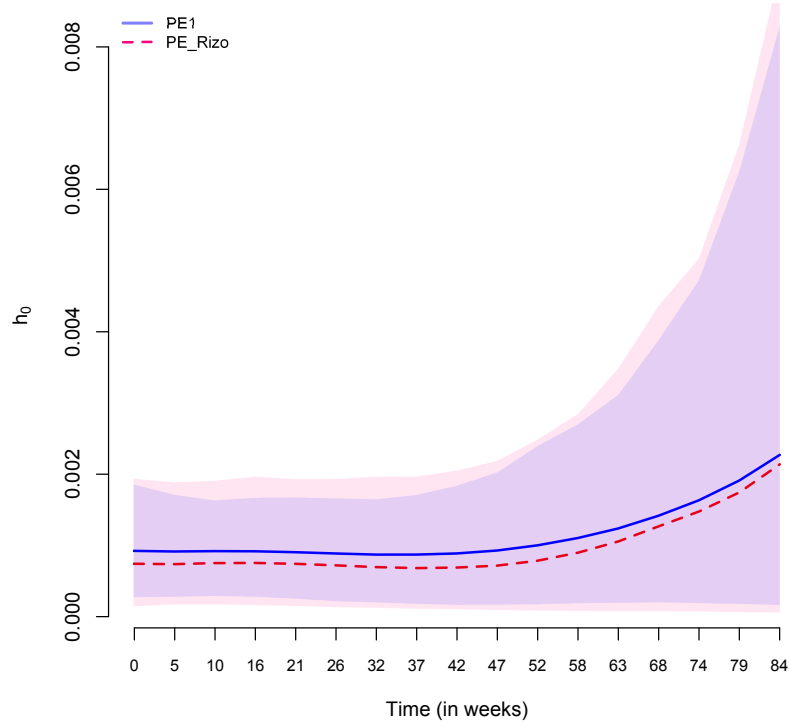


Figure 5.4: Baseline hazard function by interval k : estimated using the Rizopoulos model Rizopoulos and Ghosh (2011) under SRE approach (PE_{Rizo}), (dashed) and our proposal (solid line with credible intervals, PE_1)

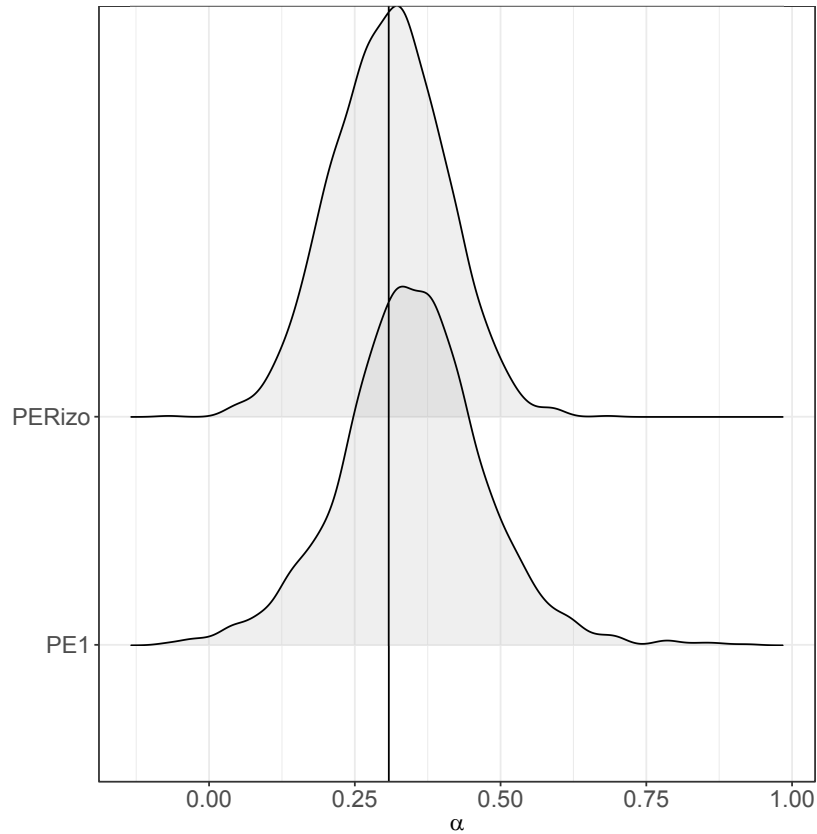


Figure 5.5: Posterior density of the effect bridge estimated using the Rizopoulos model Rizopoulos and Ghosh (2011), (dashed line, PE_{Rizo}) and our proposal (solid line, PE_1) with their credible intervals.

We see that they are very close from one another. The vertical line represents the posterior mean in the benchmark model (0.308), which shows that the marginal posterior for the parameter in the proposed model is slightly higher (0.34) than the benchmark model.

Relevant information output from the model are the transition probabilities and Markov state levels. This is shown in Figure 5.6 through the marginal posterior distributions for p_{11} , p_{22} , μ_1 and μ_2 . Remember that $p_{12} = 1 - p_{11}$ and $p_{21} = 1 - p_{22}$. We can see that the state 1, that is level close to 3 (posterior mean of μ_1), consists of most of the states. The state 2 happens more seldom with higher mean value, about 4.25 (posterior mean of μ_2), and the mean trajectory returns to state 1 with high probability. Note that, although seldom, state 2 represents higher chance of recovery, seen through effect α . That means that when the state goes to 2, then there is a $\exp(0.34 * (4.25 - 3)) \approx 1.53$ increase of the hazard function, i.e., when the state goes to 2, there is a 53% increase chance of recovery. On the other hand, when the state returns to 1, there is a 34.6%

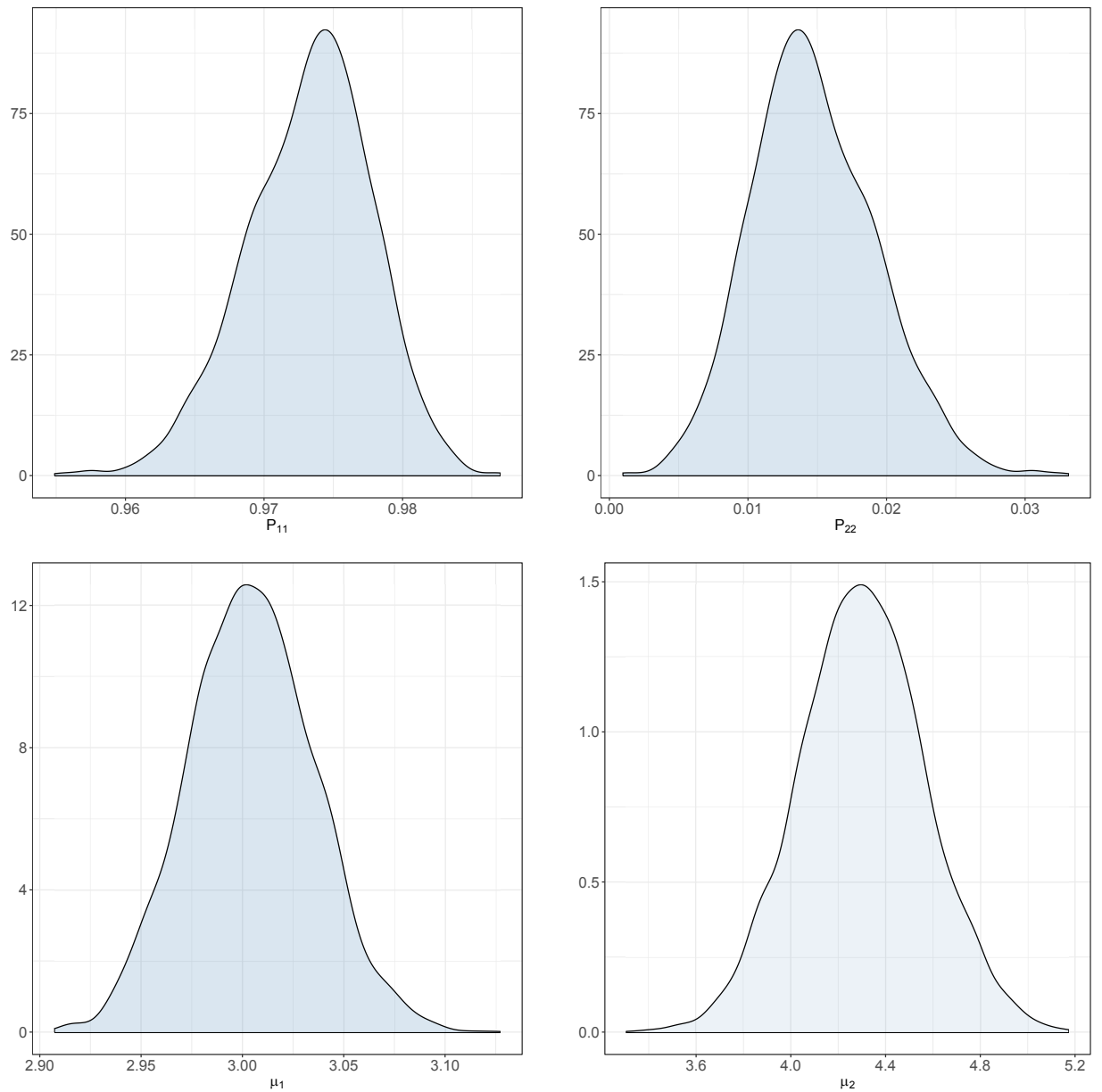


Figure 5.6: Posterior density of the parameters estimated using our proposal (PE_1). Where P_{11} and P_{22} are the transition probability. The densities for μ_1 and μ_2 represent the level trajectory.

