ANALYSIS OF LONGITUDINAL RANDOM LENGTH DATA

by

Ana-Maria Iosif

B.S. Mathematics, University of Bucharest, 1997

M.A. Applied Statistics and Optimization,

University of Bucharest, 2000

M.A. Statistics, University of Pittsburgh, 2003

Submitted to the Graduate Faculty of

Faculty of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2007

UNIVERSITY OF PITTSBURGH

FACULTY OF ARTS AND SCIENCES

This dissertation was presented

by

Ana-Maria Iosif

It was defended on

September 5, 2007

and approved by

Allan R. Sampson

Leon J. Gleser

Wesley K. Thompson

Douglas E. Williamson

Dissertation Director: Allan R. Sampson

ABSTRACT

ANALYSIS OF LONGITUDINAL RANDOM LENGTH DATA

Ana-Maria Iosif, PhD

University of Pittsburgh, 2007

In some clinical trials, data are gathered longitudinally on both the frequency of an event and its severity. Oftentimes, it is not feasible to obtain the exact time of the events, and the events are collected over fixed follow-up intervals. We refer to this type of data as longitudinal random length data, since the subjects are observed repeatedly and, at each assessment time, the data can be viewed as vectors of severities with lengths determined by the number of events experienced during the assessment.

Suppose the interest is in comparing two treatments, and the treatments are evaluated at multiple points in time. Treatment effect is reflected in simultaneous changes in both the number of events and the severity of each event. Consequently, one needs to jointly model the two outcomes to better evaluate treatment effects. The main objective of this dissertation is to introduce a framework for longitudinal random length data.

We propose two multiple population models for such data. We parameterize the models such that, at each measurement time, both the distribution of the random lengths and the distributional mean of each component of the severity vectors depend on the underlying parameter reflecting the treatment effect at that time. Given the random lengths, we assume the distribution of the severities to be multivariate normal. Conditional on the number of events, the dependence in the vector of severities recorded at a single measurement time is modeled using compound symmetry.

The first model assumes the numbers of events for a subject at different time points to be independent Poisson random variables and dependence over time is built into the severity measures. The second model generalizes the first one, by adding another layer of dependence over time. We further assume the numbers of the events experienced by a subject across time to be dependent and use a multivariate Poisson distribution to model them. For each model we describe the maximum likelihood estimation procedure and provide the asymptotic properties for the estimators. We apply both models to analyze a data set containing stressful life events in adolescents with major depressive disorder.

Keywords: longitudinal random length, repeated measurements, informative cluster size, clustered data, multivariate Poisson distribution.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

In my life I have been fortunate to meet dedicated professors. They are my models and my encounters with them shaped who I have become as a person. I am forever thankful to all of them.

I am especially grateful to my adviser Allan R. Sampson, for his guidance, support and encouragement throughout writing this dissertation. He encouraged me to pursue my research, grow as a statistician and become an independent thinker, and managed to keep his patience and sense of humor in the process.

I would like to thank Dr. Leon J. Gleser, who has been much more than a member in my dissertation committee. Always available to answer my questions and to offer advice, he kept his door open and was willing to discuss with me topics ranging from algorithms to maximize the likelihoods to (lack of) vampires in Transylvania.

I am also indebted to Dr. Douglas E. Williamson, with whom I worked in the Department of Psychiatry at the University of Pittsburgh during my graduate student researcher years. His enthusiasm and commitment to research have inspired me to become a better statistician. I thank Dr. Wesley K. Thompson, for being a member of my committee. I also would like to thank Dr. Satish Iyengar (Department of Statistics) and Dr. Neal D. Ryan (Department of Psychiatry) for financial supporting my graduate research work. I am grateful to the Department the Psychiatry for generously allowing me to use their data.

Finally, I would like to thank my family and my friends for their love and support. Although scattered around the world, they have always provided me with the encouragement I needed to complete this journey.

1.0 INTRODUCTION

In certain clinical trials, data are gathered longitudinally on both the frequency of an event and its severity. Oftentimes, it is not feasible to obtain the exact time of the events, and collecting the events over fixed follow-up intervals is how the information is obtained. A change in disease status is reflected in simultaneous changes in both the number of events and the severity of each event. Since both the frequency and the severity are important, the interest is in jointly modeling the two outcomes. This type of data can arise in many situations. A typical example of such an experiment is a clinical trial of a migraine drug; data are recorded monthly and in addition to the total number of migraines occurring during the respective month, the pain levels corresponding to each migraine are reported as well. Both the number of migraines and the pain level of each migraine at each measurement time are informative about the treatment progress. If the drug is efficacious, the patients that received the drug are expected to improve; in time they will have fewer migraines and their pain levels will be lower, as well. We refer to this type of data, when subjects are observed repeatedly and their multivariate random length measurements are recorded over time as *longitudinal random length data*.

If the data are collected only once, at the end of the follow-up, so that there are no longitudinal measurements, the obtained data are the type introduced by Barnhart [5], Barnhart and Sampson [6]. They term such data multivariate random length data, since it can be viewed as vectors of severities with lengths determined by the number of events. They treat the lengths of the random vectors as random variables and the distributional mean of each component of the random vectors depends on an underlying parameter, as do the distributions of the random length variables. For example, in diseases such as epilepsy or migraine headaches, both the number of events and the severity of each event for a patient tend to depend on that patient's overall disease status. In order to arrive at a full picture of the treatment effect, one needs to jointly model the number of events and their associated severity measures.

The overall aim of our research is to develop a methodology for dealing with *longitudinal multivariate random length data* when the length is informative. Informative length refers to a phenomenon where the expected number of observations within a follow-up interval is related to the continuous outcome of interest. Consider the case of a longitudinal trial of a new anti-epileptic drug for epileptic patients. Epilepsy is a disorder characterized by episodes of seizure activity of variable length and intensity. Anti-epileptic medication can be effective in reducing seizures. Patients are evaluated periodically during the trial. At each measurement time we record the number of observed seizures and rate them each according to severity. If the drug is effective, in time we expect the patients to have fewer seizures and we expect the corresponding severity levels to decrease.

Models for *longitudinal random length data* are necessarily complex because they must consider three types of dependence within a subject: first, between measurements on the continuous severity measures at a single time point; second, between severity measurements at different time points and third; between the number of events experienced at different time points.

1.1 MOTIVATING DATA: LIFE EVENTS AND DIFFICULTIES SCHEDULE DATA

This research was motivated in part by the LEDS data set, collected as part of a larger study at the University of Pittsburgh Western Psychiatric Institute and Clinic (WPIC). This data set contains information about stressful life events in depressed adolescents. Since the first study in 1967 by Holmes and Rahe [18], stressful life events have been a topic of great interest in psychiatric epidemiology. Although a number of studies have focused on the complex association between life events and depression onset, knowledge concerning their temporal relationship is still limited. Severe events and major difficulties have been shown to be critical in the development of depression in adults (Brown and Harris [10]). Recent research [24] has established that many of the adult forms of psychopathology, particularly depression, have their first manifestation in adolescence. In light of the number of adolescents who experience depression, with the associated serious implications for later functioning, it is important to understand the role that stressful life events play in the first onset of depression. Several reports have found that subjects who experience depression have significantly more stressful life events prior to the onset of a depressive episode than non-depressed controls [10]. However, most of this body of research is still concentrated on adults. In the effort to design effective prevention and intervention strategies, additional studies are needed that examine models of the developmental trajectories of depression across adolescence.

The objective of the larger study that generated the data we consider here was to examine the occurrence in adolescents of acute and chronic stressors prior to and during a recent depressive episode. Adolescents with major depressive disorder (MDD) were recruited through the outpatient Child and Adolescent Depression Program at the University of Pittsburgh Western Psychiatric Institute and Clinic, where they presented for treatment. Normal control (NC) adolescents were recruited from existing community controls participating in research protocols being conducted at WPIC ([9], [11]). For the current study, adolescents were classified as NC only if they had never met the criteria for any psychiatric disorder. The NCs were group matched on age, sex, and ethnicity with the MDDs resulting in comparable demographic characteristics between the two groups. Stress exposure was examined using the investigator-based Life Events and Difficulties Schedule (LEDS) [10], adapted for use with adolescents [31], via direct interview with the adolescent themselves. This interview is designed to draw out exhaustive information for acute and chronic stressors. Acute stressors are those life events occurring at one point in time (”death of a pet”, ”fight with the boyfriend”). Chronic stressors are required to last at least 4 weeks (for example ”living in an overcrowded, damp flat”). The severity of each stressor was rated on a 4-point scale (4-marked, 3-moderate, 2-some, 1-little or none).

The subjects we consider are 32 depressed (MDD) and 30 normal control (NC) subjects, all of whom are female between the ages of 13 and 18 years. We examine the occurrence of the acute stressors in the 12-month period prior to the onset of the depressive episode in depressed adolescents and during a comparable ”linked” period in normal control adolescents. At the time of the life events interview, all the MDD subjects were remitted (no longer fulfilling criteria for depression) for at least two weeks and asked to recall stressful life events experienced the year before their MDD onset. The ”linked” period refers to the 12-month period which preceded the onset of the depressive episode among the depressed adolescents. For example, if an MDD subject was remitted for one month at the time of the LEDS interview and the duration of his depressive episode was 6 months, the matched NC had to recall the stressful live events experienced during the period of time starting 19 months

Table 1: LEDS Data. Frequency of acute stressors by group. (Percentages represent subjects experiencing stress).

		Number of acute stressors									
		0	1	2	3	4	5	6	7	9	11
1-st three month period before MDD onset											
MDD[1]	$n(\%)$	6(19)	8(25)	2(6)	3(9)	7(22)	1(3)	2(6)	2(6)	-	1(3)
NC[2]	$n(\%)$	6(20)	9(30)	6(20)	6(20)	1(3)	2(7)	-	-	-	-
2-nd three month period before MDD onset											
MDD	$n(\%)$	6(19)	4(13)	7(22)	4(13)	4(13)	3(9)	1(3)	2(6)	-	1(3)
NC	$n(\%)$	6(20)	8(27)	3(10)	6(20)	4(13)	2(7)	1(3)	-	-	-
3-rd three month period before MDD onset											
MDD	$n(\%)$	8(25)	5(16)	8(25)	5(16)		2(6)	4(13)	-	-	-
NC	$n(\%)$	11(37)	3(10)	4(13)	9(30)	1(3)	1(3)	1(3)	-	-	-
4-th three month period before MDD onset[3]											
MDD	$n(\%)$	5(16)	9(28)	5(16)	7(22)	2(6)	3(9)	-	-	1(3)	-
NC	$n(\%)$	13(43)	4(13)	6(20)	5(17)	1(3)	-	1(3)	-	-	-

---

[1] MDD = major depressive disorder (sample size is 32)

[2] NC = normal control (sample size is 30)

[3] 13 subjects have the last month in the study replicated once or twice since they did not have the whole quarter available

ago and ending 7 months ago. The reason behind using such a ”linked” period in normal control adolescents is that the average length of recall for stressful life events was the same for depressed and normal control adolescents. Therefore, it is unlikely that the rates of events were artificially inflated in the normal control group. Tables 1 and 2 summarize the LEDS data.

Table 2: LEDS Data. Severity of acute stressors by group. (Percentages represent stressors).

		Severity of Acute Stressors				
		1	2	3	4	Total (acute stressors)
1-st three month period before MDD onset						
MDD[4]	$n(\%)$	41(45)	32(35)	18(20)	0(0)	91
NC[5]	$n(\%)$	27(51)	20(38)	6(11)	0(0)	53
2-nd three month period before MDD onset						
MDD	$n(\%)$	36(39)	39(42)	15(16)	2(2)	92
NC	$n(\%)$	38(59)	20(31)	3(5)	3(5)	64
3-rd three month period before MDD onset						
MDD	$n(\%)$	28(40)	21(30)	20(29)	1(1)	70
NC	$n(\%)$	38(72)	8(15)	6(11)	1(2)	53
4-th three month period before MDD onset[6]						
MDD	$n(\%)$	31(43)	26(36)	13(18)	2(3)	72
NC	$n(\%)$	22(54)	14(34)	4(10)	1(2)	41

---

[4] MDD = major depressive disorder (sample size is 32)

[5] NC = normal control (sample size is 30)

[6] 13 subjects have the last month in the study replicated once or twice since they did not have the whole quarter available

The common practice when analyzing this type of data is to reduce the dimensionality. One way of accomplishing this reduction in dimensionality is to quantify the life stress experienced by adding together the severities previously assigned to each event stated to have occurred and creating a single observation per subject at each time-point. Another way of reducing the dimensionality to a single observation is to create models for accumulation and dissipation of stress (see Surtees and Ingham [34]). Under such models, it is postulated that the life events summate in their stressful effect and that the stressful effect of a life event dissipates with time. Choosing different decay functions for the way in which life stress dissipates over time will generate different stress outcomes.

Our approach is to preserve the richness of the LEDS data and analyze it, prior to the onset of depression, by treating the acute stressors as longitudinal multivariate random length data with informative length. Each quarter of a year, data for the subjects consists of a vector of severities with length determined by the number of acute stressors experienced during that quarter of a year. We propose that as subjects draw closer in time to the onset of their episode of MDD (major depressive disorder), both the number of stressors and their

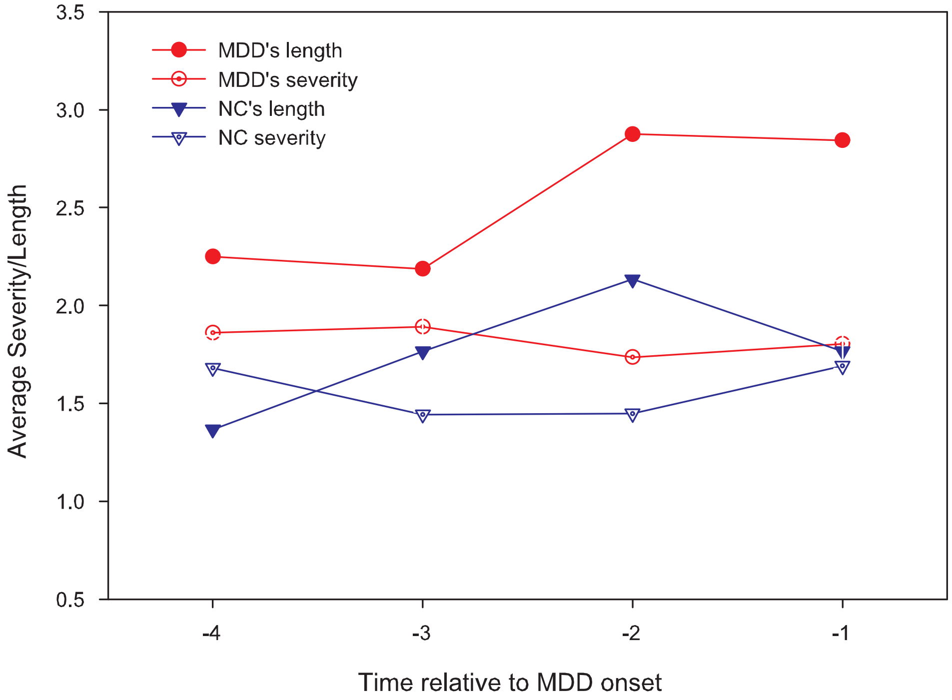

Figure 1: LEDS Data. Length and severity of acute stressors for MDD group (circles) and NC group (triangles)

severities tend to increase. Specifically, the closer that subjects in the MDD group get to the onset, the more likely we will observe a larger number of acute stressors and the more likely their severities will be higher. We would expect that those in the NC group will have relatively constant numbers of acute stressors and severities over time.

Figure 1 contains a plot of the average severity and average length of the acute stressors for the two groups, graphed starting 4 quarters before the MDD episode onset until 1 quarter before MDD episode onset. Throughout the whole period, the MDD group has more acute stressors than the NC, and the gap is larger closer to the MDD onset. The same holds for the average severity of acute stressors, but MDDs' average severity does not have the same increasing trend that their average length shows.

Although our motivation was drawn from a study examining stressful life events in adolescents, the scope of our research is broader and our methods can be used to analyze longitudinally collected data from prospective randomized studies in which the interest is combining the information from the two outcomes to better evaluate the treatment effect. For many diseases or health conditions, an individual may have repeated episodes collected over assessment intervals, together with a measure of the episodes' intensity or severity. For example, in a clinical trial of a new anti-epileptic drug, subjects are randomized to the treatment groups. If the new drug is working, we would expect the subjects in that group to improve, in that, over time they will have fewer seizures and more likely their severities will be lower.

1.2 DISSERTATION OUTLINE

This dissertation is organized in the following fashion. We start Chapter 2 by presenting in Section 2.1 relevant research dealing with random length data in the non-repeated case. Section 2.2 reviews research involving analysis of clustered data when cluster size is informative. In Chapter 3 we introduce models to analyze longitudinal random length data. We build dependence over time into the severity measures. The models are given and we describe the maximum likelihood (ML) algorithm to estimate the parameters. Before obtaining multiple population models for longitudinal random length data, we first, for methodological reasons, develop a one population model. In Section 3.1.1 we introduce the model designed to handle longitudinal multivariate random length data for only one population, along with giving a description of the ML estimation algorithm. Section 3.2 generalizes the results from the previous section to create a likelihood based multiple population model. We model jointly the number of events and their corresponding severities over time. We describe the ML estimation and the large sample behavior of the estimators. We perform a simulation study to evaluate the accuracy of the asymptotic approximation of our estimator in finite samples in Section 3.3. In Section 3.4 we illustrate the proposed methodology using the LEDS data. In Chapter 4 we construct more general models for longitudinal random length data by adding dependence over time into the random lengths. We present the models and describe the ML algorithm to estimate the parameters. In Section 4.1 we introduce a one population model, designed to handle longitudinal multivariate random length data for only one population, along with the description of the ML estimation algorithm and asymptotic properties of the estimators for the particular case when we have only two time measurements. Section 4.2 generalizes the approach from the previous section to handle multiple populations. We describe the ML estimation and the large sample properties of the estimators. We examine the small sample properties of these estimators in a simulation study in Section 4.3. In Section 4.4 we illustrate the methodology using the LEDS data. In Chapter 5 we explore potential generalizations of our research. In Sections 5.1 and 5.2 we give a description of how to generalize the models of Chapters 3 and 4, respectively, to handle more complex correlation structures between severity measurements and between random lengths at different time points within a subject. We briefly present other issues, such as introducing covariates in the model and using other distributions than Poisson to model the random lengths. We conclude with Section 5.3, in which we discuss how one might generalize one of the methods of Section 2.2 (Within Cluster Resampling) to handle longitudinal clustered data when the cluster size is informative.

2.0 LITERATURE REVIEW

2.1 RANDOM LENGTH DATA AND RELATED RESEARCH

2.1.1 Barnhart's Shared Parameter Multiple Population Model

Before embarking on modeling longitudinal random length data, we first take a look at existing approaches for the non-repeated case. This involves data collected on both the frequency of an event and its severity. Data of this type are often dealt with by two different analyses: one for the severities (ignoring the information in the lengths) and another for their number (treating the frequency of the events as the outcome). However, in order to get a full picture of a drug or treatment performance, one needs to jointly model the number of events and their associated severity measures. Properly formulated models increase the power of studies to discern treatment effects.

Barnhart [4], introduced the notion of multivariate random length data, since data on each observational unit can be viewed as vectors of severities with lengths determined by the number of events. The motivation for her research came from a coronary intervention study, where the outcomes of interest were the number of lesions and the sizes of the lesions. The dimensions of the random vectors are treated as random variables (random lengths) and the distributional mean of each of the component of these random vectors depends on an underlying parameter, as do the distributions of the random length variables. Barnhart and Sampson [6] proposed a model to deal with multivariate data without covariates. Their method employs maximum likelihood and assumes that, conditional on the number of events, the event severities are distributed as multivariate normal. The numbers of events are assumed to follow a generalized linear model. Barnhart, Kosinski and Sampson [8] extended the above approach to incorporate covariates. Barnhart [7] also proposed a probit model for multivariate random length data. Recently, Allen and Barnhart [2] introduced a model for multivariate random length ordinal data.

2.1.1.1 Description of the Multiple Population Model without Covariates First, we describe the model proposed by Barnhart and Sampson [6] to jointly model the number of occurrences of an event and their associated severities. Their model is based on the idea that the disease status affects both the number of lesions and their sizes. Thus, the parameter reflecting the underlying disease status appears in specification of both the distribution of the random lengths and the conditional distribution of the vectors of severities given the random lengths.

Suppose we have m populations, with distinct population parameters $\mu_1, ..., \mu_m$ ($\mu_{i_1} \neq \mu_{i_2}$ whenever $i_1 \neq i_2$) characterizing each population's underlying disease status. For each $i = 1, ..., m$ and $j = 1, ..., n_i$, where n_i represents the number of subjects sampled from population i, the data for the j-th subject from population i consist of the random vectors $\boldsymbol{X}_{ij}$ and their corresponding lengths K_{ij}. The model they propose has two important features:

(a) The components of the vector of severities $\boldsymbol{X}_{ij}$ given the corresponding random length are exchangeable random variables. This assumption is motivated by the coronary intervention data where the components of the random vectors, sizes of lesions, are considered to be permutation invariant, since lesion sites are non comparable across subjects.

(b) They introduce a parameter γ, whose sign and magnitude control the association of the random length with the multivariate severities. If $\gamma = 0$ there is no additional information about the number of events that is brought by knowing their severities and conversely, knowing the number of events provides no additional information about their severities. If $\gamma > 0$, $K_{i_1 j_1}$ is stochastically larger than $K_{i_2 j_2}$ for $\mu_{i_1} > \mu_{i_2}$ and any $j_1 = 1, ..., n_{i_1}, j_2 = 1, ..., n_{i_2}$ and if $\gamma < 0$, $K_{i_2 j_2}$ is stochastically larger than $K_{i_1 j_1}$ for $\mu_{i_1} > \mu_{i_2}$ and any $j_1 = 1, ..., n_{i_1}, j_2 = 1, ..., n_{i_2}$. (In other words, if $\gamma > 0$, the larger the underlying parameter μ is, the more likely it is to observe higher severities and a larger frequency of them. If $\gamma < 0$, the larger the underlying parameter μ is, the more likely it is to observe higher severities, but a smaller number of them.)

Specifically, they make the following model assumptions:

(1) The random length variable K_{ij} for population i has a discrete distribution

$$\mathbf{P}(K_{ij} = k) = g_k(\delta + \gamma\mu_i), \quad k = 0, 1, 2, ..., L \tag{2.1}$$

where $L > 0$ is assumed known (possibly infinite), $g_k(\alpha)$ is differentiable in α ($k = 0, 1, ..., L$) and the distribution determined by $g_0(\alpha), ..., g_L(\alpha)$ is stochastically increasing in α.

(2) The random vector $\boldsymbol{X}_i$ with random length K_{ij} from population i, has the conditional distribution:

$$\boldsymbol{X}_{ij}|K_{ij} = k \sim \mathbf{MVN}_k(\mu_i \boldsymbol{e}_k, \sigma^2 \mathbf{R}_k(\rho)), k = 1, 2, ... , \tag{2.2}$$

where $\boldsymbol{e}_k$ is the k-dimensional vector with all the entries 1, $\mathbf{I}_k$ is the k-dimensional identity matrix , $\mathbf{R}_k(\rho) = (1-\rho)\mathbf{I}_k + \rho \boldsymbol{e}_k \boldsymbol{e}_k^{'}$ is the intraclass correlation matrix of dimension k and $0 \leq \rho < 1$ is assumed to be positive.

(3) The observations $(K_{ij}, \boldsymbol{X}_{ij})$, $i = 1, ..., m$, $j = 1, ..., n_i$ are independent.

Note that ρ is required to be nonnegative to ensure the positive definiteness of the covariance matrix $\sigma^2 \mathbf{R}_k(\rho)$ for any $k = 1, ..., L$ and any L.

To simplify the expression of the likelihood, we transform the data using canonical reduction techniques. Given the random length $K_{ij} = k_{ij}$, the corresponding data vector $\boldsymbol{X}_{ij}$, of length greater than one, gets multiplied by a corresponding matrix $\boldsymbol{\Gamma}_{k_{ij}}$, where $\boldsymbol{\Gamma}_k$ is a $k \times k$ dimensional matrix of form

$$\boldsymbol{\Gamma}_k = \begin{pmatrix} \frac{1}{k}\boldsymbol{e}_k' \\ \mathbf{U}_k \end{pmatrix},$$

$\mathbf{U}_k$ is a $(k-1) \times k$ matrix so that $\mathbf{U}_k \mathbf{U}_k^{'} = \mathbf{I}_{k-1}$ and $\mathbf{U}_k \boldsymbol{e}_k = 0$. With the application of the appropriate transformation $\mathbf{Y}_{ij} = \boldsymbol{\Gamma}_{k_{ij}} \boldsymbol{X}_{ij}$ it follows that given K_{ij}, $\boldsymbol{Y}_{ij}$ also has a k_{ij}-dimensional multivariate normal distribution with mean $\boldsymbol{\xi}_i = (\mu_i, 0, ..., 0)$ and covariance matrix $\sigma^2 \mathbf{M}_{k_{ij}}$, where $\mathbf{M}_{k_{ij}} = \text{Diag}\left(\frac{1}{\tau_{k_{ij}}}, \frac{1}{\tau_0}, \ldots, \frac{1}{\tau_0}\right)$, $\tau_k = \frac{k}{1+(k-1)\rho}$ for $k \geq 1$ and $\tau_0 = \frac{1}{1-\rho}$.

The conditional density of the random length vector $\boldsymbol{Y}_{ij}$ is given by

$$f(\boldsymbol{y}_{ij}|k_{ij}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{k_{ij}} \left|\mathbf{M}_{k_{ij}}\right|^{-1/2} \exp\left\{-\frac{(\boldsymbol{y}_{k_{ij}} - \boldsymbol{\xi}_i)'\mathbf{M}_{k_{ij}}^{-1}(\boldsymbol{y}_{k_{ij}} - \boldsymbol{\xi}_i)}{2\sigma^2}\right\}$$

Due to the diagonal form of $\mathbf{M}_{k_{ij}}$, this can be easily be written as

$$f(\boldsymbol{y}_{ij}|k_{ij}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{k_{ij}} \left(\tau_{k_{ij}}\tau_0^{k_{ij}-1}\right)^{1/2} \exp\left\{-\frac{1}{2\sigma^2}\left[\tau_{k_{ij}}(y_{ij1} - \mu_i)^2 + \tau_0\sum_{l=2}^{k_{ij}} y_{ijl}^2\right]\right\}.$$

We can write the joint density for this one observation as:

$$f(\boldsymbol{y}_{ij}, k_{ij}) = \mathbf{P}(K_{ij} = k_{ij})f(\boldsymbol{y}_{ij}|k_{ij})^{\delta(k_{ij})},$$

where

$$\delta(k) = \begin{cases} 1 & \text{if } k \geq 1 \\ 0 & \text{if } k = 0 \end{cases}$$

Taking the logarithm of the above, we obtain:

$$\log f(\boldsymbol{y}_{ij}, k_{ij}) = \log \mathbf{P}(K_{ij} = k_{ij}) + \delta(k_{ij}) \log f(\boldsymbol{y}_{ij}|k_{ij})$$

Hence

$$\begin{aligned} \log f(\boldsymbol{y}_{ij}, k_{ij}) &= \log g_{k_{ij}}(\delta + \gamma\mu_i) \\ &\quad -\frac{\delta(k_{ij})}{2}\left[k_{ij}\log\left(\sigma^2\right) - \log\left(\tau_{k_{ij}}\tau_0^{k_{ij}-1}\right) + \frac{1}{\sigma^2}\left(\tau_{k_{ij}}(y_{ij1} - \mu_i)^2 + \tau_0\sum_{l=2}^{k_{ij}} y_{ijl}^2\right)\right]. \end{aligned}$$

Denoting the $m+4$ parameters of the model by $\boldsymbol{\theta} = (\delta, \gamma, \mu_1, ..., \mu_m, \sigma^2, \rho)'$, we can write the log-likelihood of the whole data as

$$\begin{aligned} l(\boldsymbol{\theta}) &= \sum_{i=1}^{m}\sum_{j=1}^{n_i} \log g_{k_{ij}}(\delta + \gamma\mu_i) \\ &\quad -\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\delta(k_{ij})\left[k_{ij}\log\left(\sigma^2\right) - \log\left(\tau_{k_{ij}}\tau_0^{k_{ij}-1}\right) + \frac{1}{\sigma^2}\left(\tau_{k_{ij}}(y_{ij1} - \mu_i)^2 + \tau_0\sum_{l=2}^{k_{ij}} y_{ijl}^2\right)\right]. \end{aligned}$$

It follows that the $m+4$ score equations are :

$$
\begin{aligned}
\sum_{i=1}^{m}\sum_{j=1}^{n_i}\frac{g'_{k_{ij}}(\delta+\gamma\mu_i)}{g_{k_{ij}}(\delta+\gamma\mu_i)} &= 0\\
\sum_{i=1}^{m}\sum_{j=1}^{n_i}\frac{\mu_i g'_{k_{ij}}(\delta+\gamma\mu_i)}{g_{k_{ij}}(\delta+\gamma\mu_i)} &= 0\\
\text{for}\quad i=1,...,m,\quad \sum_{j=1}^{n_i}\frac{\gamma g'_{k_{ij}}(\delta+\gamma\mu_i)}{g_{k_{ij}}(\delta+\gamma\mu_i)}+\frac{1}{\sigma^2}\sum_{j=1}^{n_i}\delta(k_{ij})\tau_{k_{ij}}(y_{ij1}-\mu_i) &= 0\\
-\frac{1}{2\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\delta(k_{ij})k_{ij}+\frac{1}{2\sigma^4}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\delta(k_{ij})\left[\tau_{k_{ij}}(y_{ij1}-\mu_i)^2+\tau_0\sum_{l=2}^{k_{ij}}y_{ijl}^2\right] &= 0\\
\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\delta(k_{ij})\left[(k_{ij}-1)\left(-\frac{\tau_{k_{ij}}}{k_{ij}}+\tau_0\right)+\frac{1}{\sigma^2}\left(\frac{k_{ij}-1}{k_{ij}}\tau^2_{k_{ij}}(y_{ij1}-\mu_i)^2-\tau_0^2\sum_{l=2}^{k_{ij}}y_{ijl}^2\right)\right] &= 0 \quad .
\end{aligned}
$$

We use the fact that $\frac{\partial\tau_k}{\partial\rho}=-\frac{k-1}{k}\tau_k^2$ and $\frac{\partial\tau_0}{\partial\rho}=\tau_0^2$.

2.1.1.2 Maximum Likelihood Estimation and Asymptotic Properties

Let us denote by $\boldsymbol{I}(\boldsymbol{\theta})$ the information matrix for $\boldsymbol{\theta}$ contained in the $n=\sum\limits_{i=1}^{m}n_i$ multivariate random length vectors $\boldsymbol{X}_{ij}$ with random lengths K_{ij}, $i=1,...,m$, $j=1,...,n_i$. $\boldsymbol{I}(\boldsymbol{\theta})$ can be computed as

$$
\boldsymbol{I}(\boldsymbol{\theta})=\boldsymbol{I}^{*}(\boldsymbol{\theta})+\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sum_{k=1}^{L}\boldsymbol{P}(K_{ij}=k)\boldsymbol{I}(\boldsymbol{\theta}|k)
$$

where $\boldsymbol{I}^{*}(\boldsymbol{\theta})$ is the information matrix about $\boldsymbol{\theta}$ contained in the random lengths K_{ij} , $i=1,...,m$, $j=1,...,n_i$ and $\boldsymbol{I}(\boldsymbol{\theta}|k)$ is the information matrix contained in $\boldsymbol{X}_{ij}|K_{ij}=k$, $k=1,...,L$.

Barnhart [4] showed that, under regularity conditions, the MLE is consistent and has an asymptotic normal distribution.

2.2 INFORMATIVE CLUSTER SIZE, WITHIN CLUSTER RESAMPLING AND RELATED RESEARCH

2.2.1 Within Cluster Resampling

An alternative way of looking at the multivariate random length data is to treat it as clustered data. Under the Barnhart model, the disease status affects both the number of events and their severities. For clustered data, this assumption translates into what is termed *informative cluster size.* Informative cluster size refers to a phenomenon where the expected number of observations within a cluster is related to the outcome of interest. In applications that involve clustered data, the number of subunits within a cluster, i.e the number of events, is often related to the outcomes measured on the individual subunits, i.e. the severities of the events. Standard analysis methods for correlated data (Liang and Zeger [29], Zeger and Liang [36]) are not appropriate for this type of data, since they rely on the assumption that the cluster size is not related to the outcome (missing at random). Analyses that ignore this dependency can lead to biased inference.

Hoffman, Sen and Weinberg [15] introduced Within Cluster Resampling (WCR) as a new procedure to analyze clustered data when the cluster size is related to the outcome. Suppose we have I independent clusters. The WCR procedure is carried out by randomly selecting one observation from each cluster. In this manner, a new data set is formed, consisting of independent univariate observations, one from each cluster. The resampled data set is then analyzed using a generalized linear model (GLM). This resampling procedure is repeated a large number of times (Q), producing Q dependent parameter estimates. Each of the resampled data sets generates $\hat{\boldsymbol{\beta}}(q)$, the maximum likelihood estimator from a GLM and $\hat{\boldsymbol{\Sigma}}(q)$, the estimate of the asymptotic variance-covariance matrix of $\sqrt{I}(\hat{\boldsymbol{\beta}}(q)-\boldsymbol{\beta})$, $q = 1, ..., Q$. Averaging over the Q resampled data sets produces the WCR parameter estimate:

$$\hat{\boldsymbol{\beta}}_{WCR} = \frac{1}{Q}\sum_{q=1}^{Q}\hat{\boldsymbol{\beta}}(q). \tag{2.3}$$

The WCR asymptotic variance-covariance estimator of $\sqrt{I}(\hat{\boldsymbol{\beta}}_{WCR} - \boldsymbol{\beta})$ is

$$\hat{\mathbf{V}}_{WCR} = \frac{1}{Q}\sum_{q=1}^{Q}\hat{\boldsymbol{\Sigma}}(q) - \frac{1}{Q}\sum_{q=1}^{Q}(\hat{\boldsymbol{\beta}}(q) - \hat{\boldsymbol{\beta}}_{WCR})(\hat{\boldsymbol{\beta}}(q) - \hat{\boldsymbol{\beta}}_{WCR})'. \tag{2.4}$$

Note that each resampled estimator $\hat{\boldsymbol{\beta}}(q)$ is the solution of a score equation

$$\boldsymbol{S}(q, \boldsymbol{\beta}) = \sum_{i=1}^{I}\sum_{j=1}^{n_i}\mathbf{U}_{ij} I[(i, j) \in r_q] = 0,$$

where $\mathbf{U}_{ij}$ is the derivative of the contribution of the j-th member of the i-th cluster to the log-likelihood and r_q is the set of indices (i, j) that are randomly resampled in the q-th data set. For a large number of clusters and a large number of resamples, Hoffman's procedure produces $\hat{\boldsymbol{\beta}}_{WCR}$, an overall WCR parameter estimator, that is approximately multivariate normal.

A method asymptotically equivalent to WCR was proposed by Williamson, Datta and Satten [35]. Note that in WCR,

$$\hat{\boldsymbol{\beta}}(q) \approx \boldsymbol{\beta} - \frac{1}{\sqrt{I}}\boldsymbol{S}(q, \boldsymbol{\beta})\mathbf{H}^{-1}(\boldsymbol{\beta}),$$

where

$$\mathbf{H}(q, \boldsymbol{\beta}) = \frac{\partial \boldsymbol{S}(q, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$$

and

$$\mathbf{H}^{-1}(\boldsymbol{\beta}) = \frac{1}{I}E(\mathbf{H}(q, \boldsymbol{\beta})).$$

Instead of averaging the $\hat{\boldsymbol{\beta}}(q)$ as in WCR, Williamson, Datta and Satten's [35] approach is to estimate $\boldsymbol{\beta}$ by solving

$$\frac{1}{Q}\sum_{q=1}^{Q}\boldsymbol{S}(q, \boldsymbol{\beta}) = 0.$$

They name this method cluster weighted generalized estimating equation (CWGEE) and show that it is asymptotically equivalent to WCR.

Dunson, Chen and Harry [12] introduced a general Bayesian framework for jointly modeling cluster size and multiple categorical and continuous subunit-level outcomes. The multiple outcomes measured for the individual subunits are assigned an underlying model that accounts for dependency between outcomes by incorporating shared latent variables within outcome-specific regression models. They allow the same latent variables to affect both cluster size and the subunit-level outcomes.

Follmann, Proschan and Leifer [13] extend WCR approach to a broader range of data. They call their method "multiple outputation" and prove its applicability to other types of clustered data. The only requirement is having a valid statistical procedure for independent data.

A major drawback of all the above methods (except Barnhart's) to analyze clustered data when the sample size is informative is their apparent failure to explicitly use the additional information in the lengths. For example, WCR does not require specifying a correlation structure among the members of the same cluster. While this method removes the bias of the estimates from GEE models applied to informative clustered data, it does not apparently use the information in the lengths. A second drawback is the incapability to handle empty clusters. Consider the case of a periodontitis study, where the participants are the clusters and each tooth is an "observation". If a person does not have any teeth, this might be highly informative since the disease could have been the cause of the missing teeth. Allen and Barnhart [2] call the phenomenon "zero length bias", and argue that subjects with at least one observation may represent a biased sample of the test population.

2.2.2 Within Cluster Resampling for the Multiple Population Model

Recall that the setting of the multiple population model without covariates proposed by Barnhart and Sampson [6] jointly models the number of events and their correlated severities when we have data from m populations. For each subject we observe a random length vector of severities. WCR randomly samples just one entry from each random length vector (provided that the length is non-zero) and applies a GLM to the resulting data set. After repeating this procedure Q times, we average the estimates.

Suppose the m populations have the separate population parameters $\mu_1, ..., \mu_m$ ($\mu_{i_1} \neq \mu_{i_2}$ for $i_1 \neq i_2$) characterizing each population's underlying disease status. The data consist of the random vectors $\boldsymbol{X}_{ij}$ and their corresponding lengths K_{ij}, with $i = 1, ..., m$ and $j = 1, ..., n_i$, where the observations $(K_{ij}, \boldsymbol{X}_{ij})$, $i = 1, ..., m$, $j = 1, ..., n_i$ are independent. Given the lengths, the distribution of the vectors of severities is multivariate normal.

For each $q = 1, ..., Q$ the data for WCR is : X_{ij}^q, $i = 1, ..., m$, $j = 1, ..., n_i$, where X_{ij}^q is the element of $\boldsymbol{X}_{ij}$ randomly sampled in the q-th data set. These observations are independent. For them we apply a one-way anova model

$$X_{ij}^q = \mu_i + \epsilon_{ij}^q \tag{2.5}$$

where ϵ_{ij}^q are independent, identically distributed normal random variables with mean zero.

Maximum likelihood estimation will generate $\hat{\boldsymbol{\mu}}(q) = (\hat{\mu}_1(q), ..., \hat{\mu}_m(q))$ and $\hat{\boldsymbol{\Sigma}}(q)$, the estimate of the asymptotic covariance matrix. The WCR estimator is

$$\hat{\boldsymbol{\mu}}_{WCR} = \frac{1}{Q} \sum_{q=1}^{Q} \hat{\boldsymbol{\mu}}(q), \tag{2.6}$$

and the WCR asymptotic covariance is

$$\hat{\mathbf{V}}_{WCR} = \frac{1}{Q} \sum_{q=1}^{Q} \hat{\boldsymbol{\Sigma}}(q) - \frac{1}{Q} \sum_{q=1}^{Q} (\hat{\boldsymbol{\mu}}(q) - \hat{\boldsymbol{\mu}}_{WCR})(\hat{\boldsymbol{\mu}}(q) - \hat{\boldsymbol{\mu}}_{WCR})'. \tag{2.7}$$

3.0 BUILDING DEPENDENCE INTO SEVERITIES. MULTIPLE POPULATION MODEL WITH INDEPENDENT POISSON LENGTHS

The aim of our research is introducing methodology for modeling and analyzing data collected longitudinally on both the frequency of an event and its severity when both the frequency and the severity are important. We model the two outcomes together and pool the information from both to discern treatment effects. A typical example of such a study is the clinical trial of a migraine drug. The two outcomes being measured on each subject at monthly clinic visits are the number of migraines experienced and the pain level of each migraine. Both the frequency of the migraines and the pain level of each migraine during the month are informative about the treatment progress. If the drug is efficacious, the patients that received the active are expected to improve in that, over time, they will have fewer migraines and their pain levels will be lower. We referred to this type of data, when subjects are observed longitudinally and their random length measurements are recorded periodically over time, as *longitudinal random length data*. The goal of this chapter is to construct a multiple population model for dealing with longitudinal random length data. For clarity of methodology, the analysis of a one population model, introduced in Section 3.1, is extensively presented. The main results of this analysis, the maximum likelihood equations and the asymptotic distributions for the MLE, are then easily generalized in the multiple population model of Section 3.2.

We assume the number of times at which measurements are taken to be the same for all subjects. Handling data from longitudinal studies requires special techniques which take into account the fact that the measurements over time within one subject are dependent. In *longitudinal random length data*, we observe repeated measurements over time on two different outcomes: the number of events, and their multivariate severities. In this chapter, we build dependence over time in only one of the two outcomes of interest. Specifically, we assume that the random lengths giving the number of events for a subject at different time points are independent random variables and we build dependence over time into the severity measures. We employ independent Poisson random variables to model the number of events experienced by a subject throughout the T time points. The Poisson distribution is used to model the random lengths since it is the natural choice for modeling occurrences of an event of interest. Given the random lengths, we assume that the distribution of the severities is multivariate normal. We parametrize the model in such a way that, at each measurement time, both the distribution of the random lengths and the distributional mean of each component of the random vectors depend on an underlying parameter, reflecting the disease status at that time. Conditional on the number of events, the dependence in the vector of severities recorded at a single measurement time on each subject is modeled using compound symmetry. Hence, we assume that any two severity measures registered at the same measurement time for a subject are equally correlated , with the correlation coefficient equal to ρ, independent of the measurement time. Furthermore, conditional on all the numbers of events experienced by a subject, we assume that any two severity measures recorded at two different measurement times within a subject have the same correlation and that correlation, ρ^*, is smaller than ρ.

3.1 ONE POPULATION MODEL WITH INDEPENDENT POISSON LENGTHS

This section introduces a one population model designed to deal with repeated measurements over time of random length data. The model is appropriate for instances in which only one treatment is involved, and the treatment is evaluated at different points in time. We assume that the population is characterized by the parameters $\mu_1, ..., \mu_T$, reflecting the underlying disease status at measurement times 1, 2 ,..., T. The subjects are followed longitudinally and their vectors of severities are recorded repeatedly for a fixed number of time periods T. Hence, at each regularly scheduled measurement time, the data for a subject are a vector of severities with random length. We assume that the random lengths for different times are independent Poisson random variables. At each measurement time, both the distribution of the random lengths and the conditional distributional of the vectors of severities given the random lengths depend on the underlying parameter reflecting the disease status at that time point. Specifically, we assume a log-linear dependence of the average number of events on the average severity.

In the LEDS data mentioned in the Introduction, the relevant quantities are the number of stressful life events that MDD subjects experience throughout the four quarters of a year, and their severity levels. For modeling purposes it makes sense to assume that there is an underlying depression status affecting both the number of events and how severe these events are. We expect that as subjects in the MDD group draw closer to their MDD onset, they will have more stressful life events and their severity levels will be higher.

3.1.1 Model Description

To formalize the above, let us introduce some notation. Each subject i, $i = 1, ..., n$ is observed T times. At each measurement time $j = 1, ..., T$, the subject i reports a random number of events K_{ij} and the corresponding measurements are recorded into the vector $\boldsymbol{X}_{ij}$. Hence all the data for subject i can be summarized by a $\sum\limits_{j=1}^{T} K_{ij}$ - dimensional vector $\boldsymbol{X}_i$, $\boldsymbol{X}_i' = \left(\boldsymbol{X}_{i1}', ..., \boldsymbol{X}_{iT}'\right)$ and the corresponding T-dimensional vector of random lengths $\boldsymbol{K}_i = (K_{i1}, ..., K_{iT})$, with $i = 1, ..., n$. Let $\boldsymbol{k}_i = (k_{i1}, ..., k_{iT})$ be a realization of the T-dimensional vector of lengths $\boldsymbol{K}_i$. Some of the the components of $\boldsymbol{k}_i$ might be zero. Let us denote by $l(\boldsymbol{k}_i)$ the number of nonzero components of $\boldsymbol{k}_i$. Using the notation from the previous chapter, we may write $l(\boldsymbol{k}_i) = \sum\limits_{j=1}^{T} \delta(k_{ij})$, where

$$\delta(k) = \begin{cases} 1 & \text{if } k \geq 1 \\ 0 & \text{if } k = 0 \end{cases} .$$

Let us denote

$$\delta(\boldsymbol{k}) = \begin{cases} 1 & \text{if } \sum\limits_{i=1}^{T} k_i \geq 1 \\ 0 & \text{if } \sum\limits_{i=1}^{T} k_i = 0 \end{cases} .$$

If the vector of random lengths $\boldsymbol{k}_i$ has at least one nonzero component, it follows that $\delta(\boldsymbol{k}_i) = 1$. If $l(\boldsymbol{k}_i) > 0$ we denote by $\widetilde{\boldsymbol{k}}_i$ the $l(\boldsymbol{k}_i)$-dimensional vector composed of the nonzero elements of $\boldsymbol{k}_i$. Hence $\widetilde{\boldsymbol{k}}_i = (k_{ir_{i1}}, k_{ir_{i2}}, ..., k_{ir_{il(\boldsymbol{k}_i)}})$, where $1 \leq r_{i1} < r_{i2} < ... < r_{il(\boldsymbol{k}_i)} \leq T$ are indices corresponding to the elements in the original vector $\boldsymbol{k}_i$. We denote this set of ordered indices by $\boldsymbol{r}_i = (r_{i1}, r_{i2}, ..., r_{il(\boldsymbol{k}_i)})$, which are the times at which a nonzero length vector X_{ij} is observed.

We make the following model assumptions. For each observational unit i, $i = 1, ..., n$,

(1) The T random length variables $K_{i1}, K_{i2}, ..., K_{iT}$ are independent, Poisson(λ_j) distributed random variables, where $\lambda_j = \exp(\delta + \gamma\mu_j)$, $j = 1, ..., T$, so that

$$\mathbf{P}(K_{ij} = k) = \frac{e^{-\lambda_j}\lambda_j^k}{k!}, \quad k = 0, 1, 2, ... \; . \tag{3.1}$$

(2) Conditional on the random lengths $\boldsymbol{K}_i = (K_{i1}, ..., K_{iT}) = (k_{i1}, ..., k_{iT})$, the distribution of $\boldsymbol{X}_i$ for the i-th subject, is a $\sum_{l=1}^{T} k_{il}$-dimensional multivariate normal

$$\boldsymbol{X}_i | \left(\boldsymbol{K}_i = (k_{i1}, ..., k_{iT})\right) \sim \mathbf{MVN}_{\sum_{t=1}^{T} k_{it}} \left(\boldsymbol{\mu}_{\boldsymbol{k}_i}, \sigma^2 \mathbf{S}_{\boldsymbol{k}_i}(\rho, \rho^*)\right), \tag{3.2}$$

for $\sum_{t=1}^{T} k_{it} = 1, 2, ...$, where

$$\boldsymbol{\mu}_{\boldsymbol{k}_i} = \boldsymbol{\mu}_{k_{i1},...,k_{iT}} = \begin{pmatrix} \mu_{r_{i1}} \boldsymbol{e}_{k_{ir_{i1}}} \\ \mu_{r_{i2}} \boldsymbol{e}_{k_{ir_{i2}}} \\ ... \\ \mu_{r_{il(\boldsymbol{k}_i)}} \boldsymbol{e}_{k_{ir_{il(\boldsymbol{k}_i)}}} \end{pmatrix} \tag{3.3}$$

and

$$\begin{aligned} \mathbf{S}_{\boldsymbol{k}_i}(\rho, \rho^*) &= \mathbf{S}_{k_{i1},...,k_{iT}}(\rho, \rho^*) \\ &= \begin{pmatrix} \mathbf{R}_{k_{ir_{i1}}}(\rho) & \rho^* \mathbf{J}_{k_{ir_{i1}}, k_{ir_{i2}}} & ... & \rho^* \mathbf{J}_{k_{ir_{i1}}, k_{ir_{il(\boldsymbol{k}_i)}}} \\ \rho^* \mathbf{J}_{k_{ir_{i2}}, k_{ir_{i1}}} & \mathbf{R}_{k_{ir_{i2}}}(\rho) & ... & \rho^* \mathbf{J}_{k_{ir_{i2}}, k_{ir_{il(\boldsymbol{k}_i)}}} \\ ... & ... & ... & \\ \rho^* \mathbf{J}_{k_{ir_{il(\boldsymbol{k}_i)}}, k_{ir_{i1}}} & \rho^* \mathbf{J}_{k_{ir_{il(\boldsymbol{k}_i)}}, k_{ir_{i2}}} & ... & \mathbf{R}_{k_{ir_{il(\boldsymbol{k}_i)}}}(\rho) \end{pmatrix}. \end{aligned} \tag{3.4}$$

We denote by $\mathbf{R}_k(\rho) = (1-\rho)\mathbf{I}_k + \rho \boldsymbol{e}_k \boldsymbol{e}_k^{'}$, the intraclass correlation matrix of dimension k; $\mathbf{J}_{k,l}$ is the $k \times l$-dimensional matrix having all entries equal to 1, $\mathbf{I}_k$ is the k-dimensional identity matrix, and $\boldsymbol{e}_k$ is the k-dimensional vector with all the entries 1. In order for the matrix $\mathbf{S}_{k_1,\ldots,k_T}(\rho, \rho^*)$ to be positive definite for all possible choices of $(k_1, ..., k_T)$, we impose the sufficient condition $0 \leq \rho^* \leq \rho < 1$.

(3) The data $(\boldsymbol{K}_i, \boldsymbol{X}_i)$ and $(\boldsymbol{K}_j, \boldsymbol{X}_j)$, for subject i and j, respectively, are independent for $i \neq j$.

As we can see from the expressions of the mean and covariance of the multivariate normal in (3.3) and (3.4), if one of the lengths for a time point is zero, it means there is no corresponding entry for that time point in both the mean vector and the covariance matrix. This is one of the main difficulties in handling the model, since not only the number of events, but also the number of blocks that constitute the mean and covariance structures in (3.3) and (3.4) can change from subject to subject. Thus, strict attentiveness and a significant amount of bookkeeping need to be conveyed in working with the conditional density functions for the multivariate severity measurements. Consider the following example of two subjects from the one population model described above, both followed for 3 time measurements. Suppose both subjects have either zero or two events recorded at every measurement; subject 1 has 0 events recorded at the first time measurement and two events recorded at each of the next two time points; subject 2 has two events recorded for each of the first two measurements and zero events collected at the last measurement. Their situation might look similar, but the structure of the means differs. Conditional on the total number of events experienced during the three measurement, subject 1's severity mean is $(\mu_2, \mu_2, \mu_3, \mu_3)$, while subject 2's is $(\mu_1, \mu_1, \mu_2, \mu_2)$. Moreover, the conditional covariance for the multivariate severities looks the same, being equal to

$$\begin{pmatrix} \mathbf{R}_2(\rho) & \rho^* \mathbf{J}_{2,2} \\ \rho^* \mathbf{J}_{2,2} & \mathbf{R}_2(\rho) \end{pmatrix}, \tag{3.5}$$

but one needs to take into account the fact that this structure corresponds to the last two time measurements for subject 1 and to the first two time measurements for subject 2.

The proposed model has the following features:

(a). The $(T{+}5)$ parameters are collectively denoted by $\boldsymbol{\theta}$, $\boldsymbol{\theta} = (\delta, \gamma, \mu_1, ..., \mu_T, \sigma^2, \rho, \rho^*)'$. The parameter space for the above model is

$$\Theta = \left\{\boldsymbol{\theta} = (\delta, \gamma, \mu_1, ..., \mu_T, \sigma^2, \rho, \rho^*)' \mid -\infty < \delta, \gamma, \mu_1, ..., \mu_T < \infty, \sigma^2 > 0, 0 \le \rho^* \le \rho < 1\right\}$$

(b). The parameter γ acts as a scaling parameter and also controls the association of μ_t with the random length. γ being zero implies that the population means $(\mu_t, t = 1, ..., T)$ have no effect on the distribution of the random lengths. Parameter $\gamma > 0$ implies that the larger the underlying parameter μ is, the larger the severity measures tend to be, and the higher their frequency tends to be. Parameter $\gamma < 0$ implies that the larger the underlying parameter μ is, the larger the severity measures tend to be, but their frequency tends to be smaller.

(c). The support of the random lengths includes zero. Any of the components of the vector $\boldsymbol{K}_i$ may be zero. If all of them are zero we observe no quantitative data for subject i. In this case, the multivariate normal distribution defined in (3.2) is not meaningful. We deal with this situation by defining the density of a vector with zero length to be equal to 1 with probability 1.

(d). We build dependence over time by means of the parameter ρ^*; therefore, any severity measure recorded for an individual at measurement time t_1 is equally correlated to any other severity measure recorded for the same individual at measurement time t_2, where $t_1 \neq t_2$. This assumption may seem unrealistic in the case of life events (an AR(1)-type covariance might be more suitable), but it is appropriate for other type of data sets and it was chosen to reduce the complexity of the model.

3.1.2 Maximum Likelihood Estimation

Consider one of these $T-$random length measurements, $\boldsymbol{X}_i$, with the random length vector $\boldsymbol{K}_i = (K_{i1}, ..., K_{iT})$. Given the random lengths $K_{i1} = k_{i1}, ..., K_{iT} = k_{iT}$, $\boldsymbol{X}_i$ has a $\sum_{t=1}^{T} k_{it}$ - dimensional multivariate normal distribution with mean $\boldsymbol{\mu}_{\boldsymbol{k}_i}$ and covariance matrix $\sigma^2 \mathbf{S}_{\boldsymbol{k}_i}(\rho, \rho^*)$

given by (3.3) and (3.4) respectively. For each subject i, $\widetilde{\boldsymbol{k}}_i = (k_{ir_1}, k_{ir_2}, ..., k_{ir_{l(\boldsymbol{k}_i)}})$ is the sub-vector of $\boldsymbol{k}_i$ composed only with nonzero elements.

We apply a matrix version of the canonical transformation technique used in the Barnhart model described in Section 2.1.1 and, given $\boldsymbol{K}_i = \boldsymbol{k}_i$, let $\boldsymbol{Y}_i = \boldsymbol{\Gamma}_{\boldsymbol{k}_i}\boldsymbol{X}_i$, where

$$\boldsymbol{\Gamma}_{\boldsymbol{k}_i} = \boldsymbol{\Gamma}_{k_{i1},...,k_{iT}} = \begin{pmatrix} \boldsymbol{\Gamma}_{k_{ir_{i1}}} & \mathbf{O}_{k_{ir_{i1}},k_{ir_{i2}}} & ... & \mathbf{O}_{k_{ir_{i1}},k_{ir_{il(\boldsymbol{k}_i)}}} \\ \mathbf{O}_{k_{ir_{i2}},k_{ir_{i1}}} & \boldsymbol{\Gamma}_{k_{ir_{i2}}} & ... & \mathbf{O}_{k_{ir_{i2}},k_{ir_{il(\boldsymbol{k}_i)}}} \\ ... & ... & ... & ... \\ \mathbf{O}_{k_{ir_{il(\boldsymbol{k}_i)}},k_{ir_{i1}}} & \mathbf{O}_{k_{ir_{il(\boldsymbol{k}_i)}},k_{ir_{i2}}} & ... & \boldsymbol{\Gamma}_{k_{ir_{il(\boldsymbol{k}_i)}}} \end{pmatrix},$$

$$\boldsymbol{X}_i = \begin{pmatrix} \boldsymbol{X}_{ir_{i1}} \\ \boldsymbol{X}_{ir_{i2}} \\ ... \\ \boldsymbol{X}_{ir_{il(\boldsymbol{k}_i)}} \end{pmatrix}.$$

and $\mathbf{O}_{k,l}$ is the $k \times l$ matrix with all the entries equal to 0.

It follows that given $\boldsymbol{K}_i = (k_{i1}, ..., k_{iT})$, $\boldsymbol{Y}_i = \left(\boldsymbol{Y}_{i1}', ..., \boldsymbol{Y}_{iT}'\right)'$ also has a multivariate normal distribution with mean $\boldsymbol{\xi}_{k_{i1},...,k_{iT}}$ and covariance matrix $\sigma^2\boldsymbol{\Delta}_{k_{i1},...,k_{iT}}$, where

$$\boldsymbol{\xi}_{k_1,...,k_T}' = ((\mu_{r_{i1}}\ 0\ ...\ 0), (\mu_{r_{i2}}\ 0\ ...\ 0), ..., (\mu_{r_{il(\boldsymbol{k}_i)}}\ 0\ ...\ 0))$$

and

$$\boldsymbol{\Delta}_{k_1,...,k_T} = \begin{pmatrix} \mathbf{M}_{k_{ir_{i1}}} & \rho^*\mathbf{E}_{k_{ir_{i1}},k_{ir_{i2}}} & ... & \rho^*\mathbf{E}_{k_{ir_{i1}},k_{ir_{il(\boldsymbol{k}_i)}}} \\ \rho^*\mathbf{E}_{k_{ir_{i2}},k_{ir_{i1}}} & \mathbf{M}_{k_{ir_{i2}}} & ... & \rho^*\mathbf{E}_{k_{ir_{i2}},k_{ir_{il(\boldsymbol{k}_i)}}} \\ ... & ... & ... & ... \\ \rho^*\mathbf{E}_{k_{ir_{il(\boldsymbol{k}_i)}},k_{ir_{i1}}} & \rho^*\mathbf{E}_{k_{ir_{il(\boldsymbol{k}_i)}},k_{ir_{i2}}} & ... & \mathbf{M}_{k_{ir_{il(\boldsymbol{k}_i)}}} \end{pmatrix}.$$

Recall that $\mathbf{M}_k = \text{Diag}\left(\frac{1}{\tau_k}, \frac{1}{\tau_0}, ..., \frac{1}{\tau_0}\right)$, $\tau_k = \frac{k}{1+(k-1)\rho}$ for $k \geq 1$, and $\tau_0 = \frac{1}{1-\rho}$. We use $\mathbf{E}_{k,l}$ to denote the $k \times l$-dimensional matrix having the $(1,1)$ entry equal to 1 and the remaining elements equal to zero.

To simplify the notation, we express all the quantities concerning the elements of the normal distribution in terms of the $\widetilde{\boldsymbol{k}}_i$, not $\boldsymbol{k}_i$, since only the nonzero length vectors do actually contribute to the multivariate normal. Thus,

$$\boldsymbol{\Delta}_{k_{i1},\ldots,k_{iT}} = \boldsymbol{\Delta}_{\widetilde{\boldsymbol{k}}_i} = \begin{pmatrix} \mathbf{M}_{\widetilde{k}_{i1}} & \rho^*\mathbf{E}_{\widetilde{k}_{i1},\widetilde{k}_{i2}} & \ldots & \rho^*\mathbf{E}_{\widetilde{k}_{i1},\widetilde{k}_{il(\boldsymbol{k}_i)}} \\ \rho^*\mathbf{E}_{\widetilde{k}_{i2},\widetilde{k}_{i1}} & \mathbf{M}_{\widetilde{k}_{i2}} & \ldots & \rho^*\mathbf{E}_{\widetilde{k}_{i2},\widetilde{k}_{il(\boldsymbol{k}_i)}} \\ \ldots & \ldots & \ldots & \ldots \\ \rho^*\mathbf{E}_{\widetilde{k}_{il(\boldsymbol{k}_i)},\widetilde{k}_{i1}} & \rho^*\mathbf{E}_{\widetilde{k}_{il(\boldsymbol{k}_i)},\widetilde{k}_{i2}} & \ldots & \mathbf{M}_{\widetilde{k}_{il(\boldsymbol{k}_i)}} \end{pmatrix}.$$

We notice that, given $\boldsymbol{K}_i$, the only correlated components in the vector $\boldsymbol{Y}_i$ are the first entries in each of the subvectors $\boldsymbol{Y}_{ij}$, $j = 1, \ldots, T$, provided that the corresponding length K_{ij} is nonzero. All the remaining entries in the $\boldsymbol{Y}_i$ vector are conditionally independent, identically distributed univariate normal random variables, with mean 0 and variance σ^2/τ_0. We have

$$\boldsymbol{Y}_i = \left(Y_{ir_{i1}1}\ Y_{ir_{i1}2}\ \ldots\ Y_{ir_{i1}k_{ir_{i1}}}\ Y_{ir_{i2}1}\ Y_{ir_{i2}2}\ \ldots\ Y_{ir_{i2}k_{ir_{i2}}} \ldots\ Y_{ir_{il(\boldsymbol{k}_i)}1}\ Y_{ir_{il(\boldsymbol{k}_i)}2}\ \ldots\ Y_{ir_{il(\boldsymbol{k}_i)}k_{ir_{il(\boldsymbol{k}_i)}}}\right)'.$$

Denote by $\boldsymbol{Z}_i$ the vector composed with the first entries in the nonzero length vectors $\boldsymbol{Y}_{ij}$

$$\boldsymbol{Z}_i = \left(Y_{ir_{i1}1}\ Y_{ir_{i2}1}\ \ldots\ Y_{ir_{il(\boldsymbol{k}_i)}1}\right)' = \left(Z_{i1}\ Z_{i2}\ \ldots\ Z_{i(\boldsymbol{k}_i)}\right)'$$

and by $\widehat{\boldsymbol{Y}}_i$ the vector containing the remaining components of $\boldsymbol{Y}_i$

$$\widehat{\boldsymbol{Y}}_i = \left(Y_{ir_{i1}2}\ \ldots\ Y_{ir_{i1}k_{ir_{i1}}}\ Y_{ir_{i2}2}\ \ldots\ Y_{ir_{i2}k_{ir_{i2}}} \ldots\ Y_{ir_{il(\boldsymbol{k}_i)}2}\ \ldots\ Y_{ir_{il(\boldsymbol{k}_i)}k_{ir_{il(\boldsymbol{k}_i)}}}\right)'.$$

Note that $\boldsymbol{Z}_i$ has dimension $l(\boldsymbol{k}_i) = \sum_{t=1}^{T} \delta(k_{it})$ and $\widehat{\boldsymbol{Y}}_i$ has dimension $\sum_{t=1}^{T} k_{it} - l(\boldsymbol{k}_i)$. Further denoting

$$\boldsymbol{\mu}_i^* = (\mu_{r_{i1}}\ \mu_{r_{i2}}\ \ldots\ \mu_{r_{il(\boldsymbol{k}_i)}})' = (\mu_1^*\ \mu_2^*\ \ldots\ \mu_{l(\boldsymbol{k}_i)}^*),$$

we can readily show that the conditional density of the random length vector $\boldsymbol{Y}_i$ can be expressed as

$$\begin{aligned} f(\boldsymbol{y}_i|\boldsymbol{k}_i) &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{l(\boldsymbol{k}_i)} |\boldsymbol{\Sigma}_{\boldsymbol{k}_i}|^{-1/2} \\ &\quad \times \exp\left\{-\frac{(\boldsymbol{z}_i - \boldsymbol{\mu}_i^*)'\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}_i^*)}{2\sigma^2}\right\} \prod_{j=1}^{l(\boldsymbol{k}_i)} \prod_{l=2}^{k_{ir_{ij}}} \frac{\exp -\frac{y_{ir_{ij}l}^2}{2\sigma^2/\tau_0}}{\sqrt{2\pi\sigma^2/\tau_0}}, \end{aligned}$$

where

$$\boldsymbol{\Sigma}_{\boldsymbol{k}_i} = \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{k}}_i} = \begin{pmatrix} \frac{1}{\tau_{\widetilde{k}_{i1}}} & \rho^* & \cdots & \rho^* \\ \rho^* & \frac{1}{\tau_{\widetilde{k}_{i2}}} & \cdots & \rho^* \\ \cdots & \cdots & \cdots & \cdots \\ \rho^* & \rho^* & \cdots & \frac{1}{\tau_{\widetilde{k}_{il(\boldsymbol{k}_i)}}} \end{pmatrix}. \tag{3.6}$$

This can be written as

$$\begin{aligned} f(\boldsymbol{y}_i|\boldsymbol{k}_i) &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{\sum\limits_{l=1}^{T} k_{il}} \tau_0^{\frac{1}{2}\left(\sum\limits_{l=1}^{T} k_{il} - l(\boldsymbol{k}_i)\right)} |\boldsymbol{\Sigma}_{\boldsymbol{k}_i}|^{-1/2} \\ &\quad \times \exp\left\{-\frac{(\boldsymbol{z}_i - \boldsymbol{\mu}_i^*)'\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}_i^*)}{2\sigma^2}\right\} \exp\left(-\frac{\sum\limits_{j=1}^{l(\boldsymbol{k}_i)} \sum\limits_{l=2}^{k_{ir_{ij}}} y_{ir_{ij}l}^2}{2\sigma^2/\tau_0}\right). \end{aligned}$$

We can now write the joint density for the i-th subject's data as

$$f(\boldsymbol{y}_i, \boldsymbol{k}_i) = \mathbf{P}(K_{i1} = k_{i1}, ..., K_{iT} = k_{iT}) f(\boldsymbol{y}_i|k_{i1}, ..., k_{iT})^{\delta(\boldsymbol{k}_i)}.$$

Taking the logarithm of the above and using the independence of the lengths, we obtain

$$\begin{aligned} \log f(\boldsymbol{y}_i, \boldsymbol{k}_i) &= \log \mathbf{P}(K_{i1} = k_{i1}, ..., K_{iT} = k_{iT}) + \delta(\boldsymbol{k}_i) \log f(\boldsymbol{y}_i|k_{i1}, ..., k_{iT}) \\ &= \sum_{j=1}^{T} \log \mathbf{P}(K_{ij} = k_{ij}) + \delta(\boldsymbol{k}_i) \log f(\boldsymbol{y}_i|k_{i1}, ..., k_{iT}). \end{aligned}$$

Hence, the contribution of the $i-$th subject to the log-likelihood is

$$\begin{aligned} \log f(\boldsymbol{y}_i, \boldsymbol{k}_i) &= \sum_{j=1}^{T}(-\lambda_j + k_{ij} \log \lambda_j) + \delta(\boldsymbol{k}_i) \Bigg[-\frac{\log \sigma^2}{2} \sum_{j=1}^{T} k_{ij} + \frac{\log \tau_0}{2} \left(\sum_{j=1}^{T} k_{ij} - \sum_{j=1}^{T} \delta(k_{ij})\right) \\ &\quad -\frac{1}{2} \log |\boldsymbol{\Sigma}_{\boldsymbol{k}_i}| - \frac{1}{2\sigma^2}(\boldsymbol{z}_i - \boldsymbol{\mu}_i^*)'\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}_i^*) - \frac{\tau_0}{2\sigma^2} \sum_{j=1}^{l(\boldsymbol{k}_i)} \sum_{l=2}^{k_{ir_{ij}}} y_{ir_{ij}l}^2 \Bigg]. \end{aligned}$$

Recalling that λ_j has the form $\lambda_j = \exp\left(\delta + \gamma\mu_j\right)$, we can write

$$\begin{aligned}\log f(\boldsymbol{y}_i, \boldsymbol{k}_i) \;&=\; \sum_{j=1}^{T}\left[-e^{\delta+\gamma\mu_j} + k_{ij}\left(\delta + \gamma\mu_j\right)\right] - \frac{\log\sigma^2}{2}\sum_{j=1}^{T} k_{ij} + \frac{\log\tau_0}{2}\left(\sum_{j=1}^{T} k_{ij} - \sum_{j=1}^{T}\delta(k_{ij})\right)\\ &\quad -\frac{1}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{k}_i}| - \frac{1}{2\sigma^2}(\boldsymbol{z}_i - \boldsymbol{\mu}_i^*)'\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}_i^*) - \frac{\tau_0}{2\sigma^2}\sum_{j=1}^{l(\boldsymbol{k}_i)}\sum_{l=2}^{k_{ir_{ij}}} y_{ir_{ij}l}^2,\end{aligned}$$

where to keep the notation simple, we give up multiplying the last 5 terms in the right hand side of the equation by $\delta(\boldsymbol{k}_i)$. Instead, we make the convention to consider these 5 terms equal to zero for the case when $\delta(\boldsymbol{k}_i)$ is zero (i.e. when all the random lengths for a subject are zero). The log-likelihood of the entire data set is

$$\begin{aligned}l(\boldsymbol{\theta}) \;&=\; \sum_{i=1}^{n}\sum_{j=1}^{T}\left[-e^{\delta+\gamma\mu_j} + k_{ij}\left(\delta + \gamma\mu_j\right)\right] - \frac{\log\sigma^2}{2}\sum_{i=1}^{n}\sum_{j=1}^{T} k_{ij} + \frac{\log\tau_0}{2}\sum_{i=1}^{n}\left(\sum_{j=1}^{T} k_{ij} - \sum_{j=1}^{T}\delta(k_{ij})\right)\\ &\quad -\frac{1}{2}\sum_{i=1}^{n}\log|\boldsymbol{\Sigma}_{\boldsymbol{k}_i}| - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\boldsymbol{z}_i - \boldsymbol{\mu}_i^*)'\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}_i^*) - \frac{\tau_0}{2\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{l(\boldsymbol{k}_i)},\sum_{l=2}^{k_{ir_{ij}}} y_{ir_{ij}l}^2 \qquad (3.7)\end{aligned}$$

where the parameter vector is $\boldsymbol{\theta} = (\delta, \gamma, \mu_1, ..., \mu_T, \sigma^2, \rho, \rho^*)'$.

From the results (A.5) - (A.8) in Appendix A, it follows that the $T + 5$ score equations are given by

$$
\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{T}\left(-e^{\delta+\gamma\mu_j}+k_{ij}\right) &= 0 && (3.8)\\
\sum_{i=1}^{n}\sum_{j=1}^{T}\left(-\mu_j e^{\delta+\gamma\mu_j}+k_{ij}\mu_j\right) &= 0 && (3.9)\\
\gamma\begin{pmatrix} -n\lambda_1+\sum\limits_{i=1}^{n}k_{i1}\\ -n\lambda_2+\sum\limits_{i=1}^{n}k_{i2}\\ \vdots\\ -n\lambda_T+\sum\limits_{i=1}^{n}k_{iT}\end{pmatrix}+\frac{1}{\sigma^2}\sum_{i=1}^{n}\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{z}_i-\boldsymbol{\mu}_i^*) &= \mathbf{0}_T && (3.10)\\
-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{T}k_{ij}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}(\boldsymbol{z}_i-\boldsymbol{\mu}_i^*)'\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{z}_i-\boldsymbol{\mu}_i^*)+\frac{\tau_0}{2\sigma^4}\sum_{i=1}^{n}\widehat{\boldsymbol{y}}_i'\widehat{\boldsymbol{y}}_i &= 0 && (3.11)\\
\frac{1}{1-\rho}\sum_{i=1}^{n}\sum_{j=1}^{T}(k_{ij}-\delta(k_{ij}))-\sum_{i=1}^{n}tr\left(\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}\left[\mathbf{I}_T-\text{Diag}\left(\frac{1}{k_{i1}},\ldots,\frac{1}{k_{iT}}\right)\right]\right) & \\
+\frac{1}{\sigma^2}\sum_{i=1}^{n}(\boldsymbol{z}_i-\boldsymbol{\mu}_i^*)'\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}\left[\mathbf{I}_T-\text{Diag}\left(\frac{1}{k_{i1}},\ldots,\frac{1}{k_{iT}}\right)\right]\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{z}_i-\boldsymbol{\mu}_i^*) & \\
-\frac{\tau_0}{\sigma^2(1-\rho)^2}\sum_{i=1}^{n}\widehat{\boldsymbol{y}}_i'\widehat{\boldsymbol{y}}_i &= 0 && (3.12)\\
-\sum_{i=1}^{n}tr\left(\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{e}_T\boldsymbol{e}_T'-\mathbf{I}_T)\right) & \\
+\frac{1}{\sigma^2}\sum_{i=1}^{n}(\boldsymbol{z}_i-\boldsymbol{\mu}_i^*)'\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{e}_T\boldsymbol{e}_T'-\mathbf{I}_T)\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{z}_i-\boldsymbol{\mu}_i^*) &= 0. && (3.13)
\end{aligned}
$$

We denote by $\mathbf{0}_T$ the T-dimensional vector with all entries equal to zero. Note that the fact that $\sum_{j=1}^{l(\boldsymbol{k}_i)}\sum_{l=2}^{k_{ir_{ij}}}y_{ir_{ij}l}^2=\widehat{\boldsymbol{y}}_i'\widehat{\boldsymbol{y}}_i$ has been used to simplify the last three score equations.

Further note that if all the observations consist of random length vectors with length zero, none of the equations (3.8) - (3.13) make sense; hence none of the parameters are estimable. If we observe only vectors with length zero or 1 then the parameter ρ is not estimable. If each subject has at most one nonzero length measurement, then parameter ρ^* is not estimable.

The maximum likelihood estimator for $\boldsymbol{\theta}$ can be obtained through a Fisher scoring algorithm or by numerical maximization of the likelihood through a standard optimization with restriction algorithm. When performing maximization, one needs to adjust for the constraints in the model. The estimates need to satisfy the restrictions $\sigma^2 > 0$ and $0 \leq \rho^* \leq \rho < 1$.

3.1.3 Asymptotic Distribution of the Maximum Likelihood Estimators

Let us denote by $\boldsymbol{I}_n(\boldsymbol{\theta})$ the information matrix for $\boldsymbol{\theta}$ contained in the n multivariate random length vectors $\boldsymbol{X}_i$ with random lengths $\boldsymbol{K}_i = (K_{i1}, ..., K_{iT})$, $i = 1, ..., n$. Conceptually, it makes sense to think of $\boldsymbol{I}_n(\boldsymbol{\theta})$ as a sum of the information about $\boldsymbol{\theta}$ contained in the lengths and the sum of information about $\boldsymbol{\theta}$ contributed by the vectors of severities, over all possible lengths. We show in Appendix A that the information about the parameter $\boldsymbol{\theta}$ contained in a single observation from the one population model, $\boldsymbol{X}$ with random lengths $\boldsymbol{K} = (K_1, ..., K_T)$ has the form

$$\boldsymbol{I}(\boldsymbol{\theta}) = \boldsymbol{I}^*(\boldsymbol{\theta}) + \sum_{\boldsymbol{k} \in \Upsilon} \mathbf{P}_{\boldsymbol{\theta}}(\boldsymbol{K} = \boldsymbol{k}) \boldsymbol{I}(\boldsymbol{\theta}|\boldsymbol{k}),$$

where $\boldsymbol{I}^*(\boldsymbol{\theta})$ is the information matrix about $\boldsymbol{\theta}$ contained in the random lengths $K = (K_1, ..., K_T)$ and $\boldsymbol{I}(\boldsymbol{\theta}|\boldsymbol{k})$ is the information matrix contained in $\boldsymbol{X}_i|\boldsymbol{K} = \boldsymbol{k}$, where $k \in \Upsilon$. We denote

$$\Upsilon = \left\{ \boldsymbol{k} = (k_1, ...k_T) |\ k_i = 0, 1, ... \text{ for } \forall i = 1, ..., T \text{ and } \sum_{i=1}^{T} k_i \geq 1 \right\}.$$

Using a general result from Barnhart [4] (Theorem A.3.1.1) we can compute the information matrix $\boldsymbol{I}_n(\boldsymbol{\theta})$ about $\boldsymbol{\theta} = (\delta, \gamma, \mu_1, ..., \mu_T, \sigma^2, \rho, \rho^*)'$ contained in the n independent observations from the one population model as

$$\boldsymbol{I}_n(\boldsymbol{\theta}) = n\boldsymbol{I}^*(\boldsymbol{\theta}) + n \sum_{\boldsymbol{k} \in \Upsilon} \mathbf{P}_{\boldsymbol{\theta}}(\boldsymbol{K} = \boldsymbol{k}) \boldsymbol{I}(\boldsymbol{\theta}|\boldsymbol{k}).$$

We show in (A.21), that $\boldsymbol{I}_n(\boldsymbol{\theta})$ has a block diagonal form

$$\boldsymbol{I}_n(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{I}_n(\delta, \gamma, \mu_1, ..., \mu_T) & \mathbf{O}_{(T+2)\times 3} \\ \mathbf{O}_{3\times(T+2)} & n \sum_{\boldsymbol{k} \in \Upsilon} \mathbf{P}_{\boldsymbol{\theta}}(\boldsymbol{K} = \boldsymbol{k}) \boldsymbol{I}_{\boldsymbol{k}}(\sigma^2, \rho, \rho^*) \end{pmatrix},$$

where

$$\boldsymbol{I}_n(\delta,\gamma,\mu_1,...,\mu_T) = n\boldsymbol{I}(\delta,\gamma,\mu_1,...,\mu_T) = n\mathbf{G} + \begin{pmatrix} \mathbf{O}_{2\times 2} & \mathbf{O}_{2\times T} \\ \mathbf{O}_{T\times 2} & n\frac{1}{\sigma^2}\sum\limits_{\boldsymbol{k}\in\Upsilon}\left[\prod\limits_{j=1}^{T} e^{-\lambda_j}\frac{\lambda_j{}^{k_j}}{k_j!}\right]\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} \end{pmatrix},$$

G denotes the matrix

$$\mathbf{G} = \begin{pmatrix} \sum\limits_{j=1}^{T}\lambda_j & \sum\limits_{j=1}^{T}\mu_j\lambda_j & \gamma\lambda_1 & \gamma\lambda_2 & ... & \gamma\lambda_T \\ \sum\limits_{j=1}^{T}\mu_j\lambda_j & \sum\limits_{j=1}^{T}\mu_j^2\lambda_j & \gamma\mu_1\lambda_1 & \gamma\mu_2\lambda_2 & ... & \gamma\mu_T\lambda_T \\ \gamma\lambda_1 & \gamma\mu_1\lambda_1 & \gamma^2\lambda_1 & 0 & ... & 0 \\ \gamma\lambda_2 & \gamma\mu_2\lambda_2 & 0 & \gamma^2\lambda_2 & ... & 0 \\ ... & ... & ... & ... & ... & ... \\ \gamma\lambda_T & \gamma\mu_T\lambda_T & 0 & 0 & ... & \gamma^2\lambda_T \end{pmatrix}, \tag{3.14}$$

the elements of the matrix $\boldsymbol{I}_{\boldsymbol{k}}(\sigma^2,\rho,\rho^*)$ are defined by (A.10) - (A.15), and $\boldsymbol{\Sigma}_{\boldsymbol{k}}$ is given by (3.6). Is is apparent from its expression that $\boldsymbol{I}_n(\delta,\gamma,\mu_1,...,\mu_T)$ is obtained as the sum of the information $n\boldsymbol{G}$ contributed by the random lengths, and the information from the multivariate severities.

As can be seen from (3.8) - (3.13), there is no closed form solution for $\hat{\boldsymbol{\theta}}_n$. Hence, the exact distribution of $\hat{\boldsymbol{\theta}}_n$ is not available. We can apply a general result on the efficiency of maximum likelihood estimators for random length data (Theorem A.3.2 in Barnhart [4]) to derive the asymptotic distribution for $\hat{\boldsymbol{\theta}}_n$, the MLE. The asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_n$, is obtained as the inverse of the information matrix $\boldsymbol{I}_n(\boldsymbol{\theta})$ and it is estimated by $\boldsymbol{I}_n^{-1}(\hat{\boldsymbol{\theta}}_n)$.

Theorem 1. *Let* $\hat{\boldsymbol{\theta}}_n = (\hat{\delta}_n,\hat{\gamma}_n,\hat{\mu}_1^{(n)},...,\hat{\mu}_T^{(n)},\hat{\sigma}_n^2,\hat{\rho}_n,\hat{\rho}_n^*)'$ *be the MLEs for a sample of size* n *from the one population model. Then, as* $n\to\infty$

(1). $\hat{\boldsymbol{\theta}}_n$ *is consistent.*

(2).

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) \xrightarrow{L} \boldsymbol{MVN}_{T+5}(\mathbf{0}, \boldsymbol{I}^{-1}(\boldsymbol{\theta})),$$

where

$$\boldsymbol{I}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{I}(\delta, \gamma, \mu_1, ..., \mu_T) & \boldsymbol{O}_{(T+2)\times 3} \\ \boldsymbol{O}_{3\times(T+2)} & \sum\limits_{\boldsymbol{k}\in\Upsilon} \boldsymbol{P}_{\boldsymbol{\theta}}(\boldsymbol{K} = \boldsymbol{k})\boldsymbol{I}_{\boldsymbol{k}}(\sigma^2, \rho, \rho^*) \end{pmatrix},$$

and

$$\boldsymbol{I}(\delta, \gamma, \mu_1, ..., \mu_T) = \boldsymbol{G} + \begin{pmatrix} \boldsymbol{O}_{2\times 2} & \boldsymbol{O}_{2\times T} \\ \boldsymbol{O}_{T\times 2} & \frac{1}{\sigma^2}\sum\limits_{\boldsymbol{k}\in\Upsilon}\left[\prod\limits_{j=1}^{T} e^{-\lambda_j}\frac{{\lambda_j}^{k_j}}{k_j!}\right]\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} \end{pmatrix},$$

where $\boldsymbol{G}$ *is given by (3.14) and the elements of the matrix* $\boldsymbol{I}_{\boldsymbol{k}}(\sigma^2, \rho, \rho^*)$ *are defined by (A.10) - (A.15).*

3.2 MULTIPLE POPULATION MODEL WITH INDEPENDENT POISSON LENGTHS

This section furthers the model introduced in Section 3.1.1 to accommodate two or more populations. The proposed model is appropriate for studies in which more than one treatment is involved. For example, in a clinical trial of a new anti-epileptic drug, patients are randomized to two treatment groups; one is the new drug and the second one is placebo (or an alternative medication). A change in disease status is reflected in simultaneous changes in both the number of seizures events and the severity of each seizure. Our proposed methodology jointly models the number of events and the vectors of severity measures. The shared parametrization exploits this notion, improving efficiency over separate parameterizations. We use the log-linear functional dependence between the two mean structures from Section 3.1.1 and allow the different populations to share the scaling parameters δ and γ, while μ's, reflecting the underlying disease status are population specific. The model is motivated by our belief that γ and δ are parameters of the process linking the underlying "disease" status to the number of events. Thus, they should remain the same, regardless of population. What changes is the underlying "disease" status. The dependence in the vectors of severities within a subject is modeled using the same parameters σ^2, ρ and ρ^*, regardless of population.

3.2.1 Model Description

To formalize the above considerations, we generalize the notation from Section 3.1. We have data from m different populations. Each subject j, $j = 1, ..., n_i$, from population i, $i = 1, ..., m$ is observed T times. At each time point $t = 1, ..., T$, subject j from population i has a random number of events K_{ijt} and the corresponding severity measurements are recorded into the vector $\boldsymbol{X}_{ijt}$. Hence all the data for this subject i can be condensed into a $\sum\limits_{t=1}^{T} K_{ijt}$-dimensional vector $\boldsymbol{X}_{ij}$, $\boldsymbol{X}_{ij}^{'} = \left(\boldsymbol{X}_{ij1}^{'}, ..., \boldsymbol{X}_{ijT}^{'}\right)$ and the corresponding T-dimensional vector of random lengths $\boldsymbol{K}_{ij} = (K_{ij1}, ..., K_{ijT})$, with $i = 1, ..., m$, $j = 1, ..., n_i$. Let $\boldsymbol{k}_{ij} = (k_{ij1}, ..., k_{ijT})$ be a realization of the T-dimensional vector of lengths $\boldsymbol{K}_{ij}$. Some of the the components of $\boldsymbol{k}_{ij}$ might be zero. Let us denote by $l(\boldsymbol{k}_{ij})$ the number of nonzero components of $\boldsymbol{k}_{ij}$. Using the notation from Chapter 2, $l(\boldsymbol{k}_{ij}) = \sum\limits_{t=1}^{T} \delta(k_{ijt})$. If $l(\boldsymbol{k}_{ij}) > 0$ we denote by $\widetilde{\boldsymbol{k}}_{ij}$ the $l(\boldsymbol{k}_{ij})$-dimensional vector composed of the nonzero elements of $\boldsymbol{k}_{ij}$. Hence $\widetilde{\boldsymbol{k}}_{ij} = (k_{ijr_{ij1}}, k_{ijr_{ij2}}, ..., k_{ijr_{ijl(\boldsymbol{k}_{ij})}})$, where $1 \leq r_{ij1} < r_{ij2} < ... < r_{ijl(\boldsymbol{k}_{ij})} \leq T$ are indices corresponding to the elements in the original vector $\boldsymbol{k}_i$. We denote this set of ordered indices by $\boldsymbol{r}_{ij} = (r_{ij1}, r_{ij2}, ..., r_{ijl(\boldsymbol{k}_{ij})})$.

To specify the model, we assume that for each population $i = 1, 2, ..., m$, and for each observational unit j, $j = 1, ..., n_i$,

(1) The random length variables $K_{ij1}, K_{ij2}, ..., K_{ijT}$ are independent and distributed Poisson(λ_{it}), where $\lambda_{it} = \exp(\delta + \gamma\mu_{it})$, $t = 1, ..., T$. Hence,

$$\mathbf{P}_{\boldsymbol{\theta}}(K_{ijt} = k) = \frac{e^{-\lambda_{it}}\lambda_{it}^{k}}{k!}, \quad k = 0, 1, 2, ... \,. \tag{3.15}$$

(2) The random vector $\boldsymbol{X}_{ij}$ with random lengths $\boldsymbol{K}_{ij} = (K_{ij1}, ..., K_{ijT})$ for the j-th subject in the i-th population, has the conditional distribution

$$\boldsymbol{X}_{ij} | \left(\boldsymbol{K}_{ij} = (k_{ij1}, ..., k_{ijT})\right) \sim \mathbf{MVN}_{\sum\limits_{l=1}^{T} k_{ijl}} \left(\boldsymbol{\mu}_{\boldsymbol{k}_{ij}}, \sigma^2 \mathbf{S}_{\boldsymbol{k}_{ij}}(\rho, \rho^*)\right), \tag{3.16}$$

for $\sum\limits_{l=1}^{T} k_{ijl} = 1, 2, ...$, where

$$\boldsymbol{\mu}_{\boldsymbol{k}_{ij}} = \boldsymbol{\mu}_{k_{ij1},...,k_{ijT}} = \begin{pmatrix} \mu_{ir_{ij1}} \boldsymbol{e}_{k_{ijr_{ij1}}} \\ \mu_{ir_{ij2}} \boldsymbol{e}_{k_{ijr_{ij2}}} \\ ... \\ \mu_{ir_{ijl(\boldsymbol{k}_{ij})}} \boldsymbol{e}_{k_{ijr_{ijl(\boldsymbol{k}_{ij})}}} \end{pmatrix} \tag{3.17}$$

and

$$\begin{aligned} \mathbf{S}_{\boldsymbol{k}_{ij}}(\rho, \rho^*) &= \mathbf{S}_{k_{ij1},...,k_{ijT}}(\rho, \rho^*) \\ &= \begin{pmatrix} \mathbf{R}_{k_{ijr_{ij1}}}(\rho) & \rho^* \mathbf{J}_{k_{ijr_{ij1}}, k_{ijr_{ij2}}} & ... & \rho^* \mathbf{J}_{k_{ijr_{ij1}}, k_{ijr_{ijl(\boldsymbol{k}_{ij})}}} \\ \rho^* \mathbf{J}_{k_{ijr_{ij2}}, k_{ijr_{ij1}}} & \mathbf{R}_{k_{ijr_{ij2}}}(\rho) & ... & \rho^* \mathbf{J}_{k_{ijr_{i2}}, k_{ijr_{ijl(\boldsymbol{k}_{ij})}}} \\ ... & ... & ... & \\ \rho^* \mathbf{J}_{k_{ijr_{ijl(\boldsymbol{k}_{ij})}}, k_{ijr_{ij1}}} & \rho^* \mathbf{J}_{k_{ijr_{ijl(\boldsymbol{k}_{ij})}}, k_{ijr_{ij2}}} & ... & \mathbf{R}_{k_{ijr_{ijl(\boldsymbol{k}_{ij})}}}(\rho) \end{pmatrix} \end{aligned} \tag{3.18}$$

In ensure that the matrix $\mathbf{S}_{k_1,...,k_T}(\rho, \rho^*)$ is positive definite for all possible lengths $(k_1, ..., k_T)$, we assume that $0 \leq \rho^* \leq \rho < 1$.

(3) The data $(\boldsymbol{K}_{i_1 j_1}, \boldsymbol{X}_{i_1 j_1})$ and $(\boldsymbol{K}_{i_2 j_2}, \boldsymbol{X}_{i_2 j_2})$, for subject j_1 from population i_1 and j_2 from population i_2, respectively, are independent for $(i_1, j_1) \neq (i_2, j_2)$.

The model exhibits the following features:

(a). The $(mT+5)$ parameters are collectively denoted by $\boldsymbol{\theta}$, where

$$\boldsymbol{\theta} = (\delta, \gamma, \mu_{11}, ..., \mu_{1T}, ..., \mu_{m1}, ..., \mu_{mT}, \sigma^2, \rho, \rho^*)'.$$

The parameter space for the above model is

$$\Theta = \Big\{ \boldsymbol{\theta} = (\delta, \gamma, \mu_{11}, ..., \mu_{mT}, \sigma^2, \rho, \rho^*)' \,| \\ -\infty < \delta, \gamma, \mu_{11}, ..., \mu_{mT} < \infty, \sigma^2 > 0, 0 \leq \rho^* \leq \rho < 1 \Big\}$$

(b). The parameter γ acts as a scaling parameter and also controls the association of μ with the random lengths and has the same interpretation as in the one population model introduced in Section 3.1.

(c). The support of the random lengths includes zero. Any of the components of the vector $\boldsymbol{K}_i$ may be zero. If all of them are zero, we observe no quantitative data. In this case, we define $f(\boldsymbol{x}_i|\boldsymbol{K}_i = \mathbf{0}) = 1$ with probability 1. As we can see from the expressions above, if one of the lengths is zero, it means there is no corresponding entry for that time point in both the mean vector and the covariance matrix defined by 3.17 and 3.18, respectively.