$(\exp(0.34 * (3 - 4.25)) - 1 \approx -0.346)$ decreased chance of recovery.

5.5 Conclusion

An extension of the model of Chapter 3 was proposed with a Markov Switching mean trajectory. The general theory was developed for any state dimension with a special case of 2 states. The latter was exemplified with a case study of diarrhea in small children. The model was developed with the intention of being a complementary tool for diagnosis, since it could detect an improvement of the patient's health not necessarily seen directly from the measurements. This could help avoid an aggressive treatment when it is not necessary. Another longitudinal measurement which the proposal intends to model is quality of life, which is known to affect patient's health. An example was not provided due to lack of available data.

Although a simulation study was not performed for this work, we are confident the model can lead to useful results due to the analysis of the example application in the text. In it, two states are clearly identifiable in the model and the longitudinal variable has a non-zero impact on the hazard function which measures the recovery rate.

Clearly the model can be improved by adding covariates in the longitudinal sub-model, as well as a more complex mean evolution. The goal of this work is only to present the Markov Switching states as a viable option in itself, but the reader can easily add more components to the model as they see fit, which is encouraged by us. This added complexity justifies a simulated study which could yield interesting proposals.

The literature of joint models has grown fast recently, and the proposed model is a small contribution to the already large array of options available. The ideas presented here are meant to be used by the reader in the most convenient form for them, although the theory is already developed for a dynamic evolution of the longitudinal marker here.

Chapter 6

Conclusions and future work

In this thesis, we worked in the context of joint models of longitudinal and time-to-event data. We note that the nature of the data is closely related to an already established field in the Statistical literature, namely (generalized) hierarchical dynamic linear models (GHDLM). We use this idea as a starting point to explore the behavior of other components of the joint model. This led to a discussion about what joins the two sub-models, namely longitudinal and survival. Since there are two clearly separate sub-models that are linked, some model comparison metrics become useful to assess each sub-model's part in the whole model. Ultimately, the proposed models are very flexible to address various research questions with different applications. Finally we have proposed an easy and elegant way of modeling the longitudinal markers and time-to-event data.

The main innovation of this work is with respect to modeling the longitudinal markers. In order for the class of GHDLMs to be used, some care must be taken about aspects of the model, such as, not all measurements are evenly distributed over time. Once these details are taken care of, then the model becomes very flexible and even recovers the mean trajectory of most benchmark models in the literature, as verified by sample applications. The general version of the model presented in equation (1.3) provides even more flexibility. In this thesis, some extensions to joint models have been adapted to the proposed GHDLM. It is clear that other extensions for other models can be as easily adapted.

Once the longitudinal structure has been proposed, some exploration of the baseline hazard function became attractive. Among the options, the semi-parametric function

brought special attention since the values evolution could be described in more than one way. A dynamic model can be argued in the log scale, but a Markovian structure can be achieved without transformation via the Gamma process.

An important discussion that emerged over the development of this thesis was about the linking component between the two sub-models, which we called the bridge component. Many choices for this component are possible and over the thesis we tried to convey the notion that the ideal choice is whatever is most natural for the problem at hand. Some example choices are present in the thesis, such as mean trajectory g -transform (i.e. the linear predictor λ_{ij}) or some quantile of the longitudinal measurements' distribution. There are likely options not portrayed in this thesis and we would like to encourage researchers to create their own, such that their work is best represented by the model. Despite the flexibility provided by the bridge term in the proposed class of models, a formal comparison metric which considers the implications of different choices has not been developed. It could be done in a future work by analyzing the information added and lost by each choice in the hazard function. Note that this metric is different from the ones developed in chapter A.2.

In the introduction of this thesis we noted the importance of this choice. A future work proposition is a thorough analysis of implications of different choices.

Lastly, but not less importantly, we base ourselves on Zhang et al. (2017) to create metrics that decompose the well known metrics DIC and $LPML$ to show the contribution of the longitudinal sub-model to the metrics. This allowed us to compare similar models with, say, different bridge choices, which showed different contribution metrics.

Chapter 2 presents the dynamic structure for the longitudinal markers, through both multivariate and hierarchical approaches. The hierarchical approach is called GHDLM and is the starting point of the rest of the thesis. It introduces a special case of the hierarchical dynamic model developed by Gamerman and Migon (1993). This case includes a locally and a globally evolving mean, which makes it easy to compare with other models as they use a globally evolving mean through a polynomial. The normality

assumption common in joint models is relaxed and replaced by an exponential family assumption. This is illustrated with the analysis of CD4 cells counts in an AIDS dataset also analyzed in Guo and Carlin (2004) through a square root transformation. We maintain the original scale of the data.

Chapter 3 focuses on the event time sub-model. Based on our proposal in the previous chapter we shift the parametric assumption of the time-to-event data into the semi-parametric assumption. It is natural to assume that the baseline hazard function is time dependent, which we model through a Markovian structure. It is compared with the standard model (Rizopoulos 2011) through a liver cirrhosis dataset application.

Chapter 4 offers a new way to think of the problem. For this we remove the parametric distribution for the longitudinal measurements and provide a tool to deal with, say, non-symmetric or heteroscedastic data. For this, quantile regression is used and we can keep the hierarchical dynamic structure of the model. A discussion is provided about the appropriate choice of the quantile used. The model is applied to the liver cirrhosis and Schizophrenia datasets which show non-symmetric longitudinal measurements. This is yet another tool in the set provided in this thesis.

Chapter 5 develops the cases where there are multiple states for the longitudinal markers. This is particularly useful when analyzing quality of life impact on event times. The theory is presented and a special case with two states is provided. An example application with data from the state of Bahia in Brazil is analyzed, where the early identification of the states may help avoid a potentially harmful over-treatment of the disease.

We used the Jags package in the R software to reduce the computational time of the MCMC in the example applications. But for the MCMC algorithms relative to the simulations a code was developed and implemented to sample parameters from the Metropolis-Hastings algorithm by block, and thus the convergence of the chain is fast. The MCMC output of multiple chains was used to assess the convergence using the Plummer et al. (2006) package. We note that the chains were very similar to one another. All calculations were done with an Intel® i5 2300 Windows PC with 4GB memory. All

code will be available in GitHub (under construction).

In the future, we can extend the simulated results, although we know that it will be a superlative effort to cover every detail that the complexity of the model can reach. The first direct extension is to use the general model as detailed in A.1. Also, some existing developments in the literature can be included in the proposal, such as Gaussian Processes for the continuous over time longitudinal trajectory as in Henderson et al. (2000).

Many technological advances have been done recently, such as the software STAN and the package Martins et al. (2013). An R package using either or both of these solutions should be done to make it easier for use of the proposed methodologies.

This proposed methodology is not restricted to the analysis of datasets in the medical area. It can also be useful for other types of areas such as education, engineering, finance, sociology. Some context can be provided for these fields in the future in order to motivate the model's use.

Also, in this thesis, a censorship process was not considered in the model, which can be done in the future.

We would also like to explore further properties of quantile regression in joint models to probe all its usefulness.

An issue which needs to be looked at, with a similar line of thought that led to the quantile regression for the longitudinal sub-model, is the residual life approach for the time-to-event data.

Despite all the future work that needs to be done, we believe that the approach presented in this thesis has great potential and can be easily used with good performance. Its strongest point is the treatment of time varying data with a time sensitive model, which, surprisingly, has not been done as carefully as in this thesis. This fact alone gives us

confidence that a meaningful contribution to science can be found in this text.

Appendix A

Appendix

A.1 The general model in detail: the longitudinal component

The framework of this thesis is the joint model, but the effort revolves mainly around the longitudinal sub-model. As was presented in the Introduction, equation (A.1) shows how the general sub-model (1.3) relates to the traditional linear mixed effects model (1.2), written in a hierarchical form in equation (A.2), with $\boldsymbol{\theta}_{ij} = (\mu'_{ij}, \boldsymbol{\beta}'_2)'$ where $\mu_{ij} = (\mu_{1ij}, \mu_{2ij}, a_{1i})'$ with $\mu_{i0} = \mu_0, \mu_0 \sim \mathcal{N}(m_0, C_0)$, $\mathbf{F}_i = (1, 0, 0, X'_{2i})'$ and

$$G = \begin{pmatrix} 1 & \delta_j & \delta_j & \mathbf{0} \\ 0 & 1 & 0 & \mathbf{0} \\ 0 & 0 & 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix}.$$

For simplicity's sake, we focus only on the evolution of the three first components of $\boldsymbol{\theta}_{ij}$, as we set

$$\mathbf{W}_j = \begin{pmatrix} W_1 & 0 & 0 & \mathbf{0} \\ 0 & W_2 & 0 & \mathbf{0} \\ 0 & 0 & 0 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Therefore the evolution equations are given by

$$\begin{aligned} \lambda_{ij} &= \mu_{1ij} + \mathbf{X}'_{2i} \boldsymbol{\beta}_2 + v_i, & v_i &\sim \mathcal{N}(0, V) \\ \mu_{1ij} &= \mu_{1ij-1} + \delta_j (\mu_{2ij} + a_{1i}) + w_{1j}, & w_{1j} &\sim \mathcal{N}(0, W_1) \\ \mu_{2ij} &= \mu_{2ij-1} + w_{2j}, & w_{2j} &\sim \mathcal{N}(0, W_2). \end{aligned} \tag{A.1}$$

Observe that $\delta_j(\mu_{2ij} + a_{1i})$ represents the linear level growth for individual i , which is basically a common linear growth $\delta_j\mu_{2ij}$ plus the random effect $\delta_j(a_{1i})$. When μ_{2ij} is the same at every interview time j , then the model has a mixed effects linear growth. For the random intercept effect, all that is needed is a random vertical offset of the trajectory per individual, achieved through the term v_i in (1.3).