3.2.2 Maximum Likelihood Estimation

Consider one of these T-random length measurements, $\boldsymbol{X}_{ij}$, with the random length vector $\boldsymbol{K}_{ij} = (K_{ij1}, ..., K_{ijT})$. Given the random lengths $K_{ij1} = k_{ij1}, ..., K_{ijT} = k_{ijT}$, $\boldsymbol{X}_{ij}$ has a $\sum\limits_{t=1}^{T} k_{ijt}$ - dimensional multivariate normal distribution with mean $\boldsymbol{\mu}_{\boldsymbol{k}_{ij}}$ and covariance matrix $\sigma^2 \mathbf{S}_{\boldsymbol{k}_{ij}}(\rho, \rho^*)$ given by (3.17) and (3.18), respectively. Each subject j, from population i can be thought as coming from a one population model of the type introduced in the previous section, with parameter vector $\boldsymbol{\theta}_i = (\delta, \gamma, \mu_{i1}, ..., \mu_{iT}, \sigma^2, \rho, \rho^*)'$. Thus, given $\boldsymbol{K}_{ij} = \boldsymbol{k}_{ij}$, we apply the appropriate transformation and define $\mathbf{Y}_{ij} = \mathbf{\Gamma}_{\boldsymbol{k}_{ij}} \boldsymbol{X}_{ij}$, where

$$\mathbf{\Gamma}_{\boldsymbol{k}_{ij}} = \mathbf{\Gamma}_{k_{ij1},...,k_{ijT}} = \begin{pmatrix} \mathbf{\Gamma}_{k_{ijr_{ij1}}} & \mathbf{O}_{k_{ijr_{ij1}},k_{ijr_{ij2}}} & ... & \mathbf{O}_{k_{ijr_{ij1}},k_{ijr_{ijl(\boldsymbol{k}_{ij})}}} \\ \mathbf{O}_{k_{ijr_{ij2}},k_{ijr_{ij1}}} & \mathbf{\Gamma}_{k_{ijr_{ij2}}} & ... & \mathbf{O}_{k_{ijr_{ij2}},k_{ijr_{ijl(\boldsymbol{k}_{ij})}}} \\ ... & ... & ... & ... \\ \mathbf{O}_{k_{ijr_{ijl(\boldsymbol{k}_{ij})}},k_{ijr_{ij1}}} & \mathbf{O}_{k_{ijr_{ijl(\boldsymbol{k}_{ij})}},k_{ijr_{ij2}}} & ... & \mathbf{\Gamma}_{k_{ijr_{ijl(\boldsymbol{k}_{ij})}}} \end{pmatrix}.$$

As before, $\mathbf{O}_{k,l}$ denotes the $k \times l$ matrix with all the entries equal to 0. We notice that, given $\boldsymbol{K}_{ij} = \boldsymbol{k}_{ij}$, the only uncorrelated entries in the vector Y_{ij} are the first entries in each of the subvectors Y_{ijt}, $t = 1, ..., T$, provided that the corresponding length K_{ijt} is nonzero. All the remaining entries in the vector Y_{ij} are conditionally independent, identically distributed univariate normal random variables, with mean 0 and variance σ^2/τ_0. We have

$$\mathbf{Y}_{ij} = \Big((Y_{ijr_{ij1}1} \; Y_{ijr_{ij2}1} \; ... \; Y_{ijr_{ijl(\boldsymbol{k}_{ij})}1}) \; (Y_{ijr_{ij1}2} \; ... \; Y_{ijr_{ij1}k_{ijr_{ij1}}} \; Y_{ijr_{ij2}2} \; ... \; Y_{ir_{i2}k_{ir_{i2}}} ... \; Y_{ir_{il(\boldsymbol{k}_i)}2} \; ... \; Y_{ir_{il(\boldsymbol{k}_i)}k_{ir_{il(\boldsymbol{k}_i)}}}) \Big)'$$

Denote by $\boldsymbol{Z}_{ij}$ the vector composed with the first entries in the nonzero length vectors $\boldsymbol{Y}_{ijt}$

$$\boldsymbol{Z}_{ij} = \Big(Y_{ijr_{ij1}1} \; Y_{ijr_{ij2}1} \; ... \; Y_{ijr_{ijl(\boldsymbol{k}_{ij})}1} \Big)' = \big(Z_{ij1} \; Z_{ij2} \; ... \; Z_{ijl(\boldsymbol{k}_{ij})} \big)',$$

and by $\widehat{Y}_{ij}$ the vector containing the remaining components of $\boldsymbol{Y}_{ij}$

$$\widehat{\boldsymbol{Y}}_{ij} = \Big(Y_{ijr_{ij1}2} \;...\; Y_{ijr_{ij1}k_{ijr_{ij1}}} \; Y_{ijr_{ij2}2} \;...\; Y_{ijr_{ij2}k_{ijr_{ij2}}} ... \; Y_{ijr_{ijl(\boldsymbol{k}_{ij})}2} \;...\; Y_{ijr_{ijl(\boldsymbol{k}_{ij})}k_{ijr_{ijl(\boldsymbol{k}_{ij})}}}\Big)' .$$

Note that $\boldsymbol{Z}_{ij}$ has dimension $l(\boldsymbol{k}_{ij}) = \sum_{t=1}^{T} \delta(k_{ijt})$ and $\widehat{\boldsymbol{Y}}_{ij}$ has dimension $\sum_{t=1}^{T} k_{ijt} - l(\boldsymbol{k}_{ij})$. Further denoting

$$\boldsymbol{\mu}_{ij}^{*} = (\mu_{ir_{ij1}} \; \mu_{ir_{ij2}} \; ... \; \mu_{ir_{ijl(\boldsymbol{k}_{ij})}})' = (\mu_{ij1}^{*} \; \mu_{ij2}^{*} \; ... \; \mu_{ijl(\boldsymbol{k}_{ij})}^{*})',$$

we can readily show that the contribution of the $j-$th subject from population i to the log-likelihood is

$$\begin{aligned} \log f(\boldsymbol{y}_{ij}, \boldsymbol{k}_{ij}) \;&=\; \sum_{t=1}^{T} (-\lambda_{it} + k_{ijt} \log \lambda_{it}) + \delta(\boldsymbol{k}_{ij}) \Bigg[-\frac{\log \sigma^2}{2} \sum_{t=1}^{T} k_{ijt} + \frac{\log \tau_0}{2} \Bigg(\sum_{t=1}^{T} k_{ijt} - \sum_{t=1}^{T} \delta(k_{ijt}) \Bigg) \\ &\quad -\frac{1}{2} \log \left| \boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}} \right| - \frac{1}{2\sigma^2} (\boldsymbol{z}_{ij} - \boldsymbol{\mu}_{ij}^{*})' \boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}^{-1} (\boldsymbol{z}_{ij} - \boldsymbol{\mu}_{ij}^{*}) - \frac{\tau_0}{2\sigma^2} \sum_{t=1}^{l(\boldsymbol{k}_{ij})} \sum_{l=2}^{k_{ijr_{ijt}}} y_{ijr_{ijt}l}^{2} \Bigg], \end{aligned}$$

where

$$\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}} = \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{k}_{ij}}} = \begin{pmatrix} \frac{1}{\tau_{\widetilde{k}_{ij1}}} & \rho^{*} & ... & \rho^{*} \\ \rho^{*} & \frac{1}{\tau_{\widetilde{k}_{ij2}}} & ... & \rho^{*} \\ ... & ... & ... & ... \\ \rho^{*} & \rho^{*} & ... & \frac{1}{\tau_{\widetilde{k}_{ijl(\boldsymbol{k}_{ij})}}} \end{pmatrix} . \tag{3.19}$$

Recalling that λ_{it} has the form $\lambda_{it} = \exp(\delta + \gamma \mu_{it})$, we can write

$$\begin{aligned} \log f(\boldsymbol{y}_{ij}, \boldsymbol{k}_{ij}) \;&=\; \sum_{t=1}^{T} \left[-e^{\delta + \gamma \mu_{it}} + k_{ijt} (\delta + \gamma \mu_{it}) \right] - \frac{\log \sigma^2}{2} \sum_{t=1}^{T} k_{ijt} + \frac{\log \tau_0}{2} \Bigg(\sum_{t=1}^{T} k_{ijt} - \sum_{t=1}^{T} \delta(k_{ijt}) \Bigg) \\ &\quad -\frac{1}{2} \log \left| \boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}} \right| - \frac{1}{2\sigma^2} (\boldsymbol{z}_{ij} - \boldsymbol{\mu}_{ij}^{*})' \boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}^{-1} (\boldsymbol{z}_{ij} - \boldsymbol{\mu}_{ij}^{*}) - \frac{\tau_0}{2\sigma^2} \sum_{t=1}^{l(\boldsymbol{k}_{ij})} \sum_{l=2}^{k_{ijr_{ijt}}} y_{ijr_{ijt}l}^{2}. \end{aligned}$$

Instead of multiplying by $\delta(\boldsymbol{k}_{ij})$ the last 5 terms in the right hand side of the above equation, we make the convention to consider them equal to zero for the case when $\delta(\boldsymbol{k}_{ij})$ is zero (i.e., when all the random lengths for a subject are zero). The log-likelihood of the entire data set is

$$\begin{aligned} l(\boldsymbol{\theta}) \;=\; & \sum_{i=1}^{m}\sum_{j=1}^{n_i}\sum_{t=1}^{T}\left[-e^{\delta+\gamma\mu_{it}} + k_{ijt}\left(\delta+\gamma\mu_{it}\right)\right] - \frac{\log\sigma^2}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sum_{t=1}^{T} k_{ijt} \\ & + \frac{\log\tau_0}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\sum_{t=1}^{T} k_{ijt} - \sum_{t=1}^{T}\delta(k_{ijt})\right) - \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\log\left|\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}\right| \\ & - \frac{1}{2\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\boldsymbol{z}_{ij}-\boldsymbol{\mu}^*_{ij})'\boldsymbol{\Sigma}^{-1}_{\boldsymbol{k}_{ij}}(\boldsymbol{z}_{ij}-\boldsymbol{\mu}^*_{ij}) - \frac{\tau_0}{2\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sum_{t=1}^{l(\boldsymbol{k}_{ij})}\sum_{l=2}^{k_{ijr_{ijt}}} y^2_{ijr_{ijt}l}, \end{aligned} \tag{3.20}$$

where the parameter vector is $\boldsymbol{\theta} = (\delta, \gamma, \mu_{11}, ..., \mu_{1T}, ..., \mu_{m1}, ..., \mu_{mT}, \sigma^2, \rho, \rho^*)'$.

Using the results (A.5) - (A.8) from Appendix A, we obtain that the $mT + 5$ score equations are given by

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sum_{t=1}^{T}\left(-e^{\delta+\gamma\mu_{it}}+k_{ijt}\right) = 0 \tag{3.21}$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sum_{t=1}^{T}\left(-\mu_{it}e^{\delta+\gamma\mu_{it}}+k_{ijt}\mu_{it}\right) = 0 \tag{3.22}$$

$$\gamma\begin{pmatrix} -n_1\lambda_{11}+\sum\limits_{j=1}^{n_1}k_{1j1} \\ -n_1\lambda_{12}+\sum\limits_{j=1}^{n_1}k_{1j2} \\ \vdots \\ -n_1\lambda_{1T}+\sum\limits_{j=1}^{n_1}k_{1jT} \end{pmatrix}+\frac{1}{\sigma^2}\sum_{j=1}^{n_1}\boldsymbol{\Sigma}_{\boldsymbol{k}_{1j}}^{-1}(\boldsymbol{z}_{1j}-\boldsymbol{\mu}_{1j}^{*}) = \mathbf{0}_T \tag{3.23}$$

$$\vdots$$

$$\gamma\begin{pmatrix} -n_m\lambda_{m1}+\sum\limits_{j=1}^{n_m}k_{mj1} \\ -n_m\lambda_{m2}+\sum\limits_{j=1}^{n_m}k_{mj2} \\ \vdots -n_m\lambda_{mT}+\sum\limits_{j=1}^{n_m}k_{mjT} \end{pmatrix}+\frac{1}{\sigma^2}\sum_{j=1}^{n_m}\boldsymbol{\Sigma}_{\boldsymbol{k}_{mj}}^{-1}(\boldsymbol{z}_{mj}-\boldsymbol{\mu}_{mj}^{*}) = \mathbf{0}_T \tag{3.24}$$

$$-\frac{1}{\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sum_{t=1}^{T}k_{ijt}+\frac{1}{\sigma^4}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\boldsymbol{z}_{ij}-\boldsymbol{\mu}_{ij}^{*})'\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}^{-1}(\boldsymbol{z}_{ij}-\boldsymbol{\mu}_{ij}^{*})$$

$$+\frac{\tau_0}{\sigma^4}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\widehat{\boldsymbol{y}}_{ij}'\widehat{\boldsymbol{y}}_{ij} = 0 \tag{3.25}$$

$$\frac{1}{1-\rho}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\sum_{t=1}^{T}k_{ijt}-\sum_{t=1}^{T}\delta(k_{ijt})\right)$$

$$-\sum_{i=1}^{m}\sum_{j=1}^{n_i}tr\left(\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}^{-1}\left[\mathbf{I}_T-\text{Diag}\left(\frac{1}{k_{ij1}},\ldots,\frac{1}{k_{ijT}}\right)\right]\right)$$

$$+\frac{1}{\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\boldsymbol{z}_{ij}-\boldsymbol{\mu}_{ij}^{*})'\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}^{-1}\left[\mathbf{I}_T-\text{Diag}\left(\frac{1}{k_{ij1}},\ldots,\frac{1}{k_{ijT}}\right)\right]\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}^{-1}(\boldsymbol{z}_{ij}-\boldsymbol{\mu}_{ij}^{*})$$

$$-\frac{\tau_0}{\sigma^2(1-\rho)^2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\widehat{\boldsymbol{y}}_{ij}'\widehat{\boldsymbol{y}}_{ij} = 0 \tag{3.26}$$

$$-\sum_{i=1}^{m}\sum_{j=1}^{n_i}tr\left(\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}^{-1}(\boldsymbol{e}_T\boldsymbol{e}_T^{'}-\mathbf{I}_T)\right)$$

$$+\frac{1}{\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\boldsymbol{z}_{ij}-\boldsymbol{\mu}_{ij}^{*})'\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}^{-1}(\boldsymbol{e}_T\boldsymbol{e}_T^{'}-\mathbf{I}_T)\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}^{-1}(\boldsymbol{z}_{ij}-\boldsymbol{\mu}_{ij}^{*}) = 0. \tag{3.27}$$

We use the fact that $\sum_{j=1}^{l(\boldsymbol{k}_i)} \sum_{l=2}^{k_{ir_{ij}}} y^2_{ir_{ij}l} = \widehat{\boldsymbol{y}}'_{ij}\widehat{\boldsymbol{y}}_{ij}$ to simplify the above expressions.

In the complete analogy with the one population model, if all the observation consist of random length vectors with length zero, none of the equations (3.21) - (3.27) make sense; hence, none of the parameters are estimable. If we observe only vectors with length zero or 1 then the parameter ρ is not estimable. If each subject has at most one nonzero length measurement, then parameter ρ^* is not estimable. As in the case of the one population model, the maximum likelihood estimator for $\boldsymbol{\theta}$ can be obtained through standard numerical techniques where one needs to adjust for the constraints in the model, since the estimates need to satisfy the restrictions $\sigma^2 > 0$ and $0 \leq \rho^* \leq \rho < 1$. We further discuss numerical estimation of the MLE for the multiple population model in Section 3.3.3.

3.2.3 Asymptotic Distribution of the Maximum Likelihood Estimators

Let us denote by $\boldsymbol{I}_n(\boldsymbol{\theta})$ the information matrix for $\boldsymbol{\theta}$ contained in the $n = \sum_{i=1}^{m} n_i$ independent observations from the multiple population model, $\boldsymbol{X}_{ij}$ with random lengths $\boldsymbol{K}_{ij} = (K_{ij1}, ..., K_{ijT})$, $i = 1, ..., m$, $j = 1, ..., n_i$. Using the same principle of adding the information about $\boldsymbol{\theta}$ contained in the lengths and the information about $\boldsymbol{\theta}$ contributed by the vectors of severities, over all possible lengths, we compute in Appendix A the information about the parameter $\boldsymbol{\theta}$ contained in a single observation $\boldsymbol{X}$ with random lengths $\boldsymbol{K} = (K_1, ..., K_T)$ as

$$\boldsymbol{I}(\boldsymbol{\theta}) = \boldsymbol{I}^*(\boldsymbol{\theta}) + \sum_{\boldsymbol{k} \in \Upsilon} \mathbf{P}_{\boldsymbol{\theta}}(\boldsymbol{K} = \boldsymbol{k})\boldsymbol{I}(\boldsymbol{\theta}|\boldsymbol{k}),$$

where $\boldsymbol{I}^*(\boldsymbol{\theta})$ is the information matrix about $\boldsymbol{\theta}$ contained in the random lengths $K = (K_1, ..., K_T)$, $\boldsymbol{I}(\boldsymbol{\theta}|\boldsymbol{k})$ is the information matrix contained in $\boldsymbol{X}_i|\boldsymbol{K} = \boldsymbol{k}$, where $\boldsymbol{k} \in \Upsilon$, and

$$\Upsilon = \left\{ \boldsymbol{k} = (k_1, ...k_T)|\ k_i = 0, 1, ... \text{ for } \forall i = 1, ..., T \text{ and } \sum_{i=1}^{T} k_i \geq 1 \right\}.$$

Using the same result from Barnhart [4](Theorem A.3.1.1) as in Section 3.1, we compute the information matrix $\boldsymbol{I}_n(\boldsymbol{\theta})$ about $\boldsymbol{\theta} = (\delta, \gamma, \mu_{11}, ..., \mu_{mT}, \sigma^2, \rho, \rho^*)'$ contained in these n independent observations from the multiple population model (see (A.23)) as

$$\boldsymbol{I}_n(\boldsymbol{\theta}) = \sum_{i=1}^{m} n_i \boldsymbol{I}^*_{\boldsymbol{i}}(\boldsymbol{\theta}) + \sum_{i=1}^{m} n_i \sum_{\boldsymbol{k} \in \Upsilon} \left[\prod_{j=1}^{T} e^{-e^{\delta+\gamma\mu_{ij}}} \frac{e^{k_j(\delta+\gamma\mu_{ij})}}{k_j!} \right] \boldsymbol{I_i}(\boldsymbol{\theta}|\boldsymbol{k_i}). \tag{3.28}$$

We show in Appendix A (A.24), that $\boldsymbol{I}_n(\boldsymbol{\theta})$ has a block diagonal form

$$\boldsymbol{I}_n(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{I}_n(\delta, \gamma, \mu_{11}, ..., \mu_{mT}) & \mathbf{O}_{(mT+2)\times 3} \\ \mathbf{O}_{3\times(mT+2)} & \sum\limits_{i=1}^{m} n_i \sum\limits_{\boldsymbol{k}\in\Upsilon} \mathbf{P}_{\boldsymbol{\theta}}(\boldsymbol{K}=\boldsymbol{k})\boldsymbol{I}_{\boldsymbol{k}}(\sigma^2, \rho, \rho^*) \end{pmatrix},$$

where

$$\begin{aligned} \boldsymbol{I}_n(\delta, \gamma, \mu_{11}, ..., \mu_{mT}) \quad &= \quad \sum_{i=1}^{m} n_i \begin{pmatrix} \mathbf{G}_{11}^{(i)} & \mathbf{O}_{2\times T(i-1)} & \mathbf{G}_{12}^{(i)} & \mathbf{O}_{2\times T(m-i)} \\ \mathbf{O}_{T(i-1)\times 2} & \mathbf{O}_{T(i-1)\times T(i-1)} & \mathbf{O}_{T(i-1)\times T} & \mathbf{O}_{T(i-1)\times T(m-i)} \\ \mathbf{G}_{12}^{(i)\prime} & \mathbf{O}_{T\times T(i-1)} & \mathbf{G}_{22}^{(i)} & \mathbf{O}_{T\times T(m-i)} \\ \mathbf{O}_{T(m-i)\times 2} & \mathbf{O}_{T(m-i)\times T(i-1)} & \mathbf{O}_{T(m-i)\times T} & \mathbf{O}_{T(m-i)\times T(m-i)} \end{pmatrix} \\ &+ \begin{pmatrix} \mathbf{O}_{2\times 2} & \mathbf{O}_{2\times mT} \\ \mathbf{O}_{mT\times 2} & \sum\limits_{i=1}^{m} n_i \sum\limits_{\boldsymbol{k}\in\Upsilon} \prod\limits_{t=1}^{T} e^{-e^{\delta+\gamma\mu_{it}}} \frac{e^{k_t(\delta+\gamma\mu_{it})}}{k_t!} \boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} \end{pmatrix}, \end{aligned}$$

the matrices $\mathbf{G}_{kl}^{(i)}$'s $(k, l = 1, 2)$ are given in (A.22) and the elements of the matrix $\boldsymbol{I}_{\boldsymbol{k}}(\sigma^2, \rho, \rho^*)$ are defined by (A.10) - (A.15).

As in the case of the one population model, the exact distribution of $\hat{\boldsymbol{\theta}}_n$ is not available and we can apply a general result on the efficiency of maximum likelihood estimators for random length data (see Theorem A.3.2 in Barnhart [4]) to derive the asymptotic distribution for $\hat{\boldsymbol{\theta}}_n$, the MLE. The asymptotic covariance matrix of the MLE $\hat{\boldsymbol{\theta}}_n$ is obtained as the inverse of the above information matrix and is estimated by $\boldsymbol{I}_n^{-1}(\hat{\boldsymbol{\theta}}_n)$.

Theorem 2. *Let* $\hat{\boldsymbol{\theta}}_n = (\hat{\delta}_n, \hat{\gamma}_n, \hat{\mu}_1^{(n)}, ..., \hat{\mu}_{11}^{(n)}, ..., \hat{\mu}_{1T}^{(n)}, ..., \hat{\mu}_{mT}^{(n)}, \hat{\sigma}_n^2, \hat{\rho}_n, \hat{\rho}_n^*)'$ *be the MLEs for a sample of size* n *from the one population model. If* $n_i/n \to \eta_i$ *with* $0 < \eta_i < 1$ *as* $n \to \infty$, *then*

(1). $\hat{\boldsymbol{\theta}}_n$ *is consistent.*

(2).

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) \xrightarrow{L} \boldsymbol{MVN}_{mT+5}(\mathbf{0}, \boldsymbol{I}^{-1}(\boldsymbol{\theta})),$$

where

$$\boldsymbol{I}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{I}(\delta, \gamma, \mu_{11}, ..., \mu_{mT}) & \boldsymbol{O}_{(mT+2)\times 3} \\ \boldsymbol{O}_{3\times(mT+2)} & \sum\limits_{i=1}^{m} \eta_i \sum\limits_{\boldsymbol{k}\in\Upsilon} \boldsymbol{P}_{\boldsymbol{\theta}}(\boldsymbol{K}=\boldsymbol{k})\boldsymbol{I}_{\boldsymbol{k}}(\sigma^2, \rho, \rho^*) \end{pmatrix},$$

and

$$\boldsymbol{I}(\delta,\gamma,\mu_{11},...,\mu_{mT}) = \sum_{i=1}^{m}\eta_i \begin{pmatrix} \boldsymbol{G}_{11}^{(i)} & \boldsymbol{O}_{2\times T(i-1)} & \boldsymbol{G}_{12}^{(i)} & \boldsymbol{O}_{2\times T(m-i)} \\ \boldsymbol{O}_{T(i-1)\times 2} & \boldsymbol{O}_{T(i-1)\times T(i-1)} & \boldsymbol{O}_{T(i-1)\times T} & \boldsymbol{O}_{T(i-1)\times T(m-i)} \\ \boldsymbol{G}_{12}^{(i)\prime} & \boldsymbol{O}_{T\times T(i-1)} & \boldsymbol{G}_{22}^{(i)} & \boldsymbol{O}_{T\times T(m-i)} \\ \boldsymbol{O}_{T(m-i)\times 2} & \boldsymbol{O}_{T(m-i)\times T(i-1)} & \boldsymbol{O}_{T(m-i)\times T} & \boldsymbol{O}_{T(m-i)\times T(m-i)} \end{pmatrix}$$
$$+ \begin{pmatrix} \boldsymbol{O}_{2\times 2} & \boldsymbol{O}_{2\times mT} \\ \boldsymbol{O}_{mT\times 2} & \sum_{i=1}^{m}\eta_i \sum_{\boldsymbol{k}\in\Upsilon}\prod_{t=1}^{T} e^{-e^{\delta+\gamma\mu_{it}}}\frac{e^{k_t(\delta+\gamma\mu_{it})}}{k_t!}\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} \end{pmatrix},$$

and the elements of the matrix $\boldsymbol{I}_{\boldsymbol{k}}(\sigma^2,\rho,\rho^*)$ *are defined by (A.10)-(A.15).*

3.2.4 Inference and Hypothesis Testing

Researchers are usually interested in describing the trend over time, and in whether there are significant differences in the trend across groups of subjects. For example, in the LEDS data one could be interested in examining the ”profiles” of the severities. We have repeated measures of severities over the four quarters of the year. Then there is a between-subjects grouping factor with two categories: MDD and NC. For the two groups, the mean severities for each of the four quarters of the years are typically termed ”profiles”. One can ask a number of questions; for instance, if the profiles are parallel (similar in shape across time); given that the profiles are parallel, if they are equal or separated; furthermore, given that the profiles are coincident, if the mean severities are the same for each of the four quarters of the year. Based on the large-sample normality of our estimators, we can perform Wald tests to answer these types of questions.

More generally, consider tests of the composite hypothesis

$$\begin{aligned} H_0 &: g(\boldsymbol{\theta}) = \mathbf{0} \quad vs \\ H_a &: g(\boldsymbol{\theta}) \neq \mathbf{0} \end{aligned}$$

where g is a vector-valued function from R^k to R^q and the MLE of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_n$ satisfies

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{d} MVN_k(\mathbf{0}, \boldsymbol{I}^{-1}(\boldsymbol{\theta})).$$

Assume g_i is differentiable and let $D = D(\boldsymbol{\theta})$ denote the $q \times k$ matrix with (i,j) entry $\frac{\partial g_i(y_1,\ldots,y_k)}{\partial y_j}$ evaluated at $\boldsymbol{\theta}$. Then, it can be shown (see Lehmann [28], or Sen [33]) that

$$n(g(\hat{\boldsymbol{\theta}}_n) - g(\boldsymbol{\theta}_0))^{'}\boldsymbol{V}^{-1}(\hat{\boldsymbol{\theta}}_n)(g(\hat{\boldsymbol{\theta}}_n) - g(\boldsymbol{\theta}_0)) \xrightarrow{d} \chi^2_q,$$

where $\boldsymbol{V}(\boldsymbol{\theta}) = D(\boldsymbol{\theta})\boldsymbol{I}^{-1}(\boldsymbol{\theta})D^{'}(\boldsymbol{\theta})$.

Because we test $H_0 : g(\boldsymbol{\theta}) = \mathbf{0}$, the level α Wald test rejects H_0 when

$$W_n = ng(\hat{\boldsymbol{\theta}}_n)^{'}\boldsymbol{V}^{-1}(\hat{\boldsymbol{\theta}}_n)g(\hat{\boldsymbol{\theta}}_n)$$

is larger than $\chi^2_q(1-\alpha)$.

3.3 SIMULATION STUDY

In this section we report the results of a simulation study conducted to evaluate the finite sample properties of our estimators. Several different scenarios are analyzed, with longitudinal random length data generated according to the multiple population model described in Section 3.2. The simulation studies allow us to compare the estimated parameters with the true underlying values and assess the accuracy of our estimators. A natural question we address is how many subjects are necessary in order for the large-sample theory to produce the desired results. In doing this, we compare the empirical standard deviations of our estimates with their theoretical asymptotic standard deviations.

3.3.1 General Framework and Quantities Computed

The simulation study was designed to mimic a hypothetical clinical trial with two different treatment groups and the same number of subjects per treatment group. Different scenarios were created to explore various parameter configurations. When choosing these configurations we took into consideration all possible combinations of two factors, each having two levels. The first factor involves the relationship between the severities and the lengths. We think the two interesting levels for the first factor are those reflecting contradictory information from lengths and severities. Thus, the first level considered for this factor mirrors the case when the average severities across time and groups are close, but the numbers of events differ. The second level corresponds to the situation in which there is little difference among the numbers of events, but the average severity is well spread across time and treatment groups. The second factor concerns the relationship between the mean severity "profiles" for the two populations. The first level of this factor corresponds to the situation in which $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are parallel and well separated (average severities in the two groups are parallel but not coincident across time) and the second one to the case when $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are not well separated (the average severities in the two groups intersect across time).

Table 3: Choice of parameters for simulation study. T=4

	$\boldsymbol{\mu}$'s close and $\boldsymbol{\lambda}$'s far	$\boldsymbol{\mu}$'s far and $\boldsymbol{\lambda}$'s close
$\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ well separated	$\boldsymbol{\mu_1} = (1.0, 1.1, 1.2, 1.3)$ $\boldsymbol{\mu_2} = (1.5, 1.6, 1.7, 1.8)$ $(\delta, \gamma) = (-1.3, 2.3)$ $(\sigma^2, \rho, \rho^*) = (1, 0.5, 0.2)$ $\boldsymbol{\lambda_1} = (2.72, 3.42, 4.31, 5.42)$ $\boldsymbol{\lambda_2} = (8.58, 10.80, 13.60, 17.12)$	$\boldsymbol{\mu_1} = (1, 2, 3, 4)$ $\boldsymbol{\mu}_2 = (1.5, 2.5, 3.5, 4.5)$ $(\delta, \gamma) = (2.5, 0.01)$ $(\sigma^2, \rho, \rho^*) = (1, 0.5, 0.2)$ $\boldsymbol{\lambda_1} = (12.30, 12.43, 12.55, 12.68)$ $\boldsymbol{\lambda}_2 = (12.37, 12.49, 12.62, 12.74)$
$\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ close	$\boldsymbol{\mu_1} = (1.3, 1.2, 1.1, 1.0)$ $\boldsymbol{\mu_2} = (1.0, 1.1, 1.2, 1.3)$ $(\delta, \gamma) = (-1, 2.5)$ $(\sigma^2, \rho, \rho^*) = (1, 0.5, 0.2)$ $\boldsymbol{\lambda_1} = (4.48, 5.75, 7.39, 9.49)$ $\boldsymbol{\lambda_2} = (9.49, 7.39, 5.75, 4.48)$	$\boldsymbol{\mu_1} = (4, 3, 2, 1)$ $\boldsymbol{\mu}_2 = (1, 2, 3, 4)$ $(\delta, \gamma) = (1.5, 0.05)$ $(\sigma^2, \rho, \rho^*) = (1, 0.5, 0.2)$ $\boldsymbol{\lambda_1} = (5.47, 5.21, 4.95, 4.71)$ $\boldsymbol{\lambda}_2 = (4.71, 4.95, 5.21, 5.47)$

The simulations try to cover four different scenarios obtained by considering all the combinations of the levels of the two factors; in the first scenario $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are parallel and not coincident, while within the same treatment group the mean severities are close and the number of events are well separated across time; the second scenario depicts the situation in which $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ intersect while within the same group the mean severities are close and the number of events are well separated across time; in the third scenario $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are parallel and not coincident, but over time the values of the mean severities are well separated, while the number of events change very little; finally, in the fourth scenario $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ intersect, while across time we have changes in the mean severities but little variation in the number of events.

Under each scenario, subjects are followed for the same number of time points, T. To match the LEDS data, the values chosen are $T = 4$ and $T = 2$. For each choice of the parameters $\delta, \gamma, \mu_{11}, ..., \mu_{2T}$, we keep the same values for σ^2, ρ and ρ^* across simulation. The choice of parameters for simulations are given in Table 3 for $T = 4$ and in Table 4 for $T = 2$.

Table 4: Choice of parameters for simulation study. T=2

	$\boldsymbol{\mu}$'s close and $\boldsymbol{\lambda}$'s far	$\boldsymbol{\mu}$'s far and $\boldsymbol{\lambda}$'s close
$\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ well separated	$\boldsymbol{\mu_1} = (1.0, 1.1)$ $\boldsymbol{\mu_2} = (1.5, 1.6)$ $(\delta, \gamma) = (-1.3, 2.3)$ $(\sigma^2, \rho, \rho^*) = (1, 0.5, 0.2)$ $\boldsymbol{\lambda_1} = (2.72, 3.42)$ $\boldsymbol{\lambda_2} = (8.58, 10.80)$	$\boldsymbol{\mu_1} = (1, 2)$ $\boldsymbol{\mu}_2 = (1.5, 2.5)$ $(\delta, \gamma = (2.5, 0.01)$ $(\sigma^2, \rho, \rho^*) = (1, 0.5, 0.2)$ $\boldsymbol{\lambda_1} = (12.30, 12.43)$ $\boldsymbol{\lambda}_2 = (12.37, 12.49)$
$\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ close	$\boldsymbol{\mu_1} = (1.3, 1.2)$ $\boldsymbol{\mu_2} = (1.2, 1.3)$ $(\delta, \gamma) = (-1, 2.5)$ $(\sigma^2, \rho, \rho^*) = (1, 0.5, 0.2)$ $\boldsymbol{\lambda_1} = (4.48, 5.75)$ $\boldsymbol{\lambda_2} = (5.75, 4.48)$	$\boldsymbol{\mu_1} = (4, 3)$ $\boldsymbol{\mu}_2 = (3, 4)$ $(\delta, \gamma = (1.5, 0.05)$ $(\sigma^2, \rho, \rho^*) = (1, 0.5, 0.2)$ $\boldsymbol{\lambda_1} = (5.47, 5.21)$ $\boldsymbol{\lambda}_2 = (5.21, 5.47)$

For each of the two choices of T, we considered $D = 1000$ independent data sets with n=20, 50 and 100, respectively, subjects per population. Thus, for example, when $T = 4$ and $n = 50$, we needed to simulate 50 subjects with 4 repeated random length vectors of severities for the first group and the same for the second group. For each simulated data set d, $d = 1, ..., D$, we fit our multiple population model. The parameter of interest is the vector $\boldsymbol{\theta} = (\delta, \gamma, \mu_{11}, ..., \mu_{1T}, \mu_{21}, ..., \mu_{2T}, \sigma^2, \rho, \rho^*)$. By maximizing the likelihood, we compute the numerical value of the $(2T+5)$ - dimensional ML estimator $\hat{\boldsymbol{\theta}}_d$ and its theoretical asymptotic variance.

Repeating the aforementioned procedure for each simulated data set provides us with D independent ML estimators, corresponding to the D simulations performed. Let θ be a generic notation for the one-dimensional parameters of interest in our models (the components of $\boldsymbol{\theta}$) and let $\hat{\theta}$ be the corresponding maximum likelihood estimator.

We compute the empirical bias, ebias, for each of the estimators by

$$ebias(\hat{\boldsymbol{\theta}}) = \overline{\hat{\boldsymbol{\theta}}} - \boldsymbol{\theta} = \frac{1}{D}\sum_{d=1}^{D} \hat{\boldsymbol{\theta}}_d - \boldsymbol{\theta}, \tag{3.29}$$

where D denotes the number of simulated data sets and $\hat{\boldsymbol{\theta}}_d$ represents the estimated parameter vector from the d-th simulation and $\boldsymbol{\theta}$ is the true value of the parameter vector.

Using the notation in 3.29, we also compute the empirical standard deviation of all the parameter estimates in the model, as

$$esd(\hat{\theta}) = \sqrt{\frac{1}{D-1}\sum_{d=1}^{D}\left(\hat{\theta}_d - \overline{\hat{\theta}}\right)^2}$$

and the square root of their MSE :

$$mse^{1/2}(\widehat{\theta}) = \sqrt{\frac{1}{D}\sum_{d=1}^{D}(\widehat{\theta}_d - \theta)^2} = \sqrt{SD(\widehat{\theta})^2 + Bias(\widehat{\theta})^2}.$$

The asymptotic covariance matrix of the MLE $\hat{\boldsymbol{\theta}}$ is obtained as the inverse of the information matrix and is estimated by its plug-in estimator, $\boldsymbol{I}_n^{-1}(\hat{\boldsymbol{\theta}})$. Ideally, we would compute the average of the estimated asymptotic standard deviations for each of the estimators in our model by

$$\overline{\widehat{Var}(\widehat{\theta})^{1/2}} = \overline{\boldsymbol{I}(\widehat{\theta})^{-1/2}} = \frac{1}{D}\sum_{d=1}^{D}\widehat{Var}(\widehat{\theta}_d)^{1/2} = \frac{1}{D}\sum_{d=1}^{D}\boldsymbol{I}(\widehat{\theta}_d)^{-1/2}$$

and compare this value with the empirical standard deviations of our estimates. However, calculating $I_n^{-1}(\hat{\boldsymbol{\theta}})$ is an extremely challenging computation. For only one data set with 100 subjects per group and $T = 4$ repeated measurements, provided that the average number of events is large, the program that calculates the information matrix needs to run overnight. Thus, we will not report this quantity. Instead we plug in the true parameters and report $I_n^{-1}(\boldsymbol{\theta})$.

3.3.2 Data Generation

In the multiple population model with independent lengths, the probability to observe a vector of random lengths $\boldsymbol{K} = (k_1, ..., k_T)$ in population i is $P(\boldsymbol{K} = \boldsymbol{k}) = \prod_{t=1}^{T} e^{-\lambda_{it}} \frac{\lambda_{it}^{k_t}}{k_t!}$, where $\lambda_{it} = e^{\delta+\gamma\mu_{it}}$. With probability $P(\boldsymbol{K} = \boldsymbol{0}) = \prod_{t=1}^{T} e^{-\lambda_{it}}$ we observe only zero-length severities at all T measurement times.

Given the vector of random lengths $\boldsymbol{K} = (k_1, ..., k_T)$, and provided that not all the components of the vector $\boldsymbol{k}$ are zero, the conditional distribution of each random vector $\boldsymbol{X}$ from population i with random lengths $\boldsymbol{K} = (k_1, ..., k_T)$ corresponding to all T measurements of severities that were recorded for the respective subject is $\mathbf{MVN}_{\sum_{i=1}^{T} k_i}(\boldsymbol{\mu}_{\boldsymbol{k}}, \sigma^2\mathbf{S}_{\boldsymbol{k}}(\rho, \rho^*))$, where the expression of $\boldsymbol{\mu}_{\boldsymbol{k}}$ and $\mathbf{S}_{\boldsymbol{k}}(\rho, \rho^*)$ are given by 3.17 and 3.18, respectively.

To generate a data set containing n observations per population from the multiple population model, we repeat the following two-step procedure, n times for each population i.

Step 1 (Generate the random lengths)

Take $\lambda_{it} = e^{\delta+\gamma\mu_{it}}$, for $i = 1, 2$ and $t = 1, ..., T$.

Generate T independent observations, each from a Poisson(λ_{it}) distribution $t = 1, ..., T$. This is the vector of random lengths $\boldsymbol{k}_i = (k_{i1}, k_{i2}, \ldots, k_{iT})$.

Step 2 (Generate the severities)

For $i = 1, 2$ and $t = 1, \ldots, T$, if $\sum_{t=1}^{T} k_{it} = 0$ we observe a zero length vector of severities. If $\sum_{t=1}^{T} k_{it} > 0$, then generate X_i from a $\mathbf{MVN}_{\sum_{t=1}^{T} k_{it}} (\boldsymbol{\mu}_{\boldsymbol{k}_i}, \sigma^2 \mathbf{S}_{\boldsymbol{k}_i}(\rho, \rho^*))$.

3.3.3 Numerical Considerations

Using a shared parameter model in the specification of the distribution of the random lengths and the conditional distribution of the vectors of severities given the random lengths increases the efficiency, because we pool information from both outcomes. The cost is an increase in the computational burden, since the shared parameters need to be estimated simultaneously from the two models.

Specifically, there are several difficulties that arise in modeling longitudinal random length data. First, the full likelihood approach to random length data is computationally complex when the number of events is large, since higher dimension matrices are involved in the conditional distribution of the multivariate severity measures.

Second, the complexity increases as the number of repeated measurements becomes larger and the number of parameters to be estimated increases. Since the number of time points with quantitative measures changes with every subject, the mean and covariance structures for the distribution of the severities change. Moreover, numerical computing of the information matrix involves summation over all possible values of the random lengths. For T measurement times that means summation of matrices over all the possible values of a T-dimensional vector of lengths. If the average number of events is large, in numerically computing the information matrix one needs to set the maximum values that the random lengths can have quite high in order to ensure that the corresponding probabilities in the right tails of the Poisson distributions are zero.

Finally, a third difficulty is associated to modeling of the slopes. We have encountered examples when the maximization procedure produced results different than the ones expected. If $\mu_{11} = \ldots \mu_{1T} = \ldots. = \mu_{mT}$, the parameters δ and γ are not identifiable. When data are generated from populations with means not well separated, the likelihood can be flat over certain regions and the R function we use to numerically maximize the log-likelihood converges to some strange solutions, particularly for the parameters δ and γ. However, even in these cases, the average event lengths are estimated correctly. We present such a case in the next section.

All the simulations were performed in R programming language (`www.r-project.org`) using PittGrid, which is the University of Pittsburgh's campus-wide computing environment. PittGrid provides the ability to access additional CPU time and memory in order to run complex calculations using existing, underutilized CPUs participating in the PittGrid network. Users are allowed to submit jobs online and monitor their jobs. Upon submission of a job, this service searches for available workstations and clusters that participate in PittGrid and that can meet the requirements of the job. Since every one of the jobs we needed to run involved complex computations for 1000 data sets, we separated each of the jobs into 20 sub-jobs, each involving only 50 data sets, which we submitted to PittGrid. After the jobs were completed, we gathered the results. The gain in efficiency was tremendous: while a normal computer needs more than five days to complete a single job involving 1000 data sets with 100 subjects per group and $T = 4$ repeated measurements, using PittGrid allowed us to run the same job divided into 20 sub-jobs in less than 24 hours.

The R function *nlm* was used to carry out the unrestricted maximization of the log-likelihood. This function carries out a minimization of the negative log-likelihood using a Newton-type algorithm. It requires specification of initial values for the parameters to be optimized over. Numerical derivatives were used in the calculation of updated parameter values. We started with good initial values and performed unrestricted maximizations. In our extended simulations we found no violations of the restrictions: $\sigma^2 > 0$, $0 \leq \rho^* \leq \rho < 1$. Alternatively, one could use the R function *optim* to perform the optimization with restrictions. This function allows specification of box constraints; that is each variable can be given a lower and/or upper bound. The initial value must satisfy the constraints. *Optim* uses a modification of the BFGS quasi-Newton method. Method "BFGS" is a quasi-Newton method (also known as a variable metric algorithm) that uses function values and gradients to build up a picture of the surface to be optimized. We conducted limited simulations and compared the results obtained from *nlm* and *optim*. We found no major differences between the results, so we decided to use *nlm*, because it has a faster convergence rate.

To obtain good initial estimates of the parameters, we adopt the following scheme. For the parameters μ_{it} reflecting the underlying disease status, the initial values are obtained using

$$\mu_{it}^{(0)} = \frac{\sum\limits_{j=1}^{n} z_{ijt}}{\sum\limits_{j=1}^{n} \delta(K_{ijt})}, \tag{3.30}$$

We find the starting values for δ and γ by fitting a Poisson regression model,

$$\log(\lambda_{ijt}) = \delta + \gamma\mu_{it}^{(0)} \quad i = 1, 2 \quad j = 1, 2, ..., n \quad T = 1, ..., 4, \tag{3.31}$$

where $\lambda_{ijt} = E(K_{ijt})$.
We solve for σ^2 in the corresponding likelihood equation

$$\sigma^2 = \sigma^2(\rho, \rho^*) = \frac{\sum\limits_{i=1}^{m} \sum\limits_{j=1}^{n_i} (z_{ij} - \mu_{ij}^{*(0)})' \boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}^{-1} (z_{ij} - \mu_{ij}^{*(*)}) + \tau_0 \sum\limits_{i=1}^{m} \widehat{\boldsymbol{y}}_{ij}' \widehat{\boldsymbol{y}}_{ij}}{\sum\limits_{i=1}^{m} \sum\limits_{j=1}^{n_i} \sum\limits_{t=1}^{2} k_{ijt}} \tag{3.32}$$

and replace this value in the expression of the conditional likelihood to get an expression that depends only on the unknown ρ and ρ^*

$$\begin{aligned} l(\rho, \rho^*) \quad = \quad & -\log \sigma^2 \sum_{i=1}^{m} \sum_{j=1}^{n_i} \sum_{t=1}^{T} k_{ijt} + \log \tau_0 \sum_{i=1}^{m} \sum_{j=1}^{n_i} \left(\sum_{t=1}^{T} k_{ijt} - \sum_{t=1}^{T} \delta(k_{ijt}) \right) - \sum_{i=1}^{m} \sum_{j=1}^{n_i} \log |\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}| \\ & - \frac{1}{2\sigma^2} \sum_{i=1}^{m} \sum_{j=1}^{n_i} (\boldsymbol{z}_{ij} - \boldsymbol{\mu}_{ij}^{*(0)})' \boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}^{-1} (\boldsymbol{z}_{ij} - \boldsymbol{\mu}_{ij}^{*(0)}) - \frac{\tau_0}{2\sigma^2} \sum_{i=1}^{m} \sum_{j=1}^{n_i} \sum_{t=1}^{l(\boldsymbol{k}_{ij})} \sum_{l=2}^{k_{ijr_{ijt}}} y_{ijr_{ijt}l}^2. \end{aligned}$$

Maximizing the above likelihood with respect to ρ and ρ^* by using *nlm* with 0 as starting values for both parameters gives us the initial values $\rho^{(0)}$ and $\rho^{*(0)}$. Plugging these values into (3.32) gives us the initial value $\sigma^{2(0)}$. If any of the initial values for σ^2, ρ and ρ^* were negative, they were assigned the value zero. Once all the initial estimates are set, the optimization procedure proceeds by using the R function *nlm*.

3.3.4 Simulation Results

We report here the results of the simulations for $T = 4$. The results of the simulations for two time points will be presented in Chapter 4, to allow for comparison with the model introduced there. Tables 6 to 17 pool the results across the 4 parameter value combinations and the 3 different choices of the sample size n and Table 5 summarizes these results. Specifically, Tables 6 to 8 show results for the first scenario, when $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are well separated and the average severities within each group are close while the average numbers of events are far apart across time. Tables 9 to 11 show results for the second scenario, when $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are close, and the average severities within each group are close while the average numbers of events are far apart across time. The results for scenario 3 are displayed in Tables 12 to 14. In this case $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are well separated, and the average severities are far apart while the average numbers of events are similar across time and groups. Finally, the results for scenario 4 are presented in Tables 15 to 17. In this scenario $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are close, and the average severities are far apart while the average numbers of events are similar across time and groups. The tables report the average parameter estimates, empirical standard deviations and square-root MSE. By comparing the standard deviations of the parameter estimates to the theoretical large-sample standard deviations, we have one assessment of the adequacy of the asymptotic covariance matrix for the sample size under consideration as approximation to the finite sample covariance.

The main parameters of interest are the μ's, representing the underlying disease status. In each of the cases we are able to estimate their true values with biases being practically insignificant. The results obtained from the simulation studies indicate that the estimates of the true μ's were unbiased for any sample size ($n = 20, 50, 100$), under all four different scenarios. The same was true for the parameters σ^2, ρ and ρ^*. The bias of these latter estimates was slightly higher for the smaller values of the sample size. In contrast, as shown by Tables 6, 9 and 10, the estimates for δ and γ were quite different from the true values for small sample sizes. For small values of n, the estimates for δ and γ are strongly biased in both scenario 1 and scenario 2. In scenarios 3 and 4 (see Tables 12 - 14 and 15 - 17), the estimation for small sample sizes works as well as in the cases with n large. The main source of this difference of results obtained from the first 2 scenarios and the last 2 is the disproportion between the average severities and the variance σ^2. In the first 2 scenarios, the theoretical severities are not well separated across time and treatment group with respect to σ^2, while in the latter scenarios they are well separated. In addition, in scenario 2 the severities for the two groups intersect across time, making the estimation more difficult.

Table 5: Summary of the simulations' results. T=4

Scenario 1 $\boldsymbol{n = 20}$ • bias only in δ and γ • asymptotic variance does not approximate the finite sample variance for δ and γ • δ and γ not normal	*Scenario 3* $\boldsymbol{n = 20}$ • no bias • asymptotic variance approximates the finite sample variance • all estimates look normally distributed
$\boldsymbol{n = 50}$ • no bias • asymptotic variance approximates the finite sample variance • all estimates look normally distributed	$\boldsymbol{n = 50}$ • no bias • asymptotic variance approximates the finite sample variance • all estimates look normally distributed
$\boldsymbol{n = 100}$ • no bias • asymptotic variance approximates the finite sample variance • all estimates look normally distributed	$\boldsymbol{n = 100}$ • no bias • asymptotic variance approximates the finite sample variance • all estimates look normally distributed
Scenario 2 $\boldsymbol{n = 20}$ • bias only in δ and γ • asymptotic variance does not approximate the finite sample variance for δ and γ • δ and γ not normal	*Scenario 4* $\boldsymbol{n = 20}$ • no bias • asymptotic variance approximates the finite sample variance • all estimates look normally distributed
$\boldsymbol{n = 50}$ • small bias for δ and γ • asymptotic variance approximates reasonably the finite sample variance for δ and γ • all estimates look normally distributed	$\boldsymbol{n = 50}$ • no bias • asymptotic variance approximates the finite sample variance • all estimates look normally distributed
$\boldsymbol{n = 100}$ • no bias • asymptotic variance approximates the finite sample variance • all estimates look normally distributed	$\boldsymbol{n = 100}$ • no bias • asymptotic variance approximates the finite sample variance • all estimates look normally distributed

QQ-plots (not presented), were generated for each simulated scenario. By examining them, we found the empirical distribution of all the MLE's except δ and γ to be symmetrical and approximately normal, as expected. Non-normal behavior was found for δ and γ in scenario 1 and scenario 2 for small sample sizes. The asymmetry decreased with increasing sample size. Investigating the behavior of the sample standard deviations produces conclusions analogous to the ones obtained from exploring the bias and normality. The empirical standard deviations are close to the large sample theoretical ones in all scenarios and for all sample sizes except for the parameters δ and γ in scenarios with small sample size ($n = 20$) and large variability in the events' severities (scenario 1 and 2). Summarizing all the above we conclude that the asymptotic results in section 3.2 are applicable for sample sizes which are greater than 50.

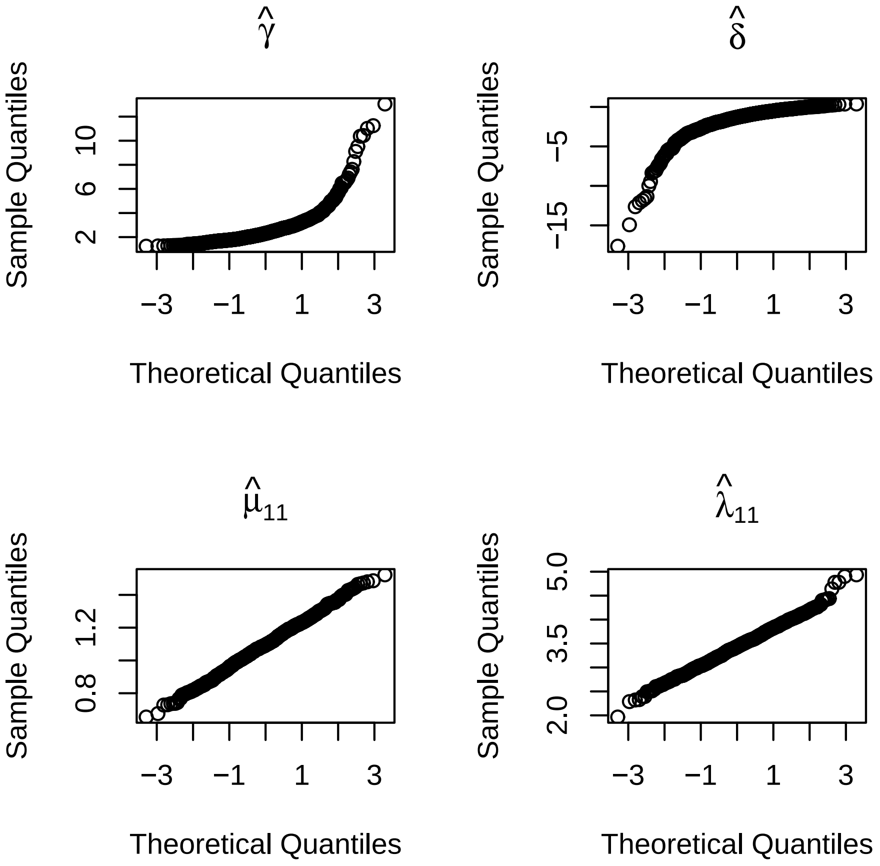

Figure 2: QQ plots for scenario 1 with $n = 20$. Although estimates for δ and γ are not normal, estimates for μ_{11} and λ_{11} are normal

Estimation results show that the algorithm gives acceptable results even for choices of n as small as 20. Figure 2 presents qq-plots for one of the "problem" scenarios, scenario 1, when the sample size is small ($n = 20$) and the average severities across time and treatment groups are close relative to their variances. We show only the qq-plots for the parameters δ and γ (not well estimated), and for one of the μ's (μ_{11}), as well as for the corresponding length, λ_{11}). Although the qq-plots for the parameters δ and γ do not look normal, as expected, we notice that both the plots for the estimated severity and length are reasonably close to a straight line, an indication of their normality. As noted before, even when the estimates for the scaling parameters are not close to the theoretical values, the method produces correct estimates for the main parameters of interest, the mean severities and lengths.

Table 6: Simulation results for the first choice of parameters, $T = 4, n = 20$

	μ_{11}	μ_{12}	μ_{13}	μ_{14}	μ_{21}	μ_{22}	μ_{23}	μ_{24}	δ	γ	σ^2	ρ	ρ^*
True(θ)	1	1.1	1.2	1.3	1.5	1.6	1.7	1.8	-1.3	2.3	1	0.5	0.2
$\lambda = e^{\delta+\gamma\mu}$	2.72	3.42	4.31	5.42	8.58	10.80	13.60	17.12					
average	0.993	1.097	1.199	1.293	1.498	1.597	1.697	1.794	-1.692	2.577	0.964	0.48	0.179
bias	-0.007	-0.003	-0.001	-0.007	-0.002	-0.003	-0.003	-0.006	-0.392	0.277	-0.036	-0.02	-0.021
sd	0.163	0.139	0.119	0.107	0.097	0.106	0.119	0.137	1.669	1.151	0.073	0.04	0.053
mse$^{1/2}$	0.163	0.139	0.119	0.107	0.097	0.106	0.119	0.138	1.714	1.184	0.081	0.045	0.057
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.151	0.131	0.114	0.101	0.095	0.104	0.119	0.138	0.95	0.652	0.079	0.04	0.058

Table 7: Simulation results for the first choice of parameters, $T = 4, n = 50$

	μ_{11}	μ_{12}	μ_{13}	μ_{14}	μ_{21}	μ_{22}	μ_{23}	μ_{24}	δ	γ	σ^2	ρ	ρ^*
True(θ)	1	1.1	1.2	1.3	1.5	1.6	1.7	1.8	-1.3	2.3	1	0.5	0.2
$\lambda = e^{\delta+\gamma\mu}$	2.72	3.42	4.31	5.42	8.58	10.80	13.60	17.12					
average	0.993	1.096	1.195	1.298	1.502	1.602	1.703	1.804	-1.371	2.35	0.975	0.488	0.188
bias	-0.007	-0.004	-0.005	-0.002	0.002	0.002	0.003	0.004	-0.071	0.05	-0.025	-0.012	-0.012
sd	0.094	0.08	0.071	0.063	0.057	0.063	0.072	0.084	0.688	0.474	0.046	0.025	0.033
mse$^{1/2}$	0.094	0.08	0.071	0.063	0.057	0.063	0.072	0.084	0.691	0.476	0.053	0.028	0.035
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.095	0.083	0.072	0.064	0.060	0.066	0.075	0.087	0.601	0.412	0.050	0.025	0.036

Table 8: Simulation results for the first choice of parameters, $T = 4, n = 100$

	μ_{11}	μ_{12}	μ_{13}	μ_{14}	μ_{21}	μ_{22}	μ_{23}	μ_{24}	δ	γ	σ^2	ρ	ρ^*
True(θ)	1	1.1	1.2	1.3	1.5	1.6	1.7	1.8	-1.3	2.3	1	0.5	0.2
$\lambda = e^{\delta+\gamma\mu}$	2.72	3.42	4.31	5.42	8.58	10.80	13.60	17.12					
average	1	1.1	1.201	1.3	1.501	1.6	1.701	1.801	-1.356	2.337	0.978	0.49	0.189
bias	0	0	0.001	0	0.001	0	0.001	0.001	-0.056	0.037	-0.022	-0.01	-0.011
sd	0.068	0.061	0.051	0.046	0.043	0.046	0.053	0.061	0.452	0.308	0.033	0.017	0.025
mse$^{1/2}$	0.068	0.061	0.051	0.046	0.043	0.046	0.053	0.061	0.455	0.311	0.04	0.02	0.027
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.067	0.059	0.051	0.045	0.042	0.046	0.053	0.062	0.425	0.292	0.035	0.018	0.026

Table 9: Simulation results for the second choice of parameters, $T = 4, n = 20$

	μ_{11}	μ_{12}	μ_{13}	μ_{14}	μ_{21}	μ_{22}	μ_{23}	μ_{24}	δ	γ	σ^2	ρ	ρ^*
True(θ)	1.3	1.2	1.1	1	1	1.1	1.2	1.3	-1.0	2.5	1	0.5	0.2
$\lambda = e^{\delta+\gamma\mu}$	4.48	5.475	7.39	9.49	9.49	7.39	5.75	4.48					
average	1.292	1.191	1.089	0.991	0.988	1.091	1.192	1.292	-2.377	3.717	0.984	0.49	0.191
bias	-0.008	-0.009	-0.011	-0.009	-0.012	-0.009	-0.008	-0.008	-1.377	1.217	-0.016	-0.01	-0.009
sd	0.11	0.095	0.097	0.119	0.119	0.098	0.095	0.11	13.21	11.552	0.077	0.041	0.056
mse$^{1/2}$	0.111	0.095	0.097	0.12	0.12	0.098	0.095	0.11	13.281	11.616	0.078	0.042	0.057
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.109	0.093	0.096	0.116	0.116	0.096	0.093	0.109	1.349	1.148	0.079	0.041	0.057

Table 10: Simulation results for the second choice of parameters, $T = 4, n = 50$

	μ_{11}	μ_{12}	μ_{13}	μ_{14}	μ_{21}	μ_{22}	μ_{23}	μ_{24}	δ	γ	σ^2	ρ	ρ^*
True(θ)	1.3	1.2	1.1	1	1	1.1	1.2	1.3	-1.0	2.5	1	0.5	0.2
$\lambda = e^{\delta+\gamma\mu}$	4.48	5.475	7.39	9.49	9.49	7.39	5.75	4.48					
average	1.301	1.2	1.099	0.997	0.997	1.099	1.201	1.299	-1.336	2.787	0.99	0.494	0.194
bias	0.001	0	-0.001	-0.003	-0.003	-0.001	0.001	-0.001	-0.336	0.287	-0.01	-0.006	-0.006
sd	0.07	0.058	0.062	0.075	0.074	0.062	0.059	0.07	1.481	1.247	0.05	0.026	0.035
mse$^{1/2}$	0.07	0.058	0.062	0.075	0.074	0.062	0.059	0.07	1.518	1.28	0.051	0.026	0.035
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.069	0.059	0.061	0.073	0.073	0.061	0.059	0.069	0.853	0.726	0.050	0.026	0.036

Table 11: Simulation results for the second choice of parameters, $T = 4, n = 100$

	μ_{11}	μ_{12}	μ_{13}	μ_{14}	μ_{21}	μ_{22}	μ_{23}	μ_{24}	δ	γ	σ^2	ρ	ρ^*
True(θ)	1.3	1.2	1.1	1	1	1.1	1.2	1.3	-1.0	2.5	1	0.5	0.2
$\lambda = e^{\delta+\gamma\mu}$	4.48	5.475	7.39	9.49	9.49	7.39	5.75	4.48					
average	1.299	1.199	1.099	0.999	0.999	1.099	1.199	1.3	-1.131	2.614	0.991	0.495	0.195
bias	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001	0	-0.131	0.114	-0.009	-0.005	-0.005
sd	0.048	0.042	0.043	0.051	0.052	0.042	0.041	0.048	0.693	0.591	0.035	0.019	0.027
mse$^{1/2}$	0.048	0.042	0.043	0.051	0.052	0.042	0.041	0.048	0.706	0.602	0.036	0.019	0.027
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.067	0.059	0.051	0.045	0.042	0.046	0.053	0.062	0.425	0.292	0.035	0.018	0.026

Table 12: Simulation results for the third choice of parameters, $T = 4, n = 20$

	μ_{11}	μ_{12}	μ_{13}	μ_{14}	μ_{21}	μ_{22}	μ_{23}	μ_{24}	δ	γ	σ^2	ρ	ρ^*
True(θ)	1	2	3	4	1.5	2.5	3.5	4.5	2.5	0.01	1	0.5	0.2
$\lambda = e^{\delta+\gamma\mu}$	12.30	12.43	12.55	12.68	12.37	12.49	12.62	12.74					
average	0.998	1.995	2.997	4.001	1.508	2.5	3.495	4.5	2.498	0.011	0.972	0.483	0.193
bias	-0.002	-0.005	-0.003	0.001	0.008	0	-0.005	0	-0.002	0.001	-0.028	-0.017	-0.007
sd	0.168	0.163	0.171	0.167	0.166	0.171	0.172	0.165	0.058	0.02	0.073	0.038	0.052
mse$^{1/2}$	0.168	0.163	0.171	0.167	0.166	0.171	0.172	0.165	0.058	0.02	0.078	0.041	0.052
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.165	0.165	0.165	0.165	0.165	0.165	0.165	0.165	0.058	0.02	0.074	0.037	0.054

Table 13: Simulation results for the third choice of parameters, $T = 4, n = 50$

	μ_{11}	μ_{12}	μ_{13}	μ_{14}	μ_{21}	μ_{22}	μ_{23}	μ_{24}	δ	γ	σ^2	ρ	ρ^*
True(θ)	1	2	3	4	1.5	2.5	3.5	4.5	2.5	0.01	1	0.5	0.2
$\lambda = e^{\delta+\gamma\mu}$	12.30	12.43	12.55	12.68	12.37	12.49	12.62	12.74					
average	1.002	2.001	3.004	4.001	1.496	2.497	3.501	4.501	2.499	0.01	0.987	0.492	0.196
bias	0.002	0.001	0.004	0.001	-0.004	-0.003	0.001	0.001	-0.001	0	-0.013	-0.008	-0.004
sd	0.104	0.104	0.103	0.107	0.103	0.106	0.102	0.103	0.038	0.013	0.046	0.024	0.034
mse$^{1/2}$	0.105	0.104	0.103	0.107	0.103	0.106	0.102	0.103	0.038	0.013	0.047	0.025	0.035
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.104	0.104	0.104	0.104	0.104	0.104	0.104	0.104	0.037	0.012	0.047	0.023	0.034

Table 14: Simulation results for the third choice of parameters, $T = 4, n = 100$

	μ_{11}	μ_{12}	μ_{13}	μ_{14}	μ_{21}	μ_{22}	μ_{23}	μ_{24}	δ	γ	σ^2	ρ	ρ^*
True(θ)	1	2	3	4	1.5	2.5	3.5	4.5	2.5	0.01	1	0.5	0.2
$\lambda = e^{\delta+\gamma\mu}$	12.30	12.43	12.55	12.68	12.37	12.49	12.62	12.74					
average	1.002	2	2.998	3.999	1.501	2.5	3.504	4.501	2.499	0.01	0.996	0.498	0.199
bias	0.002	0	-0.002	-0.001	0.001	0	0.004	0.001	-0.001	0	-0.004	-0.002	-0.001
sd	0.074	0.075	0.076	0.072	0.072	0.074	0.077	0.072	0.024	0.008	0.031	0.016	0.024
mse$^{1/2}$	0.075	0.075	0.076	0.072	0.072	0.074	0.077	0.072	0.024	0.008	0.032	0.016	0.024
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.074	0.026	0.009	0.033	0.017	0.024

Table 15: Simulation results for the fourth choice of parameters, $T = 4, n = 20$

	μ_{11}	μ_{12}	μ_{13}	μ_{14}	μ_{21}	μ_{22}	μ_{23}	μ_{24}	δ	γ	σ^2	ρ	ρ^*
True(θ)	4	3	2	1	1	2	3	4	1.5	0.05	1	0.5	0.2
$\lambda = e^{\delta+\gamma\mu}$	5.47	5.21	4.95	4.71	4.71	4.95	5.21	5.47					
average	4.004	3.002	2.002	1.006	0.997	1.996	2.995	3.993	1.504	0.049	0.96	0.478	0.188
bias	0.004	0.002	0.002	0.006	-0.003	-0.004	-0.005	-0.007	0.004	-0.001	-0.04	-0.022	-0.012
sd	0.171	0.172	0.181	0.19	0.179	0.18	0.179	0.176	0.09	0.032	0.081	0.046	0.058
mse$^{1/2}$	0.171	0.172	0.181	0.19	0.179	0.181	0.179	0.176	0.09	0.032	0.09	0.051	0.059
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.174	0.175	0.176	0.177	0.177	0.176	0.175	0.174	0.088	0.031	0.082	0.043	0.059

Table 16: Simulation results for the fourth choice of parameters, $T = 4, n = 50$

	μ_{11}	μ_{12}	μ_{13}	μ_{14}	μ_{21}	μ_{22}	μ_{23}	μ_{24}	δ	γ	σ^2	ρ	ρ^*
True(θ)	4	3	2	1	1	2	3	4	1.5	0.05	1	0.5	0.2
$\lambda = e^{\delta+\gamma\mu}$	5.47	5.21	4.95	4.71	4.71	4.95	5.21	5.47					
average	3.998	2.997	2.004	1.002	0.997	1.994	2.996	3.99	1.501	0.05	0.978	0.489	0.193
bias	-0.002	-0.003	0.004	0.002	-0.003	-0.006	-0.004	-0.01	0.001	0	-0.022	-0.011	-0.007
sd	0.113	0.113	0.113	0.112	0.115	0.114	0.113	0.111	0.055	0.02	0.051	0.028	0.038
mse$^{1/2}$	0.113	0.113	0.113	0.112	0.115	0.114	0.113	0.112	0.055	0.02	0.056	0.03	0.038
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.11	0.111	0.111	0.112	0.112	0.111	0.111	0.11	0.056	0.02	0.052	0.027	0.037

Table 17: Simulation results for the fourth choice of parameters, $T = 4, n = 100$

	μ_{11}	μ_{12}	μ_{13}	μ_{14}	μ_{21}	μ_{22}	μ_{23}	μ_{24}	δ	γ	σ^2	ρ	ρ^*
True(θ)	4	3	2	1	1	2	3	4	1.5	0.05	1	0.5	0.2
$\lambda = e^{\delta+\gamma\mu}$	5.47	5.21	4.95	4.71	4.71	4.95	5.21	5.47					
average	3.997	2.996	2	0.998	1	1.999	3.003	4	1.499	0.05	0.983	0.491	0.194
bias	-0.003	-0.004	0	-0.002	0	-0.001	0.003	0	-0.001	0	-0.017	-0.009	-0.006
sd	0.077	0.079	0.08	0.081	0.078	0.077	0.078	0.08	0.04	0.014	0.037	0.019	0.027
mse$^{1/2}$	0.078	0.079	0.08	0.081	0.078	0.077	0.078	0.08	0.04	0.014	0.041	0.021	0.027
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.078	0.078	0.079	0.079	0.079	0.079	0.078	0.078	0.039	0.014	0.037	0.019	0.026

3.4 APPLICATION TO LEDS DATA

In this section we apply the multiple population model to the LEDS data. The data were collected at the University of Pittsburgh Western Psychiatric Institute and Clinic as part of a larger study investigating stressful live events in depressed adolescents. The objective of this study was to examine the occurrence of acute and chronic stressors prior to and during a recent depressive episode in adolescents. Stress exposure was examined using the investigator-based Life Events and Difficulties Schedule (LEDS), adapted for use with adolescents. This interview is designed to exhaustively draw out information for stressors. The severity of each stressor was rated on a 4-point scale (4-marked, 3-moderate, 2-some, 1-little or none). The subjects we consider are 32 depressed (MDD) females and 30 normal control (NC) females between the ages of 13 and 18 years. The occurrence of stressors was examined quarterly in the 12-month period prior to the onset of the depressive episode in depressed adolescents and during a comparable ”linked” period in normal control adolescents.

Our model depends on the normality assumption for the distribution of the multivariate severities. In LEDS data, the severity measurements are actually ordinal (1=”little or none”, 2=”some”, 3=”moderate” and 4=”marked”) and we approximate them with continuous normal random variables. This is common use in life events data analysis. Furthermore, as many of the pain severity measurements encountered in practice are assessed by visual analogous scale, there is a interest in models for continuous severity measurements.

The two outcomes recorded for each subject are the number of stessors and the severity of each of the stessors. These outcomes are recorded for each of the four quarters of the year. Each quarter, the data for a individual subject are random vectors of event severities with the random length given by the number of events the subject experienced that quarter of year. We denote the MDD group as population 1 with mean μ_{11} reflecting the underlying depression status at the first quarter before the onset of depression, μ_{12} reflecting the underlying depression status at the second quarter before the onset of depression, μ_{13} at the third quarter before the onset of depression, and mean μ_{14} at the forth quarter before the onset of depression. Similarly, we denote the NC group as population 2 with means μ_{21} at the first quarter before the onset of depression, μ_{22} at the second quarter before the onset of depression, μ_{23} at the third quarter before the onset of depression, and mean μ_{24} at the forth quarter before the onset of depression. The multiple population model introduced in Section 3.2 can be applied now to this data set. We have $n_1 = 32$ and $n_2 = 30$. The data are summarized in Tables 1 and 2.

As in the previous section, we use the R function *nlm* to maximize the likelihood. The initial values for the parameters are computed using the technique described in Section 3.3.3. Table 18 reports the solution $\hat{\boldsymbol{\theta}}$ of the maximization procedure and the estimated standard deviations based on $\boldsymbol{I}_n^{-1}(\hat{\boldsymbol{\theta}})$, as well as the initial values for the maximization procedure. The estimated asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$ is computed using (A.24) to be

Table 18: Maximum likelihood estimates for LEDS data

	Parameter	$\hat{\theta}$	estimated SD of $\hat{\theta}$	Initial values
	δ	-16.6291	27.91	-0.1986
	γ	9.4031	15.08	0.5683
MDD				
	μ_{11}	1.8799	0.06	1.8039
	μ_{12}	1.8804	0.06	1.7364
	μ_{13}	1.8515	0.04	1.8917
	μ_{14}	1.8563	0.04	1.8605
NC				
	μ_{21}	1.8287	0.06	1.6917
	μ_{22}	1.8481	0.04	1.4479
	μ_{23}	1.8271	0.06	1.4430
	μ_{24}	1.8034	0.09	1.6814
	σ^2	0.5250	0.03	0.5324
	ρ	0.1702	0.05	0.1846
	ρ^*	0.0689	0.04	0.0787

$$\boldsymbol{I}_n^{-1}(\hat{\boldsymbol{\theta}}) = \begin{pmatrix} \boldsymbol{I}_n^{-1}(\hat{\delta}, \hat{\gamma}, \hat{\mu}_{11}, ..., \hat{\mu}_{24}) & \mathbf{O}_{10\times 3} \\ \mathbf{O}_{3\times 10} & \boldsymbol{I}_n^{-1}(\hat{\sigma}^2, \hat{\rho}, \hat{\rho}^*) \end{pmatrix},$$

where $\boldsymbol{I}_n^{-1}(\hat{\delta}, \hat{\gamma}, \hat{\mu}_{11}, ..., \hat{\mu}_{24})$ is given by

$$\begin{pmatrix} 779.2287 & -420.8131 & 1.2455 & 1.2666 & -0.0031 & 0.2079 & -1.0157 & -0.1637 & -1.088 & -2.1259 \\ -420.8131 & 227.3005 & -0.6816 & -0.693 & -0.0072 & -0.1212 & 0.5398 & 0.0796 & 0.5789 & 1.1395 \\ 1.2455 & -0.6816 & 0.0039 & 0.0038 & 0.0017 & 0.0021 & 1e-04 & 0.0015 & 0 & -0.0017 \\ 1.2666 & -0.693 & 0.0038 & 0.004 & 0.0017 & 0.0021 & 1e-04 & 0.0015 & 0 & -0.0018 \\ -0.0031 & -0.0072 & 0.0017 & 0.0017 & 0.0019 & 0.0017 & 0.0017 & 0.0017 & 0.0017 & 0.0017 \\ 0.2079 & -0.1212 & 0.0021 & 0.0021 & 0.0017 & 0.0019 & 0.0014 & 0.0017 & 0.0014 & 0.0011 \\ -1.0157 & 0.5398 & 1e-04 & 1e-04 & 0.0017 & 0.0014 & 0.0032 & 0.0019 & 0.0031 & 0.0044 \\ -0.1637 & 0.0796 & 0.0015 & 0.0015 & 0.0017 & 0.0017 & 0.0019 & 0.0019 & 0.0019 & 0.0021 \\ -1.088 & 0.5789 & 0 & 0 & 0.0017 & 0.0014 & 0.0031 & 0.0019 & 0.0034 & 0.0046 \\ -2.1259 & 1.1395 & -0.0017 & -0.0018 & 0.0017 & 0.0011 & 0.0044 & 0.0021 & 0.0046 & 0.0077 \end{pmatrix}$$

and

$$\boldsymbol{I}_n^{-1}(\hat{\sigma}^2, \hat{\rho}, \hat{\rho}^*) = \begin{pmatrix} 0.0011 & 0.0005 & 0.0003 \\ 0.0005 & 0.0026 & 0.0006 \\ 0.0003 & 0.0006 & 0.0015 \end{pmatrix}.$$

The estimated parameter γ has a positive sign indicating a positive relationship between the average number of events and the average severity (larger severities and higher number of events). Its estimated standard deviation is large, indicating that γ is not significantly different than zero. We conclude that there is no additional information about the severities supplied by knowing the lengths, and similarly, knowing the severities would not provide any information about the average lengths. Another question of interest is testing if the profiles of the two groups are parallel. This is equivalent to testing that there is no interaction between time and group. This composite hypothesis can be written as

$$\begin{aligned} H_0: \quad & (\mu_{12} - \mu_{11}) - (\mu_{22} - \mu_{21}) = 0 \\ & (\mu_{13} - \mu_{12}) - (\mu_{23} - \mu_{22}) = 0 \\ & (\mu_{14} - \mu_{13}) - (\mu_{24} - \mu_{23}) = 0. \end{aligned}$$

We set up the corresponding matrix

$$\boldsymbol{C} = \begin{pmatrix} -1 & 1 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 1 & -1 \end{pmatrix},$$

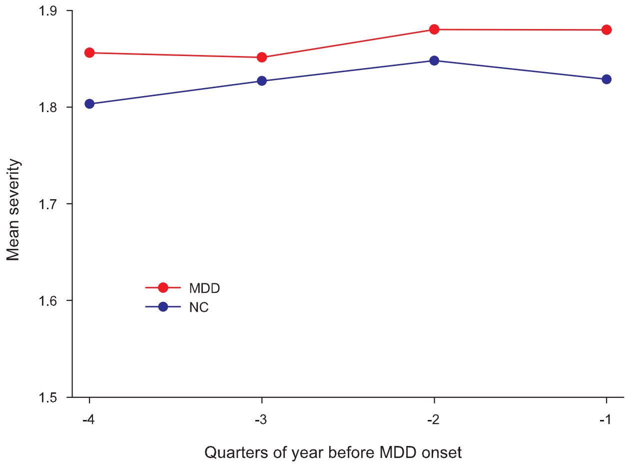

Figure 3: Severity of events profile plot for LEDS data

write the hypothesis in matrix form

$$H_0: \quad \boldsymbol{C\mu} = \mathbf{0},$$

where $\boldsymbol{\mu} = (\mu_{11}, ..., \mu_{14}, \mu_{21}, ..., \mu_{24})$. It follows that the value of the test statistic is

$$(\boldsymbol{C\hat{\mu}})' \left[\boldsymbol{C I}^{-1}(\boldsymbol{\hat{\mu}})\boldsymbol{C}'\right]^{-1} \boldsymbol{C\hat{\mu}} = 0.3237,$$

which is not significant with respect to a chi-square distribution with 3 degrees of freedom. This means that the two profiles are not significantly different in shape and we conclude that the profiles are parallel. Given that the profiles are parallel, we are interested to see if they are also coincident. The corresponding null hypothesis of equal treatment effects is

$$H_0: \quad \mu_{11} + \mu_{12} + \mu_{13} + \mu_{14} = \mu_{21} + \mu_{22} + \mu_{23} + \mu_{24}.$$

We can state this hypothesis in matrix form as

$$H_0: \quad \boldsymbol{c\mu} = \mathbf{0},$$

where $\boldsymbol{c} = (1 \;\; 1 \;\; 1 \;\; 1 \;\; -1 \;\; -1 \;\; -1 \;\; -1)$. The value of the test statistic can be found as

$$(\boldsymbol{c\hat{\mu}})' \left[\boldsymbol{c I}^{-1}(\boldsymbol{\hat{\mu}})\boldsymbol{c}'\right]^{-1} \boldsymbol{c\hat{\mu}} = 0.3904,$$

which is not significant with respect to a chi-square distribution with 1 degree of freedom. We conclude that the two profiles are coincident and we proceed to test the hypothesis of equal response effects

$$H_0: \quad \mu_{11} + \mu_{21} = \mu_{12} + \mu_{22} = \mu_{13} + \mu_{23} = \mu_{14} + \mu_{24}.$$

In matrix form, we state the hypothesis as

$$H_0: \quad \boldsymbol{D}\boldsymbol{\mu} = \mathbf{0},$$

where

$$\boldsymbol{D} = \begin{pmatrix} -1 & 1 & 0 & 0 & -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & -1 & 1 \end{pmatrix}.$$

We obtain the value of the test statistic as

$$(\boldsymbol{D}\hat{\boldsymbol{\mu}})' \left[\boldsymbol{D}\boldsymbol{I}^{-1}(\hat{\boldsymbol{\mu}})\boldsymbol{D}' \right]^{-1} \boldsymbol{D}\hat{\boldsymbol{\mu}} = 0.3243,$$

which is not significant with respect to a chi-square distribution with 3 degrees of freedom.

4.0 BUILDING DEPENDENCE INTO SEVERITIES AND LENGTHS. MULTIPLE POPULATION MODEL WITH DEPENDENT POISSON LENGTHS

An individual's disease condition may affect both the frequency and the severity of events that occur repeatedly and are collected over assessment intervals. We termed this type of data as *longitudinal random length data* and the overall aim of our research is to develop new methods for dealing with this kind of data, methods which incorporate information from both the frequencies of the events and their severities and allow their joint modeling. One of the biggest challenges when constructing models for *longitudinal random length data* is related to the problem of accounting for the dependence among the outcomes for a given subject. Within a subject, not only are the repeated severity measures recorded at different times correlated, so are the severity measures recorded within a single time point measurement, as well as the number of events reported by a subject at different time points.

Dependence over time can be built in two ways: into the vectors of severities and in the random lengths. In Chapter 3 we address this problem by considering that the number of events, i.e. the random lengths are independent across time and set up dependence over time into the severities. Specifically, we assumed that within a subject, conditional on all the number of events experienced by the subject, any two severity measures recorded at two different measurement times have the same correlation, ρ^*, independent of the measurement times. In this chapter we generalize the model of Chapter 3, by adding another layer of dependence over time, built into the random lengths. Note that we are still assuming that within an assessment period, conditional on the number of events experienced, the severities are correlated, and model this dependence in the vector of severities using compound symmetry. Furthermore, we consider that, conditional on the total number of events experienced by an individual the severity measurements taken at different times are dependent, and, additionally, the number of these reported severities are dependent. This assumption is motivated by our belief that in certain scenarios, it is likely that the number of events experienced by a subject over time are dependent. In the context of the LEDS data, it makes sense to assume the number of these stressful life events collected at the end of each quarter of year would be dependent.

We employ a multivariate Poisson distribution to model the lengths dependency. While the univariate Poisson distribution has been widely used as a modeling approach for numbers of events, the use of its multivariate counterpart has been rather limited in the literature. The main obstacle that appears to limit the usage of the multivariate Poisson distribution in practice is the difficulty of calculating its probability function and the complexity of the likelihood (see Johnson et al.,[19]). This has led to the use of a simplified model with just one covariance term for all pairs of variables (see Karlis [20]). Even for this simplified version, likelihood inference is quite complex. We model the joint distribution of the number of events recorded throughout the follow-up using this simplified multivariate Poisson distribution.

Given all the events experienced by a subject throughout the repeated measurements, we assume that the vector of severities has a multivariate normal distribution. We parametrize the model in such a way that, at each measurement time, both the distribution of the number of events and the conditional distribution of the severities depend on an underlying parameter, reflecting the disease status at that time point.

As in Chapter 3, because the notation becomes cumbersome and to help cement the concepts, the analysis of a one population model, introduced in Section 4.1, is extensively presented. The main results of this analysis, the maximum likelihood equations and the asymptotic distributions for the MLE, are then easily generalized to the corresponding multiple population case in Section 4.2.

4.1 ONE POPULATION MODEL WITH DEPENDENT POISSON LENGTHS

This chapter generalizes the model introduced in Section 3.1.1 by adding dependency in the random lengths. The model is appropriate for instances in which only one treatment is involved, and the treatment is evaluated at different points in time. The subjects are followed longitudinally and their vectors of severities are recorded repeatedly for a fixed number of time periods T. The disease status at measurement times 1, 2 ,..., T is reflected by the parameters $\mu_1, ..., \mu_T$. The model we introduce here is similar in structure to the one of Chapter 3 . However it differs from it in that it has an additional parameter reflecting dependence over time. This parameter, λ_0 appears in the specification of the distribution of the total number of events experienced by subjects throughout the repeated measurement times.

In the LEDS data, the two outcomes of interest are the number of stressful life events that subjects experience throughout the four quarters of a year and their respective severity levels. Conceptually, it makes sense to assume that there is an underlying subject depression status affecting both the number of events and how severe these events are. We expect that, as they draw closer to their MDD onset, the subjects in the MDD group will have more stressful life events and their severity levels will be higher.

4.1.1 Model Description

The notation introduced in Chapter 3 is used, with each subject i, $i = 1, ..., n$ being observed T times. At each measurement time $j = 1, ..., T$, each subject i has a random number of events K_{ij} and the corresponding measurements are recorded into the vector $\boldsymbol{X}_{ij}$. Hence all the data for subject i can be condensed into the $\sum\limits_{j=1}^{T} K_{ij}$-dimensional vector $\boldsymbol{X}_i$, $\boldsymbol{X}_i' = \left(\boldsymbol{X}_{i1}', ..., \boldsymbol{X}_{iT}'\right)$ and the corresponding T-dimensional vector of random lengths $\boldsymbol{K}_i = (K_{i1}, ..., K_{iT})$, with $i = 1, ..., n$. If $\boldsymbol{k}_i = (k_{i1}, ..., k_{iT})$ is a realization of the T-dimensional vector of lengths $\boldsymbol{K}_i$, we denote by $l(\boldsymbol{k}_i)$ the number of non-zero components of $\boldsymbol{k}_i$. As in

Chapter 3, if $l(\boldsymbol{k}_i) > 0$ we denote by $\widetilde{\boldsymbol{k}}_i$ the $l(\boldsymbol{k}_i)$-dimensional vector composed of the non-zero elements of $\boldsymbol{k}_i$. Hence $\widetilde{\boldsymbol{k}}_i = (k_{ir_{i1}}, k_{ir_{i2}}, ..., k_{ir_{il(\boldsymbol{k}_i)}})$, where $1 \leq r_{i1} < r_{i2} < ... < r_{il(\boldsymbol{k}_i)} \leq T$ are indices corresponding to the elements in the original vector $\boldsymbol{k}_i$. We denote this set of ordered indices by $r(\boldsymbol{k}_i) = (r_{i1}, r_{i2}, ..., r_{il(\boldsymbol{k}_i)})$. Furthermore, both the distribution of the random lengths and the conditional density of the severities given the random lengths we allow the parameter Specifically, we make the following model assumptions. For each observational unit i, $i = 1, ..., n$,

(1) The T-dimensional vector of length variables $\boldsymbol{K}_i = (K_{i1}, K_{i2}, ..., K_{iT})$ has a multivariate Poisson$(\lambda_0, \lambda_1, ..., \lambda_T)$ distribution, as described in Appendix B, where $\lambda_j = \exp(\delta + \gamma\mu_j)$, $j = 1, ..., T$. Hence,

$$\mathbf{P}(K_{i1} = k_{i1}, K_{i2} = k_{i2}, ..., K_{iT} = k_{iT}) = e^{-\sum\limits_{j=0}^{T} \lambda_j} \prod_{j=1}^{T} \frac{\lambda_j^{k_{ij}}}{k_{ij}!} \left[\sum_{l=0}^{s} \prod_{j=1}^{T} \binom{k_{ij}}{l} l! \left(\frac{\lambda_0}{\prod\limits_{k=1}^{T} \lambda_k} \right)^{l} \right], \tag{4.1}$$

where $s = \min(k_{i1}, k_{i2}, ..., k_{iT})$.

(2) The random vector $\boldsymbol{X}_i$ with random lengths $\boldsymbol{K}_i = (K_{i1}, ..., K_{iT})$ for the i-th subject, has the conditional distribution:

$$\boldsymbol{X}_i | \left(\boldsymbol{K}_i = (k_{i1}, ..., k_{iT}) \right) \sim \mathbf{MVN}_{\sum\limits_{l=1}^{T} k_{il}} \left(\boldsymbol{\mu}_{\boldsymbol{k}_i}, \sigma^2 \mathbf{S}_{\boldsymbol{k}_i}(\rho, \rho^*) \right), \tag{4.2}$$

for $\sum\limits_{l=1}^{T} k_{il} = 1, 2, ...,$ where

$$\boldsymbol{\mu}_{\boldsymbol{k}_i} = \boldsymbol{\mu}_{k_{i1},...,k_{iT}} = \begin{pmatrix} \mu_{r_{i1}} \boldsymbol{e}_{k_{ir_{i1}}} \\ \mu_{r_{i2}} \boldsymbol{e}_{k_{ir_{i2}}} \\ ... \\ \mu_{r_{il(\boldsymbol{k}_i)}} \boldsymbol{e}_{k_{ir_{il(\boldsymbol{k}_i)}}} \end{pmatrix} \tag{4.3}$$

and

$$\boldsymbol{S}_{\boldsymbol{k}_i}(\rho, \rho^*) = \boldsymbol{S}_{k_{i1},\dots,k_{iT}}(\rho) = \begin{pmatrix} \mathbf{R}_{k_{ir_{i1}}}(\rho, \rho^*) & \rho^* \mathbf{J}_{k_{ir_{i1}}, k_{ir_{i2}}} & \dots & \rho^* \mathbf{J}_{k_{ir_{i1}}, k_{ir_{il(\boldsymbol{k}_i)}}} \\ \rho^* \mathbf{J}_{k_{ir_{i2}}, k_{ir_{i1}}} & \mathbf{R}_{k_{ir_{i2}}}(\rho) & \dots & \rho^* \mathbf{J}_{k_{ir_{i2}}, k_{ir_{il(\boldsymbol{k}_i)}}} \\ \dots & \dots & \dots & \\ \rho^* \mathbf{J}_{k_{ir_{il(\boldsymbol{k}_i)}}, k_{ir_{i1}}} & \rho^* \mathbf{J}_{k_{ir_{il(\boldsymbol{k}_i)}}, k_{ir_{i2}}} & \dots & \mathbf{R}_{k_{ir_{il(\boldsymbol{k}_i)}}}(\rho) \end{pmatrix}. \quad (4.4)$$

As before, we denote by $\mathbf{R}_k(\rho) = (1-\rho)\mathbf{I}_k + \rho \boldsymbol{e}_k \boldsymbol{e}_k^{'}$, the intraclass correlation matrix of dimension k; $\mathbf{I}_k$ is the k-dimensional identity matrix, $\boldsymbol{e}_k$ is the k-dimensional vector with all the entries 1 and $\mathbf{J}_{k \times l}$ is a $k \times l$-dimensional matrix of ones.

In order for the matrix $\mathbf{S}_{k_1,\dots,k_T}(\rho, \rho^*)$ to be positive definite, for all possible combinations of lengths $(k_1, \dots, k_T)$, we impose $0 \le \rho^* \le \rho < 1$.

(3) The data $(\boldsymbol{K}_i, \boldsymbol{X}_i)$ and $(\boldsymbol{K}_j, \boldsymbol{X}_j)$, for subject i and j, respectively, are independent for $i \neq j$.

The main features of the proposed model are summarized by:

(a). The $(T{+}6)$ parameters are collectively denoted by $\boldsymbol{\theta}$, $\boldsymbol{\theta} = (\delta, \gamma, \lambda_0, \mu_1, \dots, \mu_T, \sigma^2, \rho, \rho^*)'$. The parameter space for the above model is

$$\Theta = \Big\{ \boldsymbol{\theta} = (\delta, \gamma, \lambda_0, \mu_1, \dots, \mu_T, \sigma^2, \rho, \rho^*)' \,|\, -\infty < \delta, \gamma, \mu_1, \dots, \mu_T < \infty, \\ \sigma^2, \lambda_0 > 0, \ 0 \le \rho^* \le \rho < 1 \Big\}$$

(b). The support of the random lengths includes zero. Any of the components of the vector $\boldsymbol{K}_i$ may be zero. If all of them are zero, we observe no quantitative data. We deal with this situation as in Chapter 3 by defining the density of a zero-length vector to be equal to 1 with probability 1. As we can see from the expressions of the mean and covariance of the multivariate normal in (4.3) and (4.4), if one of the lengths is zero, it means there is no corresponding entry for that time point in both the mean vector and the covariance matrix.

(c). The parameter γ has the same interpretation as in Chapter 3, acting as a scaling parameter and also controlling the association of μ with the random length. The parameter γ being zero implies that the population means $\mu_t, t = 1, ..., T$ have no effect on the distribution of the random lengths. Parameter $\gamma > 0$ implies that, the larger the underlying parameter μ_t is, the larger the corresponding observed severity measures tend to be, and the higher their frequency tends to be. Parameter $\gamma < 0$ implies that, the larger the underlying parameter μ_t is, the larger the severity measures observed tend to be, but their frequency tends to be smaller.

(d). Marginally each K_{ij} is a Poisson random variable with $E(K_{ij}) = \lambda_j + \lambda_0$. Moreover, $Cov(K_{ij_1}, K_{ij_2}) = \lambda_0$ for any $j_1 \neq j_2$. Hence λ_0 measures the dependence between any pair of random lengths for a subject. If $\lambda_0 = 0$ then the random lengths are independent and the multivariate Poisson distribution reduces to the product of T independent Poisson distributions and this model reduces to the model of Section 3.1.1.

4.1.2 Maximum Likelihood Estimation

Consider the data corresponding to the T repeated measurements of random length vectors of severities on the same subject, $\boldsymbol{X}_i$, with the vector of random lengths $\boldsymbol{K}_i = (K_{i1}, ..., K_{iT})$. Given the random lengths $K_{i1} = k_{i1}, ..., K_{iT} = k_{iT}$, $\boldsymbol{X}_i$ has a $\sum_{l=1}^{T} k_{il}$ - dimensional multivariate normal distribution with mean $\boldsymbol{\mu}_{\boldsymbol{k}_i}$ and covariance matrix $\sigma^2 \mathbf{S}_{\boldsymbol{k}_i}(\rho, \rho^*)$ given by (4.3) and (4.4), respectively. For each subject i, $\widetilde{\boldsymbol{k}}_i = (k_{ir_1}, k_{ir_2}, ..., k_{ir_{l(\boldsymbol{k}_i)}})$ is the subvector of $\boldsymbol{k}_i$ composed only with non-zero elements.

We employ the same matrix version of the canonical transformation technique used in Chapter 3 and define $\boldsymbol{Y}_i = \boldsymbol{\Gamma}_{\boldsymbol{k}_i} \boldsymbol{X}_i$, where

$$\boldsymbol{\Gamma}_{\boldsymbol{k}_i} = \boldsymbol{\Gamma}_{k_{i1}, ..., k_{iT}} = \begin{pmatrix} \boldsymbol{\Gamma}_{k_{ir_{i1}}} & \mathbf{O}_{k_{ir_{i1}}, k_{ir_{i2}}} & ... & \mathbf{O}_{k_{ir_{i1}}, k_{ir_{il(\boldsymbol{k}_i)}}} \\ \mathbf{O}_{k_{ir_{i2}}, k_{ir_{i1}}} & \boldsymbol{\Gamma}_{k_{ir_{i2}}} & ... & \mathbf{O}_{k_{ir_{i2}}, k_{ir_{il(\boldsymbol{k}_i)}}} \\ ... & ... & ... & ... \\ \mathbf{O}_{k_{ir_{il(\boldsymbol{k}_i)}}, k_{ir_{i1}}} & \mathbf{O}_{k_{ir_{il(\boldsymbol{k}_i)}}, k_{ir_{i2}}} & ... & \boldsymbol{\Gamma}_{k_{ir_{il(\boldsymbol{k}_i)}}} \end{pmatrix}, \boldsymbol{X}_i = \begin{pmatrix} \boldsymbol{X}_{ir_{i1}} \\ \boldsymbol{X}_{ir_{i2}} \\ ... \\ \boldsymbol{X}_{ir_{il(\boldsymbol{k}_i)}} \end{pmatrix},$$

and $\mathbf{O}_{k,l}$ is the $k \times l$ matrix with all the entries equal to 0.

As described in Chapter 3, it follows that given $\boldsymbol{K}_i = (k_{i1}, ..., k_{iT})$, $\boldsymbol{Y}_i = \left(\boldsymbol{Y}_{i1}', ..., \boldsymbol{Y}_{iT}'\right)'$ also has a multivariate normal distribution, but with simpler mean and covariance structures. The only correlated components in the vector $\boldsymbol{Y}_i$ are the first entries in each of the subvectors $\boldsymbol{Y}_{ij}$, $j = 1, ..., T$, provided that the corresponding length K_{ij} is nonzero. All the remaining entries in the $\boldsymbol{Y}_i$ vector are conditionally independent, identically distributed univariate normal random variables, with mean 0 and variance σ^2/τ_0. We have

$$\boldsymbol{Y}_i = \left(Y_{ir_{i1}1}\ Y_{ir_{i1}2}\ ...\ Y_{ir_{i1}k_{ir_{i1}}}\ Y_{ir_{i2}1}\ Y_{ir_{i2}2}\ ...\ Y_{ir_{i2}k_{ir_{i2}}} ...\ Y_{ir_{il(\boldsymbol{k}_i)}1}\ Y_{ir_{il(\boldsymbol{k}_i)}2}\ ...\ Y_{ir_{il(\boldsymbol{k}_i)}k_{ir_{il(\boldsymbol{k}_i)}}}\right)'.$$

Denote by $\boldsymbol{Z}_i$ the vector composed with the first entries in the non-zero length vectors $\boldsymbol{Y}_{ij}$

$$\boldsymbol{Z}_i = \left(Y_{ir_{i1}1}\ Y_{ir_{i2}1}\ ...\ Y_{ir_{il(\boldsymbol{k}_i)}1}\right)' = \left(Z_{i1}\ Z_{i2}\ ...\ Z_{i(\boldsymbol{k}_i)}\right)'$$

and by $\widehat{\boldsymbol{Y}}_i$ the vector containing the remaining components of $\boldsymbol{Y}_i$, which are conditionally independent, identically distributed univariate normal random variables, with mean 0 and variance σ^2/τ_0.

$$\widehat{\boldsymbol{Y}}_i = \left(Y_{ir_{i1}2}\ ...\ Y_{ir_{i1}k_{ir_{i1}}}\ Y_{ir_{i2}2}\ ...\ Y_{ir_{i2}k_{ir_{i2}}} ...\ Y_{ir_{il(\boldsymbol{k}_i)}2}\ ...\ Y_{ir_{il(\boldsymbol{k}_i)}k_{ir_{il(\boldsymbol{k}_i)}}}\right)'.$$

Note that $\boldsymbol{Z}_i$ has dimension $l(\boldsymbol{k}_i) = \sum_{l=1}^{T} \delta(k_{il})$ and $\widehat{\boldsymbol{Y}}_i$ has dimension $\sum_{l=1}^{T} k_{il} - l(\boldsymbol{k}_i)$. Further denoting

$$\boldsymbol{\mu}_i^* = \left(\mu_{r_{i1}}\ \mu_{r_{i2}}\ ...\ \mu_{r_{il(\boldsymbol{k}_i)}}\right)' = \left(\mu_{i1}^*\ \mu_{i2}^*\ ...\ \mu_{il(\boldsymbol{k}_i)}^*\right)',$$

we can write the conditional density of the random length vector $\boldsymbol{Y}_i$ as

$$\begin{aligned} f(\boldsymbol{y}_i|\boldsymbol{k}_i) &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{l(\boldsymbol{k}_i)} |\boldsymbol{\Sigma}_{\boldsymbol{k}_i}|^{-1/2} \\ &\times \exp\left\{-\frac{(\boldsymbol{z}_i - \boldsymbol{\mu}_i^*)'\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}_i^*)}{2\sigma^2}\right\} \prod_{j=1}^{l(\boldsymbol{k}_i)} \prod_{l=2}^{k_{ir_{ij}}} \frac{\exp\left(-\frac{y_{ir_{ij}l}^2}{2\sigma^2/\tau_0}\right)}{\sqrt{2\pi\sigma^2/\tau_0}}, \end{aligned}$$

where

$$\boldsymbol{\Sigma}_{\boldsymbol{k}_i} = \boldsymbol{\Sigma}_{\widetilde{\boldsymbol{k}}_i} = \begin{pmatrix} \frac{1}{\tau_{\widetilde{k}_{i1}}} & \rho^* & \cdots & \rho^* \\ \rho^* & \frac{1}{\tau_{\widetilde{k}_{i2}}} & \cdots & \rho^* \\ \cdots & \cdots & \cdots & \cdots \\ \rho^* & \rho^* & \cdots & \frac{1}{\tau_{\widetilde{k}_{il(\boldsymbol{k}_i)}}} \end{pmatrix}. \tag{4.5}$$

This can be written as

$$\begin{aligned} f(\boldsymbol{y}_i|\boldsymbol{k}_i) &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{\sum\limits_{l=1}^{T} k_{il}} \tau_0^{\frac{1}{2}\left(\sum\limits_{l=1}^{T} k_{il} - l(\boldsymbol{k}_i)\right)} |\boldsymbol{\Sigma}_{\boldsymbol{k}_i}|^{-1/2} \\ &\quad \times \exp\left\{-\frac{(\boldsymbol{z}_i - \boldsymbol{\mu}_i^*)'\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}_i^*)}{2\sigma^2}\right\} \exp\left(-\frac{\sum\limits_{j=1}^{l(\boldsymbol{k}_i)} \sum\limits_{l=2}^{k_{ir_{ij}}} y_{ir_{ij}l}^2}{2\sigma^2/\tau_0}\right). \end{aligned}$$

We can write the joint density for this one observation as

$$f(\boldsymbol{y}_i, \boldsymbol{k}_i) = \mathbf{P}(K_{i1} = k_{i1}, ..., K_{iT} = k_{iT}) f(\boldsymbol{y}_i | k_{i1}, ..., k_{iT})^{\delta(\boldsymbol{k}_i)}.$$

Taking the logarithm of the above, we obtain

$$\log f(\boldsymbol{y}_i, \boldsymbol{k}_i) = \log \mathbf{P}(K_{i1} = k_{i1}, ..., K_{iT} = k_{iT}) + \delta(\boldsymbol{k}_i) \log f(\boldsymbol{y}_i | k_{i1}, ..., k_{iT}).$$

The contribution of the $i-$th subject to the log-likelihood is given by

$$\begin{aligned} \log f(\boldsymbol{y}_i, \boldsymbol{k}_i) &= -\lambda_0 + \sum_{j=1}^{T}(-\lambda_j + k_{ij}\log\lambda_j) + \log\left(\sum_{l=0}^{s}\prod_{j=1}^{T}\binom{k_{ij}}{l} l! \left(\frac{\lambda_0}{\prod\limits_{k=1}^{T}\lambda_k}\right)^l\right) \\ &\quad + \delta(\boldsymbol{k}_i)\left[-\frac{\log\sigma^2}{2}\sum_{j=1}^{T}k_{ij} + \frac{\log\tau_0}{2}\left(\sum_{j=1}^{T}k_{ij} - \sum_{j=1}^{T}\delta(k_{ij})\right)\right. \\ &\quad \left. -\frac{1}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{k}_i}| - \frac{1}{2\sigma^2}(\boldsymbol{z}_i - \boldsymbol{\mu}_i^*)'\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}_i^*) - \frac{\tau_0}{2\sigma^2}\sum_{j=1}^{l(\boldsymbol{k}_i)}\sum_{l=2}^{k_{ir_{ij}}} y_{ir_{ij}l}^2\right]. \end{aligned}$$

Recalling that λ_j has the form $\lambda_j = \exp(\delta + \gamma\mu_j)$, we can write

$$\begin{aligned} \log f(\boldsymbol{y}_i, \boldsymbol{k}_i) &= -\lambda_0 + \sum_{j=1}^{T}\left[-e^{\delta+\gamma\mu_j} + k_{ij}(\delta + \gamma\mu_j)\right] + \log\left(\sum_{l=0}^{s}\prod_{j=1}^{T}\binom{k_{ij}}{l} l! \left(\frac{\lambda_0}{\prod\limits_{k=1}^{T}\lambda_k}\right)^l\right) \\ &\quad -\frac{\log\sigma^2}{2}\sum_{j=1}^{T}k_{ij} + \frac{\log\tau_0}{2}\left(\sum_{j=1}^{T}k_{ij} - l(\boldsymbol{k}_i)\right) \\ &\quad -\frac{1}{2}\log|\boldsymbol{\Sigma}_{\boldsymbol{k}_i}| - \frac{1}{2\sigma^2}(\boldsymbol{z}_i - \boldsymbol{\mu}_i^*)'\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}_i^*) - \frac{\tau_0}{2\sigma^2}\sum_{j=1}^{l(\boldsymbol{k}_i)}\sum_{l=2}^{k_{ir_{ij}}} y_{ir_{ij}l}^2. \end{aligned}$$

To keep the notation simple, we give up multiplying the last 5 terms in the right hand side of the above equation by $\delta(\boldsymbol{k}_i)$. Instead, we make the convention to consider these five terms equal to zero for the case when $\delta(\boldsymbol{k}_i)$ is zero (i.e. when all the random lengths for a subject are zero). The log-likelihood of the entire data set is

$$\begin{aligned} l(\boldsymbol{\theta}) &= -n\lambda_0 + \sum_{i=1}^{n}\sum_{j=1}^{T}\left[-e^{\delta+\gamma\mu_j} + k_{ij}\left(\delta+\gamma\mu_j\right)\right] + \sum_{i=1}^{n}\log\left(\sum_{l=0}^{s}\prod_{j=1}^{T}\binom{k_{ij}}{l}l!\left(\frac{\lambda_0}{\prod\limits_{k=1}^{T}\lambda_k}\right)^{l}\right) \\ &\quad -\frac{\log\sigma^2}{2}\sum_{i=1}^{n}\sum_{j=1}^{T}k_{ij} + \frac{\log\tau_0}{2}\sum_{i=1}^{n}\left(\sum_{j=1}^{T}k_{ij} - l(\boldsymbol{k}_i)\right) \\ &\quad -\frac{1}{2}\sum_{i=1}^{n}\log|\boldsymbol{\Sigma}_{\boldsymbol{k}_i}| - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(\boldsymbol{z}_i-\boldsymbol{\mu}_i^*)'\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{z}_i-\boldsymbol{\mu}_i^*) - \frac{\tau_0}{2\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{l(\boldsymbol{k}_i)}\sum_{l=2}^{k_{ir_{ij}}}y^2_{ir_{ij}l}, \end{aligned} \tag{4.6}$$

where the parameter vector is $\boldsymbol{\theta} = (\delta, \gamma, \lambda_0, \mu_1, ..., \mu_T, \sigma^2, \rho, \rho^*)'$.