For reference, the standard true trajectory model that is recovered by the description above is

$$\begin{aligned}
y_{ij} &= m_i(t) + \mathbf{X}'_{2i}\boldsymbol{\beta}_2 + \epsilon_{ij} \\
m_i(t) &= (a_0 + a_{0i}) + (a_1 + a_{1i})t \\
\begin{pmatrix} a_{0i} \\ a_{1i} \end{pmatrix} &\sim \mathcal{N}\left(\mathbf{0}, \begin{pmatrix} \Sigma_{a_{0i}} & 0 \\ 0 & \Sigma_{a_{1i}} \end{pmatrix}\right)
\end{aligned} \tag{A.2}$$

Another option for a general linear growth model is

Observation equation:

$$y_{ij} \sim Ef(\eta_{ij}, \psi), \quad E(y_{ij} | \eta_{ij}, \psi) = g^{-1}(\lambda_{ij}),$$

Structural equations:

$$\begin{aligned}
\lambda_{ij} &= \mathbf{F}'_{1i}\boldsymbol{\theta}_{ij} + v_{1i}, \quad v_{1i} \sim \mathcal{N}(0, V_1) \\
\boldsymbol{\theta}_{ij} &= \mathbf{F}'_{2i}\boldsymbol{\theta}_j + v_{2ij}
\end{aligned} \tag{A.3}$$

System equation:

$$\boldsymbol{\theta}_j = \mathbf{G}_j\boldsymbol{\theta}_{j-1} + \mathbf{w}_j, \quad \mathbf{w}_j \sim \mathcal{N}(\mathbf{0}, \mathbf{W})$$

Initial information:

$$\boldsymbol{\theta}_0 | \mathcal{D}_0 \sim \mathcal{N}(\mathbf{m}_0, \mathbf{C}_0),$$

where $v_{2ij} = \delta_j v'_{2i}$ and $v'_{2i} \sim \mathcal{N}(0, V_2)$. Then, if $\mathbf{F}'_{1i} = 1, \forall i$ then $\lambda_{ij} = \mathbf{F}'_{2i}\boldsymbol{\theta}_j + \delta_j v'_{2i} + v_{1i}$ which constitutes the equivalence with the mixed effects linear growth model, if we choose $\mathbf{F}'_{2i} = (1, 0)'$, $\boldsymbol{\theta}_j = (\mu_{1j}, \mu_{2j})'$ and $\mathbf{G}_j = \begin{pmatrix} 1 & \delta_j \\ 0 & 1 \end{pmatrix}$.

This alternative represents the linear growth model but another polynomial growth form may require more hierarchy levels.

A.2 Supplementary material for chapter 2

A.2.1 Revisiting Some Special Cases in the Literature

Some well-known models in the current literature are special cases of our general structure. The first special case is the model developed by Guo and Carlin (2004), which consists of a Bayesian version of the joint model proposed by Henderson et al. (2000). The stochastic dependence is described by a latent stationary zero mean bivariate Gaussian process shared by both the longitudinal and time-to-event processes.

Specifically, Guo and Carlin (2004) define the vector $\mathbf{u}'_i = (u_{1i}, u_{2i})$ following a Gaussian process, i.e., $\mathbf{u}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{U})$. They apply the mixed linear normal model to the longitudinal component with latent part given by $W_{1i}(t) = u_{1i} + u_{2i}t$. To complete the joint modeling, the hazard risk for time-to-event model is given by $h_i(t) = h_0(t) \exp(\mathbf{X}'_{1i}\boldsymbol{\gamma}_1 + W_{2i}(t))$, where $W_{2i}(t) = \zeta_1 u_{1i} + \zeta_2 u_{2i} + \zeta_3 W_{1i}(t) + u_{3i}$, with u_{3i} denoting an independent frailty term, modeled as iid $\mathcal{N}(0, U_3)$ variables, independent of \mathbf{u}_i , for $i = 1, \dots, N$.

Based on our proposal under the regular situation, consider the normal univariate second-order dynamic polynomial model with known $\mathbf{F}'_i = (1, 0, \mathbf{X}_{2i})$, transition matrix \mathbf{G}_j as $L_2(\delta_j)$ (note that $\delta_{ij} = \delta_j, \forall i$, which means that we can write $L_p(\delta_{ij})$ as $L_p(\delta_j)$), where the elements are based on δ_j for all individuals. If $\mathbf{W} = \mathbf{0}$, then the latent parameter related to the evolution equation is $\theta_j = m_1 + m_2j$ and the disturbance term is $v_i(j) = u_{1i} + u_{2i}j$. Here we define v_i as a function of the follow-up counter j instead of using j as an index since, if we consider that $\lambda_{ij} = \mathbf{F}'_i\boldsymbol{\theta}_j + v_i(j)$, then the stochastic increment between two successive interviews is $v_i(j) - v_i(j-1) = u_{2i}$, which does not depend on j , that is, we remain in the proposed class of models. It follows that the structure equation, which is the equation for λ_{ij} in (2.3) is given by $\lambda_{ij} = m_1 + u_{1i} + (m_2 + u_{2i})j$. This term is widely known as the mean trajectory and this situation is comparable to equation (1) for the longitudinal data model in Guo and Carlin (2004). We can interpret this as a static model in the sense that the coefficient is parameterized as a fixed linear function of time, where the term $v_i(j)$ becomes equivalent to $W_{1i}(t)$, so that a global linear trend component is obtained.

Finally, take our survival sub-model in (1) where the components $\mathbf{Z}_i = (\mathbf{X}'_{1i}, \boldsymbol{\xi}'_i)'$, and $\boldsymbol{\xi}'_i = (u_{1i}, u_{2i}, v_i(1), \dots, v_i(J), 1)$ and $\boldsymbol{\gamma}'_i = (\boldsymbol{\beta}'_1, \zeta_1, \zeta_2, \boldsymbol{\zeta}'_3, u_{3i})$, where $\boldsymbol{\zeta}_3$ is a J -dimensional vector with components $v_i(1), \dots, v_i(J)$. Furthermore, we need to make the disturbance component $\boldsymbol{\Sigma}_\zeta = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{u_{3i}} \end{bmatrix}$ to lose the full hierarchy considered in our model and to describe the frequently used regression model to model the time-to-event data with an exception of the term of frailty u_{3i} . In this way, we realize that the model in Guo and Carlin (2004) is a special case of our proposal.

The second special case is related to the *SRE* class of models developed by Rizopoulos (2012, Chapter 4), who also considers the mixed regression model for the longitudinal part. The time-to-event sub-model is defined through the relative risk $h_i(t) = h_0(t) \exp(\mathbf{W}'_i \boldsymbol{\gamma}_R + m_i(t) \alpha_R)$, where $m_i(t)$ refers to the mean trajectory and this term is included as the linear predictor in the relative risk function. The subscript “ R ” is included to indicate parameters of Rizopoulos’ model. To model the longitudinal component, consider the latent structure equation $\boldsymbol{\xi}_i = \boldsymbol{\lambda}_i$, equivalent to $m_i(t)$. Then, we get a mixed linear function of time based on the second-order polynomial DLM. For the time-to-event model, consider the term $\mathbf{Z}'_i = (\mathbf{X}'_{1i}, \boldsymbol{\xi}'_i)$ with $\mathbf{X}_{1i} = \mathbf{W}_i$ and $\boldsymbol{\beta}_{1i} = \boldsymbol{\gamma}_R$, while $\boldsymbol{\alpha}$ in (1) is a univariate effect of the link component α_R . Note that the hierarchical feature is not considered in Rizopoulos’ survival sub-model, which means that we need to set the disturbance component $\boldsymbol{\Sigma}_\zeta = \mathbf{0}$ because our model encompasses a hierarchical regression to model the time to event. Therefore, we conclude that the standard model developed in Rizopoulos (2012, Chapter 4) can be rewritten as a special case of our proposal.

Once we define $h_i(t) = h_0(t) \exp(\mathbf{X}'_{1i} \boldsymbol{\beta}_{1i} + \boldsymbol{\xi}'_i \boldsymbol{\alpha})$ and set $\mathbf{Z}'_i = (\mathbf{X}_{1i}, \boldsymbol{\xi}_i)$, where $\boldsymbol{\xi}_i = \mathbf{u}_i$ is a vector of random effects as in the first special case and $\boldsymbol{\alpha}$ is a two-dimensional vector capturing the effect of the $\boldsymbol{\xi}_i$, we get a special case equivalent to the model developed in Philipson et al. (2017). In this context, this model is identical to the model in the *stjm* Stata package based on Crowther et al. (2013). Furthermore, in the three presented cases the longitudinal component is modeled using the mixture random effects model.