Denoting by

$$Q_{\boldsymbol{k}_i}(\lambda_0, \lambda_1, ..., \lambda_T) = \sum_{l=0}^{\min(k_{i1},...,k_{iT})}\prod_{j=1}^{T}\binom{k_{ij}}{l}l!\left(\frac{\lambda_0}{\prod\limits_{k=1}^{T}\lambda_k}\right)^{l} \tag{4.7}$$

and using the results (A.5) - (A.8) from Appendix A, it follows that the $T+6$ score equations are given by

$$\begin{aligned}
\sum_{i=1}^{n}\sum_{j=1}^{T}\left(-e^{\delta+\gamma\mu_j}+k_{ij}\right)+\sum_{i=1}^{n}\frac{\partial}{\partial\delta}\log Q_{\boldsymbol{k}_i}(\lambda_0,\lambda_1,...,\lambda_T) &= 0 \quad (4.8)\\
\sum_{i=1}^{n}\sum_{j=1}^{T}\left(-\mu_j e^{\delta+\gamma\mu_j}+k_{ij}\mu_j\right)+\sum_{i=1}^{n}\frac{\partial}{\partial\gamma}\log Q_{\boldsymbol{k}_i}(\lambda_0,\lambda_1,...,\lambda_T) &= 0 \quad (4.9)\\
-n+\sum_{i=1}^{n}\frac{\partial}{\partial\lambda_0}\log Q_{\boldsymbol{k}_i}(\lambda_0,\lambda_1,...,\lambda_T) &= 0 \quad (4.10)\\
\gamma\begin{pmatrix} -n\lambda_1+\sum_{i=1}^{n}k_{i1}\\ -n\lambda_2+\sum_{i=1}^{n}k_{i2}\\ \vdots\\ -n\lambda_T+\sum_{i=1}^{n}k_{iT}\end{pmatrix}+\begin{pmatrix} \sum_{i=1}^{n}\frac{\partial}{\partial\mu_1}\log Q_{\boldsymbol{k}_i}(\lambda_0,\lambda_1,...,\lambda_T)\\ \sum_{i=1}^{n}\frac{\partial}{\partial\mu_2}\log Q_{\boldsymbol{k}_i}(\lambda_0,\lambda_1,...,\lambda_T)\\ \vdots\\ \sum_{i=1}^{n}\frac{\partial}{\partial\mu_T}\log Q_{\boldsymbol{k}_i}(\lambda_0,\lambda_1,...,\lambda_T)\end{pmatrix}+\frac{1}{\sigma^2}\sum_{i=1}^{n}\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{z}_i-\boldsymbol{\mu}_i^*) &= \mathbf{0}_T \quad (4.11)\\
-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{T}k_{ij}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}(\boldsymbol{z}_i-\boldsymbol{\mu}_i^*)^{'}\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{z}_i-\boldsymbol{\mu}_i^*)+\frac{\tau_0}{2\sigma^4}\sum_{i=1}^{n}\widehat{\boldsymbol{y}}_i^{'}\widehat{\boldsymbol{y}}_i &= 0 \quad (4.12)\\
\frac{1}{1-\rho}\sum_{i=1}^{n}\sum_{j=1}^{T}(k_{ij}-\delta(k_{ij}))-\sum_{i=1}^{n}tr\left(\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}\left[\mathbf{I}_T-\mathrm{Diag}\left(\frac{1}{k_{i1}},...,\frac{1}{k_{iT}}\right)\right]\right)&\\
+\frac{1}{\sigma^2}\sum_{i=1}^{n}(\boldsymbol{z}_i-\boldsymbol{\mu}_i^*)^{'}\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}\left[\mathbf{I}_T-\mathrm{Diag}\left(\frac{1}{k_{i1}},...,\frac{1}{k_{iT}}\right)\right]\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{z}_i-\boldsymbol{\mu}_i^*)-\frac{\tau_0}{\sigma^2(1-\rho)^2}\sum_{i=1}^{n}\widehat{\boldsymbol{y}}_i^{'}\widehat{\boldsymbol{y}}_i &= 0 \quad (4.13)\\
-\sum_{i=1}^{n}tr\left(\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{e}_T\boldsymbol{e}_T^{'}-\mathbf{I}_T)\right)+\frac{1}{\sigma^2}\sum_{i=1}^{n}(\boldsymbol{z}_i-\boldsymbol{\mu}_i^*)^{'}\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{e}_T\boldsymbol{e}_T^{'}-\mathbf{I}_T)\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{z}_i-\boldsymbol{\mu}_i^*) &= 0. \quad (4.14)
\end{aligned}$$

We use the fact that $\sum_{j=1}^{l(\boldsymbol{k}_i)}\sum_{l=2}^{k_{ir_{ij}}}y_{ir_{ij}l}^2=\widehat{\boldsymbol{y}}_i^{'}\widehat{\boldsymbol{y}}_i$ in the above expressions.

As in Chapter 3, if all the observations consist of random length vectors with length zero, none of the equations (4.8) - (4.14) make sense; hence, none of the parameters are estimable. If we observe only vectors with length zero or 1 then the parameter ρ is not estimable. Due to the complexity of the expression $Q_{\boldsymbol{k}_i}(\lambda_0,\lambda_1,...,\lambda_T)$ and to the computational burden of the multivariate Poisson probability mass function, we do not present the explicit expressions for the score equations for the general case, but only for the simpler scenario where we have two measurements for each subject.

4.1.2.1 Maximum Likelihood Estimation for 2 Time Points

Let us assume each subject was observed only $T = 2$ times. Thus, the random lengths are distributed jointly as bivariate Poisson$(\lambda_0, \lambda_1, \lambda_2)$ random variables, where this joint bivariate probability mass function is given by

$$\begin{aligned} f(r,s) &= \mathbf{P}(K_1 = r, K_2 = s) \\ &= e^{-\sum\limits_{i=0}^{2}\lambda_i} \frac{\lambda_1^r}{r!}\frac{\lambda_2^s}{s!} \sum_{i=0}^{\min(r,s)} \binom{r}{i}\binom{s}{i} i! \left(\frac{\lambda_0}{\lambda_1\lambda_2}\right)^i . \end{aligned} \tag{4.15}$$

Taking into account the expression of the score equations for the vector of lengths in (C.5) - (C.9) and using f to denote the pmf in (4.15), we may explicitly write the score equations for the one population with 2 time points model as

$$-\sum_{i=1}^{n}\sum_{j=1}^{2} e^{\delta+\gamma\mu_j} + \sum_{i=1}^{n}\left[\lambda_1 \frac{f(k_{i1}-1,k_{i2})}{f(k_{i1},k_{i2})} + \lambda_2 \frac{f(k_{i1},k_{i2}-1)}{f(k_{i1},k_{i2})}\right] = 0 \tag{4.16}$$

$$-\sum_{i=1}^{n}\sum_{j=1}^{T} \mu_j e^{\delta+\gamma\mu_j} + \sum_{i=1}^{n}\left[\lambda_1 \frac{f(k_{i1}-1,k_{i2})}{f(k_{i1},k_{i2})} + \lambda_2 \frac{f(k_{i1},k_{i2}-1)}{f(k_{i1},k_{i2})}\right] = 0 \tag{4.17}$$

$$+\sum_{i=1}^{n} \frac{f(k_{i1},k_{i2}) - f(k_{i1}-1,k_{i2}) - f(k_{i1},k_{i2}-1) + f(k_{i1}-1,k_{i2}) - 1}{f(k_{i1},k_{i2})} = 0 \tag{4.18}$$

$$\gamma \begin{pmatrix} -n\lambda_1 + \lambda_1 \sum\limits_{i=1}^{n} \frac{f(k_{i1}-1,k_{i2})}{f(k_{i1},k_{i2})} \\ -n\lambda_2 + \lambda_2 \sum\limits_{i=1}^{n} \frac{f(k_{i1},k_{i2}-1)}{f(k_{i1},k_{i2})} \end{pmatrix} + \frac{1}{\sigma^2}\sum_{i=1}^{n} \boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}_i^*) = \mathbf{0}_2 \tag{4.19}$$

$$-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{T} k_{ij} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(\boldsymbol{z}_i - \boldsymbol{\mu}_i^*)'\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}_i^*) + \frac{\tau_0}{2\sigma^4}\sum_{i=1}^{n}\widehat{\boldsymbol{y}}_i'\widehat{\boldsymbol{y}}_i = 0 \tag{4.20}$$

$$\begin{aligned} \frac{1}{1-\rho}\sum_{i=1}^{n}\sum_{j=1}^{T}(k_{ij} - \delta(k_{ij})) - \sum_{i=1}^{n} tr\left(\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}\text{Diag}\left(1-\frac{1}{k_{i1}}, 1-\frac{1}{k_{i2}}\right)\right) & \\ +\frac{1}{\sigma^2}\sum_{i=1}^{n}(\boldsymbol{z}_i - \boldsymbol{\mu}_i^*)'\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}\text{Diag}\left(1-\frac{1}{k_{i1}}, 1-\frac{1}{k_{i2}}\right)\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}_i^*) & \\ -\frac{\tau_0}{\sigma^2(1-\rho)^2}\sum_{i=1}^{n}\widehat{\boldsymbol{y}}_i'\widehat{\boldsymbol{y}}_i &= 0 \end{aligned}$$

$$\sum_{i=1}^{n} tr\left(\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{e}_2\boldsymbol{e}_2' - \mathbf{I}_2)\right) - \frac{1}{\sigma^2}\sum_{i=1}^{n}(\boldsymbol{z}_i - \boldsymbol{\mu}_i^*)'\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{e}_2\boldsymbol{e}_2' - \mathbf{I}_2)\boldsymbol{\Sigma}_{\boldsymbol{k}_i}^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}_i^*) = 0. \tag{4.21}$$

We discus in additional detail how to numerically obtain the MLE's in Section 4.3.3.

4.1.3 Asymptotic Distribution of the Maximum Likelihood Estimates for $T = 2$

Let us denote by $\boldsymbol{I}_n(\boldsymbol{\theta})$ the information matrix for $\boldsymbol{\theta}$ contained in the n multivariate random length vectors $\boldsymbol{X}_i$ with random lengths $\boldsymbol{K}_i = (K_{i1}, K_{i2})$, $i = 1, ..., n$. As in the one population model of Chapter 3, $\boldsymbol{I}_n(\boldsymbol{\theta})$ can be computed as a sum of the information about $\boldsymbol{\theta}$ contained in the lengths and the sum of information about $\boldsymbol{\theta}$ contributed by the vectors of severities over all possible lengths. In Appendix C, we show that the information about the parameter $\boldsymbol{\theta}$ contained in a single observation from the one population model, $\boldsymbol{X}$ with random lengths $\boldsymbol{K} = (K_1, ..., K_T)$ has the form

$$\boldsymbol{I}(\boldsymbol{\theta}) = \boldsymbol{I}^*(\boldsymbol{\theta}) + \sum_{\boldsymbol{k}\in\Upsilon} \mathbf{P}_{\boldsymbol{\theta}}(\boldsymbol{K} = \boldsymbol{k})\boldsymbol{I}(\boldsymbol{\theta}|\boldsymbol{k}),$$

where $\boldsymbol{I}^*(\boldsymbol{\theta})$ is the information matrix about $\boldsymbol{\theta}$ contained in the random lengths $K = (K_1, ..., K_T)$ and $\boldsymbol{I}(\boldsymbol{\theta}|\boldsymbol{k})$ is the information matrix contained in $\boldsymbol{X}_i|\boldsymbol{K} = \boldsymbol{k}$, $\boldsymbol{k} \in \Upsilon$, where we denote

$$\Upsilon = \left\{ \boldsymbol{k} = (k_1, k_2)|\ k_i = 0, 1, ... \text{ for } \forall i = 1, 2 \text{ and } \sum_{i=1}^{2} k_i \geq 1 \right\}.$$

Using the independence of the subjects, we can compute the information matrix $\boldsymbol{I}_n(\boldsymbol{\theta})$ about $\boldsymbol{\theta} = (\delta, \gamma, \lambda_0, \mu_1, \mu_2, \sigma^2, \rho, \rho^*)'$ contained in the n independent observations from the one population model as

$$\boldsymbol{I}_n(\boldsymbol{\theta}) = n\boldsymbol{I}^*(\boldsymbol{\theta}) + n \sum_{\boldsymbol{k}\in\Upsilon} \mathbf{P}_{\boldsymbol{\theta}}(\boldsymbol{K} = \boldsymbol{k})\boldsymbol{I}(\boldsymbol{\theta}|\boldsymbol{k}).$$

As in Appendix C, denote

$$\delta_1 = \lambda_0 \left[1 - \lambda_0(\tau - 1)\right] \tag{4.22}$$

$$\delta_2 = -(\lambda_1 + \lambda_2) + \left[\lambda_1^* \lambda_2^* - \lambda_0^2\right](\tau - 1) \tag{4.23}$$

$$\delta_3 = \left[(\lambda_1^* \lambda_2^* - \lambda_0^2\right] \left[\tau - 1 - (\lambda_1 + \lambda_2)\right] \tag{4.24}$$

$$\tau = \sum_{r,s=1}^{\infty} \frac{f^2(r-1, s-1)}{f(r, s)}, \tag{4.25}$$

where $\lambda_i^* = \lambda_i + \lambda_0$, $i = 1, 2$ and f is the probability mass function for the bivariate distribution from (4.15). We show in (C.21), that $\boldsymbol{I}_n(\boldsymbol{\theta})$ has a block diagonal form

$$\boldsymbol{I}_n(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{I}_n(\delta,\gamma,\lambda_0,\mu_1,\mu_2) & \mathbf{O}_{5\times 3} \\ \mathbf{O}_{3\times 5} & n\sum\limits_{\boldsymbol{k}\in\Upsilon}\left[e^{-\sum\limits_{i=0}^{2}\lambda_i}\frac{\lambda_1^{k_1}}{k_1!}\frac{\lambda_2^{k_2}}{k_2!}\sum\limits_{i=0}^{\min(k_1,k_2)}\binom{k_1}{i}\binom{k_2}{i}i!\left(\frac{\lambda_0}{\lambda_1\lambda_2}\right)^i\right]\boldsymbol{I}_{\boldsymbol{k}}(\sigma^2,\rho,\rho^*) \end{pmatrix} \tag{4.26}$$

where

$$\begin{aligned} \boldsymbol{I}_n(\delta,\gamma,\lambda_0,\mu_1,\mu_2) &= n\boldsymbol{I}(\delta,\gamma,\lambda_0,\mu_1,\mu_2) \\ &= n\mathbf{H} + \begin{pmatrix} \mathbf{O}_{3\times 3} & \mathbf{O}_{3\times 2} \\ \mathbf{O}_{2\times 3} & n\frac{1}{\sigma^2}\sum\limits_{\boldsymbol{k}\in\Upsilon}\left[e^{-\sum\limits_{i=0}^{2}\lambda_i}\frac{\lambda_1^{k_1}}{k_1!}\frac{\lambda_2^{k_2}}{k_2!}\sum\limits_{i=0}^{\min(k_1,k_2)}\binom{k_1}{i}\binom{k_2}{i}i!\left(\frac{\lambda_0}{\lambda_1\lambda_2}\right)^i\right]\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} \end{pmatrix}, \end{aligned}$$

and the elements of the matrix $\boldsymbol{I}_{\boldsymbol{k}}(\sigma^2,\rho,\rho^*)$ are defined by (A.10) - (A.15). $\mathbf{H}$ denotes the matrix

$$\mathbf{H} = \begin{pmatrix} \lambda_1+\lambda_2-4\delta_1 & \sum\limits_{i=1}^{2}\mu_i(\lambda_i-2\delta_1) & -\frac{2\delta_2\lambda_0}{\lambda_1\lambda_2} & \gamma(\lambda_1-2\delta_1) & \gamma(\lambda_2-2\delta_1) \\ \sum\limits_{i=1}^{2}\mu_i(\lambda_i-2\delta_1) & \sum\limits_{i=1}^{2}\mu_i^2\lambda_i-\delta_1(\sum\limits_{i=1}^{2}\mu_i)^2 & -\frac{\delta_2\lambda_0(\mu_1+\mu_2)}{\lambda_1\lambda_2} & \gamma(\mu_1\lambda_1-\delta_1\sum\limits_{i=1}^{2}\mu_i) & \gamma(\mu_2\lambda_2-\delta_1\sum\limits_{i=1}^{2}\mu_i) \\ -\frac{2\delta_2\lambda_0}{\lambda_1\lambda_2} & -\frac{\delta_2\lambda_0(\mu_1+\mu_2)}{\lambda_1\lambda_2} & \frac{\delta_3}{\lambda_1^2\lambda_2^2} & -\frac{\gamma\delta_2\lambda_0}{\lambda_1\lambda_2} & -\frac{\gamma\delta_2\lambda_0}{\lambda_1\lambda_2} \\ \gamma(\lambda_1-2\delta_1) & \gamma(\mu_1\lambda_1-\delta_1\sum\limits_{i=1}^{2}\mu_i) & -\frac{\gamma\delta_2\lambda_0}{\lambda_1\lambda_2} & \gamma^2(\lambda_1-\delta_1) & -\gamma^2\delta_1 \\ \gamma(\lambda_2-2\delta_1) & \gamma(\mu_2\lambda_2-\delta_1\sum\limits_{i=1}^{2}\mu_i) & -\frac{\gamma\delta_2\lambda_0}{\lambda_1\lambda_2} & -\gamma^2\delta_1 & \gamma^2(\lambda_2-\delta_1) \end{pmatrix}.$$

From the expression of the score equations, it is clear that there is no closed form solution for $\hat{\boldsymbol{\theta}}_n$, and asymptotic results are required for distributional results. We can apply a general result on the efficiency of maximum likelihood estimators for random length data (see Theorem A.3.2 in Barnhart [4]) to derive the asymptotic distribution for $\hat{\boldsymbol{\theta}}_n$, the MLE. The asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}_n$, is obtained as the inverse of the above information matrix and is estimated by $\boldsymbol{I}_n^{-1}(\hat{\boldsymbol{\theta}}_n)$.

Theorem 3. *Let* $\hat{\boldsymbol{\theta}}_n = (\hat{\delta}_n, \hat{\gamma}_n, \hat{\lambda}_0^{(n)}, \hat{\mu}_1^{(n)}, \hat{\mu}_2^{(n)}, \hat{\sigma}_n^2, \hat{\rho}_n, \hat{\rho}_n^*)'$ *be the MLEs for a sample of size* n *from the one population model. Then, as* $n \to \infty$

(1). $\hat{\boldsymbol{\theta}}_n$ *is consistent.*

(2).

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n-\boldsymbol{\theta}\right)\xrightarrow{L}\boldsymbol{MVN}_8(\mathbf{0},\boldsymbol{I}^{-1}(\boldsymbol{\theta})),$$

where

$$\boldsymbol{I}(\boldsymbol{\theta})=\begin{pmatrix}\boldsymbol{I}(\delta,\gamma,\lambda_0,\mu_1,\mu_2) & \boldsymbol{O}_{5\times 3}\\ \boldsymbol{O}_{3\times 5} & n\sum\limits_{\boldsymbol{k}\in\Upsilon}\left[e^{-\sum_{i=0}^{2}\lambda_i}\frac{\lambda_1^{k_1}}{k_1!}\frac{\lambda_2^{k_2}}{k_2!}\sum\limits_{i=0}^{\min(k_1,k_2)}\binom{k_1}{i}\binom{k_2}{i}i!\left(\frac{\lambda_0}{\lambda_1\lambda_2}\right)^i\right]\boldsymbol{I}_{\boldsymbol{k}}(\sigma^2,\rho,\rho^*)\end{pmatrix}\tag{4.27}$$

and

$$\boldsymbol{I}(\delta,\gamma,\lambda_0,\mu_1,\mu_2)=\boldsymbol{H}+\begin{pmatrix}\boldsymbol{O}_{3\times 3} & \boldsymbol{O}_{3\times 2}\\ \boldsymbol{O}_{2\times 3} & \frac{1}{\sigma^2}\sum\limits_{\boldsymbol{k}\in\Upsilon}\left[e^{-\sum_{i=0}^{2}\lambda_i}\frac{\lambda_1^{k_1}}{k_1!}\frac{\lambda_2^{k_2}}{k_2!}\sum\limits_{i=0}^{\min(k_1,k_2)}\binom{k_1}{i}\binom{k_2}{i}i!\left(\frac{\lambda_0}{\lambda_1\lambda_2}\right)^i\right]\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\end{pmatrix}.$$

The matrix $\boldsymbol{H}$ is defined above and the elements of the matrix $\boldsymbol{I}_{\boldsymbol{k}}(\sigma^2,\rho,\rho^*)$ are given by (A.10) - (A.15). We denote

$$\Upsilon=\left\{\boldsymbol{k}=(k_1,k_2)|\ k_i=0,1,...\ \textit{for}\ \forall i=1,2\ \textit{and}\ \sum_{i=1}^{2}k_i\geq 1\right\}.$$

4.2 MULTIPLE POPULATION MODEL

This section generalizes the model introduced in Section 4.1 to accommodate two or more populations. We allow the different populations to share the scaling parameters δ and γ and the parameter λ_0, while μ's are population specific. The model is motivated by our belief that γ and δ are parameters of the process linking the the underlying "disease" status to the length, while λ_0 is a parameter describing the dependency of the number of events over time. Thus, they should remain the same, regardless of population. What changes is the underlying "disease" status. For example, in the context of the migraine example, is the disease status that is affected by the drug, and not the relationship between the disease status and severity of migraines, or the relationship between the number of events experienced every quarter of a year.

4.2.1 Model Description

Using the notation of Chapter 3, suppose that we have data from m different populations. Each subject j, $j = 1, ..., n_i$, from population i, $i = 1, ..., m$ is observed T times. At each time point $t = 1, ..., T$, each subject j from population i has a random number of events K_{ijt} and the corresponding severity measurements are recorded into the vector $\boldsymbol{X}_{ijt}$. Hence all the data for this subject i can be condensed into a $\sum\limits_{t=1}^{T} K_{ijt}$-dimensional vector $\boldsymbol{X}_{ij}$, $\boldsymbol{X}_{ij}^{'} = \left(\boldsymbol{X}_{ij1}^{'}, ..., \boldsymbol{X}_{ijT}^{'}\right)$ and the corresponding T-dimensional vector of random lengths $\boldsymbol{K}_{ij} = (K_{ij1}, ..., K_{ijT})$, with $i = 1, ..., m$, $j = 1, ..., n_1$. Let $\boldsymbol{k}_{ij} = (k_{ij1}, ..., k_{ijT})$ be a realization of the T-dimensional vector of lengths $\boldsymbol{K}_{ij}$. Some of the the components of $\boldsymbol{k}_{ij}$ might be zero. Let us denote by $l(\boldsymbol{k}_{ij})$ the number of non-zero components of $\boldsymbol{k}_{ij}$. We have $l(\boldsymbol{k}_{ij}) = \sum\limits_{t=1}^{T} \delta(k_{ijt})$. If $l(\boldsymbol{k}_{ij}) > 0$ we denote by $\widetilde{\boldsymbol{k}}_{ij}$ the $l(\boldsymbol{k}_{ij})$-dimensional vector composed of the non-zero elements of $\boldsymbol{k}_{ij}$. Hence $\widetilde{\boldsymbol{k}}_{ij} = (k_{ijr_{ij1}}, k_{ijr_{ij2}}, ..., k_{ijr_{ijl(\boldsymbol{k}_{ij})}})$, where $1 \leq r_{ij1} < r_{ij2} < ... < r_{ijl(\boldsymbol{k}_{ij})} \leq T$ are indices corresponding to the elements in the original vector $\boldsymbol{k}_i$. We denote this set of ordered indices by $\boldsymbol{r}_{ij} = (r_{ij1}, r_{ij2}, ..., r_{ijl(\boldsymbol{k}_{ij})})$.

We make the following model assumptions. For each population $i = 1, 2, ..., m$, and each observational unit j, $j = 1, ..., n_i$,

(1) The random length variables $(K_{ij1}, K_{ij2}, ..., K_{ijT})$ are dependent and follow a multivariate Poisson$(\lambda_0, \lambda_{i1}, ..., \lambda_{iT})$ distribution, where $\lambda_{it} = \exp(\delta + \gamma\mu_{it})$, $t = 1, ..., T$. Hence,

$$\mathbf{P}_{\boldsymbol{\theta}}(K_{ij1} = k_{ij1}, ..., K_{ijT} = k_{ijT}) = e^{-\sum\limits_{t=0}^{T} \lambda_{it}} \prod_{t=1}^{T} \frac{\lambda_{it}^{k_{ijt}}}{k_{ijt}!} \sum_{l=0}^{\min(k_{ij1}, ..., k_{ijT})} \prod_{t=1}^{T} \binom{k_{ijt}}{l} l! \left(\frac{\lambda_0}{\prod\limits_{k=1}^{T} \lambda_{ik}} \right)^{l} \tag{4.28}$$

(2) The random vector $\boldsymbol{X}_{ij}$ with random lengths $\boldsymbol{K}_{ij} = (K_{ij1}, ..., K_{ijT})$ for the j-th subject in the i-th population, has the conditional distribution

$$\boldsymbol{X}_i | \left(\boldsymbol{K}_{ij} = (k_{ij1}, ..., k_{ijT})\right) \sim \mathbf{MVN}_{\sum\limits_{l=1}^{T} k_{ijl}} \left(\boldsymbol{\mu}_{\boldsymbol{k}_{ij}}, \sigma^2 \mathbf{S}_{\boldsymbol{k}_{ij}}(\rho, \rho^*)\right), \tag{4.29}$$

for $\sum_{t=1}^{T} k_{ijt} = 1, 2, ...$, where

$$\boldsymbol{\mu}_{\boldsymbol{k}_{ij}} = \boldsymbol{\mu}_{k_{ij1},...,k_{ijT}} = \begin{pmatrix} \mu_{ir_{ij1}} \boldsymbol{e}_{k_{ijr_{ij1}}} \\ \mu_{ir_{ij2}} \boldsymbol{e}_{k_{ijr_{ij2}}} \\ ... \\ \mu_{ir_{ijl(\boldsymbol{k}_{ij})}} \boldsymbol{e}_{k_{ijr_{ijl(\boldsymbol{k}_{ij})}}} \end{pmatrix} \tag{4.30}$$

and

$$\begin{aligned} \mathbf{S}_{\boldsymbol{k}_{ij}}(\rho, \rho^*) &= \mathbf{S}_{k_{ij1},...,k_{ijT}}(\rho, \rho^*) \\ &= \begin{pmatrix} \mathbf{R}_{k_{ijr_{ij1}}}(\rho) & \rho^* \mathbf{J}_{k_{ijr_{ij1}}, k_{ijr_{ij2}}} & ... & \rho^* \mathbf{J}_{k_{ijr_{ij1}}, k_{ijr_{ijl(\boldsymbol{k}_{ij})}}} \\ \rho^* \mathbf{J}_{k_{ijr_{ij2}}, k_{ijr_{ij1}}} & \mathbf{R}_{k_{ijr_{ij2}}}(\rho) & ... & \rho^* \mathbf{J}_{k_{ijr_{i2}}, k_{ijr_{ijl(\boldsymbol{k}_{ij})}}} \\ ... & ... & ... & \\ \rho^* \mathbf{J}_{k_{ijr_{ijl(\boldsymbol{k}_{ij})}}, k_{ijr_{ij1}}} & \rho^* \mathbf{J}_{k_{ijr_{ijl(\boldsymbol{k}_{ij})}}, k_{ijr_{ij2}}} & ... & \mathbf{R}_{k_{ijr_{ijl(\boldsymbol{k}_{ij})}}}(\rho) \end{pmatrix} \end{aligned} \tag{4.31}$$

As before, $\mathbf{J}_{k,l}$ denotes the $k \times l$ matrix with all the entries equal to 1. In order for the the matrix $\mathbf{S}_{k_1,...,k_T}(\rho, \rho^*)$ to be positive definite for all combinations of random lengths, we impose that $0 \leq \rho^* \leq \rho < 1$.

(3) The data $(\boldsymbol{K}_{i_1 j_1}, \boldsymbol{X}_{i_1 j_1})$ and $(\boldsymbol{K}_{i_2 j_2}, \boldsymbol{X}_{i_2 j_2})$, for subject j_1 from population i_1 and j_2 from population i_2, respectively, are independent for $(i_1, j_1) \neq (i_2, j_2)$.

The covariance structure used to model the dependency of the random lengths mirrors the one used in Chapter 3 to model the dependency of the severities over time. There we assumed that, conditional on the number of events experienced throughout the follow-up, any two severities recorded at different measurement times have the same correlation, ρ^*. In this model, we further assume that any pair of random lengths recorded at different measurement times have the same covariance, λ_0. Having the covariance of any pair of random lengths for any population equal to λ_0 is parallel in concept with the assumption of Chapter 3 that any two severities recorded at different measurement times have the same correlation, ρ^*, regardless of population and the measurement times.

The multiple population model has the following features:

(a). The $(mT+6)$ parameters are collectively denoted by $\boldsymbol{\theta} = (\delta, \gamma, \lambda_0, \mu_1, ..., \mu_{mT}, \sigma^2, \rho, \rho^*)'$. The parameter space for the above model is

$$\Theta = \Big\{\boldsymbol{\theta} = (\delta, \gamma, \lambda_0, \mu_{11}, ..., \mu_{mT}, \sigma^2, \rho)' \,|\, -\infty < \delta, \gamma, \mu_{11}, ..., \mu_{mT} < \infty, \\ \sigma^2, \lambda_0 > 0,\ 0 \leq \rho^* \leq \rho < 1\Big\}$$

(b). The support of the random lengths includes zero. Any of the components of the vector $\boldsymbol{K}_i$ may be zero. If all of them are zero, we observe no quantitative data. In this case, the multivariate normal distribution defined in (4.29) does not exist. We deal with this situation by defining the density of a vector with zero length to be equal to 1 with probability 1. As we can see from the expressions of the mean and covariance of the multivariate normal in (4.30) and (4.31), if one of the lengths is zero, it means there is no corresponding entry for that time point in both the mean vector and the covariance matrix.

(c). The parameter γ has the same interpretation as in Section 4.1, controlling the association of μ with the random lengths.

(d). Marginally each K_{ijt} is a Poisson random variable with $E(K_{ijt}) = \lambda_{it} + \lambda_0$. Moreover, $Cov(K_{ijt_1}, K_{ijt_2}) = \lambda_0$ for any $t_1 \neq t_2$. Hence λ_0 is a measure of dependence between any pair of random lengths for a subject. If $\lambda_0 = 0$ then the random lengths are independent and the multivariate Poisson distribution reduces to the product of T independent Poisson distributions.

4.2.2 Maximum Likelihood Estimation

Consider one of these T-random length measurements, $\boldsymbol{X}_{ij}$, with the random length vector $\boldsymbol{K}_{ij} = (K_{ij1}, ..., K_{ijT})$. Given the vector of random lengths $\boldsymbol{K}_{ij} = (K_{ij1}, ..., K_{ijT}) = (k_{ij1}, ..., = k_{ijT})$, $\boldsymbol{X}_{ij}$ has a $\sum\limits_{t=1}^{T} k_{ijt}$ - dimensional multivariate normal distribution with mean $\boldsymbol{\mu}_{\boldsymbol{k}_{ij}}$ and covariance matrix $\sigma^2 \mathbf{S}_{\boldsymbol{k}_{ij}}(\rho, \rho^*)$ given by (4.30) and (4.31) respectively. Each subject j, from population i can be thought as coming from a one population model of the type introduced in the previous section, with parameter vector $\boldsymbol{\theta}_i = (\delta, \gamma, \lambda_0, \mu_{i1}, ..., \mu_{iT}, \sigma^2, \rho, \rho^*)'$.

Thus, we apply the appropriate transformation from Section 4.1 and define $\boldsymbol{Y}_{ij} = \boldsymbol{\Gamma}_{\boldsymbol{k}_{ij}} \boldsymbol{X}_{ij}$, where

$$\boldsymbol{\Gamma}_{\boldsymbol{k}_{ij}} = \boldsymbol{\Gamma}_{k_{ij1},\ldots,k_{ijT}} = \begin{pmatrix} \boldsymbol{\Gamma}_{k_{ijr_{ij1}}} & \mathbf{O}_{k_{ijr_{ij1}},k_{ijr_{ij2}}} & \ldots & \mathbf{O}_{k_{ijr_{ij1}},k_{ijr_{ijl(\boldsymbol{k}_{ij})}}} \\ \mathbf{O}_{k_{ijr_{ij2}},k_{ijr_{ij1}}} & \boldsymbol{\Gamma}_{k_{ijr_{ij2}}} & \ldots & \mathbf{O}_{k_{ijr_{ij2}},k_{ijr_{ijl(\boldsymbol{k}_{ij})}}} \\ \ldots & \ldots & \ldots & \ldots \\ \mathbf{O}_{k_{ijr_{ijl(\boldsymbol{k}_{ij})}},k_{ijr_{ij1}}} & \mathbf{O}_{k_{ijr_{ijl(\boldsymbol{k}_{ij})}},k_{ijr_{ij2}}} & \ldots & \boldsymbol{\Gamma}_{k_{ijr_{ijl(\boldsymbol{k}_{ij})}}} \end{pmatrix}.$$

As before, $\mathbf{O}_{k,l}$ denotes the $k \times l$ matrix with all the entries equal to 0. We notice that the entries in the vector Y_{ij} are uncorrelated; furthermore, except for the first entries in each of the subvectors Y_{ijt}, $t = 1, \ldots, T$, provided that the corresponding length K_{ijt} is nonzero, all entries have the same mean 0 and variance σ^2/τ_0. We have

$$\boldsymbol{Y}_{ij} = \Big((Y_{ijr_{ij1}1}\ Y_{ijr_{ij2}1}\ \ldots\ Y_{ijr_{ijl(\boldsymbol{k}_{ij})}1})\ (Y_{ijr_{ij1}2}\ \ldots\ Y_{ijr_{ij1}k_{ijr_{ij1}}}\ Y_{ijr_{ij2}2}\ \ldots\ Y_{ir_{i2}k_{ir_{i2}}} \ldots\ Y_{ir_{il(\boldsymbol{k}_i)}2}\ \ldots\ Y_{ir_{il(\boldsymbol{k}_i)}k_{ir_{il(\boldsymbol{k}_i)}}}) \Big)'$$

Denote by $\boldsymbol{Z}_{ij}$ the vector composed with the first entries in the non-zero length vectors $\boldsymbol{Y}_{ijt}$

$$\boldsymbol{Z}_{ij} = \Big(Y_{ijr_{ij1}1}\ Y_{ijr_{ij2}1}\ \ldots\ Y_{ijr_{ijl(\boldsymbol{k}_{ij})}1} \Big)' = \big(Z_{ij1}\ Z_{ij2}\ \ldots\ Z_{ijl(\boldsymbol{k}_{ij})} \big)',$$

and by $\widehat{Y}_{ij}$ the vector containing the remaining components of $\boldsymbol{Y}_{ij}$, which are conditionally independent, identically distributed univariate normal random variables, with mean 0 and variance σ^2/τ_0

$$\widehat{\boldsymbol{Y}}_{ij} = \Big(Y_{ijr_{ij1}2}\ \ldots\ Y_{ijr_{ij1}k_{ijr_{ij1}}}\ Y_{ijr_{ij2}2}\ \ldots\ Y_{ijr_{ij2}k_{ijr_{ij2}}} \ldots\ Y_{ijr_{ijl(\boldsymbol{k}_{ij})}2}\ \ldots\ Y_{ijr_{ijl(\boldsymbol{k}_{ij})}k_{ijr_{ijl(\boldsymbol{k}_{ij})}}} \Big)'.$$

Note that $\boldsymbol{Z}_{ij}$ has dimension $l(\boldsymbol{k}_{ij}) = \sum_{t=1}^{T} \delta(k_{ijt})$ and $\widehat{\boldsymbol{Y}}_{ij}$ has dimension $\sum_{t=1}^{T} k_{ijt} - l(\boldsymbol{k}_{ij})$. Denoting

$$\boldsymbol{\mu}^*_{ij} = \boldsymbol{\mu}_{ij}(\boldsymbol{k}_{ij}) = (\mu_{ir_{ij1}}\ \mu_{ir_{ij2}}\ \ldots\ \mu_{ir_{ijl(\boldsymbol{k}_{ij})}})' = (\mu^*_{ij1}\ \mu^*_{ij2}\ \ldots\ \mu^*_{ijl(\boldsymbol{k}_{ij})})',$$

it is straightforward to show that the contribution of the $j-$th subject from population i to the log-likelihood is:

$$\begin{aligned}\log f(\boldsymbol{y}_{ij},\boldsymbol{k}_{ij}) &= -\lambda_0+\sum_{t=1}^{T}(-\lambda_{it}+k_{ijt}\log\lambda_{it})+\log\left(\sum_{l=0}^{s}\prod_{t=1}^{T}\binom{k_{ijt}}{l}l!\left(\frac{\lambda_0}{\prod\limits_{k=1}^{T}\lambda_{ik}}\right)^{l}\right)\\ &\quad+\delta(\boldsymbol{k}_{ij})\left[-\frac{\log\sigma^2}{2}\sum_{t=1}^{T}k_{ijt}+\frac{\log\tau_0}{2}\left(\sum_{t=1}^{T}k_{ijt}-\sum_{t=1}^{T}\delta(k_{ijt})\right)\right.\\ &\quad\left.-\frac{1}{2}\log\left|\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}\right|-\frac{1}{2\sigma^2}(\boldsymbol{z}_{ij}-\boldsymbol{\mu}^*_{ij})'\boldsymbol{\Sigma}^{-1}_{\boldsymbol{k}_{ij}}(\boldsymbol{z}_{ij}-\boldsymbol{\mu}^*_{ij})-\frac{\tau_0}{2\sigma^2}\sum_{t=1}^{l(\boldsymbol{k}_{ij})}\sum_{l=2}^{k_{ijr_{ijt}}}y^2_{ijr_{ijt}l}\right],\end{aligned}$$

where

$$\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}=\boldsymbol{\Sigma}_{\tilde{\boldsymbol{k}}_{ij}}=\begin{pmatrix}\frac{1}{\tau_{\tilde{k}_{ij1}}} & \rho^* & ... & \rho^*\\ \rho^* & \frac{1}{\tau_{\tilde{k}_{ij2}}} & ... & \rho^*\\ ... & ... & ... & ...\\ \rho^* & \rho^* & ... & \frac{1}{\tau_{\tilde{k}_{ijl(\boldsymbol{k}_{ij})}}}\end{pmatrix} \tag{4.32}$$

Recalling that λ_{it} has the form $\lambda_{it}=\exp{(\delta+\gamma\mu_{it})}$, we can write

$$\begin{aligned}\log f(\boldsymbol{y}_{ij},\boldsymbol{k}_{ij}) &= -\lambda_0+\sum_{t=1}^{T}(-\exp{(\delta+\gamma\mu_{it})}+k_{ijt}\log\lambda_{it})+\log\left(\sum_{l=0}^{s}\prod_{t=1}^{T}\binom{k_{ijt}}{l}l!\left(\frac{\lambda_0}{\prod\limits_{k=1}^{T}\lambda_{ik}}\right)^{l}\right)\\ &\quad-\frac{\log\sigma^2}{2}\sum_{t=1}^{T}k_{ijt}+\frac{\log\tau_0}{2}\left(\sum_{t=1}^{T}k_{ijt}-\sum_{t=1}^{T}\delta(k_{ijt})\right)\\ &\quad-\frac{1}{2}\log\left|\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}\right|-\frac{1}{2\sigma^2}(\boldsymbol{z}_{ij}-\boldsymbol{\mu}^*_{ij})'\boldsymbol{\Sigma}^{-1}_{\boldsymbol{k}_{ij}}(\boldsymbol{z}_{ij}-\boldsymbol{\mu}^*_{ij})-\frac{\tau_0}{2\sigma^2}\sum_{t=1}^{l(\boldsymbol{k}_{ij})}\sum_{l=2}^{k_{ijr_{ijt}}}y^2_{ijr_{ijt}l}.\end{aligned}$$

Instead of multiplying by $\delta(\boldsymbol{k}_{ij})$ the last 5 terms in the right hand side of the above equation, we make the convention to consider them equal to zero for the case when $\delta(\boldsymbol{k}_{ij})$ is zero (i.e.

when all the random lengths for a subject are zero). The log-likelihood of the entire data set is

$$\begin{aligned}
l(\boldsymbol{\theta}) &= -\lambda_0 \sum_{i=1}^{m} n_i + \sum_{i=1}^{m}\sum_{j=1}^{n_i}\sum_{t=1}^{T}\left[-e^{\delta+\gamma\mu_{it}} + k_{ijt}\left(\delta+\gamma\mu_{it}\right)\right] \\
&+\sum_{i=1}^{m}\sum_{j=1}^{n_i}\log\left(\sum_{l=0}^{s}\prod_{t=1}^{T}\binom{k_{ijt}}{l}l!\left(\frac{\lambda_0}{\prod_{k=1}^{T}\lambda_{ik}}\right)^{l}\right) \\
&-\frac{\log\sigma^2}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sum_{t=1}^{T}k_{ijt} \\
&+\frac{\log\tau_0}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\sum_{t=1}^{T}k_{ijt}-\sum_{t=1}^{T}\delta(k_{ijt})\right)-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\log\left|\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}\right| \\
&-\frac{1}{2\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\boldsymbol{z}_{ij}-\boldsymbol{\mu}^*_{ij})'\boldsymbol{\Sigma}^{-1}_{\boldsymbol{k}_{ij}}(\boldsymbol{z}_{ij}-\boldsymbol{\mu}^*_{ij})-\frac{\tau_0}{2\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sum_{t=1}^{l(\boldsymbol{k}_{ij})}\sum_{l=2}^{k_{ijr_{ijt}}}y^2_{ijr_{ijt}l}, \quad (4.33)
\end{aligned}$$

where the parameter vector is $\boldsymbol{\theta} = (\delta, \gamma, \lambda_0, \mu_{11}, ..., \mu_{mT}, \sigma^2, \rho, \rho^*)'$.
Denoting by

$$Q_{\boldsymbol{k}_{ij}}(\lambda_0, \lambda_{i1}, ..., \lambda_{iT}) = \sum_{l=0}^{\min(k_{ij1},...,k_{ijT})}\prod_{t=1}^{T}\binom{k_{ijt}}{l}l!\left(\frac{\lambda_0}{\prod_{k=1}^{T}\lambda_{ik}}\right)^{l} \quad (4.34)$$

and using the results (A.5) - (A.8) from Appendix A, it follows that the $mT + 6$ score equations are given by

$$
\begin{aligned}
\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sum_{j=1}^{T}\left(-e^{\delta+\gamma\mu_j}+k_{ijt}\right)+\sum_{i=1}^{m}\sum_{j=1}^{n_i}\frac{\partial}{\partial\delta}\log Q_{\boldsymbol{k}_{ij}}(\lambda_0,\lambda_{i1},...,\lambda_{iT}) &= 0\\
\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sum_{t=1}^{T}\left(-\mu_{it}e^{\delta+\gamma\mu_{it}}+k_{ijt}\mu_{it}\right)+\sum_{i=1}^{m}\sum_{j=1}^{n_i}\frac{\partial}{\partial\gamma}\log Q_{\boldsymbol{k}_{ij}}(\lambda_0,\lambda_{i1},...,\lambda_{iT}) &= 0\\
-\sum_{j=1}^{m}n_i+\sum_{i=1}^{m}\sum_{j=1}^{n_i}\frac{\partial}{\partial\lambda_0}\log Q_{\boldsymbol{k}_{1j}}(\lambda_0,\lambda_{i1},...,\lambda_{iT}) &= 0\\
\gamma\begin{pmatrix} -n_1\lambda_{11}+\sum\limits_{j=1}^{n_1}k_{1j1}\\ -n_1\lambda_{12}+\sum\limits_{j=1}^{n_1}k_{1j2}\\ \vdots\\ -n_1\lambda_{1T}+\sum\limits_{j=1}^{n_1}k_{1jT}\end{pmatrix}+\begin{pmatrix}\sum\limits_{j=1}^{n_1}\frac{\partial}{\partial\mu_{11}}\log Q_{\boldsymbol{k}_{1j}}(\lambda_0,\lambda_{11},...,\lambda_{1T})\\ \sum\limits_{j=1}^{n_1}\frac{\partial}{\partial\mu_{12}}\log Q_{\boldsymbol{k}_{1j}}(\lambda_0,\lambda_{11},...,\lambda_{1T})\\ \vdots\\ \sum\limits_{j=1}^{n_1}\frac{\partial}{\partial\mu_{1T}}\log Q_{\boldsymbol{k}_{1j}}(\lambda_0,\lambda_{11},...,\lambda_{1T})\end{pmatrix}+\frac{1}{\sigma^2}\sum_{j=1}^{n_1}\boldsymbol{\Sigma}_{\boldsymbol{k}_{1j}}^{-1}(\boldsymbol{z}_{1j}-\boldsymbol{\mu}_{1j}^{*}) &= \mathbf{0}_T\\
&\ \ \vdots\\
\gamma\begin{pmatrix} -n_m\lambda_{m1}+\sum\limits_{j=1}^{n_m}k_{mj1}\\ -n_m\lambda_{m2}+\sum\limits_{j=1}^{n_m}k_{mj2}\\ \vdots\\ -n_m\lambda_{mT}+\sum\limits_{j=1}^{n_m}k_{mjT}\end{pmatrix}+\begin{pmatrix}\sum\limits_{j=1}^{n_m}\frac{\partial}{\partial\mu_{m1}}\log Q_{\boldsymbol{k}_{mj}}(\lambda_0,\lambda_{m1},...,\lambda_{mT})\\ \sum\limits_{j=1}^{n_m}\frac{\partial}{\partial\mu_{m2}}\log Q_{\boldsymbol{k}_{mj}}(\lambda_0,\lambda_{m1},...,\lambda_{mT})\\ \vdots\\ \sum\limits_{j=1}^{n_m}\frac{\partial}{\partial\mu_{mT}}\log Q_{\boldsymbol{k}_{mj}}(\lambda_0,\lambda_{m1},...,\lambda_{mT})\end{pmatrix}+\frac{1}{\sigma^2}\sum_{j=1}^{n_m}\boldsymbol{\Sigma}_{\boldsymbol{k}_{mj}}^{-1}(\boldsymbol{z}_{mj}-\boldsymbol{\mu}_{mj}^{*}) &= \mathbf{0}_T\\
-\frac{1}{\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sum_{t=1}^{T}k_{ijt}+\frac{1}{\sigma^4}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\boldsymbol{z}_{ij}-\boldsymbol{\mu}_{ij}^{*})'\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}^{-1}(\boldsymbol{z}_{ij}-\boldsymbol{\mu}_{ij}^{*})+\frac{\tau_0}{\sigma^4}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\widehat{\boldsymbol{y}}_{ij}^{'}\widehat{\boldsymbol{y}}_{ij} &= 0\\
\frac{1}{1-\rho}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\sum_{t=1}^{T}k_{ijt}-\sum_{t=1}^{T}\delta(k_{ijt})\right)-\sum_{i=1}^{m}\sum_{j=1}^{n_i}tr\left(\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}^{-1}\left[\mathbf{I}_T-\mathrm{Diag}\left(\frac{1}{k_{ij1}},...,\frac{1}{k_{ijT}}\right)\right]\right)&\\
+\frac{1}{\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\boldsymbol{z}_{ij}-\boldsymbol{\mu}_{ij}^{*})^{'}\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}^{-1}\left[\mathbf{I}_T-\mathrm{Diag}\left(\frac{1}{k_{ij1}},...,\frac{1}{k_{ijT}}\right)\right]\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}^{-1}(\boldsymbol{z}_{ij}-\boldsymbol{\mu}_{ij}^{*})&\\
-\frac{\tau_0}{\sigma^2(1-\rho)^2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\widehat{\boldsymbol{y}}_{ij}^{'}\widehat{\boldsymbol{y}}_{ij} &= 0\\
\sum_{i=1}^{m}\sum_{j=1}^{n_i}tr\left(\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}^{-1}(\boldsymbol{e}_T\boldsymbol{e}_T^{'}-\mathbf{I}_T)\right)-\frac{1}{\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\boldsymbol{z}_{ij}-\boldsymbol{\mu}_{ij}^{*})^{'}\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}^{-1}(\boldsymbol{e}_T\boldsymbol{e}_T^{'}-\mathbf{I}_T)\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}^{-1}(\boldsymbol{z}_{ij}-\boldsymbol{\mu}_{ij}^{*}) &= 0.
\end{aligned}
$$

We use the fact that $\sum\limits_{t=1}^{l(\boldsymbol{k}_{ij})}\sum\limits_{l=2}^{k_{ijr_{ijt}}}y_{ijr_{ijt}l}^2=\widehat{\boldsymbol{y}}_{ij}^{'}\widehat{\boldsymbol{y}}_{ij}$ to simplify the above expressions. $\mathbf{0}_T$ denotes the T-dimensional null vector.

As noted before, if all the observations consist of random length vectors with length zero, none of the equations above make sense; hence none of the parameters are estimable. If we observe only vectors with length zero or 1 then the parameter ρ is not estimable.

Similar to the approach taken in the one population model, the explicit expressions for the score equations are presented only for the case $T = 2$.

4.2.2.1 Maximum Likelihood Estimation for 2 Time points Let us assume each subject was observed only $T = 2$ times. Thus, the random lengths for a subject from population i are distributed jointly as bivariate Poisson$(\lambda_0, \lambda_{i1}, \lambda_{i2})$ random variables. Let us denote this joint probability mass function by

$$\begin{aligned} f_i(r,s) &= \mathbf{P}_{\boldsymbol{\theta}}(K_{i1} = r, K_{i2} = s) \\ &= e^{-\sum\limits_{i=0}^{2}\lambda_i} \frac{\lambda_{i1}^r}{r!}\frac{\lambda_{i2}^s}{s!} \sum_{i=0}^{\min(r,s)} \binom{r}{i}\binom{s}{i} i! \left(\frac{\lambda_0}{\lambda_{i1}\lambda_{i2}}\right)^i . \end{aligned} \tag{4.35}$$

Taking into account the expression of the score equations for the vector of lengths in (C.5) - (C.9), we may explicitly write the score equations for the multiple population model for $T = 2$ time points as

$$-\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sum_{t=1}^{2} e^{\delta+\gamma\mu_{it}} + \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left[\lambda_{i1}\frac{f_i(k_{ij1}-1,k_{ij2})}{f_i(k_{ij1},k_{ij2})} + \lambda_{i2}\frac{f_i(k_{ij1},k_{ij2}-1)}{f_i(k_{ij1},k_{ij2})}\right] = 0 \quad (4.36)$$

$$-\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sum_{t=1}^{2} \mu_{it} e^{\delta+\gamma\mu_{it}} + \sum_{i=1}^{m}\sum_{j=1}^{n_i}\left[\lambda_{i1}\frac{f_i(k_{ij1}-1,k_{ij2})}{f_i(k_{ij1},k_{ij2})} + \lambda_{i2}\frac{f_i(k_{ij1},k_{ij2}-1)}{f_i(k_{ij1},k_{ij2})}\right] = 0 \quad (4.37)$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}\frac{f_i(k_{ij1},k_{ij2}) - f_i(k_{ij1}-1,k_{ij2}) - f_i(k_{ij1},k_{ij2}-1) + f_i(k_{ij1}-1,k_{ij2}) - 1}{f(k_{ij1},k_{ij2})} = 0 \quad (4.38)$$

$$\gamma\begin{pmatrix} -n_1\lambda_{11} + \lambda_{11}\sum_{j=1}^{n_1}\frac{f(k_{1j1}-1,k_{1j2})}{f(k_{1j1},k_{1j2})} \\ -n_1\lambda_{12} + \lambda_{12}\sum_{j=1}^{n_1}\frac{f(k_{1j1},k_{1j2}-1)}{f(k_{1j1},k_{1j2})} \end{pmatrix} + \frac{1}{\sigma^2}\sum_{j=1}^{n_1}\boldsymbol{\Sigma}_{\boldsymbol{k}_{1j}}^{-1}(\boldsymbol{z}_{1j} - \boldsymbol{\mu}_{1j}^*) = \mathbf{0}_2 \quad (4.39)$$

$$\vdots$$

$$\gamma\begin{pmatrix} -n_m\lambda_{m1} + \lambda_{m1}\sum_{j=1}^{n_m}\frac{f(k_{mj1}-1,k_{mj2})}{f(k_{mj1},k_{mj2})} \\ -n_m\lambda_{m2} + \sum_{j=1}^{n_m}\frac{f(k_{mj1},k_{mj2}-1)}{f(k_{mj1},k_{mj2})} \end{pmatrix} + \frac{1}{\sigma^2}\sum_{j=1}^{n_m}\boldsymbol{\Sigma}_{\boldsymbol{k}_{mj}}^{-1}(\boldsymbol{z}_{mj} - \boldsymbol{\mu}_{mj}^*) = \mathbf{0}_2 \quad (4.40)$$

$$-\frac{1}{\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\sum_{t=1}^{2} k_{ijt} + \frac{1}{\sigma^4}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\boldsymbol{z}_{ij} - \boldsymbol{\mu}_{ij}^*)'\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}^{-1}(\boldsymbol{z}_{ij} - \boldsymbol{\mu}_{ij}^*) + \frac{\tau_0}{\sigma^4}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\widehat{\boldsymbol{y}}_{ij}'\widehat{\boldsymbol{y}}_{ij} = 0 \quad (4.41)$$

$$\begin{aligned}\frac{1}{1-\rho}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(\sum_{t=1}^{2} k_{ijt} - l(\boldsymbol{k}_{ij})\right) - \sum_{i=1}^{m}\sum_{j=1}^{n_i} tr\left(\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}^{-1}\text{Diag}\left(1-\frac{1}{k_{ij1}}, 1-\frac{1}{k_{ij2}}\right)\right) \\ +\frac{1}{\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\boldsymbol{z}_{ij} - \boldsymbol{\mu}_{ij}^*)'\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}^{-1}\text{Diag}\left(1-\frac{1}{k_{ij1}}, 1-\frac{1}{k_{ij2}}\right)\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}^{-1}(\boldsymbol{z}_{ij} - \boldsymbol{\mu}_{ij}^*) \\ -\frac{\tau_0}{\sigma^2(1-\rho)^2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\widehat{\boldsymbol{y}}_{ij}'\widehat{\boldsymbol{y}}_{ij} &= 0 \quad (4.42)\end{aligned}$$

$$\begin{aligned}\sum_{i=1}^{m}\sum_{j=1}^{n_i} tr\left(\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}^{-1}(\boldsymbol{e}_2\boldsymbol{e}_2{}' - \mathbf{I}_2)\right) \\ -\frac{1}{\sigma^2}\sum_{i=1}^{m}\sum_{j=1}^{n_i}(\boldsymbol{z}_{ij} - \boldsymbol{\mu}_{ij}^*)'\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}^{-1}(\boldsymbol{e}_2\boldsymbol{e}_2{}' - \mathbf{I}_2)\boldsymbol{\Sigma}_{\boldsymbol{k}_{ij}}^{-1}(\boldsymbol{z}_{ij} - \boldsymbol{\mu}_{ij}^*) &= 0. \quad (4.43)\end{aligned}$$

We discuss numerical estimation of the MLE's in Section 4.3.3, along with considerations about the technical difficulties we encountered.

4.2.3 Asymptotic Distribution of the MLE for $T = 2$

Let us denote by $\boldsymbol{I}_n(\boldsymbol{\theta})$ the information matrix for $\boldsymbol{\theta}$ contained in the $n = \sum_{i=1}^{m} n_i$ independent observations from the multiple population model, $\boldsymbol{X}_{ij}$ with random lengths $\boldsymbol{K}_{ij} = (K_{ij1}, ..., K_{ijT})$, $i = 1, ..., m$, $j = 1, ..., n_i$. It makes sense to think of $\boldsymbol{I}_n(\boldsymbol{\theta})$ as a sum of the information about $\boldsymbol{\theta}$ contained in the lengths and the sum of information about $\boldsymbol{\theta}$ contributed by the vectors of severities, over all possible lengths. Using the results from Appendix C, we compute the information about the parameter $\boldsymbol{\theta}$ contained in a single observation from the one population model, $\boldsymbol{X}$ with random lengths $\boldsymbol{K} = (K_1, ..., K_T)$ as

$$\boldsymbol{I}(\boldsymbol{\theta}) = \boldsymbol{I}^*(\boldsymbol{\theta}) + \sum_{\boldsymbol{k} \in \Upsilon} \mathbf{P}_{\boldsymbol{\theta}}(\boldsymbol{K} = \boldsymbol{k}) \boldsymbol{I}(\boldsymbol{\theta}|\boldsymbol{k}),$$

where $\boldsymbol{I}^*(\boldsymbol{\theta})$ is the information matrix about $\boldsymbol{\theta}$ contained in the random lengths $K = (K_1, ..., K_T)$ and $\boldsymbol{I}(\boldsymbol{\theta}|\boldsymbol{k})$ is the information matrix contained in $\boldsymbol{X}_i|\boldsymbol{K} = \boldsymbol{k}$, where $k \in \Upsilon$. We denote

$$\Upsilon = \left\{ \boldsymbol{k} = (k_1, ...k_T)| \; k_i = 0, 1, ... \text{ for } \forall i = 1, ..., T \text{ and } \sum_{i=1}^{T} k_i \geq 1 \right\}.$$

We compute the information matrix $\boldsymbol{I}_n(\boldsymbol{\theta})$ about $\boldsymbol{\theta} = (\delta, \gamma, \lambda_0, \mu_{11}, ..., \mu_{mT}, \sigma^2, \rho, \rho^*)'$ contained in these $n = \sum\limits_{i=1}^{m} n_i$ independent observations from the multiple population model (see C.27) as

$$\boldsymbol{I}_n(\boldsymbol{\theta}) = \sum_{i=1}^{m} n_i \boldsymbol{I}_{\boldsymbol{i}}^*(\boldsymbol{\theta}) + \sum_{i=1}^{m} n_i \sum_{\boldsymbol{k} \in \Upsilon} \left[e^{-\lambda_0 - \sum\limits_{t=1}^{2} \lambda_{it}} \frac{\lambda_{i1}^{k_1}}{k_1!} \frac{\lambda_{i2}^{k_2}}{k_2!} \sum_{i=0}^{\min(k_1,k_2)} \binom{k_1}{i} \binom{k_2}{i} i! \left(\frac{\lambda_0}{\lambda_{i1}\lambda_{i2}} \right)^i \right] \boldsymbol{I}_{\boldsymbol{i}}(\boldsymbol{\theta}|\boldsymbol{k}_{\boldsymbol{i}}). \quad (4.44)$$

We show in Appendix C, relation C.27, that $\boldsymbol{I}_n(\boldsymbol{\theta})$ has a block diagonal form

$$\boldsymbol{I}_n(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{I}_n(\delta, \gamma, \lambda_0, \mu_{11}, ..., \mu_{m2}) & \mathbf{O}_{(2m+3)\times 3} \\ \mathbf{O}_{3\times(2m+3)} & \sum\limits_{i=1}^{m} n_i \sum\limits_{\boldsymbol{k} \in \Upsilon} f_i(k_1, k_2) \boldsymbol{I}_{\boldsymbol{k}}(\sigma^2, \rho, \rho^*) \end{pmatrix}, \quad (4.45)$$

where

$$f_i(k_1, k_2) = e^{-\lambda_0 - \sum\limits_{t=1}^{2} \lambda_{it}} \frac{\lambda_{i1}^{k_1}}{k_1!} \frac{\lambda_{i2}^{k_2}}{k_2!} \sum_{i=0}^{\min(k_1,k_2)} \binom{k_1}{i} \binom{k_2}{i} i! \left(\frac{\lambda_0}{\lambda_{i1}\lambda_{i2}} \right)^i,$$

$$\boldsymbol{I}_n(\delta, \gamma, \lambda_0, \mu_{11}, ..., \mu_{m2}) = \sum_{i=1}^{m} n_i \begin{pmatrix} \mathbf{H}_{11}^{(i)} & \mathbf{O}_{3\times 2(i-1)} & \mathbf{H}_{12}^{(i)} & \mathbf{O}_{3\times 2(m-i)} \\ \mathbf{O}_{2(i-1)\times 3} & \mathbf{O}_{2(i-1)\times 2(i-1)} & \mathbf{O}_{2(i-1)\times 2} & \mathbf{O}_{2(i-1)\times 2(m-i)} \\ \mathbf{H}_{12}^{(i)'} & \mathbf{O}_{2\times 2(i-1)} & \mathbf{H}_{22}^{(i)} & \mathbf{O}_{2\times 2(m-i)} \\ \mathbf{O}_{2(m-i)\times 3} & \mathbf{O}_{2(m-i)\times 2(i-1)} & \mathbf{O}_{2(m-i)\times 2} & \mathbf{O}_{2(m-i)\times 2(m-i)} \end{pmatrix}$$

$$+ \begin{pmatrix} \mathbf{O}_{3\times 3} & \mathbf{O}_{3\times 2m} \\ \mathbf{O}_{2m\times 3} & \sum\limits_{i=1}^{m} n_i \sum\limits_{\boldsymbol{k} \in \Upsilon} f_i(k_1, k_2) \boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} \end{pmatrix},$$

the matrices $\mathbf{H}_{kl}^{(i)}$'s $(k, l = 1, 2)$ are given in (C.26) and the elements of the matrix $\boldsymbol{I}_{\boldsymbol{k}}(\sigma^2, \rho, \rho^*)$ are defined by (A.10) - (A.15).

From the expression of the score equations it is apparent that there is no closed form solution for $\hat{\boldsymbol{\theta}}_n$. Hence, the exact distribution of $\hat{\boldsymbol{\theta}}_n$ is not available. We can apply a general result on the efficiency of maximum likelihood estimators for random length data (see Theorem A.3.2 in Barnhart [4]) to derive the asymptotic distribution for $\hat{\boldsymbol{\theta}}_n$, the MLE. We obtain its asymptotic covariance matrix as the inverse of the above information matrix and estimate it by $\boldsymbol{I}_n^{-1}(\hat{\boldsymbol{\theta}}_n)$.

Theorem 4. *Let* $\hat{\boldsymbol{\theta}}_n = (\hat{\delta}_n, \hat{\gamma}_n, \hat{\lambda}_0^{(n)}, \hat{\mu}_{11}^{(n)}, \hat{\mu}_{12}^{(n)}, ..., \hat{\mu}_{m1}^{(n)}, \hat{\mu}_{m2}^{(n)}, \hat{\sigma}_n^2, \hat{\rho}_n, \hat{\rho}_n^*)'$ *be the MLEs for a sample of size* $n = \sum\limits_{i=1}^{m} n_1$ *from the multiple population model. If* $n_i/n \to \eta_i$ *with* $0 < \eta_i < 1$ *as* $n \to \infty$*, then*

(1). $\hat{\boldsymbol{\theta}}_n$ *is consistent.*

(2).

$$\sqrt{n}\left(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\right) \xrightarrow{L} \boldsymbol{MVN}_{2m+6}(\mathbf{0}, \boldsymbol{I}^{-1}(\boldsymbol{\theta})),$$

where

$$\boldsymbol{I}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{I}(\delta, \gamma, \lambda_0, \mu_{11}, ..., \mu_{m2}) & \boldsymbol{O}_{(2m+3)\times 3} \\ \boldsymbol{O}_{3\times(2m+3)} & \sum\limits_{i=1}^{m} \eta_i \sum\limits_{\boldsymbol{k}\in\Upsilon} f_i(k_1, k_2) \boldsymbol{I}_{\boldsymbol{k}}(\sigma^2, \rho, \rho^*) \end{pmatrix}, \tag{4.46}$$

where

$$f_i(k_1, k_2) = e^{-\lambda_0 - \sum\limits_{t=1}^{2} \lambda_{it}} \frac{\lambda_{i1}^{k_1}}{k_1!} \frac{\lambda_{i2}^{k_2}}{k_2!} \sum_{i=0}^{\min(k_1,k_2)} \binom{k_1}{i} \binom{k_2}{i} i! \left(\frac{\lambda_0}{\lambda_{i1}\lambda_{i2}}\right)^i,$$

$$\begin{aligned} \boldsymbol{I}(\delta, \gamma, \lambda_0, \mu_{11}, ..., \mu_{m2}) \quad = \quad & \sum_{i=1}^{m} \eta_i \begin{pmatrix} \boldsymbol{H}_{11}^{(i)} & \boldsymbol{O}_{3\times 2(i-1)} & \boldsymbol{H}_{12}^{(i)} & \boldsymbol{O}_{3\times 2(m-i)} \\ \boldsymbol{O}_{2(i-1)\times 3} & \boldsymbol{O}_{2(i-1)\times 2(i-1)} & \boldsymbol{O}_{2(i-1)\times 2} & \boldsymbol{O}_{2(i-1)\times 2(m-i)} \\ \boldsymbol{H}_{12}^{(i)\prime} & \boldsymbol{O}_{2\times 2(i-1)} & \boldsymbol{H}_{22}^{(i)} & \boldsymbol{O}_{2\times 2(m-i)} \\ \boldsymbol{O}_{2(m-i)\times 3} & \boldsymbol{O}_{2(m-i)\times 2(i-1)} & \boldsymbol{O}_{2(m-i)\times 2} & \boldsymbol{O}_{2(m-i)\times 2(m-i)} \end{pmatrix} \\ & + \begin{pmatrix} \boldsymbol{O}_{3\times 3} & \boldsymbol{O}_{3\times 2m} \\ \boldsymbol{O}_{2m\times 3} & \sum\limits_{i=1}^{m} \eta_i \sum\limits_{\boldsymbol{k}\in\Upsilon} f_i(k_1, k_2) \boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} \end{pmatrix}, \end{aligned}$$

the matrices $\boldsymbol{H}_{kl}^{(i)}$*'s* $(k, l = 1, 2)$ *are given in (C.26), the elements of* $\boldsymbol{I}_{\boldsymbol{k}}(\sigma^2, \rho, \rho^*)$ *are defined by (A.10) - (A.15), and* $\Upsilon = \left\{ \boldsymbol{k} = (k_1, k_2) | \; k_i = 0, 1, \ldots \quad \forall i = 1, 2 \text{ and } \sum_{i=1}^{2} k_i \geq 1 \right\}$.

4.3 SIMULATION STUDY

In this section we report the results of a simulation study conducted to explore the behavior of the proposed model for $T = 2$ and to evaluate the finite sample properties of our estimators. Several different scenarios are analyzed, with longitudinal random length data generated according to the multiple population model described in Section 4.2. We compare the estimated parameters with the true underlying values, investigate how close the asymptotic variance approximates the finite sample variance and examine the normality of the estimators. In addition we examine how large the groups need to be in order for the large-sample theory to hold.

4.3.1 Description of the Simulations

The simulation study was designed to resemble the LEDS data; we create a hypothetical trial with patients divided evenly between a treatment and a control group and followed up for the same number of time periods $T = 2$. To allow for comparisons, we find it useful to report here the results of the simulations for $T = 2$ from Chapter 3.

In order to be able to compare the results from the two multiple population models, we keep the same parameter configurations for $\delta, \gamma, \mu_{11}, ..., \mu_{22}$, σ^2, ρ and ρ^* as in the simulation study from Chapter 3 and set the value of λ_0 to 1 across all simulations. We try to cover the same four different scenarios obtained by considering all possible combinations of the two factors taken into account; the first factor involves the relationship between the severities and the lengths and the second one concerns the relationship between the mean severity "profiles" for the two populations. In the first scenario $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are parallel and not coincident, while within the same treatment group the mean severities are close but the number of events are well separated across time; the second scenario depicts the situation in which $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ intersect while within the same group the mean severities are close and the number of events are well separated across time; in the third scenario $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are parallel and not coincident, but over time the values of the mean severities are well separated, while the number of events change very little; finally, in the fourth scenario $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ intersect, while across time we have changes in the mean severities but little variation in the number of events. The explicit parameter configurations are shown in Table 19.

Table 19: Choice of parameters for simulation study. T=2

	μ's close and λ's far	μ's far and λ's close
$\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ well separated	$\boldsymbol{\mu_1} = (1.0, 1.1)$ $\boldsymbol{\mu_2} = (1.5, 1.6)$ $(\delta, \gamma, \lambda_0) = (-1.3, 2.3, 1)$ $(\sigma^2, \rho, \rho^*) = (1, 0.5, 0.2)$ $\boldsymbol{\lambda_1} = (2.72, 3.42)$ $\boldsymbol{\lambda_2} = (8.58, 10.80)$	$\boldsymbol{\mu_1} = (1, 2)$ $\boldsymbol{\mu}_2 = (1.5, 2.5)$ $(\delta, \gamma, \lambda_0) = (2.5, 0.01, 1)$ $(\sigma^2, \rho, \rho^*) = (1, 0.5, 0.2)$ $\boldsymbol{\lambda_1} = (12.30, 12.43)$ $\boldsymbol{\lambda}_2 = (12.37, 12.49)$
$\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ close	$\boldsymbol{\mu_1} = (1.3, 1.2)$ $\boldsymbol{\mu_2} = (1.2, 1.3)$ $(\delta, \gamma, \lambda_0) = (-1, 2.5, 1)$ $(\sigma^2, \rho, \rho^*) = (1, 0.5, 0.2)$ $\boldsymbol{\lambda_1} = (4.48, 5.75)$ $\boldsymbol{\lambda_2} = (5.75, 4.48)$	$\boldsymbol{\mu_1} = (4, 3)$ $\boldsymbol{\mu}_2 = (3, 4)$ $(\delta, \gamma, \lambda_0) = (1.5, 0.05, 1)$ $(\sigma^2, \rho, \rho^*) = (1, 0.5, 0.2)$ $\boldsymbol{\lambda_1} = (5.47, 5.21)$ $\boldsymbol{\lambda}_2 = (5.21, 5.47)$

As in the case of the simulation study from Chapter 3, the data are generated and the analyses are performed using the R programming language and PittGrid's computational framework.

In each scenario, we generate $D = 1000$ independent data sets under the model introduced in Section 4.2 (Model 2), with two different populations and $T = 2$ time points. For each population i, $i = 1, 2$ we simulate the same number of subjects $n = 20, 50$ and 100.

For each data set d, $d = 1, ..., D$, we fit our multiple population model with $T = 2$ and, by maximizing the likelihood, we compute the numerical value of the 10 - dimensional ML estimator $\hat{\boldsymbol{\theta}}$, where $\boldsymbol{\theta} = (\delta, \gamma, \lambda_0, \mu_{11}, \mu_{12}, \mu_{21}, \mu_{22}, \sigma^2, \rho, \rho^*)$. For each of the 10 parameters, we compute the empirical bias, standard deviation and square-root of MSE, as described in Section 3.3.1. As in Chapter 3, we are not able to report estimated asymptotic variances for the parameters, because calculating them is extremely computationally challenging. Thus, we report instead the theoretical value $\boldsymbol{I}_n^{-1}(\boldsymbol{\theta})$.

4.3.2 Data Generation

The bivariate Poisson distribution described in Appendix B is employed to model the distribution of the random lengths. Thus, the random lengths of a subject from population i have the joint pmf given by

$$P(K_{i1} = k_1, K_{i2} = k_2) = e^{-\lambda_0 - \sum\limits_{t=1}^{2} \lambda_{it}} \frac{\lambda_{i1}^{k_1}}{k_1!} \frac{\lambda_{i2}^{k_2}}{k_2!} \sum_{i=0}^{\min(k_1,k_2)} \binom{k_1}{i} \binom{k_2}{i} i! \left(\frac{\lambda_0}{\lambda_{i1}\lambda_{i2}} \right)^i .$$

Given the random lengths $\boldsymbol{K}_i = (k_{i1}, k_{i2})$, the distribution of the vector of severities $\boldsymbol{X}_i$, corresponding to the two measurement times is $\mathbf{MVN}_{\sum\limits_{t=1}^{2} k_{it}} (\boldsymbol{\mu}_{\boldsymbol{k}}, \sigma^2 \mathbf{S}_{\boldsymbol{k}}(\rho, \rho^*))$. With probability $P(\boldsymbol{K} = \mathbf{0}) = e^{-\lambda_0 - \lambda_{i1} - \lambda_{i2}}$ we observe a zero-length vector for each of the two time measurements.

To generate a data set containing n observations per population from the multiple population model introduced in Section 4.2, we repeat n times the following two-step procedure for each of the populations $i = 1, 2$.

Step 1 (Generate the random lengths)

Generate 3 independent observations: one from a Poisson(λ_{it}) distribution, $t = 1, 2$ and one observation from a Poisson(λ_0) distribution. Adding the value generated from the Poisson(λ_0) distribution to each of the other two values generated from Poisson(λ_{it}) distributions produces the vector of random lengths $\boldsymbol{k}_i = (k_{i1}, k_{i2})$.

Step 2 (Generate the severities)

If $\sum_{t=1}^{2} k_{it} = 0$ we observe two zero-length vectors of severities.

If $\sum_{t=1}^{2} k_{it} > 0$, then generate $\boldsymbol{X}_i$ from a $\mathbf{MVN}_{\sum_{t=1}^{T} k_{it}}(\boldsymbol{\mu}_{\boldsymbol{k}_i}, \sigma^2 \mathbf{S}_{\boldsymbol{k}_i}(\rho, \rho^*))$.

4.3.3 Numerical Considerations

The difficulties that arose in this simulation study are the same as in the case of $T = 4$ and described in Section 3.4. Having $T = 2$ alleviates to some extent the computational complexity because there are fewer parameters to estimate. However, when the number of events is large, the algorithm is quite slow. In the case of the model with dependent lengths, an additional burden of complexity is brought by the probability mass function of the multivariate Poisson distribution. The complicated structure of the likelihood for the multivariate Poisson distribution is the main reason for us not being able to present the score equations and asymptotic distribution for the general model introduced in Chapter 4. Although there is a rich recent body of research involving the multivariate Poisson (see Karlis, [20]-[23]), numerical methods are implemented only for the bivariate Poisson case (Karlis [21]). We performed the simulation using our own program to compute the probability mass function for the bivariate Poisson distribution. We validated the program when $T = 2$, showing that the results were the same as those obtained by using the R package *bivpois*, implemented by Karlis [21], which allows efficient calculation of the bivariate Poisson probabilities. We did write a general program, capable of handling cases with $T > 2$, but for reasons we do not fully understand, while the program produces results, in some cases they appear to be meaningless.

As described in Section 3.3.4, the number of time points with quantitative measures changes with every subject. As the number of time points with quantitative measures changes, the mean and covariance structures for the distribution of the severities change. Furthermore, numerical computing of the information matrix involves summation over all possible values of the random lengths. For $T = 2$ measurement times that means summation of matrices over all the possible values of a bi-dimensional vector of lengths. In numerical computation of the information matrix, one needs to set threshold values for the summation indices. These thresholds are chosen such that the bivariate Poisson probability is negligible beyond them. The larger the average number of events is, the larger the thresholds become and making the summation more difficult. Furthermore, these matrices are weighted by a bivariate Poisson probability mass function. Having to compute this quantity for every combination of bivariate vector of random lengths adds another layer of complexity to the computation of the information matrix.

As in the case of the model of Chapter 3, the third complicating element is related to the difficulty associated with modeling slopes. If $\mu_{11} = ... = \mu_{22}$ the parameters δ and γ are not identifiable. When data comes from populations with poorly separated means, the likelihood can be flat over certain regions and *nlm* sometimes converges to some strange solutions, particularly for δ and γ. This was the case for the sample size $n = 20$ in scenario 2, when the theoretical means were close $\mu_{11} = \mu_{22} = 1.3$ and $\mu_{12} = \mu_{21} = 1.2$. However, even in these cases, the average severities and event lengths are estimated correctly.

The R function *nlm* was used to carry out the unrestricted maximization of the log-likelihood. This function requires specification of initial values for the parameters. We obtain the initial values of the parameters in a similar manner as in Chapter 3. For λ_0, we start with the method of moments estimator.

$$\lambda_0^{(0)} = \frac{1}{2n}\sum_{i=1}^{2}\sum_{j=1}^{n}(k_{ij1} - \overline{k}_1)(k_{ij2} - \overline{k}_2), \tag{4.47}$$

where

$$\overline{k}_1 = \frac{1}{2n}\sum_{i=1}^{2}\sum_{j=1}^{n} k_{ij1}$$
$$\overline{k}_2 = \frac{1}{2n}\sum_{i=1}^{2}\sum_{j=1}^{n} k_{ij2}.$$

If the above $\lambda_0^{(0)} < 0$ we assign 0 as an initial value for λ_0. For all the other parameters in the model, we assign the initial values according to the algorithm described in Section 3.3.3. Having all the initial estimates set, the optimization procedure proceeds by using the R function *nlm*.

4.3.4 Simulation Results

Appendix D contains the results of the simulations for the different scenarios and choices of the sample size n for the multiple population model of Chapter 3 (Model 1) and 4 (Model 2), respectively. Tables 24 to 35 show the results of the simulations, and Table 20 summarizes these results.

Table 20: Summary of the simulations' results. T=2

Scenario 1 **$n = 20$** • bias only in λ_0, δ and γ • asymptotic variance does not approximate the finite sample variance for λ_0, δ and γ • λ_0, δ and γ not normal	*Scenario 3* **$n = 20$** • no bias • asymptotic variance approximates the finite sample variance • all estimates look normally distributed
$n = 50$ • no bias • asymptotic variance approximates the finite sample variance • all estimates look normally distributed	**$n = 50$** • no bias • asymptotic variance approximates the finite sample variance • all estimates look normally distributed
$n = 100$ • no bias • asymptotic variance approximates the finite sample variance • all estimates look normally distributed	**$n = 100$** • no bias • asymptotic variance approximates the finite sample variance • all estimates look normally distributed
Scenario 2 **$n = 20$** • bias only in λ_0, δ and γ • asymptotic variance does not approximate the finite sample variance for λ_0, δ and γ • λ_0, δ and γ not normal	*Scenario 4* **$n = 20$** • bias only in λ_0 • asymptotic variance does not approximate the finite sample variance for λ_0 • λ_0 not normal
$n = 50$ • small bias for δ and γ • asymptotic variance approximates reasonably the finite sample variance for δ and γ • some evidence of non-normality	**$n = 50$** • no bias • asymptotic variance approximates the finite sample variance • all estimates look normally distributed
$n = 100$ • slight bias for δ and γ • asymptotic variance approximates the finite sample variance • all estimates look normally distributed	**$n = 100$** • no bias • asymptotic variance approximates the finite sample variance • all estimates look normally distributed

A quick glance at the tables tells us that the results from the two models are very similar.

Furthermore, the pattern of bias in the estimators is similar to the one in Chapter 3. The main parameters of interest are the μ's, representing the underlying disease status. For each of the four parameter combinations and for each of the sample sizes, biases in most of the estimates are very small. Specifically, as in the case of the simulations with $T = 4$, the results obtained from the simulation studies suggest that the estimates of the true μ's are unbiased for all choices of $n = 20, 50, 100$, under all four different scenarios and for both Model 1 and Model 2. The same is true for the parameters σ^2, ρ and ρ^*. This is not always the case for λ_0, δ and γ. For small samples ($n = 20$), the estimates for λ_0 are far from the true value in all scenarios except scenario 3 (see Tables 24, 27, and 33). For sample sizes larger than $n = 50$, the estimates for λ_0 appear unbiased, regardless of the parameter configuration. For small values of n, the averages of the estimates for δ and γ are strongly biased in both scenario 1 and scenario 2. Scenario 2 actually produces average estimates for δ and γ that have incorrect signs. (see Tables 27 and 28). The bias decreases with increasing the sample size, but even for $n = 100$, scenario 2 produces slightly biased estimates for δ and γ. In the remaining two scenarios (see Tables 30 - 35), the estimation works well, even for n small. This is due mainly to the fact that the severities are generated from distributions with well separated means with respect to σ^2.

We generated qq-plots (not presented), for each simulated scenario. By examining them, we found the empirical distribution of the MLEs to be symmetrical and approximately normal, as expected. The exceptions from this normal behavior parallel the findings from investigating the bias. Evidence of non-normal behavior was exhibited by the estimates of λ_0 when the sample size is small ($n = 20$) in scenarios 1, 2 and 4, and by estimates of δ and γ for $n = 20$ in scenario 1 and for $n = 20, 50$ in scenario 2.

Inspecting the behavior of the sample standard deviations produces the same type of conclusions as the ones from investigating the bias and normality of the estimators. The asymptotic variances approximate quite well the finite sample variances in most instances, even for sample sizes as small as $n = 20$. Different comportment is shown by the estimates of δ and γ in scenarios with small sample size and large variability in the events' severities (scenario 1, $n = 20$ and 2, $n = 20, 50$), and by estimates of λ_0 in scenarios with small sample size (scenario 1, 2 and 4 with $n = 20$). An interesting fact is that in these "problem" scenarios, the finite sample variances for δ and γ are smaller in Model 2 than in Model 1.

To summarize our findings, we conclude that the asymptotic results in section 3.2 and 4.2 for $T = 2$ are applicable for sample sizes which are greater than 50. Estimation results show that the algorithm gives acceptable results even for choices of n as small as 20, provided that the theoretical values of the mean severities are not extremely poorly separated relative to their variance. Furthermore, even for "problem" scenarios, in which δ and γ are not well estimated, we found the parameter of interest (μ's and λ's) to be well estimated.

4.4 APPLICATION TO LEDS DATA

In this section, we apply the method introduced in Section 4.2 to the LEDS data. Since the score equations and information matrix are available only for the particular case when $T = 2$, we analyze the data divided semi-annually. We also apply the multiple population model of 3.2 with $T = 2$ to this data and compare the results from the two models. Thus we obtain comparisons between dependence caused only by the event severities in different time periods and dependence built in both the number of events and severities across the two time periods.

As described in Section 3.4, LEDS data refers to stressful life events in 62 subjects, out of which 30 are normal controls (NC) and 32 had a major depressive disorder episode (MDD). The two outcomes recorded for each subject are the number of stessors and the severity of each of the stessors. These outcomes are recorded for each of the 2 halves of the year. Each half year, the data for an individual subject are random vectors of event severities with the random length given by the number of events the subject experienced that half year. Table 21 gives a description of the number of events experienced by the subjects in the LEDS data and Table 22 presents the severities of the events stratified by group. We denote the MDD group as population 1 with mean μ_{11} reflecting the underlying depression status at the first half year before the onset of depression and μ_{12} at the second half year before the onset of depression. Similarly, denote the subjects in the NC group as population 2 with means μ_{21} at measurement 1 (underlying disease severity at time 1) and μ_{22} at measurement 2 (underlying disease severity at time 2). The multiple population models described in Section 4.2 and 3.2 with $m = 2$ and $T = 2$ are applicable to this data. We have $n_1 = 32$ and $n_2 = 30$.

Table 21: LEDS Data. Frequency of acute stressors by group.

	Number of acute stressors														
	0	1	2	3	4	5	6	7	8	9	10	11	12	15	16
1-st half year before MDD onset															
MDD[1]	2	3	2	6	1	3	2	4	2	1	2	1	1	1	1
NC[2]	2	3	4	8	2	2	3	3	2	1	0	0	0	0	0
2-nd half year before MDD onset[3]															
MDD	2	2	6	5	5	2	2	2	2	1	1	2	0	0	0
NC	8	1	4	5	4	3	2	1	0	1	0	1	0	0	0

---

[1] MDD = major depressive disorder (sample size is 32)

[2] NC = normal control (sample size is 30)

[3] 13 subjects have the last month in the study replicated once or twice since they did not have the whole quarter available

As seen when examining the data quarterly, a quick look at the Tables 21 and 22 suggests that even from a semi-annual viewpoint, the MDDs have more events than the NCs and the severity of the events experienced by the MDDs is higher. Moreover, both the number of events and their severity seem to increase as the MDDs draw closer to their episode of depression. We observe that there are subjects in both groups and during both periods with no events, which, nonetheless our models handle. The data for these subjects at the corresponding measurement time are treated as zero-length random length vectors. We notice that overall the MDD group experienced more events than the NC group (183 during the first half year before the onset of depression and 142 during the second half year before the onset of depression as compared to 117 and 94, respectively). Furthermore, from Table 21, the largest number of stressors the NCs experienced is 11, while the MDD group contains subjects that undergo up to 16 events during a half of a year. During the half of the year

Table 22: LEDS Data. Severity of acute stressors by group. (Percentages represent stressors).

		Severity of Acute Stressors				
		1	2	3	4	Total (acute stressors)
1-st half year before MDD onset						
MDD	$n(\%)$	77(42)	71(39)	33(18)	2(1)	183
NC	$n(\%)$	65(56)	40(34)	9(8)	3(2)	117
2-nd half year before MDD onset						
MDD	$n(\%)$	59(42)	47(33)	33(14)	3(1)	142
NC	$n(\%)$	60(64)	22(23)	10(11)	2(2)	94

immediately preceding the onset of depression the number of events experienced by MDD group has more spread than during the prior half year.

To ensure that the constraints of the model are verified we use the R function *optim* to maximize the likelihood. This function is similar to *nlm*, but includes an option for box-constrained optimization. The initial values for the parameters are computed using the techniques described in Sections 3.3.3 and 4.3.3, respectively. We denote by $\hat{\boldsymbol{\theta}} = (\hat{\delta}, \hat{\gamma}, \hat{\mu}_{11}, \hat{\mu}_{12}, \hat{\mu}_{21}, \hat{\mu}_{22}, \hat{\sigma}^2, \hat{\rho}, \hat{\rho}^*)$ the MLE for the multiple population model with independent lengths (Model 1) and $\tilde{\boldsymbol{\theta}} = (\tilde{\delta}, \tilde{\gamma}, \tilde{\lambda}_0, \tilde{\mu}_{11}, \tilde{\mu}_{12}, \tilde{\mu}_{21}, \tilde{\mu}_{22}, \tilde{\sigma}^2, \tilde{\rho}, \tilde{\rho}^*)$ the MLE for for the multiple population model with dependent lengths (Model 2). Table 23 gives the solutions $\hat{\boldsymbol{\theta}}$ of the maximization procedure for the models in Chapter 3 (Model 1) and 4 (Model 2), respectively, and their corresponding estimated standard deviations based on $\boldsymbol{I}_n^{-1}(\hat{\boldsymbol{\theta}})$ and $\boldsymbol{I}_n^{-1}(\tilde{\boldsymbol{\theta}})$, respectively.

The first thing to notice is that the estimates from the two models are essentially identical. This is a result of the fact that Model 2 produces an estimate equal to zero for the parameter reflecting dependence over time, λ_0. Note that when the parameter λ_0 is equal to zero, the two models are identical. In addition, the model estimates $\rho^* = 0$.

The estimated parameter γ has a positive sign indicating a positive relationship between the average number of events and the average severity (larger severities and higher number of events). Its estimated standard deviation is small and the Wald tests for both models have p-values smaller than 0.001, indicating that γ is significantly different than zero. This result is different from the one in Section 3.4, where we analyzed the LEDS data quarterly, and it is probably due to the fact that there is a lot more variation in the quarterly data.

Table 23: Maximum likelihood solution for the two models

		Model 1		Model 2	
	Parameter	$\hat{\theta}$	estimated SD of $\hat{\theta}$	$\tilde{\theta}$	estimated SD of $\tilde{\theta}$
	δ	-1.0964	1.024	-1.0963	1.024
	γ	1.5195	0.604	1.5195	1.604
MDD					
	μ_{11}	1.8332	0.066	1.8332	0.066
	μ_{12}	1.7433	0.057	1.7433	0.058
NC					
	μ_{21}	1.5849	0.066	1.5849	0.065
	μ_{22}	1.5150	0.077	1.5150	0.077
	σ^2	0.5854	0.037	0.5854	0.037
	ρ	0.1241	0.041	0.1241	0.041
	ρ^*	0.0000	0.044	0.0000	0.044
	λ_0	-	-	0.0000	0.791

Further, we want to test if the profiles of the two groups are parallel. This is equivalent to testing that there is no interaction between time and group. This composite hypothesis can be written as

$$H_0: \quad (\mu_{12} - \mu_{11}) - (\mu_{22} - \mu_{21}) = 0$$

We set up the corresponding matrix

$$\boldsymbol{C} = \begin{pmatrix} -1 & 1 & 1 & -1 \end{pmatrix},$$

and write the hypothesis in matrix form

$$H_0: \quad \boldsymbol{C}\boldsymbol{\mu} = \boldsymbol{0},$$

where $\boldsymbol{\mu} = (\mu_{11}, \mu_{12}, \mu_{21}, \mu_{22})$. It follows that the value of the test statistic is

$$(\boldsymbol{C}\hat{\boldsymbol{\mu}})' \left[\boldsymbol{C}\boldsymbol{I}^{-1}(\hat{\boldsymbol{\mu}})\boldsymbol{C}'\right]^{-1} \boldsymbol{C}\hat{\boldsymbol{\mu}} = 0.0441$$

which is not significant with respect to a chi-square distribution with 1 degree of freedom. This means that the two profiles are not significantly different in shape and we conclude that the profiles are parallel. Given that the profiles are parallel, we are interested to see if they are also coincident. The corresponding null hypothesis of equal treatment effects is

$$H_0: \quad \mu_{11} + \mu_{12} = \mu_{21} + \mu_{22}.$$

We can state this hypothesis in matrix form as

$$H_0: \quad \boldsymbol{c}\boldsymbol{\mu} = \boldsymbol{0},$$

where $\boldsymbol{c} = (1 \;\; 1 \;\; -1 \;\; -1)$. The value of the test can be found as

$$(\boldsymbol{c}\hat{\boldsymbol{\mu}})' \left[\boldsymbol{c}\boldsymbol{I}^{-1}(\hat{\boldsymbol{\mu}})\boldsymbol{c}'\right]^{-1} \boldsymbol{c}\hat{\boldsymbol{\mu}} = 9.1031,$$

which is significant with respect to a chi-square distribution with 1 degree of freedom. We conclude that the two profiles are not coincident. -

Overall, the two models are providing us with the same insight into the LEDS data ; while there is some correlation between the severities within a time measurement, there is no dependence over time, neither in the severity measures, nor in the random lengths. Model 1 estimates $\rho^* = 0$ as an estimate for the correlation between two severities recorded at different halves of the year. In addition to estimating $\rho^* = 0$, Model 2 estimates $\lambda_0 = 0$ for the covariance between the random number of events recorded during the first half year before the onset of depression and the random number of events recorded during the second half year before the onset of depression. Thus, these semi-annual LEDS data could have been analyzed with a simple non-repeated four population model, in which we treat every half year of data as an independent population.