A.2.2 Dynamic Hierarchical Joint Model for the Schizophrenia Dataset

The observations y_{ij} are given by the PANSS score and they describe the usual Gaussian DLM form, with mean trajectory λ_{ij} for each patient with evolution term θ_j and disturbance term v_i , for $i = 1, \dots, 150$ and $j = 1, \dots, 6$. We use a $p = 2$ degree polynomial. The time-to-event parametric model (1) is used and we assume right-censored times with $\mathbf{X}'_{1i} = (1, Treat_i)$ and the latent term $\boldsymbol{\xi}'_i = (\lambda_{i1}, \dots, \lambda_{iJ})$. For the longitudinal sub-model, $\mathbf{F}'_i = (1, 0, Treat_i)$ and with follow-up times $\boldsymbol{\delta}' = (0, 1, 1, 2, 2, 2)$. The prior distributions are $r \sim Exp(0.1)$, $1/\sigma^2 \sim IG(0.01, 0.01)$, $1/V \sim IG(0.01, 0.01)$, $1/W_1 \sim IG(0.01, 0.01)$, $1/W_2 \sim IG(0.01, 0.01)$, $\boldsymbol{\beta}_1 \sim \mathcal{N}(\mathbf{m}0, 100\mathbf{I}_2)$, $\boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{m}0, 100\mathbf{I}_6)$, and $\beta_2 \sim \mathcal{N}(0, 100)$, with $\boldsymbol{\Sigma}_\zeta = \mathbf{0}$, that is, without hierarchical effect in the survival component.

The term λ_{ij} denotes the structure equation to describe the filter term related to the PANSS score and follows a hierarchical DLM structure for each patient and follow-up time generated from a normal second-order polynomial DLM.

Note that by substituting the mean structure λ_{ij} into the observation equation, we obtain that for each patient and interview and for each score $PANSS_{ij}$, the mean structure regarding patient i is decomposed as the common mean $\mathbf{F}'_i\boldsymbol{\theta}_j = \mu_{j1} + Treat_i\beta_2$ determining the borrowed information and showing the strength among patients belonging to the same schedule j to better learn about the covariance structure of the elements of λ_{ij} .

Conditional on v_i , the elements of the score $PANSS_{ij}$ are independent and integrating the distribution of $PANSS_{ij}$ with respect to v_i , we obtain $y_{ij} \sim \mathcal{N}(\mathbf{F}'_i\boldsymbol{\theta}_j, \sigma^2 + V)$. It is directly calculable assuming that all processes errors ϵ_{ij} , v_i , and \mathbf{w}_j are mutually independent for all interviews.

On the other hand, the hazard function for patient i is $h_i(t) = rt^{r-1} \exp(\beta_{11} + Treat_i\beta_{12} + \sum_{j=1}^J \lambda_{ij}\xi_{ij})$. This function includes the effect of the latent true score λ_{ij} through $\boldsymbol{\alpha}' = (\alpha_1, \dots, \alpha_J)$. Furthermore, note that the two sub-models are linked by the term $\boldsymbol{\lambda}_i$ and

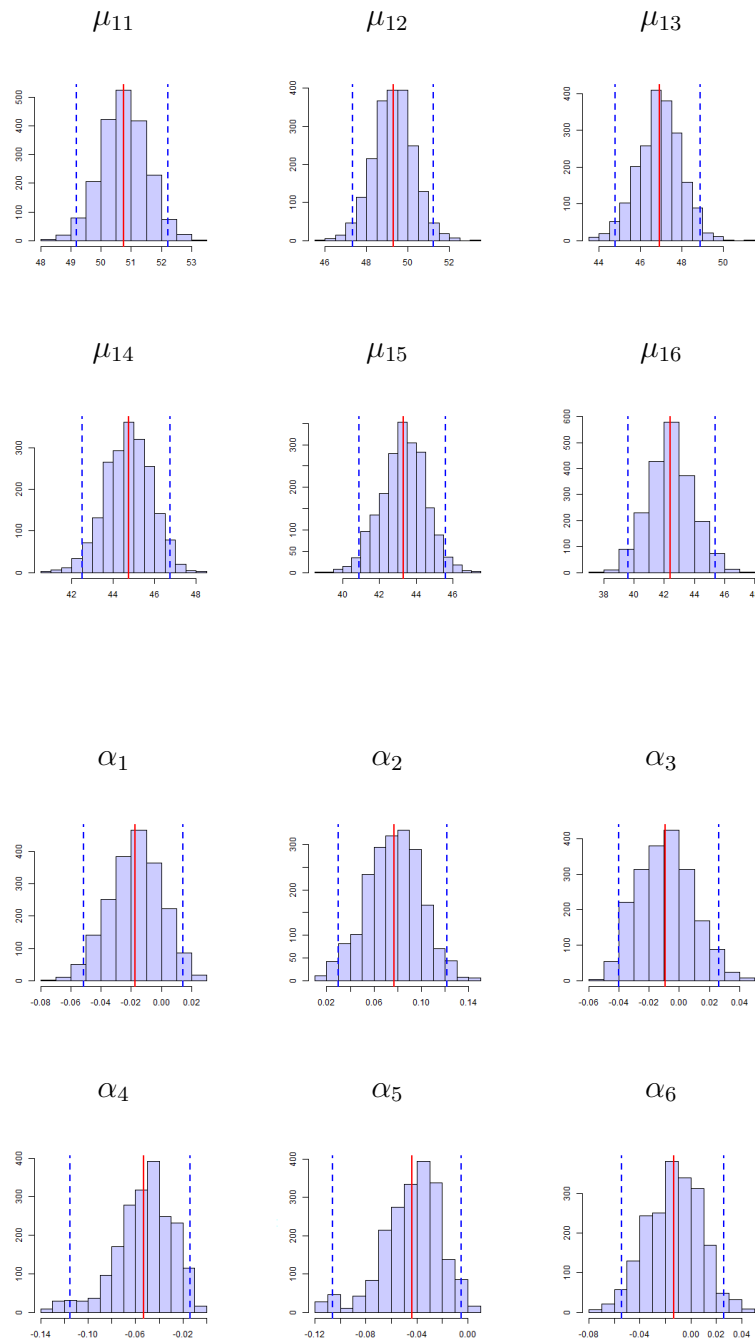
α quantifies the multiplicative effect of this term on the hazard function.

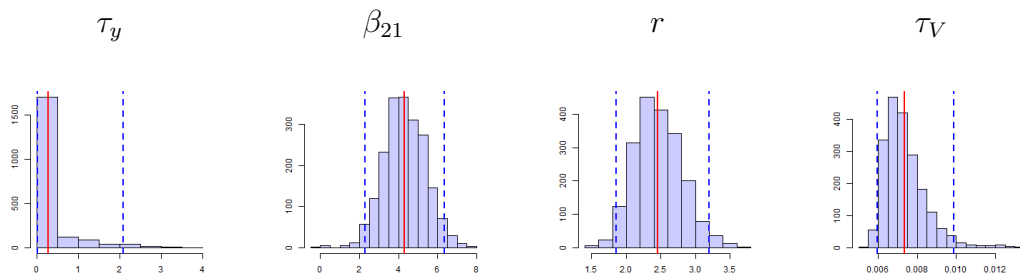
A.2.3 Decomposition of the Comparison Criteria

In order to evaluate the performance of the models, this section shows the comparison criteria to assess the contribution of the longitudinal data to the fit of the survival data. Specifically, we adapted the methodology in Zhang et al. (2017) to our proposal. Using the factorization $[t_i, d_i, \mathbf{y}_i \mid \boldsymbol{\xi}_i] = [t_i, d_i \mid \boldsymbol{\xi}_i][\mathbf{y}_i \mid \boldsymbol{\xi}_i]$, for $i = 1, \dots, N$, it is easy to decompose the Deviance Information Criterion as $DIC = DIC_{Surv, Long} = DIC_{Surv \mid Long} + DIC_{Long}$. The smaller the value of DIC, the better is the fit of the model. We adopt the model comparison criterion $\Delta DIC_{Surv} = DIC_{Surv \mid Long} - DIC_{Surv, 0}$, where $DIC_{Surv, 0}$ is the DIC value calculated ignoring the uncertainty in the longitudinal component. We see that ΔDIC_{Surv} captures the improvement of the fit in the survival component due to the longitudinal data with a penalty for the additional parameters in the survival component of the joint model. A model with a small value of ΔDIC_{Surv} is preferred, noticing that ΔDIC_{Surv} can be negative. The same procedure is used to compute the decomposition of the comparison criteria pD , then a small value of ΔpD_{Surv} is preferred.