5.0 DISCUSSION AND DIRECTIONS FOR FUTURE RESEARCH

Our goal is to build models that allow us to analyze data gathered longitudinally on both the frequency of an event and its severity when both the frequency and the severity are important for the experimenter and the interest is in modeling the two outcomes together to draw inference about the treatment effect.

We refer to this type of data as *longitudinal random length data*. Building models for such data is a complex task. For instance, not only does the number of events that a subject experiences change over time; the number of time points with observed quantitative measures changes with every subject. Another complicating issue is that the mean and covariance structures for the distribution of the severities change with the change of the number of time points with observed quantitative measures. When the number of repeated time measurements increases, the number of parameters that need to be estimated increases. As the number of recorded events experienced by subjects and time measurements with quantitative measure increases, the difficulty of numerically estimating the parameters in the model increases, as well.

In this dissertation we propose two types of models to deal with *longitudinal random length data*, one with dependence over time built into the severity measures and a more complex second one with two layers of dependence over time. Although our motivation was drawn from a study examining stressful life events in adolescents, the methods appear to be more broadly useful. For many diseases or health conditions, an individual may have repeated episodes collected over assessment intervals, together with a measure of each episode's intensity or severity. Since the data that motivated our research refers to life events, the natural distribution to be considered in modeling the number of events is Poisson. However, other discrete distributions may be applicable and provide better fits, for example, the negative binomial distribution or, more generally, a family of discrete distribution with appropriate behavior.

In the proposed models, we treat the severity measures as continuous random variables. However, many of the severity measures encountered in practice are categorical (e.g. in LEDS data, 1=”little or none”, 2=”some”, 3=”moderate” and 4=”marked”). Thus, there is interest in developing models for longitudinal ordinal random length data.

We did make a first step into analyzing *longitudinal random length data*, but there are a number of interesting directions we see to further this research. The following sections describe some of the ideas that can be used to generalize our methods.

5.1 BUILDING DEPENDENCE INTO SEVERITIES

5.1.1 Introducing Covariates

The model introduced in Chapter 3 accommodates multiple populations but does not include covariates. Because both the number of events recorded at a measurement time and their severities both reflect the depression status, it is reasonable to assume that certain covariates that could impact the depression status may affect both the number of events and their severities in a similar way. We plan to develop models to accommodate covariates and account for their influence on both the lengths and the severities. For example, in the LEDS data, age, socio-economical status, and race could all be considered as covariates.

5.1.2 Using Different Covariance Structures to Model Severities’ Dependence over Time

In Chapter 3 we considered a simple covariance structure; any two severity measurements recorded at different time measurements have the same correlation coefficient, independent of the measurement times. Possible extensions of the multiple population model incorporate more general correlation structures for modeling the dependence between severity measurements at different time points within a subject. For example, one simple assumption is that correlation between severities observed at time t_1 and t_2 has the form $\rho^{|t_1-t_2|}$, i.e., decays geometrically with $|t_1 - t_2|$; this makes sense in that the degree of correlation may tend to be greater for observations that are closer in time than for severities that are far apart. Thus, one way to further our research is by exploring other covariance structures.

5.2 BUILDING DEPENDENCE INTO LENGTHS WITH MULTIVARIATE POISSON

5.2.1 Using More Complex Structures to Model the Dependence of the Lengths over Time

The multivariate Poisson model that we considered for modeling the vector of random lengths assumes one common positive covariance term for all pairs of random lengths. Using the models introduced by Karlis [20], we could relax the assumption of equal covariance among all pairs of random lengths and propose models with different covariances for pairs of random lengths observed at different measurement times.

Furthermore, in the multivariate Poisson model we use, the marginal mean and variance of each random length coincide, an assumption that is not appropriate for overdispersed number of events. As an alternative way of modeling the vector of random lengths, we could consider finite multivariate Poisson mixtures (see Karlis [22], [23]), which allow for both negative correlations and overdispersion.

5.2.2 Building Dependence into both Lengths and Severities. Extending the Supermodel

The main drawback of the model of Section 4.2 is the availability of the score equations and information matrix for only the case $T = 2$. There are possible approaches to more efficiently estimate the parameters of the multivariate Poisson distribution, like those proposed by Karlis [20], [21]. We plan on implementing them to simplify the most computational part of our estimation algorithm and extend the method to values of T larger than 2.

We first introduced models with only one layer of dependence over time. The multiple population model of Chapter 3 treats the random lengths as independent random variables and assumes that, given the total number of events, the vectors of severities are correlated over time. We then generalized these models in Chapter 4, by adding another layer of dependence over time. While still assuming the vectors of severities as dependent over time, we add dependence among the random lengths by using the multivariate Poisson distribution. The natural question that arises is which of the two models fits a certain data set better? Unfortunately, we can not give an answer to this question yet. To do so, we plan to extend the supermodel from Chapter 4, designed to build time dependence into both the severity measures and the lengths. This extension refers to solving the score equations and deriving the asymptotic distribution for the general case when $T > 2$. This supermodel will allow us to test the goodness-of-fit of the models introduced in Chapter 3 and Chapter 4.

5.3 GENERALIZING HOFFMAN'S APPROACH FOR LONGITUDINAL CLUSTERED DATA

Before building models for longitudinal random length data, we reviewed the existing approaches for dealing with longitudinal and clustered data. We decided to move forward by means of joint modeling of the severities and numbers of events, because this approach made efficient use of the information in the two outcomes. However, another provocative research idea is to develop a Within Cluster Resampling-like approach for longitudinal data and contrast it with the previous methods proposed in Chapter 4. In this setting, each subject has a cluster of severities at each of the T time points. The sizes of the clusters may be informative: in the context of the LEDS data, for example, as they get closer to the onset of their MDD episode, the subjects have more events and the severities increase. Thus, the sizes of the clusters are correlated to the outcome. Since standard methods ([36],[29]) for analyzing clustered data usually assume that the size is non-informative, thus producing biased estimates when the size is actually informative, it might be of interest to develop a method that debiases the GEE-type estimators.

In our attempt to model longitudinal random length data, we first compared the existing approaches for the non-repeated case. We simulated data under the Barnhart paradigm and then fitted both the model proposed by Barnhart and WCR. While the likelihood based method performed better, as expected, the results of the estimation for WCR were unbiased. Thus, we are confident that a method that debiases the GEE-type estimators when the cluster size is informative based on WCR technique will produce reasonable results. Roughly, the idea is to apply WCR to random sample a observation from each cluster, at every time. For this T-dimensional vector, we plan to apply a GEE-type analysis and repeat the procedure Q-times. Averaging these Q results will produce a WCR-type estimator. We plan to perform simulation studies to explore the finite-sample behavior of this estimator and prove that for a large number of clusters and a large number of resampling Q this estimator is asymptotically normal.

APPENDIX A

DERIVATION OF THE SCORE EQUATIONS AND INFORMATION MATRIX FOR THE MODEL IN CHAPTER 3

A.1 PRELIMINARY RESULT

Consider the following p-dimensional vector of lengths, $\boldsymbol{k} = (k_1, ..., k_p)$. To simplify things, we assume all components of $\boldsymbol{k}$ are non-zero. We first need to find $\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} = \boldsymbol{\Sigma}_{k_1,...,k_p}$, for $\boldsymbol{\Sigma}_{\boldsymbol{k}}$ as defined in 4.5. Recall that $\frac{\partial \tau_k}{\partial \rho} = -\frac{k-1}{k}\tau_k^2$ and $\frac{\partial \tau_0}{\partial \rho} = \tau_0^2$. Furthermore, $\boldsymbol{\Sigma}_{\boldsymbol{k}}$ can be written as

$$\boldsymbol{\Sigma}_{\boldsymbol{k}} = \text{Diag}\left(\frac{1}{\tau_{k_1}} - \rho^*, \frac{1}{\tau_{k_2}} - \rho^*, ..., \frac{1}{\tau_{k_p}} - \rho^*\right) + \rho^* \boldsymbol{e}_p \boldsymbol{e}_p^{'} = \boldsymbol{\Omega}_{\boldsymbol{k}} + \rho^* \boldsymbol{e}_p \boldsymbol{e}_p^{'},$$

where $\boldsymbol{e}_p$ is the p-dimensional vector with all entries 1 and $\boldsymbol{\Sigma}_{\boldsymbol{k}}$ is a matrix of dimensions $p \times p$. Using a result from Rao [32](pg 33, eg 2.8) we can write

$$\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} = \boldsymbol{\Omega}_{k}^{-1} - \frac{\rho^* \boldsymbol{\Omega}_{\boldsymbol{k}}^{-1} \boldsymbol{e}_p \boldsymbol{e}_p^{'} \boldsymbol{\Omega}_{\boldsymbol{k}}^{-1}}{1 + \rho^* \boldsymbol{e}_p^{'} \boldsymbol{\Omega}_{\boldsymbol{k}}^{-1} \boldsymbol{e}_p} \tag{A.1}$$

Since,

$$\boldsymbol{\Omega}_{\boldsymbol{k}}^{-1} = \text{Diag}\left(\frac{1}{1/\tau_{k_1} - \rho^*}, \frac{1}{1/\tau_{k_2} - \rho^*}, ..., \frac{1}{1/\tau_{k_p} - \rho^*}\right)$$

(A.1) can be written as

$$\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} = \boldsymbol{\Omega}_{\boldsymbol{k}}^{-1} - \frac{\rho^*}{1 + \rho^* \sum 1/\omega_i} \Delta_{\boldsymbol{k}},$$

where $\boldsymbol{\Delta}_{\boldsymbol{k}}$ has the form

$$\boldsymbol{\Delta}_{\boldsymbol{k}} = [\delta_{ij}]_{1 \le i,j \le p} = \left[\frac{1}{\omega_i}\frac{1}{\omega_j}\right]_{1 \le i,j \le p}.$$

A.2 DERIVATION OF THE SCORE EQUATIONS AND INFORMATION MATRIX FOR ONE OBSERVATION FROM THE ONE POPULATION MODEL

Let us consider one observation from the one population model introduced in Section 3.1. The data are condensed into a $\sum_{j=1}^{T} K_j$ - dimensional vector $\boldsymbol{X}$, $\boldsymbol{X}^{'} = \left(\boldsymbol{X}_1^{'}, ..., \boldsymbol{X}_T^{'}\right)$ and the corresponding T - dimensional vector of random lengths $\boldsymbol{K} = (K_1, ..., K_T)$. Let $\boldsymbol{k} = (k_1, ..., k_T)$ be a realization of the T-dimensional vector of lengths $\boldsymbol{K}$. Some of the the components of $\boldsymbol{k}$ might be zero. Let us denote by $l(\boldsymbol{k})$ the number of non-zero components of $\boldsymbol{k}$. $\widetilde{\boldsymbol{k}}$ the $l(\boldsymbol{k})$-dimensional vector composed of the non-zero elements of $\boldsymbol{k}$. Hence $\widetilde{\boldsymbol{k}} = (k_{r_1}, k_{r_2}, ..., k_{r_{l(\boldsymbol{k})}})$, where $1 \leq r_1 < r_2 < ... < r_{l(\boldsymbol{k})} \leq T$ are indices corresponding to the elements in the original vector $\boldsymbol{k}$. We denote this set of ordered indices by $\boldsymbol{r} = (r_1, r_2, ..., r_{l(\boldsymbol{k})})$.

The parameter vector is $\boldsymbol{\theta} = (\delta, \gamma, \mu_1, ..., \mu_T, \sigma^2, \rho, \rho^*)^{'}$. Let us denote by $\boldsymbol{I}(\boldsymbol{\theta})$ the information matrix for $\boldsymbol{\theta}$ contained in the one multivariate random length vector $\boldsymbol{X}$ with random lengths $\boldsymbol{K} = (K_1, ..., K_T)$. Conceptually, it makes sense to think of $\boldsymbol{I}(\boldsymbol{\theta})$ as a sum of the information about $\boldsymbol{\theta}$ contained in the lengths and the sum of information about $\boldsymbol{\theta}$ contributed by the vectors of severities, over all possible lengths. Using a general result from Barnhart [4](Theorem A.3.1.1) we can compute $\boldsymbol{I}(\boldsymbol{\theta})$ as:

$$\boldsymbol{I}(\boldsymbol{\theta}) = \boldsymbol{I}^*(\boldsymbol{\theta}) + \sum_{\boldsymbol{k} \in \Upsilon} \mathbf{P}_{\boldsymbol{\theta}}(\boldsymbol{K} = \boldsymbol{k}) \boldsymbol{I}(\boldsymbol{\theta}|\boldsymbol{k}),$$

where $\boldsymbol{I}^*(\boldsymbol{\theta})$ is the information matrix about $\boldsymbol{\theta}$ contained in the random lengths $\boldsymbol{K} = (K_1, ..., K_T)$ and $\boldsymbol{I}(\boldsymbol{\theta}|\boldsymbol{k})$ is the information matrix contained in $\boldsymbol{X}|\boldsymbol{K} = \boldsymbol{k}$, $\boldsymbol{k} \in \Upsilon$. We denote

$$\Upsilon = \left\{ \boldsymbol{k} = (k_1, ..., k_T) | k_i = 0, 1, ... \text{ for } \forall i = 1, ..., T \text{ and} \sum_{i=1}^{T} k_i \geq 1 \right\}.$$

Recall that after applying the appropriate transformation for the severities, we can write the log-likelihood for one observation from the one population model described in Section 3, $\log f(\boldsymbol{y}, \boldsymbol{k})$ as

$$\log f(\boldsymbol{y}, \boldsymbol{k}) = \log \mathbf{P}_{\boldsymbol{\theta}}(K_1 = k_1, ..., K_T = k_T) + \delta(\boldsymbol{k}) \log f(\boldsymbol{y}|\boldsymbol{k}).$$

Provided that $\boldsymbol{k}$ has at least one non-zero component (i.e. $\delta(\boldsymbol{k}) = 1$), the conditional distribution of $f(\boldsymbol{y})$ given $\boldsymbol{k}$, $f(\boldsymbol{y}|\boldsymbol{k})$ is a multivariate normal. Thus, we will use a result from McCulloch and Searle [30] who give the expressions of the score function and information matrix for the general model under the multivariate normality assumption, $\boldsymbol{Y} \sim \mathbf{MVN}(\boldsymbol{\mu}, \mathbf{V})$ with $E(\boldsymbol{Y}) = \boldsymbol{\mu}$ and $Var(\boldsymbol{Y}) = \mathbf{V}$.

Consider a general parametrization of $\boldsymbol{\mu}$ and $\mathbf{V}$ such that each element of $\boldsymbol{\mu}$ is a function of elements of a parameter vector $\boldsymbol{\beta}$ and each element of $\mathbf{V}$ is a function of the elements of a d - dimensional parameter vector $\boldsymbol{\varphi}$, unrelated to $\boldsymbol{\beta}$. Thus,

$$\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\beta}) \text{ and } \mathbf{V} = \mathbf{V}(\boldsymbol{\varphi}) \ .$$

It follows that the first order derivatives are

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \frac{\partial \boldsymbol{\mu}'}{\partial \boldsymbol{\beta}} \mathbf{V}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) \tag{A.2}$$

$$\frac{\partial l}{\partial \varphi_k} = -\frac{1}{2}\left[tr\left(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \varphi_k}\right) - (\boldsymbol{y} - \boldsymbol{\mu})'\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \varphi_k}\mathbf{V}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})\right], \tag{A.3}$$

for $k = 1, 2, ..., d$, where φ_k is the k-th element of the d - dimensional parameter vector $\boldsymbol{\varphi}$. Equating the expressions in A.2-A.3 to zero gives the score equations.
Furthermore, the information matrix is given by

$$-E\left[\begin{array}{cc} \frac{\partial^2 l}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} & \frac{\partial^2 l}{\partial \boldsymbol{\beta}\partial \boldsymbol{\varphi}'} \\ \left(\frac{\partial^2 l}{\partial \boldsymbol{\beta}\partial \boldsymbol{\varphi}'}\right)' & \frac{\partial^2 l}{\partial \boldsymbol{\varphi}\partial \boldsymbol{\varphi}'} \end{array}\right] = \left[\begin{array}{cc} \frac{\partial \boldsymbol{\mu}'}{\partial \boldsymbol{\beta}}\mathbf{V}^{-1}\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}} & \mathbf{O} \\ \mathbf{O}' & \frac{1}{2}\left\{tr\left(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \varphi_k}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \varphi_s}\right)\right\}_{1\le k,s\le d} \end{array}\right], \tag{A.4}$$

where d is, as before, the dimension of the parameter vector $\boldsymbol{\varphi}$.

We are going to use the above result to find the score function and information matrix generated by the conditional density part of the log-likelihood $\log f(\boldsymbol{y}|\boldsymbol{k})$, where

$$\log f(\boldsymbol{y}, \boldsymbol{k}) = \log \mathbf{P}_{\boldsymbol{\theta}}(K_1 = k_1, ..., K_T = k_T) + \log f(\boldsymbol{y}|\boldsymbol{k})$$

The conditional distribution $f(\boldsymbol{y}|\boldsymbol{k})$ given $\boldsymbol{k}$ is a multivariate normal of the typed described above, with $\boldsymbol{\mu} = (\mu_1, \mu_2, ..., \mu_T, 0, ..., 0)'$ and

$$\mathbf{V} = \sigma^2 \left[\begin{array}{cc} \boldsymbol{\Sigma}_{\boldsymbol{k}} & \mathbf{O} \\ \mathbf{O}' & \frac{1}{\tau_0}\mathbf{I}_{\boldsymbol{k}_+ - T} \end{array}\right],$$

where $\boldsymbol{k}_+ = \sum_{i=1}^{T} k_i$ and $\mathbf{O}$ is the $T \times (\boldsymbol{k}_+ - T)$ - dimensional matrix with all the entries equal to 0. To keep the notation simple, we drop the indices representing the dimensions of the matrix $\mathbf{O}$. For now, let us assume that all the components of $\boldsymbol{k}$ are nonzero.

Following the notation in McCulloch and Searle [30] we have $\boldsymbol{\beta} = (\mu_1, \mu_2, ..., \mu_T)$ and $\boldsymbol{\varphi} = (\sigma^2, \rho, \rho^*)$, where $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\beta})$ and $\mathbf{V} = \mathbf{V}(\boldsymbol{\varphi})$. It follows easily that

$$\mathbf{V}^{-1} = \frac{1}{\sigma^2} \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} & \mathbf{O} \\ \mathbf{O}^{'} & \tau_0 \mathbf{I}_{\boldsymbol{k}_+ - T} \end{bmatrix}.$$

As described in Section A.1, we have

$$\boldsymbol{\Sigma}_{\boldsymbol{k}} = \text{Diag}\left(\frac{1}{\tau_{k_1}} - \rho^*, \frac{1}{\tau_{k_2}} - \rho^*, ..., \frac{1}{\tau_{k_T}} - \rho^*\right) + \rho^* \boldsymbol{e}_T \boldsymbol{e}_T^{'} = \boldsymbol{\Omega}_{\boldsymbol{k}} + \rho^* \boldsymbol{e}_T \boldsymbol{e}_T^{'},$$

and

$$\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} = \boldsymbol{\Omega}_{\boldsymbol{k}}^{-1} - \frac{\rho^*}{1 + \rho^* \sum 1/\omega_i} \boldsymbol{\Delta}_{\boldsymbol{k}},$$

where

$$\boldsymbol{\Delta}_{\boldsymbol{k}} = \boldsymbol{\Omega}_{\boldsymbol{k}}^{-1} \boldsymbol{e}_T \boldsymbol{e}_T^{'} \boldsymbol{\Omega}_{\boldsymbol{k}}^{-1} = [\delta_{ij}]_{1 \le i,j \le T} = \left[\frac{1}{\omega_i} \frac{1}{\omega_j}\right]_{1 \le i,j \le T}$$

and ω_i is the i-th diagonal element of the matrix $\boldsymbol{\Omega}$.

Noting that

$$\frac{\partial}{\partial \rho}\left(\frac{1}{\tau_{k_i}}\right) = \frac{k_i - 1}{k_i} = 1 - \frac{1}{k_i},$$

it follows that

$$\frac{\partial \boldsymbol{\Sigma}_{\boldsymbol{k}}}{\partial \rho} = \text{Diag}\left(1 - \frac{1}{k_1}, ..., 1 - \frac{1}{k_T}\right) = \mathbf{I}_T - \text{Diag}\left(\frac{1}{k_1}, ..., \frac{1}{k_T}\right)$$

and

$$\frac{\partial \boldsymbol{\Sigma}_{\boldsymbol{k}}}{\partial \rho^*} = \text{Diag}\left(-1, ..., -1\right) + \boldsymbol{e}_T \boldsymbol{e}_T^{'} = \boldsymbol{e}_T \boldsymbol{e}_T^{'} - \mathbf{I}_T.$$

Thus, the components of $\frac{\partial \mathbf{V}}{\partial \boldsymbol{\varphi}}$ are

$$\begin{aligned}
\frac{\partial \mathbf{V}}{\partial \sigma^2} &= \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{k}} & \mathbf{O} \\ \mathbf{O}' & \frac{1}{\tau_0}\mathbf{I}_{\boldsymbol{k}_+ - T} \end{bmatrix} = \frac{1}{\sigma^2}\mathbf{V} \\
\frac{\partial \mathbf{V}}{\partial \rho} &= \sigma^2 \begin{bmatrix} \mathbf{I}_T - \mathrm{Diag}\left(\frac{1}{k_1}, ..., \frac{1}{k_T}\right) & \mathbf{O} \\ \mathbf{O}' & -\mathbf{I}_{\boldsymbol{k}_+ - T} \end{bmatrix} \\
\frac{\partial \mathbf{V}}{\partial \rho^*} &= \sigma^2 \begin{bmatrix} \boldsymbol{e}_T \boldsymbol{e}_T' - \mathbf{I}_T & \mathbf{O} \\ \mathbf{O}' & \mathbf{O} \end{bmatrix}.
\end{aligned}$$

As before, to keep the notation simple, we use $\mathbf{O}$ to denote a matrix with all the entries equal to 0 and we drop the indices that give the dimensions of the matrix. It follows that

$$\begin{aligned}
\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \sigma^2} &= \frac{1}{\sigma^2}\mathbf{V}^{-1}\mathbf{V} = \frac{1}{\sigma^2}\mathbf{I}_{\boldsymbol{k}_+} \\
\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \rho} &= \sigma^2 \mathbf{V}^{-1} \begin{bmatrix} \mathbf{I}_T - \mathrm{Diag}\left(\frac{1}{k_1}, ..., \frac{1}{k_T}\right) & \mathbf{O} \\ \mathbf{O}' & -\mathbf{I}_{\boldsymbol{k}_+ - T} \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\left(\mathbf{I}_T - \mathrm{Diag}\left(\frac{1}{k_1}, ..., \frac{1}{k_T}\right)\right) & \mathbf{O} \\ \mathbf{O}' & -\tau_0\mathbf{I}_{\boldsymbol{k}_+ - T} \end{bmatrix} \\
\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \rho^*} &= \sigma^2 \mathbf{V}^{-1} \begin{bmatrix} \boldsymbol{e}_T \boldsymbol{e}_T' - \mathbf{I}_T & \mathbf{O} \\ \mathbf{O}' & \mathbf{O} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}(\boldsymbol{e}_T \boldsymbol{e}_T' - \mathbf{I}_T) & \mathbf{O} \\ \mathbf{O}' & \mathbf{O} \end{bmatrix}
\end{aligned}$$

Applying (A.3) and taking into account the above expressions, we get

$$\begin{aligned}
\frac{\partial l}{\partial \sigma^2} &= -\frac{1}{2}\left[tr\left(\frac{1}{\sigma^2}\mathbf{I}_{\boldsymbol{k}_+}\right) - \frac{1}{\sigma^2}(\boldsymbol{y}-\boldsymbol{\mu})'\mathbf{V}^{-1}(\boldsymbol{y}-\boldsymbol{\mu})\right] \\
\frac{\partial l}{\partial \rho} &= -\frac{1}{2}\left[tr\left(\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\left(\mathbf{I}_T - \mathrm{Diag}\left(\frac{1}{k_1}, ..., \frac{1}{k_T}\right)\right)\right) - \tau_0 tr\left(\mathbf{I}_{\boldsymbol{k}_+ - T}\right)\right] \\
&\quad + \frac{1}{2\sigma^2}(\boldsymbol{y}-\boldsymbol{\mu})' \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}(\mathbf{I}_T - \mathrm{Diag}\left(\frac{1}{k_1}, ..., \frac{1}{k_T}\right))\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} & \mathbf{O} \\ \mathbf{O}' & -\tau_0^2\mathbf{I}_{\boldsymbol{k}_+ - T} \end{bmatrix} (\boldsymbol{y}-\boldsymbol{\mu}) \\
\frac{\partial l}{\partial \rho^*} &= -\frac{1}{2}\left[tr\left(\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}(\boldsymbol{e}_T \boldsymbol{e}_T' - \mathbf{I}_T)\right) - \frac{1}{\sigma^2}(\boldsymbol{y}-\boldsymbol{\mu})' \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}(\boldsymbol{e}_T \boldsymbol{e}_T' - \mathbf{I}_T)\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} & \mathbf{O} \\ \mathbf{O}' & \mathbf{O} \end{bmatrix} (\boldsymbol{y}-\boldsymbol{\mu})\right].
\end{aligned}$$

Recalling that we separate the elements of $\boldsymbol{y}$ into $\boldsymbol{z}$ and $\widehat{\boldsymbol{y}}$, we may write

$$
\begin{aligned}
\frac{\partial l}{\partial \sigma^2} &= -\frac{1}{2}\left[\frac{\boldsymbol{k}_+}{\sigma^2} - \frac{1}{\sigma^4}(\boldsymbol{z}-\widetilde{\boldsymbol{\mu}})'\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}(\boldsymbol{z}-\boldsymbol{\mu}^*) - \frac{\tau_0}{\sigma^4}\widehat{\boldsymbol{y}}'\widehat{\boldsymbol{y}}\right] &\text{(A.5)}\\
\frac{\partial l}{\partial \rho} &= -\frac{1}{2}\left[tr\left(\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\left[\mathbf{I}_T - \text{Diag}\left(\frac{1}{k_1},\ldots,\frac{1}{k_T}\right)\right]\right) - \tau_0(\boldsymbol{k}_+ - T)\right] &\\
&\quad + \frac{1}{2\sigma^2}(\boldsymbol{z}-\boldsymbol{\mu}^*)'\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\left[\mathbf{I}_T - \text{Diag}\left(\frac{1}{k_1},\ldots,\frac{1}{k_T}\right)\right]\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}(\boldsymbol{z}-\boldsymbol{\mu}^*) &\\
&\quad - \frac{\tau_0^2}{2\sigma^2}\widehat{\boldsymbol{y}}'\widehat{\boldsymbol{y}} &\text{(A.6)}\\
\frac{\partial l}{\partial \rho^*} &= -\frac{1}{2}\left[tr\left(\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}(\boldsymbol{e}_T\boldsymbol{e}_T' - \mathbf{I}_T)\right) - \frac{1}{\sigma^2}(\boldsymbol{z}-\boldsymbol{\mu}^*)'\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}(\boldsymbol{e}_T\boldsymbol{e}_T' - \mathbf{I}_T)\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}(\boldsymbol{z}-\boldsymbol{\mu}^*)\right] &\text{(A.7)}
\end{aligned}
$$

Straightforward computations lead to

$$
\begin{aligned}
\frac{\partial l}{\partial \boldsymbol{\beta}} &= \frac{\partial \boldsymbol{\mu}'}{\partial \boldsymbol{\beta}}\mathbf{V}^{-1}(\boldsymbol{y}-\boldsymbol{\mu}) = \begin{bmatrix} \mathbf{I}_T & \mathbf{O} \end{bmatrix}\frac{1}{\sigma^2}\begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} & \mathbf{O} \\ \mathbf{O}' & \tau_0\mathbf{I}_{\boldsymbol{k}_+ - T} \end{bmatrix}(\boldsymbol{y}-\boldsymbol{\mu})\\
&= \frac{1}{\sigma^2}\begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} & \mathbf{O} \end{bmatrix}(\boldsymbol{y}-\boldsymbol{\mu}) = \frac{1}{\sigma^2}\begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} & \mathbf{O} \end{bmatrix}\begin{pmatrix} \boldsymbol{z}-\boldsymbol{\mu}^* \\ \widehat{\boldsymbol{y}} \end{pmatrix}\\
&= \frac{1}{\sigma^2}\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}(\boldsymbol{z}-\boldsymbol{\mu}^*). \qquad \text{(A.8)}
\end{aligned}
$$

Equating to zero the expressions in (A.5)-(A.7) and (A.8) gives us the score equations. Having completed computing the score equations, we proceed to find the information matrix. To do that, we need the two matrices in the expression of the information matrix in (A.4). First, we compute the upper left corner matrix,

$$
\begin{aligned}
-E\left[\frac{\partial^2 l}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}\right] &= \frac{\partial \boldsymbol{\mu}'}{\partial \boldsymbol{\beta}}\mathbf{V}^{-1}\frac{\partial \boldsymbol{\mu}'}{\partial \boldsymbol{\beta}} = \begin{bmatrix} \mathbf{I}_T & \mathbf{O} \end{bmatrix}\frac{1}{\sigma^2}\begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} & \mathbf{O} \\ \mathbf{O}' & \tau_0\mathbf{I}_{\boldsymbol{k}_+ - T} \end{bmatrix}\begin{bmatrix} \mathbf{I}_T \\ \mathbf{O} \end{bmatrix}\\
&= \tfrac{1}{\sigma^2}\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}. \qquad \text{(A.9)}
\end{aligned}
$$

Second, let us denote the matrix in the right lower corner by

$$
I_{\boldsymbol{k}}(\boldsymbol{\varphi}) = \frac{1}{2}\left\{tr\left(\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \varphi_t}\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \varphi_s}\right)\right\}_{1\le t,s\le 3},
$$

where $\boldsymbol{\varphi} = (\sigma^2, \rho, \rho^*)$. We will compute each element of this matrix individually.

$$
\begin{aligned}
I_{\boldsymbol{k}}(\boldsymbol{\varphi})_{11} &= \frac{1}{2} tr(\frac{1}{\sigma^4}\mathbf{I}_{\boldsymbol{k}_+}) = \frac{\boldsymbol{k}_+}{2\sigma^4} && \text{(A.10)}\\
I_{\boldsymbol{k}}(\boldsymbol{\varphi})_{12} &= \frac{1}{2} tr\left(\mathbf{V}^{-1}\left[\begin{array}{cc} \mathbf{I}_T - \text{Diag}\left(\frac{1}{k_1}, ..., \frac{1}{k_T}\right) & \mathbf{O} \\ \mathbf{O}' & -\mathbf{I}_{\boldsymbol{k}_+ - T} \end{array}\right]\right) \\
&= \frac{1}{2\sigma^2}\left[tr\left(\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\left(\mathbf{I}_T - \text{Diag}\left(\frac{1}{k_1}, ..., \frac{1}{k_T}\right)\right)\right) - \tau_0 tr(\mathbf{I}_{\boldsymbol{k}_+ - T})\right] && \text{(A.11)}\\
I_{\boldsymbol{k}}(\boldsymbol{\varphi})_{13} &= \frac{1}{2} tr\left(\mathbf{V}^{-1}\left[\begin{array}{cc} \boldsymbol{e}_T\boldsymbol{e}_T' - \mathbf{I}_T & \mathbf{O} \\ \mathbf{O}' & \mathbf{O}_{\boldsymbol{k}_+ - T}) \end{array}\right]\right) = \frac{1}{2\sigma^2} tr\left(\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}(\boldsymbol{e}_T\boldsymbol{e}_T' - \mathbf{I}_T)\right) && \text{(A.12)}\\
I_{\boldsymbol{k}}(\boldsymbol{\varphi})_{22} &= \frac{1}{2} tr\left(\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\left(\mathbf{I}_T - \text{Diag}\left(\frac{1}{k_1}, ..., \frac{1}{k_T}\right)\right)\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\left(\mathbf{I}_T - \text{Diag}\left(\frac{1}{k_1}, ..., \frac{1}{k_T}\right)\right)\right) \\
&\quad + \frac{\tau_0^2}{2} tr(\mathbf{I}_{\boldsymbol{k}_+ - T})) && \text{(A.13)}\\
I_{\boldsymbol{k}}(\boldsymbol{\varphi})_{23} &= \frac{1}{2} tr\left(\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\left(\mathbf{I}_T - \text{Diag}\left(\frac{1}{k_1}, ..., \frac{1}{k_T}\right)\right)\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}(\boldsymbol{e}_T\boldsymbol{e}_T' - \mathbf{I}_T)\right) && \text{(A.14)}\\
I_{\boldsymbol{k}}(\boldsymbol{\varphi})_{33} &= \frac{1}{2} tr\left(\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}(\boldsymbol{e}_T\boldsymbol{e}_T' - \mathbf{I}_T)\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}(\boldsymbol{e}_T\boldsymbol{e}_T' - \mathbf{I}_T)\right). && \text{(A.15)}
\end{aligned}
$$

We assumed all the components of $\boldsymbol{k}$ to be nonzero. However, our model allows for zero-length vectors. In this instance, there is no contribution to the mean vector $\widetilde{\boldsymbol{\mu}}$ brought by the zero-length vectors. Thus, the corresponding entries in the matrix $\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}$ are zero and all the above computations have to be carried out replacing $\boldsymbol{k}$ with its subvector $\widetilde{\boldsymbol{k}}$ containing only nonzero components. The corresponding matrix $\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}$ is actually obtained by applying the equations above for the vector of nonzero lengths $\widetilde{\boldsymbol{k}}$ and filling in the corresponding spots with zero so that we obtain a $T \times T$ matrix and $I_{\boldsymbol{k}}(\varphi)$ is actually $I_{\widetilde{\boldsymbol{k}}}(\varphi)$. Thus, the information matrix about $(\boldsymbol{\beta}, \boldsymbol{\varphi}) \equiv (\mu_1, ..., \mu_T, \sigma^2, \rho, \rho^*)'$ contained in $\boldsymbol{X}|\boldsymbol{K} = \boldsymbol{k}$ is given by

$$
-E\left[\begin{array}{cc} \frac{\partial^2 l}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} & \frac{\partial^2 l}{\partial\boldsymbol{\beta}\partial\boldsymbol{\varphi}'} \\ \left(\frac{\partial^2 l}{\partial\boldsymbol{\beta}\partial\boldsymbol{\varphi}'}\right)' & \frac{\partial^2}{\partial\boldsymbol{\varphi}\partial\boldsymbol{\varphi}'} \end{array}\right] = \left[\begin{array}{cc} \frac{1}{\sigma^2}\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} & \mathbf{O} \\ \mathbf{O}' & I_{\boldsymbol{k}}(\sigma^2, \rho, \rho^*) \end{array}\right]. \qquad \text{(A.16)}
$$

Since the expression of $f(\boldsymbol{y}|\boldsymbol{k})$ involves neither δ nor γ, it follows that the information about $\boldsymbol{\theta} = (\delta, \gamma, \mu_1, ..., \mu_T, \sigma^2, \rho, \rho^*)'$ contributed by the vectors of severities for one subject is given by the expression

$$\boldsymbol{I}(\boldsymbol{\theta}|\boldsymbol{k}) = \begin{bmatrix} \mathbf{O} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \frac{1}{\sigma^2}\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{O}' & I_{\boldsymbol{k}}(\sigma^2, \rho, \rho^*) \end{bmatrix}. \tag{A.17}$$

Recall that

$$\boldsymbol{I}(\boldsymbol{\theta}) = \boldsymbol{I}^*(\boldsymbol{\theta}) + \sum_{\boldsymbol{k}\in\Upsilon} \mathrm{P}_{\boldsymbol{\theta}}(\boldsymbol{K} = \boldsymbol{k})\boldsymbol{I}(\boldsymbol{\theta}|\boldsymbol{k}),$$

where $\boldsymbol{I}^*(\boldsymbol{\theta})$ is the information matrix about $\boldsymbol{\theta}$ contained in the random lengths $\boldsymbol{K} = (K_1, ..., K_T)$. We need to compute it. We have

$$\log f(\boldsymbol{k}) \;=\; \sum_{j=1}^{T} \left[-e^{\delta+\gamma\mu_j} + k_j\left(\delta + \gamma\mu_j\right)\right].$$

It follows easily that

$$\begin{aligned}
\frac{\partial \log f(\boldsymbol{k})}{\partial \delta} &= \sum_{j=1}^{T} \left(-e^{\delta+\gamma\mu_j} + k_j\right) \\
\frac{\partial \log f(\boldsymbol{k})}{\partial \gamma} &= \sum_{j=1}^{T} \left(-\mu_j e^{\delta+\gamma\mu_j} + k_j\mu_j\right) \\
\frac{\partial \log f(\boldsymbol{k})}{\partial \mu_j} &= -\gamma e^{\delta+\gamma\mu_j} + \gamma k_j
\end{aligned}$$

and

$$\begin{aligned}
-\frac{\partial^2 \log f(\boldsymbol{k})}{\partial \delta^2} &= \sum_{j=1}^{T} e^{\delta+\gamma\mu_j} = \sum_{j=1}^{T} \lambda_j \\
-\frac{\partial^2 \log f(\boldsymbol{k})}{\partial \delta \partial \gamma} &= \sum_{j=1}^{T} \mu_j e^{\delta+\gamma\mu_j} = \sum_{j=1}^{T} \mu_j\lambda_j \\
-\frac{\partial^2 \log f(\boldsymbol{k})}{\partial \delta \partial \mu_j} &= \gamma e^{\delta+\gamma\mu_j} = \gamma\lambda_j \\
-\frac{\partial^2 \log f(\boldsymbol{k})}{\partial \gamma^2} &= \sum_{j=1}^{T} \mu_j^2 e^{\delta+\gamma\mu_j} = \sum_{j=1}^{T} \mu_j^2\lambda_j \\
-\frac{\partial^2 \log f(\boldsymbol{k})}{\partial \gamma \partial \mu_j} &= \sum_{j=1}^{T} \left((\gamma\mu_j + 1)e^{\delta+\gamma\mu_j} - k_j\right) = \sum_{j=1}^{T} \left(\lambda_j(\gamma\mu_j + 1) - k_j\right) \\
-\frac{\partial^2 \log f(\boldsymbol{k})}{\partial \mu_j^2} &= \gamma^2 e^{\delta+\gamma\mu_j} = \gamma^2\lambda_j.
\end{aligned}$$

Thus, the information matrix for $\boldsymbol{\theta}$ contained in the random lengths for one subject has the expression

$$\boldsymbol{I}^{*}(\boldsymbol{\theta}) = \begin{pmatrix} \sum_{j=1}^{T}\lambda_j & \sum_{j=1}^{T}\mu_j\lambda_j & \gamma\lambda_1 & ... & \gamma\lambda_T & 0 & 0 & 0 \\ \sum_{j=1}^{T}\mu_j\lambda_j & \sum_{j=1}^{T}\mu_j^2\lambda_j & \gamma\mu_1\lambda_1 & ... & \gamma\mu_T\lambda_T & 0 & 0 & 0 \\ \gamma\lambda_1 & \gamma\mu_1\lambda_1 & \gamma^2\lambda_1 & ... & 0 & 0 & 0 & 0 \\ ... & ... & ... & ... & ... & ... & ... & ... \\ \gamma\lambda_T & \gamma\mu_T\lambda_T & 0 & ... & \gamma^2\lambda_T & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Recall from (A.17) that information contributed by the vectors of severities given the lengths $\boldsymbol{I}(\boldsymbol{\theta}|\boldsymbol{k})$ can be computed as

$$\boldsymbol{I}(\boldsymbol{\theta}|\boldsymbol{k}) = \begin{pmatrix} \mathbf{O}_{2\times 2} & \mathbf{O}_{2\times T} & \mathbf{O}_{2\times 3} \\ \mathbf{O}_{T\times 2} & \frac{1}{\sigma^2}\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} & \mathbf{O}_{T\times 3} \\ \mathbf{O}_{3\times 2} & \mathbf{O}_{3\times T} & \boldsymbol{I}_{\boldsymbol{k}}(\sigma^2, \rho, \rho^*) \end{pmatrix}.$$

So, adding the corresponding pieces gives us

$$\boldsymbol{I}(\boldsymbol{\theta}) = \boldsymbol{I}^{*}(\boldsymbol{\theta}) + \sum_{\boldsymbol{k}\in\Upsilon}\left[\prod_{j=1}^{T} e^{-\lambda_j}\frac{{\lambda_j}^{k_j}}{k_j!}\right]\boldsymbol{I}(\boldsymbol{\theta}|\boldsymbol{k}). \tag{A.18}$$

Denoting

$$\mathbf{G} = \begin{pmatrix} \sum_{j=1}^{T}\lambda_j & \sum_{j=1}^{T}\mu_j\lambda_j & \gamma\lambda_1 & ... & \gamma\lambda_T \\ \sum_{j=1}^{T}\mu_j\lambda_j & \sum_{j=1}^{T}\mu_j^2\lambda_j & \gamma\mu_1\lambda_1 & ... & \gamma\mu_T\lambda_T \\ \gamma\lambda_1 & \gamma\mu_1\lambda_1 & \gamma^2\lambda_1 & ... & 0 \\ ... & ... & ... & ... & ... \\ \gamma\lambda_T & \gamma\mu_T\lambda_T & 0 & ... & \gamma^2\lambda_T \end{pmatrix},$$

we can further write this as the block diagonal matrix

$$\boldsymbol{I}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{I}(\delta, \gamma, \mu_1, ..., \mu_T) & \mathbf{O}_{(T+2)\times 3} \\ \mathbf{O}_{3\times(T+2)} & \sum\limits_{\boldsymbol{k}\in\Upsilon} \left[\prod\limits_{j=1}^{T} e^{-\lambda_j} \frac{{\lambda_j}^{k_j}}{k_j!}\right] \boldsymbol{I}_{\boldsymbol{k}}(\sigma^2, \rho, \rho^*) \end{pmatrix}, \tag{A.19}$$

where

$$\boldsymbol{I}(\delta, \gamma, \mu_1, ..., \mu_T) = \mathbf{G} + \begin{pmatrix} \mathbf{O}_{2\times 2} & \mathbf{O}_{2\times T} \\ \mathbf{O}_{T\times 2} & \frac{1}{\sigma^2} \sum\limits_{\boldsymbol{k}\in\Upsilon} \left[\prod\limits_{j=1}^{T} e^{-\lambda_j} \frac{{\lambda_j}^{k_j}}{k_j!}\right] \mathbf{\Sigma}_{\boldsymbol{k}}^{-1} \end{pmatrix},$$

and the elements of the matrix $\boldsymbol{I}_{\boldsymbol{k}}(\sigma^2, \rho, \rho^*)$ are defined (A.10)-(A.15).

A.3 DERIVATION OF THE INFORMATION MATRIX FOR THE ONE POPULATION MODEL

Suppose now that we have n independent observations from the one population model. The resulting information about the parameter $\boldsymbol{\theta}$ from all the data can be computed as the sum of the information contained in the n independent observations, found using (A.18). Hence, we get

$$\boldsymbol{I}_n(\boldsymbol{\theta}) = n\boldsymbol{I}^*(\boldsymbol{\theta}) + n \sum_{\boldsymbol{k}\in\Upsilon} \left[\prod_{j=1}^{T} e^{-\lambda_j} \frac{{\lambda_j}^{k_j}}{k_j!}\right] \boldsymbol{I}(\boldsymbol{\theta}|\boldsymbol{k}). \tag{A.20}$$

Using the expression in (A.19), it easy to show that $\boldsymbol{I}_n(\boldsymbol{\theta})$ also has a block diagonal form

$$\boldsymbol{I}_n(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{I}_n(\delta, \gamma, \mu_1, ..., \mu_T) & \mathbf{O}_{(T+2)\times 3} \\ \mathbf{O}_{3\times(T+2)} & n\sum\limits_{\boldsymbol{k}\in\Upsilon} \left[\prod\limits_{j=1}^{T} e^{-\lambda_j} \frac{{\lambda_j}^{k_j}}{k_j!}\right] \boldsymbol{I}_{\boldsymbol{k}}(\sigma^2, \rho, \rho^*) \end{pmatrix}, \tag{A.21}$$

where

$$\boldsymbol{I}_n(\delta, \gamma, \mu_1, ..., \mu_T) = n\boldsymbol{I}(\delta, \gamma, \mu_1, ..., \mu_T) = n\mathbf{G} + \begin{pmatrix} \mathbf{O}_{2\times 2} & \mathbf{O}_{2\times T} \\ \mathbf{O}_{T\times 2} & n\frac{1}{\sigma^2} \sum\limits_{\boldsymbol{k}\in\Upsilon} \left[\prod\limits_{j=1}^{T} e^{-\lambda_j} \frac{{\lambda_j}^{k_j}}{k_j!}\right] \mathbf{\Sigma}_{\boldsymbol{k}}^{-1} \end{pmatrix},$$

and the elements of the matrix $\boldsymbol{I}_{\boldsymbol{k}}(\sigma^2, \rho, \rho^*)$ are defined by (A.10) - (A.15). As before, we denote

$$\Upsilon = \left\{ \boldsymbol{k} = (k_1, ...k_T) |\ k_i = 0, 1, ... \text{ for } \forall i = 1, ..., T \text{ and } \sum_{i=1}^{T} k_i \geq 1 \right\}.$$

A.4 DERIVATION OF THE INFORMATION MATRIX FOR THE MULTIPLE POPULATION MODEL

Let us now consider the multiple population model introduced in Chapter 3, Section 3.2. The parameter vector is $\boldsymbol{\theta} = (\delta, \gamma, \mu_{11}, ..., \mu_{1T}, ..., \mu_{m1}, ..., \mu_{mT}, \sigma^2, \rho, \rho^*)'$. Let us denote by $\boldsymbol{I_i}(\boldsymbol{\theta})$ the information matrix for $\boldsymbol{\theta}$ contained in the one multivariate random length vector $\boldsymbol{X}_i$ from population i with random lengths $\boldsymbol{K} = (K_1, ..., K_T)$. This vector belongs to a one population model with parameter vector $\boldsymbol{\theta_i} = (\delta, \gamma, \mu_{i1}, ..., \mu_{iT}, \sigma^2, \rho, \rho^*)'$. We can apply the results from Section A.2 to find the information matrix about $\boldsymbol{\theta_i}$.