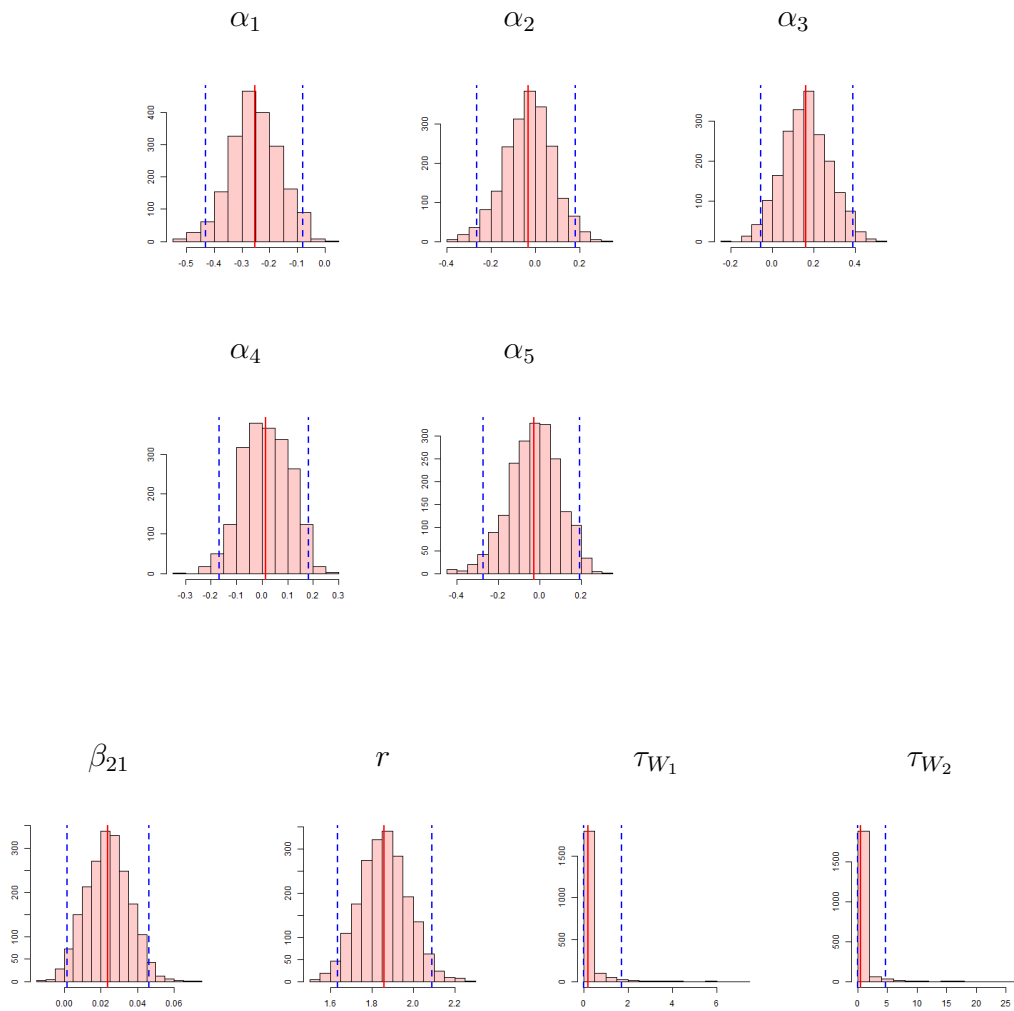
We also use the logarithm of the pseudo-marginal likelihood ($LPML$) as a criterion, which is based on the conditional predictive ordinate (CPO), as defined in Gelfand and Mallick (1995). Larger values of $LPML$ indicate a preferable model. We define the $LPML$ decomposition as $LPML = LPML_{Surv, Long} = LPML_{Surv \mid Long} + LPML_{Long}$. We present the model comparison criterion $\Delta LPML_{Surv} = LPML_{Surv \mid Long} - LPML_{Surv, 0}$, so that $\Delta LPML_{Surv}$ quantifies the gain in fit in the survival component due to the longitudinal data with a penalty for the additional parameters in the survival component of the joint model. A model with a large value of $\Delta LPML_{Surv}$ is preferred.

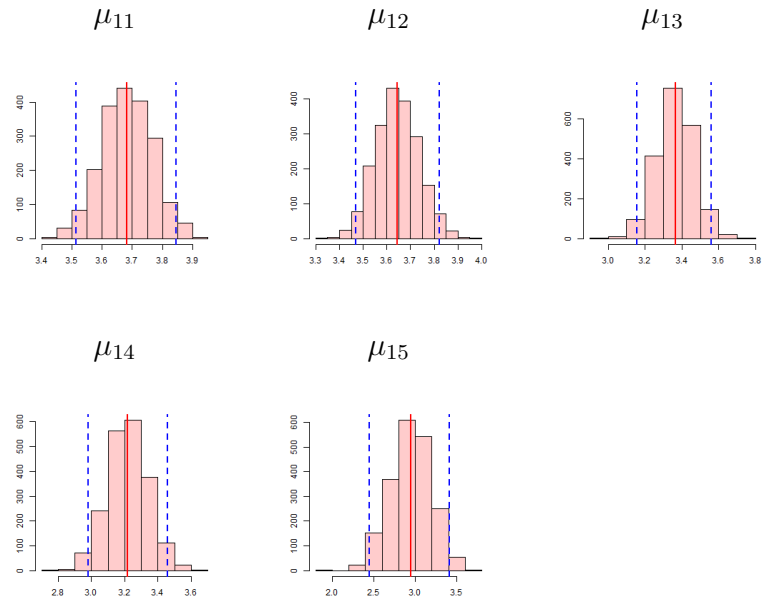
A.2.4 Posterior histogram for the parameters for the Schizophrenia Dataset model





A.2.5 Posterior histogram for the parameters for the Aids Dataset model





A.2.6 Simulated data generator for chapter 2

```

Sim.Nor.Weib <- function(N=N,J=J,p=p,x10=x10,x20=x20,W1=W1,W2=W2,V=V,sigma2=sigma2,
r=r,gamma=gamma,alpha=c(a1,a2,a3),mean.Cens=mean.Cens){
# longitudinal markers .....
x0 = c(x10,x20)      # initial values

y      = matrix(0,N,J)
mu     = matrix(0,N,J)
lambda = matrix(0,N,J)
v      = matrix(0,N,J)
x      = matrix(0,J,p)

w = cbind(rnorm(J,0,sqrt(W1)),rnorm(J,0,sqrt(W2)))

w[1,1] = W1
w[1,2] = W2

for(j in 1:J){
v[,j] = rnorm(N,0,sqrt(V))

```

```

}

# evolution
# .....
x[1,1] = x0[1] + x0[2] + w[1,1]
x[1,2] = x0[2] + w[1,2]

# structure
# .....
mu[,1] = x[1,1] + v[,1]
lambda[,1] = mu[,1]

# observation
# .....
y[,1] = rnorm(N,lambda[,1])

for (j in 2:J){
x[j,1] = x[(j-1),1] + x[(j-1),2] + w[j,1]
x[j,2] = x[(j-1),2] + w[j,2]

mu[,j] = x[j,1] + v[,j]

lambda[,j] = mu[,j]

y[,j] = rnorm(N,lambda[,j])
}

# simulate event times .....
gammas <- c(gamma,alpha)

Z <- cbind(1,mu)

```



```

eta.t <- as.vector(Z %*% gammas)
invS <- function (t, u, i) {
h <- function (s) {
exp(log(r) + (r - 1) * log(s) + eta.t[i])
}
integrate(h, lower = 0, upper = t)$value + log(u)
}
u <- runif(N)
trueTimes <- numeric(N)
for (i in 1:N) {
Up <- 50
tries <- 5
Root <- try(uniroot(invS, interval = c(1e-05, Up), u = u[i], i = i)$root, TRUE)
while(inherits(Root, "try-error") && tries > 0) {
tries <- tries - 1
Up <- Up + 200
Root <- try(uniroot(invS, interval = c(1e-05, Up), u = u[i], i = i)$root, TRUE)
}
trueTimes[i] <- if (!inherits(Root, "try-error")) Root else NA
}
na.ind <- !is.na(trueTimes)
trueTimes <- trueTimes[na.ind]

# simulate censoring times from an exponential distribution,
# and calculate the observed event times, i.e.,
# min(true event times, censoring times)
Ctimes <- runif(N, 0, 2 * mean.Cens)
Time <- pmin(trueTimes, Ctimes)
event <- as.numeric(trueTimes <= Ctimes) # event indicator

True.par <- as.data.frame(c(sigma2,V,W1,W2,r,gammas))

```

```

row.names(True.par)
<-c("sig2", "V", "W1", "W2", "r", "gamma", "alpha1", "alpha2", "alpha3")
names(True.par) <-c("Value")

Data.Sim <- as.data.frame(cbind(1:N,y,mu,trueTimes,Ctimes,Time,event))
names(Data.Sim)
<- c("id", "y1", "y2", "y3", "mu1", "mu2", "mu3", "trueTimes", "Ctimes", "Time", "event")

Simulado <- list(True.par=True.par,Data.Sim=Data.Sim)
return(Simulado)
}

Define
N          # num observations
J          # num interview
p          # grau do polynomio

x10
x20

W1          # var evolution level
W2          # var evolution incresing
V          # var structure
sigma2      # var observation

alpha      = c(a1,a2,a3)
mean.Cens # mean of the exponential distribution for the censoring mechanism

simulado.dat <- Sim.Nor.Weil(N=N,J=J,p=p,x10=x10,x20=x20,W1=W1,W2=W2,
V=V,sigma2=sigma2, r=r,gamma=gamma,alpha=c(a1,a2,a3),mean.Cens=mean.Cens)

```

A.3 Supplementary material for chapter 3

A.3.1 Bayesian Inference

For clarity of exposition, we consider the inference procedure for the dynamic hierarchical sub-model with a piecewise exponential hazard function for $h_0(t)$ defined in Chapter (3). In particular we detail the prior specification and posterior computation.

$$\begin{aligned}
 T_i \mid \boldsymbol{\xi}_i &\sim PE(\mathbf{h}_0(\cdot)) \\
 h_i(t) &= h_0(t) \exp\{\mathbf{Z}'_i \boldsymbol{\beta}_1 + \boldsymbol{\xi}'_i \boldsymbol{\alpha}\}, \\
 h_0(t) &= \begin{cases} h_{01} & \text{if } t \in [\tau_0, \tau_1) \\ \vdots \\ h_{0K} & \text{if } t \in [\tau_{K-1}, \tau_K) \end{cases}
 \end{aligned} \tag{A.4}$$

The likelihood function is

$$\begin{aligned}
 &L(\mathbf{h}_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\alpha}, \psi, \boldsymbol{\Lambda}, \boldsymbol{\theta}, V, \mathbf{W} \mid \mathbf{t}, \mathbf{d}, \mathbf{y}) \\
 &= \prod_{i=1}^N \left[\prod_{k=1}^K \exp(d_{ik}(\log h_{0k} + \mathbf{X}'_{1i} \boldsymbol{\beta}_1 + \boldsymbol{\xi}'_i \boldsymbol{\alpha})) \right] \\
 &\times \exp \left(- \exp(\mathbf{X}'_{1i} \boldsymbol{\beta}_1 + \boldsymbol{\xi}'_i \boldsymbol{\alpha}) \sum_{\substack{k=1 \\ \tau_k \leq t_i}}^K h_{0k}(\min\{\tau_k, t_i\} - \tau_{k-1}) \right) \\
 &\times \prod_{i=1}^N \mathcal{N}(\boldsymbol{\theta}_0 \mid \mathbf{m}_0, \mathbf{C}_0) \prod_{j=1}^J b(y_{ij}, \psi) \exp \left\{ \phi(y_{ij} g^{-1}(\lambda_{ij})) - a(g^{-1}(\lambda_{ij})) \right\} \exp \left\{ \frac{-(\lambda_{ij} - \mathbf{F}'_i \boldsymbol{\theta}_j)^2}{2V} \right\} \\
 &\times V^{1/2} \mathcal{N}(\boldsymbol{\theta}_j \mid \boldsymbol{\theta}_{j-1}, \mathbf{W}) f_h(\mathbf{h}_0)
 \end{aligned} \tag{A.5}$$

Write $\mathbf{h}_0 = (h_{01}, \dots, h_{0K})'$, f_h represents the time dependence structure chosen for the piecewise baseline hazard function and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_J)$. For the simulated data presented in Section 3.4, where the responses y_{ij} are Poisson distributed with rate λ_{ij} , then the

likelihood function reduces to:

$$\begin{aligned}
& L(\mathbf{h}_0, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\alpha}, \psi, \boldsymbol{\Lambda}, \boldsymbol{\theta}, V, \mathbf{W} \mid \mathbf{t}, \mathbf{d}, \mathbf{y}) \\
&= \prod_{i=1}^N \left[\prod_{k=1}^K \exp(d_{ik}(\log h_{0k} + \mathbf{X}'_{1i}\boldsymbol{\beta}_1 + \boldsymbol{\xi}'_i\boldsymbol{\alpha})) \right] \\
&\times \exp \left(- \exp(\mathbf{X}'_{1i}\boldsymbol{\beta}_1 + \boldsymbol{\xi}'_i\boldsymbol{\alpha}) \sum_{\substack{k=1 \\ \tau_k \leq t_i}}^K h_{0k}(\min\{\tau_k, t_i\} - \tau_{k-1}) \right) \\
&\times \prod_{i=1}^N \mathcal{N}(\boldsymbol{\theta}_0 \mid \mathbf{m}_0, \mathbf{C}_0) \prod_{j=1}^J \exp\{-\lambda_{ij}\} \frac{\lambda_{ij}^{y_{ij}}}{y_{ij}!} \exp\left\{-\frac{-(\lambda_{ij} - \mathbf{F}'_i\boldsymbol{\theta}_j)^2}{2V}\right\} \\
&\times V^{1/2} \mathcal{N}(\boldsymbol{\theta}_j \mid \boldsymbol{\theta}_{j-1}, \mathbf{W}) f_h(\mathbf{h}_0)
\end{aligned} \tag{A.6}$$

We assume independent normal priors for $\boldsymbol{\beta}_2$, $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}_1$, that is $\mathcal{N}(\boldsymbol{\mu}_p, \boldsymbol{\Sigma}_p)$, where $\boldsymbol{\mu}_p$ and $\boldsymbol{\Sigma}_p$ (a positive definite matrix) for $p = \boldsymbol{\beta}_2, \boldsymbol{\alpha}, \boldsymbol{\beta}_1$, are prespecified hyperparameters. When the distribution of the exponential family is Gaussian then ψ represents the distribution variance. Then an inverse gamma (IG) prior is specified for ψ , i.e., $\psi \sim \text{IG}(a_0, b_0)$, where $a_0 \geq 0$ and $b_0 \geq 0$ are prespecified. For \mathbf{W} , we take an Inverse-Wishart (IW) prior with known parameters \mathbf{R}, g , where \mathbf{R} is a $(2+q_2) \times (2+q_2)$ positive definite matrix and g is the degrees of freedom. For the parameter $V \sim \text{IG}(a_1, b_1)$, where a_1, b_1 are prespecified as well.

By the Bayes theorem we obtain the posterior distribution which is not known in closed form. We use Markov Chain Monte Carlo techniques to draw samples from the posterior distribution.

We use the Metropolis within Gibbs algorithm. The parameters vector $\boldsymbol{\beta}_2, \boldsymbol{\theta}_j$ and V can be sampled from its full conditional distribution, namely inverse gamma for V and Normal for the other two.

For all other parameters a Metropolis step (Metropolis et al. 1953) is used. All proposal distributions are either Cauchy centered on the previous step or log-Cauchy also centered on the previous step, depending on whether it is real or positive valued.

A.3.2 Simulated data generator function for chapter 3

Define: K

```

dados <- array(NA,dim=c(amostras,n,(K+3)))
a <- array(NA,dim=c(amostras,14,5))
temp.com <- array(NA,dim=c(amostras,3))

for(f in 1:amostras){
# Simulating 2nd order Poisson DLM
# .....
# 2nd order

n          # num observations
J          # num interview
p          # grau do polynomio

W1         # var evolution level
W2         # var evolution increasing

V          # var structure

verdTRUE1 = c(V,W1,W2)

x10
x20

x0 = c(x10,x20)      # initial values

y         = matrix(0,n,J)
mu        = matrix(0,n,J)
lambda   = matrix(0,n,J)
v         = matrix(0,n,J)
x         = matrix(0,J,p)

```

```

w = cbind(rnorm(J,0,sqrt(W1)),rnorm(J,0,sqrt(W2)))
for(j in 1:J){
v[,j] = rnorm(n,0,sqrt(V))
}

r1
r2
w[1,2] = r2
w[1,1] = r1

# evolution
# .....
x[1,2] = x0[2] + w[1,2]
x[1,1] = x0[1] + x0[2] + w[1,1]

# structure
# .....
mu[,1] = x[1,1] + v[,1]
lambda[,1] = exp(mu[,1])

# observation
# .....
y[,1] = rpois(n,lambda[,1])

for (j in 2:J){
x[j,2] = x[j-1,2] + w[j,2]
x[j,1] = x[j-1,1] + x[j-1,2] + w[j,1]

mu[,j] = x[j,1] + v[,j]

lambda[,j] = exp(mu[,j])
}

```

```

y[,j] = rpois(n,lambda[,j])
}

y2 = cbind(1:n,y)
x2 = x

y2 = as.data.frame(y2)

vi1      = rnorm(n,0,U1)
vi2      = rnorm(n,0,U2)
vi3      = rnorm(n,0,U3)

betai1   = b1 + vi1
betai2   = b2 + vi2
betai3   = b3 + vi3

x1       = mu[,1]
x2       = mu[,2]
x3       = mu[,3]

mean     = betai1*x1 + betai2*x2 + betai3*x3

## baseline function Gompertz h0 = alpha0 + alpha0*t ( linear )
m.cen    = 0.7 # (10;90%) (5;70) (2;60) (1;50)
cens.time = rexp(n,1/m.cen)

# h0     = a0 + a1*seq(0,max(cens.time),length=n)
# U0     = 0.1
# vi0    = rnorm(n,0,U0)

# h0i    = h0 + vi0

```

```

#   h   = exp(h0i + mean)

verdTRUE2 = c(U1,U2,U3,a0,a1,b1,b2,b3)

u       = runif(n)

cc <- exp(a0+mean)

true.time <- (log( (- a1*log(u)/cc) +1) )/a1

Time     = pmin(true.time, cens.time)*10
event    = as.numeric(true.time <= cens.time) # event indicator

cat(paste0(round(100 * sum(event) / n, 1), "% experienced event\n"))

surv.data = as.data.frame(cbind(Time,event))

par(mfrow=c(1,2))
require(survival)
plot(survfit(Surv(Time, event)~1, data = surv.data), conf.int = FALSE,
mark.time = TRUE, ylab = "Proportion surviving", xlab = "Time")

plot( seq(0,max(Time),length=length(Time)), exp(a0+a1*sort(Time/10)),
ylab=expression(h[0]== exp (a[0]+a[1]*T)),xlab="Time",type = "l" )

names(y2) = c("id","y.1","y.2","y.3")

# To Balanced dataset
# .....

require(joiner)

```



```

Counts <- to.unbalanced(y2, id.col = 1, times = 1:J,
Y.col = 2:(J+1))
names(Counts)
names(Counts)[3] <- "Y"

Counts$time <- as.character(Counts$time)

require(ggplot2)

# Plotting 2nd order Poisson DLMS
# .....

h1 <- ggplot(na.omit(Counts), aes(Y))

h1+ geom_histogram(breaks=seq(0,200,5), alpha=0.5,
position="identity", lwd=0.2)+
ggtitle("Hist Poisson data simulation")

id.sort <- sort(Time, index.return = TRUE)$ix

y <- y[id.sort,]
Time <- Time[id.sort]
event <- event[id.sort]

K <- J
Q <- K
W <- cbind(1); ncW <- ncol(W)
qs <- seq(0,max(Time)+.5,length=K+1)
ind <- findInterval(Time, qs, rightmost.closed = TRUE)
D <- matrix(0, length(ind), Q)

```

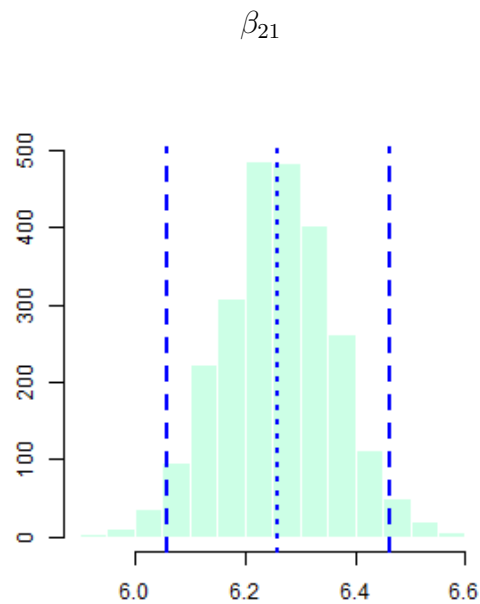
```

D[cbind(seq_along(ind), ind)] <- 1
D <- D * event
Tiq <- outer(Time, qs, pmin)
Lo <- Tiq[, 1:Q]
Up <- Tiq[, 2:(Q+1)]
T <- Up - Lo
P <- T / 2
P1 <- (Up + Lo) / 2
X1 <- cbind(1); ncX1 <- ncol(X1)

dados[f,,] <- cbind(y,Time,event,c(qs,rep(0,dim(y)[1]-length(qs))))

```

A.3.3 Posterior histogram for the parameters for the Liver Dataset.



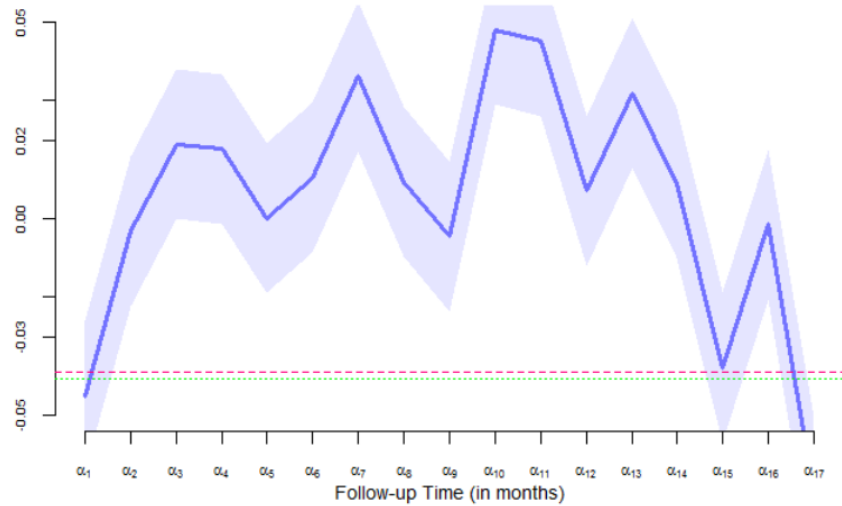


Figure A.1: The α_k effect by interval k : estimated using the package Philipson et al. (2017) based on Henderson et al. (2000) SRE approach (M3), Rizopoulos and Ghosh (2011) (dashed and dotted, respectively, M2) and our proposal (hard line with credible intervals, M1)

A.4 Jags code for Chapter 4

```

{
sigma ~ dgamma(0.01,0.01)
u ~ dgamma(1,sigma)
tauyi <- (1/b.tau)*(1/u)*sigma
mu1[1] ~ dnorm(0, 1)
mu2[1] ~ dnorm(0, 1)
tauyi[1] <- (1/b.tau)*(1/u[1])*sigma
for (j in 2:J) {
mu1[j] ~ dnorm(mu1[(j - 1)] + delta[j]*mu2[(j - 1)], tauw1j)
mu2[j] ~ dnorm(mu2[(j - 1)], tauw2j)
}
tauv ~ dgamma(0.01,0.01)
beta21 ~ dnorm(0,0.01)
for (i in 1:N) {
for (j in 1:J) {
y[i, j] ~ dnorm(lambda[i, j] + beta21*treat[i] + a.tau*u, tauyi)
}
}
}

```

```

lambda[i, j] ~ dnorm(mu1[j], tauv)
ll[i, j] <- log(sqrt(tauyi)) +
(-tauyi/2) * (y[i,j] - lambda[i, j]-beta21*treat[i] - a.tau*u) *
(y[i, j] - lambda[i, j])
}
surt.cen[i] ~ dinterval(surt.time[i],cen[i])
surt.time[i] ~ dweib(r, muh[i])
log(muh[i]) <- beta1[1] + beta1[2]*treat[i] + alpha*mean(lambda[i,1:J])
ls[i] <- log(muh[i]) - surt.time[i]*muh[i]
}
r ~ dexp(10)
beta1[1:2] ~ dmnorm(mbeta1[], vbeta1[,])
alpha ~ dnorm(0, 0.001)
}

```

A.5 Jags code for Chapter 5

```

data {
for (k in 1:(K+1)) {
a[k] <- N*(k-1)/K; # partition the time axis
}
for (i in 1:N) {
for (k in 1:K) { # indicates event-time in interval k
d[i,k] <- (1 - event[i])*step(surv.time[i] - a[k])*step(a[k+1] - surv.time[i]);
}
}
}
{
for( i in 1 : N ) {
for( j in 1 : J ) {
y[i,j] ~ dnorm(lambda[i,j], tau)
lambda[i,j] <- mu[T[i,j]]
T[i,j] ~ dcat(P[])
}
}
}

```

```

ll[i,j]<-log(sqrt(tau))+(-tau/2)*(y[i,j]-lambda[i,j])*(y[i,j]-lambda[i,j])
}

for (k in 1:K) {
# length of overlap of surv.time[i] with interval k
deltah[i,k] <- (min(surv.time[i], a[k+1]) - a[k])*step(surv.time[i] - a[k]);

# the piecewise exponential hazard rate lambda[k]
zeta[i,k] <- exp(h0[k] + alpha*mean(lambda[i,1:K]) + beta1*treat[i])
h[i,k] <- deltah[i,k]*zeta[i,k];
# define the likelihood
d[i,k] ~ dpois(h[i,k]);
}
}

h0[1] ~ dnorm(0.0, kappa);
for (k in 2:K) {
h0[k] ~ dnorm(h0[(k-1)], kappa);
}
kappa ~ dgamma(0.1,0.1)
tau ~ dgamma(0.1,0.1)

#sig ~ dunif(0.01,100) # vague Gelman prior for sigma
#kappa <- 1/(sig*sig)
#sigma ~ dunif(0.01,100) # vague Gelman prior for sigma
#tau<- 1/(sigma*sigma)

P[1] ~ dbeta(a.p1,b.p1)
P[2] ~ dbeta(a.p2,b.p2)
theta ~ dnorm(0.0, 1.0E-6)T(0.0, )
mu[2] <- mu[1] + theta

```

```
mu[1] ~ dnorm(0.0, 1.0E-6)
alpha ~ dnorm(0.0, 1.0E-2)
beta1 ~ dnorm(0.0, 1.0E-2)
}
```

Bibliography

- Andersen, P., Borgan, O., Gill, R., and Kieding, N. (1993). *Statistical Models Based on Counting Processes*. New York: Springer.
- Barreto, L., Farenzena, G., Fiaccone, L., Santos, P., Assis, O., and Araújo, N. and Santos, B. (1994). Effect of vitamin a supplementation on diarrhoea and acute lower-respiratory-tract infections in young children in brazil. *Lancet*, pages 228–231.
- Biller, C. and Fahrmeir, L. (1997). Bayesian spline-type smoothing in generalized regression models. *Computational Statistics*, 12:135–151.
- Brown, E., Ibrahim, J. G., and DeGruttola, V. (2005). A flexible B-spline model for multiple longitudinal biomarkers and survival. *Biometrics*, 61:64–73.
- Carvalho, M., Andreozzi, V., Codeço, C., Campos, D., Bardosa, M., and Shimakura, S. (2012). *Análise de sobrevivência. Teoria e aplicações em saúde*. Fiocruz.
- Chen, P. and Sen, P. (2001). Quality adjusted survival estimation with periodic observations. *Biometrics*, pages 868–874.
- Chi, Y. and Ibrahim, J. G. (2006). Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*, 62:432–445.
- Chokethaworn, K., Yamaka, W., and Maneejuk, P. (2017). Maximum entropy quantile regression with unknown quantile. *Thai Journal of Mathematics: Special Issue on Entropy in Econometrics*, page 107–119.
- Cox, D. (1972). Regression model and life-tables (with discussion). *Journal of the Royal Statistical Society*, pages 187–220.
- Crowther, M. J., Abrams, K. R., and Lambert, P. C. (2013). Joint modeling of longitudinal and survival data. *Stata Journal*, 13:165–184.

- DeGruttola, V. and Tu, X. (1994). Modelling progression of cd4-lymphocyte count and its relationship to survival time. *Biometrics*, pages 1003–1014.
- Diggle, P., Heagerty, P., Liang, K., and Zeger, S. (1998). *Analysis of Longitudinal Data*. Oxford University Press.
- Farcomeni, A. and Viviani, S. (2015). Longitudinal quantile regression in the presence of informative dropout through longitudinal-survival joint modeling. *Statistical Methods in Medical Research*, 34:1199–1213.
- Faucett, C. and Thomas, D. (1996). Simultaneously modelling censored survival data and repeatedly measured covariates: A gibbs sampling approach. *Statistics in Medicine*, pages 1663–1685.
- Fiteni, F., Anota, A., Westeel, V., and Bonnetain, F. (2015). La qualité de vie relative à la santé dans les essais cliniques de phase III en oncologie: de l’analyse statistique. *Bulletin du Cancer*, pages 360–366.
- Gamerman, D. (1991). Dynamic bayesian models for survival data. *Journal of Applied Statistics*, pages 63–79.
- Gamerman, D. and Lopes, H. (2006). *Markov Chain Monte Carlo - Stochastic Simulation for Bayesian Inference*. London: Chapman & Hall/CRC Press.
- Gamerman, D. and Migon, H. S. (1993). Dynamic hierarchical models. *Journal of the Royal Statistical Society B*, 55:629–642.
- Gelber, R., Gelman, R., and Goldhrisch, A. (1989). A quality of life oriented endpoint for comparing therapies. *Biometrics*, pages 781–795.
- Gelfand, A. and Mallick, B. (1995). Bayesian analysis of proportional hazard models built from monotone functions. *Biometrics*, 51:843–852.
- Ghosh, S. and Mukhopadhyay, P. (2007). Bayesian analysis of quality adjusted lifetime (qal) data. *Journal of Statistical Theory and Practice*, 1:233–251.
- Gonçalves, K., Migon, H., and Bastos, L. (2018). Dynamic quantile linear models: a bayesian approach.

- Guo, X. and Carlin, B. (2004). Separate and joint modeling of longitudinal and event time data using standard computer packages. *The American Statistician*, 58:1–9.
- Hamilton, J. and Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics, Elsevier*, 64:307–333.
- Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series subject to change in regime. *Econometrica*, 57:357–384.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton Univ. Press, Princeton.
- Harrell, F. E. (2001). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Springer-Verlag New York.
- Henderson, R. and Diggle, P., D. A. (2002). Identification and efficacy of longitudinal markers for survival. *Biostatistics*, pages 33–50.
- Henderson, R., Diggle, P., and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics*, 1:465–480.
- Ibrahim, J., Chen, M., and Sinha, D. (2004). Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials. *Statistica Sinica*, pages 863–883.
- Ibrahim, J. G., Chu, H., and Chen, L. (2010). Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology*, 28:2796–2801.
- Kaplan, E. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, pages 457–481.
- Kim, C. and Nelson, C. (1999). *State Space Models with Regime Switching*. The MIT Press.
- Kim, C. J. (1994). Dynamic linear models with markov switching. *Journal of Econometrics*, 60:1–22.
- Klein, J. P., van Houwelingen, H. C., Ibrahim, J. G., and Scheike, T. H., editors (2014). *Handbook of Survival Analysis*, Boca Raton. Chapman & Hall/CRC.
- Koenker, G. and Bassett, G. J. (1978). Regression quantiles. *Econometrica*, 46:33–50.

- Kotz, S., Kozubowski, T., and Podgórski, K. (2012). *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*. Boston, MA: Birkhäuser Boston – Springer Science & Business Media.
- Leng, C. and Zhang, W. (2014). Smoothing combined estimating equations in quantile regression for longitudinal data. *Statistics and Computing*, 24:123–136.
- Martins, T., Simpson, D., Lindgren, F., and Rue, H. (2013). *Bayesian computing with INLA: new features*.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21:1087–1092.
- Migon, H. S., Gamerman, D., Lopes, H., and Ferreira, M. (2005). Dynamic models. In Dey, D. K. and Rao, C. R., editors, *Handbook of Statistics 25*, pages 557–592. Elsevier, Amsterdam.
- Nieto-Barajas, L. Walker, S. (2002). Markov beta and gamma processes for modelling hazard rates. *Scandinavian Journal of Statistics*, pages 413–424.
- Paez, M. and Gamerman, D. (2011). Technical report 244, dynamic hierarchical model, statistical laboratory, universidade federal do rio de janeiro. <http://www.dme.im.ufrj.br/arquivos/publicacoes/arquivo244.pf>.
- Philipson, P., Sousa, I., Diggle, P. J., Williamson, P., Kolamunnage-Dona, R., Henderson, R., and Hickey, G. L. (2017). *joiner: Joint Modelling of Repeated Measurements and Time-to-Event Data*. R package version 1.1.0.
- Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). Coda: Convergence diagnosis and output analysis for mcmc. *R News*, 6:7–11.
- Proust-Lima, C., Séne, M., Taylor, J. M. G., and Jacqmin-Gadda, H. (2014). Joint latent class models for longitudinal and time-to-event data: a review. *Statistical Methods in Medical Research*, 23:74–90.
- R Development Core Team (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

- Rizopoulos, D. (2011). Dynamic predictions and prospective accuracy in joint models for longitudinal and time to event data. *Biometrics*, 67:819–829.
- Rizopoulos, D. (2012). *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. Chapman & Hall/CRC, Boca Raton.
- Rizopoulos, D. (2016). The R package JMbayes for fitting joint models for longitudinal and time-to-event data using mcmc. *Journal of Statistical Software*, 72(7):1–45.
- Rizopoulos, D. and Ghosh, P. (2011). A bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time to event. *Statistics in Medicine*, 30:1366–1380.
- Rue, H., Martino, S., and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models using integrated nested laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, 71:319–392.
- Silva, G., Lima, A., and Sen, P. (2009). A semi markov multistate model for estimation of the mean quality adjusted survival for non progressive processes. *Lifetime Data Anal*, pages 216–240.
- So, M., Lam, K., and Li, W. (1998). A stochastic volatility model with markov switching. *J. Business Econom. Statist.*, 16:244–253.
- Su, Y.-S. and Yajima, M. (2015). *R2jags: Using R to Run JAGS*. R package version 0.5-7.
- Tanner, M. and Wong, W. (1987). The calculation of posterior distributions by data augmentation (with dicussion). *Journal of American Statistical Association*, pages 528–550.
- Taylor, J., Park, Y., Ankerst, D., Proust-Lima, C., Williams, S., Kestin, L., Bae, K., Pickles, T., and Sandler, H. (2013). Real-time individual predictions of prostate cancer recurrence using joint models. *Biometrics*, 69:206–213.
- van den Hout, A. and Muniz-Terrera, G. (2016). Joint models for discrete longitudinal outcomes in aging research. *Journal of the Royal Statistical Society C*, 65:167–186.

- Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society B*, 40:364–372.
- Wang, Y. and Taylor, J. (2001). Jointly modeling longitudinal and event time data with application to acquired immunodeficiency syndrome. *Journal of the American Statistical Association*, 96:895–905.
- Wen, Y. and Menggang, Y. (2014). Joint models of longitudinal and survival data. In Klein, J. P., van Houwelingen, H. C., Ibrahim, J. G., and Scheike, T. H., editors, *Handbook of Survival Analysis*, pages 523–547. Chapman & Hall/CRC, Boca Raton.
- West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models*. Springer, New York, 2nd edition.
- West, M., Harrison, J., and Migon, H. (1985). Dynamic generalized linear models and bayesian forecasting. *Journal of the American Statistical Association*, 80-389:73–83.
- Wulfsohn, M. and Tsiatis, A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, pages 330–339.
- Yang, M., Luo, S., and DeSantis, S. (2018). Bayesian quantile regression joint models: Inference and dynamic predictions. *Statistical Methods in Medical Research*, pages 0–14.
- Yu, K., Lu, Z., and Stander, J. (2003). Quantile regression: Applications and current research areas. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 52:331–350.
- Yu, K. and Moyeed, R. (2001). Bayesian quantile regression. *Statistics & Probability Letters*, 54:437–447.
- Zhang, D., Chen, M., Ibrahim, J. G., Boye, M. E., and Shen, W. (2017). Bayesian model assessment in joint modelling of longitudinal and survival data with applications to cancer clinical trials. *Journal of Computational and Graphical Statistics*, 26:121–133.