It follows that the information about $\boldsymbol{\theta_i} = (\delta, \gamma, \mu_{i1}, ..., \mu_{iT}, \sigma^2, \rho, \rho^*)'$ contained in the random lengths for one subject has the expression

$$\boldsymbol{I_i^*}(\boldsymbol{\theta_i}) = \begin{pmatrix} \sum_{j=1}^{T} \lambda_{ij} & \sum_{j=1}^{T} \mu_{ij}\lambda_{ij} & \gamma\lambda_{i1} & ... & \gamma\lambda_{iT} & 0 & 0 & 0 \\ \sum_{j=1}^{T} \mu_{ij}\lambda_{ij} & \sum_{j=1}^{T} \mu_{ij}^2\lambda_{ij} & \gamma\mu_{i1}\lambda_{i1} & ... & \gamma\mu_{iT}\lambda_{iT} & 0 & 0 & 0 \\ \gamma\lambda_{i1} & \gamma\mu_{i1}\lambda_{i1} & \gamma^2\lambda_{i1} & ... & 0 & 0 & 0 & 0 \\ ... & ... & ... & ... & ... & ... & ... & ... \\ \gamma\lambda_{iT} & \gamma\mu_{iT}\lambda_{iT} & 0 & ... & \gamma^2\lambda_{iT} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

To simplify the notation, let

$$\mathbf{G}_i = \begin{pmatrix} \sum_{t=1}^{T} \lambda_{it} & \sum_{t=1}^{T} \mu_{it}\lambda_{it} & \gamma\lambda_{i1} & ... & \gamma\lambda_{iT} \\ \sum_{t=1}^{T} \mu_{it}\lambda_{it} & \sum_{t=1}^{T} \mu_{it}^2\lambda_{it} & \gamma\mu_{i1}\lambda_{i1} & ... & \gamma\mu_{iT}\lambda_{iT} \\ \gamma\lambda_{i1} & \gamma\mu_{i1}\lambda_{i1} & \gamma^2\lambda_{i1} & ... & 0 \\ ... & ... & ... & ... & ... \\ \gamma\lambda_{iT} & \gamma\mu_{iT}\lambda_{iT} & 0 & ... & \gamma^2\lambda_{iT} \end{pmatrix},$$

and partition it into

$$\mathbf{G}_i = \begin{pmatrix} \mathbf{G}_{11}^{(i)} & \mathbf{G}_{12}^{(i)} \\ \mathbf{G}_{12}^{(i)\prime} & \mathbf{G}_{22}^{(i)} \end{pmatrix}, \tag{A.22}$$

where $\mathbf{G}_{11}^{(i)}$ is the upper left-corner 2×2 submatrix of $\mathbf{G}_i$. Since there is no information in these lengths about the other μ's, the information about $\boldsymbol{\theta}$ contained in the random lengths for one subject in population i has the expression

$$\boldsymbol{I}_{\boldsymbol{i}}^{*}(\boldsymbol{\theta}) = \begin{pmatrix} \sum_{j=1}^{T} \lambda_{ij} & \sum_{j=1}^{T} \mu_{ij}\lambda_{ij} & \mathbf{0}'_{(i-1)T} & \gamma\lambda_{i1} & \cdots & \gamma\lambda_{iT} & \mathbf{0}'_{(m-i)T} & 0 & 0 & 0 \\ \sum_{j=1}^{T} \mu_{ij}\lambda_{ij} & \sum_{j=1}^{T} \mu_{ij}^{2}\lambda_{ij} & \mathbf{0}'_{(i-1)T} & \gamma\mu_{i1}\lambda_{i1} & \cdots & \gamma\mu_{iT}\lambda_{iT} & \mathbf{0}'_{(m-i)T} & 0 & 0 & 0 \\ \mathbf{0}_{(i-1)T} & \mathbf{0}_{(i-1)T} & \mathbf{O} & 0 & \cdots & 0 & \mathbf{O} & 0 & 0 & 0 \\ \gamma\lambda_{i1} & \gamma\mu_{i1}\lambda_{i1} & \mathbf{0}'_{(i-1)T} & \gamma^{2}\lambda_{i1} & \cdots & 0 & \mathbf{0}'_{(m-i)T} & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \gamma\lambda_{iT} & \gamma\mu_{iT}\lambda_{iT} & \mathbf{0}'_{(i-1)T} & 0 & \cdots & \gamma^{2}\lambda_{iT} & \mathbf{0}'_{(m-i)T} & 0 & 0 & 0 \\ \mathbf{0}_{(m-i)T} & \mathbf{0}_{(m-i)T} & \mathbf{O} & 0 & \cdots & 0 & \mathbf{O} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{0}'_{(i-1)T} & 0 & \cdots & 0 & \mathbf{0}'_{(m-i)T} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{0}'_{(i-1)T} & 0 & \cdots & 0 & \mathbf{0}'_{(m-i)T} & 0 & 0 & 0 \\ 0 & 0 & \mathbf{0}'_{(i-1)T} & 0 & \cdots & 0 & \mathbf{0}'_{(m-i)T} & 0 & 0 & 0 \end{pmatrix}.$$

A simpler way of writing the above matrix is

$$\boldsymbol{I}_{\boldsymbol{i}}^{*}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{G}_{11}^{(i)} & \mathbf{O}_{2\times T(i-1)} & \mathbf{G}_{12}^{(i)} & \mathbf{O}_{2\times T(m-i)} & \mathbf{O}_{2\times 3} \\ \mathbf{O}_{T(i-1)\times 2} & \mathbf{O}_{T(i-1)\times T(i-1)} & \mathbf{O}_{T(i-1)\times T} & \mathbf{O}_{T(i-1)\times T(m-i)} & \mathbf{O}_{T(i-1)\times 3} \\ \mathbf{G}_{12}^{(i)\prime} & \mathbf{O}_{T\times T(i-1)} & \mathbf{G}_{22}^{(i)} & \mathbf{O}_{T\times T(m-i)} & \mathbf{O}_{T\times 3} \\ \mathbf{O}_{T(m-i)\times 2} & \mathbf{O}_{T(m-i)\times T(i-1)} & \mathbf{O}_{T(m-i)\times T} & \mathbf{O}_{T(m-i)\times T(m-i)} & \mathbf{O}_{T(m-i)\times 3} \\ \mathbf{O}_{2\times 2} & \mathbf{O}_{2\times T(i-1)} & \mathbf{O}_{2\times T} & \mathbf{O}_{2\times T(m-i)} & \mathbf{O}_{2\times 3} \end{pmatrix}.$$

Using (A.17), the information contributed by the vectors of severities given the lengths $\boldsymbol{I}_{\boldsymbol{i}}(\boldsymbol{\theta}_{\boldsymbol{i}}|\boldsymbol{k})$ can be computed as

$$\boldsymbol{I}_{\boldsymbol{i}}(\boldsymbol{\theta}_{\boldsymbol{i}}|\boldsymbol{k}) = \begin{pmatrix} \mathbf{O}_{2\times 2} & \mathbf{O}_{2\times T} & \mathbf{O}_{2\times 3} \\ \mathbf{O}_{T\times 2} & \frac{1}{\sigma^{2}}\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} & \mathbf{O}_{T\times 3} \\ \mathbf{O}_{3\times 2} & \mathbf{O}_{3\times T} & \boldsymbol{I}_{\boldsymbol{k}}(\sigma^{2}, \rho, \rho^{*}) \end{pmatrix}.$$

Similarly to the above, since $\boldsymbol{\theta_i}$ involves only μ_{ij}'s, $j = 1, ..., T$ we may write the information about $\boldsymbol{\theta}$ contributed by the vectors of severities given the lengths $\boldsymbol{I_i}(\boldsymbol{\theta}|\boldsymbol{k})$ as

$$\boldsymbol{I_i}(\boldsymbol{\theta}|\boldsymbol{k}) = \begin{pmatrix} \mathbf{O}_{2\times 2} & \mathbf{O}_{2\times(i-1)T} & \mathbf{O}_{2\times T} & \mathbf{O}_{2\times(m-i)T} & \mathbf{O}_{2\times 3} \\ \mathbf{O}_{(i-1)T\times 2} & \mathbf{O}_{(i-1)T\times(i-1)T} & \mathbf{O}_{(i-1)T\times T} & \mathbf{O}_{(i-1)T\times(m-i)T} & \mathbf{O}_{(i-1)T\times 3} \\ \mathbf{O}_{T\times 2} & \mathbf{O}_{T\times(i-1)T} & \frac{1}{\sigma^2}\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} & \mathbf{O}_{T\times(m-i)T} & \mathbf{O}_{T\times 3} \\ \mathbf{O}_{(m-i)T\times 2} & \mathbf{O}_{(m-i)T\times(i-1)T} & \mathbf{O}_{(m-i)T\times T} & \mathbf{O}_{(m-i)T\times(m-i)T} & \mathbf{O}_{(m-i)T\times 3} \\ \mathbf{O}_{3\times 2} & \mathbf{O}_{3\times(i-1)T} & \mathbf{O}_{3\times T} & \mathbf{O}_{3\times(m-i)T} & \boldsymbol{I_k}(\sigma^2, \rho, \rho^*) \end{pmatrix}.$$

Finally, adding all the corresponding pieces gives us the information matrix for the multiple population model from Section 3.2,

$$\boldsymbol{I}_n(\boldsymbol{\theta}) = \sum_{i=1}^{m} n_i \boldsymbol{I_i^*}(\boldsymbol{\theta}) + \sum_{i=1}^{m} n_i \sum_{\boldsymbol{k}\in\Upsilon} \left[\prod_{j=1}^{T} e^{-\lambda_{ij}} \frac{\lambda_{ij}{}^{k_j}}{k_j!}\right] \boldsymbol{I_i}(\boldsymbol{\theta}|\boldsymbol{k_i}). \tag{A.23}$$

As before, we denote

$$\Upsilon = \left\{ \boldsymbol{k} = (k_1, ..., k_T)|\ k_j = 0, 1, ... \text{ for } \forall j = 1, ..., T \text{ and } \sum_{j=1}^{T} k \geq 1 \right\}.$$

It can easily be shown that the information matrix $\boldsymbol{I}_n(\boldsymbol{\theta})$ about the parameter vector $\boldsymbol{\theta}$, $\boldsymbol{\theta} = (\delta, \gamma, \mu_{11}, ..., \mu_{1T}, ..., \mu_{m1}, ..., \mu_{mT}, \sigma^2, \rho, \rho^*)'$, contained in the $n = \sum_{i=1}^{m} n_i$ independent observations from the multiple population model has a block diagonal form

$$\boldsymbol{I}_n(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{I}_n(\delta, \gamma, \mu_{11}, ..., \mu_{mT}) & \mathbf{O}_{(mT+2)\times 3} \\ \mathbf{O}_{3\times(mT+2)} & \sum_{i=1}^{m} n_i \sum_{\boldsymbol{k}\in\Upsilon} \left[\prod_{j=1}^{T} e^{-\lambda_{ij}} \frac{\lambda_{ij}{}^{k_j}}{k_j!}\right] \boldsymbol{I_k}(\sigma^2, \rho, \rho^*) \end{pmatrix}, \tag{A.24}$$

where

$$\begin{aligned} \boldsymbol{I}_n(\delta, \gamma, \mu_{11}, ..., \mu_{mT}) \ &= \ \sum_{i=1}^{m} n_i \begin{pmatrix} \mathbf{G}_{11}^{(i)} & \mathbf{O}_{2\times T(i-1)} & \mathbf{G}_{12}^{(i)} & \mathbf{O}_{2\times T(m-i)} \\ \mathbf{O}_{T(i-1)\times 2} & \mathbf{O}_{T(i-1)\times T(i-1)} & \mathbf{O}_{T(i-1)\times T} & \mathbf{O}_{T(i-1)\times T(m-i)} \\ \mathbf{G}_{12}^{(i)\prime} & \mathbf{O}_{T\times T(i-1)} & \mathbf{G}_{22}^{(i)} & \mathbf{O}_{T\times T(m-i)} \\ \mathbf{O}_{T(m-i)\times 2} & \mathbf{O}_{T(m-i)\times T(i-1)} & \mathbf{O}_{T(m-i)\times T} & \mathbf{O}_{T(m-i)\times T(m-i)} \end{pmatrix} \\ &+ \begin{pmatrix} \mathbf{O}_{2\times 2} & \mathbf{O}_{2\times mT} \\ \mathbf{O}_{mT\times 2} & \sum_{i=1}^{m} n_i \sum_{\boldsymbol{k}\in\Upsilon} \prod_{t=1}^{T} e^{-\lambda_{it}} \frac{\lambda_{it}{}^{k_t}}{k_t!} \boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} \end{pmatrix}, \end{aligned}$$

the matrices $\mathbf{G}_{kl}^{(i)}$'s $(k, l = 1, 2)$ are given in (A.22) and the elements of the matrix $\boldsymbol{I_k}(\sigma^2, \rho, \rho^*)$ are defined by (A.10) - (A.15).

APPENDIX B

MULTIVARIATE POISSON DISTRIBUTION

We present here a multivariate Poisson distribution that arises naturally as a multivariate extension of the univariate Poisson distribution. The marginals of this multivariate Poisson distribution are univariate Poisson. Such multivariate generalizations are not unique in the sense that different multivariate distributions may have marginal distributions of the same family. The generalization we employ is the one used by Karlis [20] and introduced by Holgate [17].

Suppose that Y_i are independent Poisson random variables with mean θ_i for , $i = 0, ..., m$. Define the new random variables

$$\begin{aligned} X_1 &= Y_1 + Y_0 \\ X_2 &= Y_2 + Y_0 \\ \cdots & \\ X_m &= Y_m + Y_0. \end{aligned} \tag{B.1}$$

Then the random variables $(X_1, ..., X_m)$ are said to follow jointly an m-variate Poisson distribution, where m denotes the dimension of the distribution. The joint probability function is given by

$$\begin{aligned} \mathbf{P}(\boldsymbol{X}) &= \mathbf{P}(X_1 = x_1, ..., X_m = x_m) \\ &= \exp\left(-\sum_{i=0}^{m} \theta_i\right) \prod_{i=1}^{m} \frac{\theta_i^{x_i}}{x_i!} \sum_{i=0}^{s} \prod_{j=1}^{m} \binom{x_j}{i} i! \left(\frac{\theta_0}{\prod_{k=1}^{m} \theta_k}\right)^i, \end{aligned} \tag{B.2}$$

where $s = \min(x_1, x_2, ..., x_m)$.

We will denote this distribution MVPoisson$(\theta_0, \theta_1, \theta_2, ..., \theta_m)$. Marginally, each of the X_i's follows a Poisson distribution with parameter $\theta_0 + \theta_i$. The parameter θ_0 is the covariance between all the pairs of random variables. If $\theta_0 = 0$, then the variables are independent and the multivariate Poisson distribution reduces to the product of independent Poisson distributions.

APPENDIX C

DERIVATION OF THE SCORE EQUATIONS AND INFORMATION MATRIX FOR THE MODEL IN CHAPTER 4 AND TWO TIME MEASUREMENTS

C.1 DERIVATION OF THE INFORMATION MATRIX FOR ONE POPULATION MODEL

Let us consider one observation from the one population model introduced in Section 4.1 for the particular case $T = 2$. Hence the data is condensed into a $\sum_{j=1}^{2} K_j$ - dimensional vector $\boldsymbol{X}$, $\boldsymbol{X}' = \left(\boldsymbol{X}_1', \boldsymbol{X}_2'\right)$ and the corresponding bi-dimensional vector of random lengths $\boldsymbol{K} = (K_1, K_2)$. Let $\boldsymbol{k} = (k_1, k_2)$ be a realization of the bivariate vector of lengths $\boldsymbol{K}$. Some of the the components of $\boldsymbol{k}$ might be zero.

The parameter vector is $\boldsymbol{\theta} = (\delta, \gamma, \lambda_0, \mu_1, \mu_2, \sigma^2, \rho^{'}, \rho^*)$. If we denote by $\boldsymbol{I}(\boldsymbol{\theta})$ the information matrix for $\boldsymbol{\theta}$ contained in the one multivariate random length vector $\boldsymbol{X}$ with random lengths $\boldsymbol{K} = (K_1, K_2)$, it can be computed as a sum of the information about $\boldsymbol{\theta}$ contained in the lengths and the sum of information about $\boldsymbol{\theta}$ contributed by the vectors of severities, over all possible lengths. Using a result similar to the one in Barnhart [4](Theorem 3.3.1) we obtain $\boldsymbol{I}(\boldsymbol{\theta})$ as

$$\boldsymbol{I}(\boldsymbol{\theta}) = \boldsymbol{I}^*(\boldsymbol{\theta}) + \sum_{\boldsymbol{k} \in \Upsilon} \mathrm{P}_{\boldsymbol{\theta}}(\boldsymbol{K} = \boldsymbol{k}) \boldsymbol{I}(\boldsymbol{\theta} | \boldsymbol{k}),$$

where $\boldsymbol{I}^*(\boldsymbol{\theta})$ is the information matrix about $\boldsymbol{\theta}$ contained in the random lengths $\boldsymbol{K} = (K_1, K_2)$ and $\boldsymbol{I}(\boldsymbol{\theta}|\boldsymbol{k})$ is the information matrix contained in $\boldsymbol{X}|\boldsymbol{K} = \boldsymbol{k}$, $\boldsymbol{k} \in \Upsilon$. We denote

$$\Upsilon = \left\{ \boldsymbol{k} = (k_1, k_2) | k_i = 0, 1, ... \text{ for } \forall i = 1, 2 \text{ and } \sum_{t=1}^{2} k_t \geq 1 \right\}.$$

Recall that after applying the appropriate transformation for the severities, we can write the log-likelihood for one observation from the one population model described in Chapter 4, $\log f(\boldsymbol{y}, \boldsymbol{k})$ as

$$\log f(\boldsymbol{y}, \boldsymbol{k}) = \log \mathbf{P}_{\boldsymbol{\theta}}(K_1 = k_1, K_2 = k_2) + \log f(\boldsymbol{y}|\boldsymbol{k}).$$

Provided that $\boldsymbol{k}$ has at least one non-zero component, the conditional distribution of $f(\boldsymbol{y})$ given $\boldsymbol{k}$, $f(\boldsymbol{y}|\boldsymbol{k})$ is a multivariate normal. We will use the results from Appendix A to give the expressions of the score function and information matrix for this conditional distribution.

C.1.1 Derivation of the Score Equations and Information Matrix for a Vector of Bivariate Random Lengths from the One Population Model

The vector of random lengths $\boldsymbol{K} = (K_1, K_2)$ has a bivariate Poisson distribution that arises naturally as a multivariate extension of the univariate Poisson distribution. The the marginals are univariate Poisson random variables. The generalization we employ is the one used by Holgate [17] and Kocherlakota and Kocherlakota [25]. We obtain the bivariate vector of lengths $\boldsymbol{K} = (K_1, K_2)$ from the independent Poisson random variables G_i, with mean λ_i for $i = 0, 1, 2$, by defining

$$\begin{aligned} K_1 &= G_1 + G_0 \\ K_2 &= G_2 + G_0. \end{aligned}$$

The resulting random variables (K_1, K_2) follow jointly a bivariate Poisson distribution. The joint probability function is given by

$$\begin{aligned} f(r, s) &= \mathbf{P}_{\boldsymbol{\theta}}(K_1 = r, K_2 = s) \\ &= e^{-\sum_{i=0}^{2} \lambda_i} \frac{\lambda_1^r}{r!} \frac{\lambda_2^s}{s!} \sum_{i=0}^{\min(r,s)} \binom{r}{i} \binom{s}{i} i! \left(\frac{\lambda_0}{\lambda_1 \lambda_2} \right)^i. \end{aligned}$$

Using the reparametrization

$$\begin{aligned}\lambda_1^* &= \lambda_1 + \lambda_0 \\ \lambda_2^* &= \lambda_2 + \lambda_0\end{aligned}$$

and the recurrence relations

$$\begin{aligned}rf(r,s) &= (\lambda_1^* - \lambda_0)f(r-1,s) + \lambda_0 f(r-1,s-1) \\ sf(r,s) &= (\lambda_2^* - \lambda_0)f(r,s-1) + \lambda_0 f(r-1,s-1),\end{aligned}$$

Kocherlakota and Kocherlakota [25] find the expression of the derivatives of the probability function and information matrix as

$$\begin{aligned}\frac{\partial f(r,s)}{\partial \lambda_1^*} &= f(r-1,s) - f(r,s) \\ \frac{\partial f(r,s)}{\partial \lambda_2^*} &= f(r,s-1) - f(r,s) \\ \frac{\partial f(r,s)}{\partial \lambda_0} &= f(r,s) - f(r-1,s) - f(r,s-1) + f(r-1,s-1),\end{aligned}$$

and

$$\boldsymbol{I} = \begin{pmatrix} \frac{\lambda_1-\delta_1}{\lambda_1^2} & -\frac{\delta_1}{\lambda_1\lambda_2} & -\frac{\delta_2\lambda_0}{\lambda_1^2\lambda_2} \\ -\frac{\delta_1}{\lambda_1\lambda_2} & \frac{\lambda_2-\delta_1}{\lambda_2^2} & -\frac{\delta_2\lambda_0}{\lambda_1\lambda_2^2} \\ -\frac{\delta_2\lambda_0}{\lambda_1^2\lambda_2} & -\frac{\delta_2\lambda_0}{\lambda_1\lambda_2^2} & \frac{\delta_3}{\lambda_1^2\lambda_2^2} \end{pmatrix}.$$

We denote

$$\delta_1 = \lambda_0\left[1 - \lambda_0(\tau - 1)\right] \tag{C.1}$$

$$\delta_2 = -(\lambda_1 + \lambda_2) + \left[\lambda_1^*\lambda_2^* - \lambda_0^2\right](\tau - 1) \tag{C.2}$$

$$\delta_3 = \left[\lambda_1^*\lambda_2^* - \lambda_0^2\right]\left[\tau - 1 - (\lambda_1 + \lambda_2)\right] \tag{C.3}$$

$$\tau = \sum_{r,s=1}^{\infty} \frac{f^2(r-1,s-1)}{f(r,s)}. \tag{C.4}$$

Now, taking into consideration that

$$\begin{aligned}\lambda_1^* &= \lambda_1 + \lambda_0 = \exp(\delta + \gamma\mu_1) + \lambda_0 \\ \lambda_2^* &= \lambda_2 + \lambda_0 = \exp(\delta + \gamma\mu_2) + \lambda_0\end{aligned}$$

one can easily show that

$$\begin{aligned}\frac{\partial \lambda_1^*}{\partial \mu_1} &= \frac{\partial \lambda_1}{\partial \mu_1} = \gamma\lambda_1 \\ \frac{\partial \lambda_2^*}{\partial \mu_2} &= \frac{\partial \lambda_2}{\partial \mu_2} = \gamma\lambda_2 \\ \frac{\partial \lambda_i^*}{\partial \delta} &= \frac{\partial \lambda_i}{\partial \delta} = \lambda_i \\ \frac{\partial \lambda_i^*}{\partial \gamma} &= \frac{\partial \lambda_i}{\partial \gamma} = \mu_i\lambda_i.\end{aligned}$$

It follows that

$$\begin{aligned}\frac{\partial f(r,s)}{\partial \mu_1} &= \gamma\lambda_1 \left[f(r-1,s) - f(r,s)\right] \\ \frac{\partial f(r,s)}{\partial \mu_2} &= \gamma\lambda_2 \left[f(r,s-1) - f(r,s)\right] \\ \frac{\partial f(r,s)}{\partial \delta} &= \lambda_1 f(r-1,s) + \lambda_2 f(r,s-1) - (\lambda_1 + \lambda_2) f(r,s) \\ \frac{\partial f(r,s)}{\partial \gamma} &= \mu_1\lambda_1 f(r-1,s) + \mu_2\lambda_2 f(r,s-1) - (\mu_1\lambda_1 + \mu_2\lambda_2) f(r,s) \\ \frac{\partial f(r,s)}{\partial \lambda_0} &= f(r,s) - f(r-1,s) - f(r,s-1) + f(r-1,s-1).\end{aligned}$$

Furthermore, one can show that

$$\begin{aligned}\frac{\partial \log f(r,s)}{\partial \lambda_1^*} &= \frac{f(r-1,s)}{f(r,s)} - 1 \\ \frac{\partial \log f(r,s)}{\partial \lambda_2^*} &= \frac{f(r,s-1)}{f(r,s)} - 1 \\ \frac{\partial \log f(r,s)}{\partial \lambda_0} &= \frac{f(r,s) - f(r-1,s) - f(r,s-1) + f(r-1,s-1)}{f(r,s)}\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial \log f(r,s)}{\partial \mu_1} &= \gamma\lambda_1\left[\frac{f(r-1,s)}{f(r,s)}-1\right] = \gamma\lambda_1\frac{\partial \log f(r,s)}{\partial \lambda_1^*} \\
\frac{\partial \log f(r,s)}{\partial \mu_2} &= \gamma\lambda_2\left[\frac{f(r,s-1)}{f(r,s)}-1\right] = \gamma\lambda_2\frac{\partial \log f(r,s)}{\partial \lambda_2^*} \\
\frac{\partial \log f(r,s)}{\partial \delta} &= \lambda_1\frac{f(r-1,s)}{f(r,s)} + \lambda_2\frac{f(r,s-1)}{f(r,s)} - (\lambda_1+\lambda_2) \\
&= \lambda_1\frac{\partial \log f(r,s)}{\partial \lambda_1^*} + \lambda_2\frac{\partial \log f(r,s)}{\partial \lambda_2^*} \\
\frac{\partial \log f(r,s)}{\partial \gamma} &= \mu_1\lambda_1\frac{f(r-1,s)}{f(r,s)} + \mu_2\lambda_2\frac{f(r,s-1)}{f(r,s)} - (\mu_1\lambda_1+\mu_2\lambda_2) \\
&= \mu_1\lambda_1\frac{\partial \log f(r,s)}{\partial \lambda_1^*} + \mu_2\lambda_2\frac{\partial \log f(r,s)}{\partial \lambda_2^*} \\
\frac{\partial \log f(r,s)}{\partial \lambda_0} &= \frac{f(r,s)-f(r-1,s)-f(r,s-1)+f(r-1,s-1)}{f(r,s)}.
\end{aligned}$$

We can write the score equations for the vector of lengths belonging to one observation from the one population model with two time points as

$$\begin{aligned}
\frac{\partial \log f(r,s)}{\partial \mu_1} &= \gamma\lambda_1\left[\frac{f(r-1,s)}{f(r,s)}-1\right] = 0 && \text{(C.5)} \\
\frac{\partial \log f(r,s)}{\partial \mu_2} &= \gamma\lambda_2\left[\frac{f(r,s-1)}{f(r,s)}-1\right] = 0 && \text{(C.6)} \\
\frac{\partial \log f(r,s)}{\partial \delta} &= \lambda_1\frac{f(r-1,s)}{f(r,s)} + \lambda_2\frac{f(r,s-1)}{f(r,s)} - (\lambda_1+\lambda_2) = 0 && \text{(C.7)} \\
\frac{\partial \log f(r,s)}{\partial \gamma} &= \mu_1\lambda_1\frac{f(r-1,s)}{f(r,s)} + \mu_2\lambda_2\frac{f(r,s-1)}{f(r,s)} - (\mu_1\lambda_1+\mu_2\lambda_2) = 0 && \text{(C.8)} \\
\frac{\partial \log f(r,s)}{\partial \lambda_0} &= \frac{f(r,s)-f(r-1,s)-f(r,s-1)+f(r-1,s-1)}{f(r,s)} = 0. && \text{(C.9)}
\end{aligned}$$

We can now proceed to find the expression of the information matrix about the parameter vector $\boldsymbol{\theta}_1 = (\delta, \gamma, \lambda_0, \mu_1, \mu_2)$. Simple use of the formulas above gives

$$\begin{aligned}
E\left[\frac{\partial \log f(r,s)}{\partial \mu_1}\right]^2 &= (\gamma\lambda_1)^2 E\left[\frac{\partial \log f(r,s)}{\partial \lambda_1^*}\right]^2 = \gamma^2(\lambda_1 - \delta_1) \\
E\left[\frac{\partial \log f(r,s)}{\partial \mu_2}\right]^2 &= (\gamma\lambda_2)^2 E\left[\frac{\partial \log f(r,s)}{\partial \lambda_2^*}\right]^2 = \gamma^2(\lambda_2 - \delta_1) \\
E\left[\frac{\partial \log f(r,s)}{\partial \mu_1}\frac{\partial \log f(r,s)}{\partial \mu_2}\right] &= \gamma^2\lambda_1\lambda_2 E\left[\frac{\partial \log f(r,s)}{\partial \lambda_1^*}\frac{\partial \log f(r,s)}{\partial \lambda_2^*}\right] = -\gamma^2\delta_1 \\
E\left[\frac{\partial \log f(r,s)}{\partial \mu_1}\frac{\partial \log f(r,s)}{\partial \lambda_0}\right] &= \gamma\lambda_1 E\left[\frac{\partial \log f(r,s)}{\partial \lambda_1^*}\frac{\partial \log f(r,s)}{\partial \lambda_0}\right] = -\frac{\gamma\delta_2\lambda_0}{\lambda_1\lambda_2} \\
E\left[\frac{\partial \log f(r,s)}{\partial \mu_2}\frac{\partial \log f(r,s)}{\partial \lambda_0}\right] &= \gamma\lambda_2 E\left[\frac{\partial \log f(r,s)}{\partial \lambda_2^*}\frac{\partial \log f(r,s)}{\partial \lambda_0}\right] = -\frac{\gamma\delta_2\lambda_0}{\lambda_1\lambda_2}.
\end{aligned}$$

Observing that

$$\frac{\partial \log f(r,s)}{\partial \mu_1}\frac{\partial \log f(r,s)}{\partial \delta} = \gamma\lambda_1\left[\frac{f(r-1,s)}{f(r,s)} - 1\right]\left[\lambda_1\left(\frac{f(r-1,s)}{f(r,s)} - 1\right) + \lambda_2\left(\frac{f(r,s-1)}{f(r,s)} - 1\right)\right],$$

we get

$$\begin{aligned}
E\left[\frac{\partial \log f(r,s)}{\partial \mu_1}\frac{\partial \log f(r,s)}{\partial \delta}\right] &= \gamma\lambda_1^2 E\left[\frac{\partial \log f(r,s)}{\partial \lambda_1^*}\right]^2 + \gamma\lambda_1\lambda_2 E\left[\frac{\partial \log f(r,s)}{\partial \lambda_1^*}\frac{\partial \log f(r,s)}{\partial \lambda_2^*}\right] \\
&= \gamma\lambda_1^2\frac{\lambda_1 - \delta_1}{\lambda_1^2} + \gamma\lambda_1\lambda_2\frac{-\delta_1}{\lambda_1\lambda_2} = \gamma(\lambda_1 - 2\delta_1).
\end{aligned}$$

Similarly, we obtain

$$E\left[\frac{\partial \log f(r,s)}{\partial \mu_2}\frac{\partial \log f(r,s)}{\partial \delta}\right] = \gamma(\lambda_2 - 2\delta_1).$$

Applying similar computation we have that

$$\begin{aligned}
E\left[\frac{\partial \log f(r,s)}{\partial \delta}\frac{\partial \log f(r,s)}{\partial \lambda_0}\right] &= \lambda_1 E\left[\frac{\partial \log f(r,s)}{\partial \lambda_1^*}\frac{\partial \log f(r,s)}{\partial \lambda_0}\right] + \lambda_2 E\left[\frac{\partial \log f(r,s)}{\partial \lambda_2^*}\frac{\partial \log f(r,s)}{\partial \lambda_0}\right] \\
&= \lambda_1\frac{-\delta_2\lambda_0}{\lambda_1^2\lambda_2} + \lambda_2\frac{-\delta_2\lambda_0}{\lambda_1\lambda_2^2} = -\frac{2\delta_2\lambda_0}{\lambda_1\lambda_2} \\
E\left[\frac{\partial \log f(r,s)}{\partial \gamma}\frac{\partial \log f(r,s)}{\partial \lambda_0}\right] &= \mu_1\lambda_1 E\left[\frac{\partial \log f(r,s)}{\partial \lambda_1^*}\frac{\partial \log f(r,s)}{\partial \lambda_0}\right] + \mu_2\lambda_2 E\left[\frac{\partial \log f(r,s)}{\partial \lambda_2^*}\frac{\partial \log f(r,s)}{\partial \lambda_0}\right] \\
&= \mu_1\lambda_1\frac{-\delta_2\lambda_0}{\lambda_1^2\lambda_2} + \mu_2\lambda_2\frac{-\delta_2\lambda_0}{\lambda_1\lambda_2^2} = -\frac{\delta_2\lambda_0(\mu_1 + \mu_2)}{\lambda_1\lambda_2}
\end{aligned}$$

and

$$\begin{aligned}
E\left[\frac{\partial \log f(r,s)}{\partial \delta}\right]^2 &= \lambda_1^2 E\left[\frac{\partial \log f(r,s)}{\partial \lambda_1^*}\right]^2 + \lambda_2^2 E\left[\frac{\partial \log f(r,s)}{\partial \lambda_2^*}\right]^2 + 2\lambda_1\lambda_2 E\left[\frac{\partial \log f(r,s)}{\partial \lambda_1^*}\frac{\partial \log f(r,s)}{\partial \lambda_2^*}\right] \\
&= \lambda_1^2 \frac{\lambda_1-\delta_1}{\lambda_1^2} + \lambda_2^2 \frac{\lambda_2-\delta_1}{\lambda_2^2} + 2\lambda_1\lambda_2 \frac{-\delta_1}{\lambda_1\lambda_2} = \lambda_1+\lambda_2-4\delta_1 \\
E\left[\frac{\partial \log f(r,s)}{\partial \gamma}\right]^2 &= (\mu_1\lambda_1)^2 E\left[\frac{\partial \log f(r,s)}{\partial \lambda_1^*}\right]^2 + (\mu_2\lambda_2)^2 E\left[\frac{\partial \log f(r,s)}{\partial \lambda_2^*}\right]^2 \\
&\quad +2\mu_1\lambda_1\mu_2\lambda_2 E\left[\frac{\partial \log f(r,s)}{\partial \lambda_1^*}\frac{\partial \log f(r,s)}{\partial \lambda_2^*}\right] \\
&= \mu_1^2\lambda_1^2 \frac{\lambda_1-\delta_1}{\lambda_1^2} + \mu_2^2\lambda_2^2 \frac{\lambda_2-\delta_1}{\lambda_2^2} + 2\mu_1\lambda_1\mu_2\lambda_2 \frac{-\delta_1}{\lambda_1\lambda_2} = \mu_1^2\lambda_1 + \mu_2^2\lambda_2 - \delta_1(\mu_1+\mu_2)^2.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
E\left[\frac{\partial \log f(r,s)}{\partial \mu_1}\frac{\partial \log f(r,s)}{\partial \gamma}\right] &= \gamma\mu_1\lambda_1^2 E\left[\frac{\partial \log f(r,s)}{\partial \lambda_1^*}\right]^2 + \gamma\lambda_1\mu_2\lambda_2 E\left[\frac{\partial \log f(r,s)}{\partial \lambda_1^*}\frac{\partial \log f(r,s)}{\partial \lambda_2^*}\right] \\
&= \gamma\mu_1\lambda_1^2 \frac{\lambda_1-\delta_1}{\lambda_1^2} + \gamma\lambda_1\mu_2\lambda_2 \frac{-\delta_1}{\lambda_1\lambda_2} = \gamma\mu_1\lambda_1 - \gamma\delta_1(\mu_1+\mu_2)
\end{aligned}$$

and similarly,

$$E\left[\frac{\partial \log f(r,s)}{\partial \mu_2}\frac{\partial \log f(r,s)}{\partial \gamma}\right] = \gamma\mu_2\lambda_2 - \gamma\delta_1(\mu_1+\mu_2).$$

Finally,

$$\begin{aligned}
E\left[\frac{\partial \log f(r,s)}{\partial \delta}\frac{\partial \log f(r,s)}{\partial \gamma}\right] &= \mu_1\lambda_1^2 E\left[\frac{\partial \log f(r,s)}{\partial \lambda_1^*}\right]^2 + \mu_2\lambda_2^2 E\left[\frac{\partial \log f(r,s)}{\partial \lambda_2^*}\right]^2 \\
&\quad +\lambda_1\lambda_2(\mu_1+\mu_2) E\left[\frac{\partial \log f(r,s)}{\partial \lambda_1^*}\frac{\partial \log f(r,s)}{\partial \lambda_2^*}\right] \\
&= \mu_1\lambda_1^2 \frac{\lambda_1-\delta_1}{\lambda_1^2} + \mu_2\lambda_2^2 \frac{\lambda_2-\delta_1}{\lambda_2^2} + 2\lambda_1\lambda_2(\mu_1+\mu_2) \frac{-\delta_1}{\lambda_1\lambda_2} \\
&= \mu_1\lambda_1 + \mu_2\lambda_2 - 2\delta_1(\mu_1+\mu_2).
\end{aligned}$$

We put together the pieces above and find that the expression of the information matrix about $\boldsymbol{\theta}_1 = (\delta, \gamma, \lambda_0, \mu_1, \mu_2)$ is given by

$$\boldsymbol{I^*(\theta_1)} = \begin{pmatrix} \lambda_1 + \lambda_2 - 4\delta_1 & \sum_{i=1}^{2} \mu_i(\lambda_i - 2\delta_1) & -\frac{2\delta_2\lambda_0}{\lambda_1\lambda_2} & \gamma(\lambda_1 - 2\delta_1) & \gamma(\lambda_2 - 2\delta_1) \\ \sum_{i=1}^{2} \mu_i(\lambda_i - 2\delta_1) & \sum_{i=1}^{2} \mu_i^2\lambda_i - \delta_1(\sum_{i=1}^{2} \mu_i)^2 & -\frac{\delta_2\lambda_0 \sum_{i=1}^{2} \mu_i}{\lambda_1\lambda_2} & \gamma(\mu_1\lambda_1 - \delta_1 \sum_{i=1}^{2} \mu_i) & \gamma(\mu_2\lambda_2 - \delta_1 \sum_{i=1}^{2} \mu_i) \\ -\frac{2\delta_2\lambda_0}{\lambda_1\lambda_2} & -\frac{\delta_2\lambda_0(\mu_1+\mu_2)}{\lambda_1\lambda_2} & \frac{\delta_3}{\lambda_1^2\lambda_2^2} & -\frac{\gamma\delta_2\lambda_0}{\lambda_1\lambda_2} & -\frac{\gamma\delta_2\lambda_0}{\lambda_1\lambda_2} \\ \gamma(\lambda_1 - 2\delta_1) & \gamma(\mu_1\lambda_1 - \delta_1 \sum_{i=1}^{2} \mu_i) & -\frac{\gamma\delta_2\lambda_0}{\lambda_1\lambda_2} & \gamma^2(\lambda_1 - \delta_1) & -\gamma^2\delta_1 \\ \gamma(\lambda_2 - 2\delta_1) & \gamma(\mu_2\lambda_2 - \delta_1 \sum_{i=1}^{2} \mu_i) & -\frac{\gamma\delta_2\lambda_0}{\lambda_1\lambda_2} & -\gamma^2\delta_1 & \gamma^2(\lambda_2 - \delta_1) \end{pmatrix}.$$

It follows easily that the information about $\boldsymbol{\theta} = (\delta, \gamma, \lambda_0, \mu_1, \mu_2, \sigma^2, \rho, \rho^*)'$ contributed by the bivariate vector of random lengths is given by

$$\boldsymbol{I}^*(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{I}^*(\boldsymbol{\theta_1}) & \mathbf{0}_5 & \mathbf{0}_5 & \mathbf{0}_5 \\ \mathbf{0}_5^{'} & 0 & 0 & \\ \mathbf{0}_5^{'} & 0 & 0 & \\ \mathbf{0}_5^{'} & 0 & 0 & \end{pmatrix}. \tag{C.10}$$

C.1.2 Derivation of the Score Equations and Information Matrix for the Vector of Severities Given the Bivariate Vector of Random Lengths from the One Population Model

Provided that $\boldsymbol{k}$ has at least one non-zero component, the conditional distribution of $f(\boldsymbol{y})$ given $\boldsymbol{k}$, $f(\boldsymbol{y}|\boldsymbol{k})$ is a multivariate normal. We will use the results from Appendix A to give the expressions of the score function and information matrix for this conditional distribution. Following the notations in Appendix A, we have $\boldsymbol{\beta} = (\mu_1, \mu_2)$, $\boldsymbol{\varphi} = (\sigma^2, \rho, \rho^*)$ and

$$\frac{\partial l}{\partial \sigma^2} = -\frac{1}{2}\left[\frac{\boldsymbol{k}_+}{\sigma^2} - \frac{1}{\sigma^4}(\boldsymbol{z} - \widetilde{\boldsymbol{\mu}})'\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}(\boldsymbol{z} - \boldsymbol{\mu}^*) - \frac{\tau_0}{\sigma^4}\widehat{\boldsymbol{y}}'\widehat{\boldsymbol{y}}\right] \tag{C.11}$$

$$\begin{aligned} \frac{\partial l}{\partial \rho} = & -\frac{1}{2}\left[tr\left(\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\left[\mathbf{I}_T - \text{Diag}\left(\frac{1}{k_1}, ..., \frac{1}{k_T}\right)\right]\right) - \tau_0(\boldsymbol{k}_+ - T)\right] \\ & + \frac{1}{2\sigma^2}(\boldsymbol{z} - \boldsymbol{\mu}^*)'\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\left[\mathbf{I}_T - \text{Diag}\left(\frac{1}{k_1}, ..., \frac{1}{k_T}\right)\right]\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}(\boldsymbol{z} - \boldsymbol{\mu}^*) \\ & - \frac{\tau_0^2}{2\sigma^2}\widehat{\boldsymbol{y}}'\widehat{\boldsymbol{y}} \end{aligned} \tag{C.12}$$

$$\frac{\partial l}{\partial \rho^*} = -\frac{1}{2}\left[tr\left(\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}(\boldsymbol{e}_T\boldsymbol{e}_T^{'} - \mathbf{I}_T)\right) - \frac{1}{\sigma^2}(\boldsymbol{z} - \boldsymbol{\mu}^*)'\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}(\boldsymbol{e}_T\boldsymbol{e}_T^{'} - \mathbf{I}_T)\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}(\boldsymbol{z} - \boldsymbol{\mu}^*)\right] \tag{C.13}$$

where $\boldsymbol{z}$ is the vector obtained with the first entries in the transformed data corresponding to first and second time measurements and $\widehat{\boldsymbol{y}}$ contains the rest of the transformed data. As is Appendix A, we denote

$$\boldsymbol{k}_{+} = \sum_{t=1}^{2} k_t. \tag{C.14}$$

Furthermore,

$$\frac{\partial l}{\partial \boldsymbol{\beta}} = \frac{1}{\sigma^2} \boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} (\boldsymbol{z} - \boldsymbol{\mu}^*). \tag{C.15}$$

Equating to zero the expressions in (C.11) - (C.13) and (C.15) gives us the score equations. The information matrix about $(\boldsymbol{\beta}, \boldsymbol{\varphi}) = (\mu_1, \mu_2, \sigma^2, \rho, \rho^*)'$ contained in $\boldsymbol{X}|\boldsymbol{K} = \boldsymbol{k}$ is given by

$$-E\left[\begin{array}{cc} \frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} & \frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\varphi}'} \\ \left(\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\varphi}'}\right)' & \frac{\partial^2}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}'} \end{array}\right] = \left[\begin{array}{cc} \frac{1}{\sigma^2} \boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} & \mathbf{O} \\ \mathbf{O}' & I_{\boldsymbol{k}}(\sigma^2, \rho, \rho^*) \end{array}\right], \tag{C.16}$$

where the elements of $I_{\boldsymbol{k}}(\sigma^2, \rho, \rho^*)$ are given by (A.10) - (A.15) and one needs to account for the possibility of observing zero-length vectors. Since the expression of $f(\boldsymbol{y}|\boldsymbol{k})$ does not involve any of the parameters δ , γ and λ_0, it follows that the information about $\boldsymbol{\theta} = (\delta, \gamma, \lambda_0, \mu_1, \mu_2, \sigma^2, \rho', \rho^*$ contributed by the the vectors of severities for one subject is given by the expression

$$\boldsymbol{I}(\boldsymbol{\theta}|\boldsymbol{k}) = \begin{pmatrix} \mathbf{O}_{3\times 3} & \mathbf{O}_{3\times 2} & \mathbf{O}_{3\times 3} \\ \mathbf{O}_{2\times 3} & \frac{1}{\sigma^2} \boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} & \mathbf{O}_{2\times 3} \\ \mathbf{O}_{3\times 3} & \mathbf{O}_{3\times 2} & \boldsymbol{I}_{\boldsymbol{k}}(\sigma^2, \rho, \rho^*) \end{pmatrix}. \tag{C.17}$$

Recall that

$$\boldsymbol{I}(\boldsymbol{\theta}) = \boldsymbol{I}^*(\boldsymbol{\theta}) + \sum_{\boldsymbol{k} \in \Upsilon} \mathbf{P}_{\boldsymbol{\theta}}(\boldsymbol{K} = \boldsymbol{k}) \mathbf{I}(\boldsymbol{\theta}|\boldsymbol{k}),$$

where $\boldsymbol{I}^*(\boldsymbol{\theta})$ is the information matrix about $\boldsymbol{\theta}$ contained in the bivariate vector of random lengths $\boldsymbol{K} = (K_1, K_2)$ and $\Upsilon = \{(k_1, k_2)|\, k_1, k_2 = 0, 1, ..., k_1 + k_2 > 0\}$. Adding the corresponding pieces gives us

$$\boldsymbol{I}(\boldsymbol{\theta}) = \boldsymbol{I}^*(\boldsymbol{\theta}) + \sum_{\boldsymbol{k} \in \Upsilon} \left[e^{-\sum_{i=0}^{2} \lambda_i} \frac{\lambda_1^{k_1}}{k_1!} \frac{\lambda_2^{k_2}}{k_2!} \sum_{i=0}^{\min(k_1, k_2)} \binom{k_1}{i} \binom{k_2}{i} i! \left(\frac{\lambda_0}{\lambda_1 \lambda_2} \right)^i \right] \boldsymbol{I}(\boldsymbol{\theta}|\boldsymbol{k}). \tag{C.18}$$

Denoting

$$\mathbf{H} = \begin{pmatrix} \lambda_1+\lambda_2-4\delta_1 & \sum\limits_{i=1}^{2}\mu_i(\lambda_i-2\delta_1) & -\frac{2\delta_2\lambda_0}{\lambda_1\lambda_2} & \gamma(\lambda_1-2\delta_1) & \gamma(\lambda_2-2\delta_1) \\ \sum\limits_{i=1}^{2}\mu_i(\lambda_i-2\delta_1) & \sum\limits_{i=1}^{2}\mu_i^2\lambda_i-\delta_1(\sum\limits_{i=1}^{2}\mu_i)^2 & -\frac{\delta_2\lambda_0(\mu_1+\mu_2)}{\lambda_1\lambda_2} & \gamma(\mu_1\lambda_1-\delta_1\sum\limits_{i=1}^{2}\mu_i) & \gamma(\mu_2\lambda_2-\delta_1\sum\limits_{i=1}^{2}\mu_i) \\ -\frac{2\delta_2\lambda_0}{\lambda_1\lambda_2} & -\frac{\delta_2\lambda_0(\mu_1+\mu_2)}{\lambda_1\lambda_2} & \frac{\delta_3}{\lambda_1^2\lambda_2^2} & -\frac{\gamma\delta_2\lambda_0}{\lambda_1\lambda_2} & -\frac{\gamma\delta_2\lambda_0}{\lambda_1\lambda_2} \\ \gamma(\lambda_1-2\delta_1) & \gamma(\mu_1\lambda_1-\delta_1\sum\limits_{i=1}^{2}\mu_i) & -\frac{\gamma\delta_2\lambda_0}{\lambda_1\lambda_2} & \gamma^2(\lambda_1-\delta_1) & -\gamma^2\delta_1 \\ \gamma(\lambda_2-2\delta_1) & \gamma(\mu_2\lambda_2-\delta_1\sum\limits_{i=1}^{2}\mu_i) & -\frac{\gamma\delta_2\lambda_0}{\lambda_1\lambda_2} & -\gamma^2\delta_1 & \gamma^2(\lambda_2-\delta_1) \end{pmatrix},$$

we can further write this as the block diagonal matrix

$$\boldsymbol{I}(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{I}(\delta,\gamma,\lambda_0,\mu_1,\mu_2) & \mathbf{O}_{5\times 3} \\ \mathbf{O}_{3\times 5} & \sum\limits_{\boldsymbol{k}\in\Upsilon}\left[e^{-\sum\limits_{i=0}^{2}\lambda_i}\frac{\lambda_1^{k_1}}{k_1!}\frac{\lambda_2^{k_2}}{k_2!}\sum\limits_{i=0}^{\min(k_1,k_2)}\binom{k_1}{i}\binom{k_2}{i}i!\left(\frac{\lambda_0}{\lambda_1\lambda_2}\right)^i\right]\boldsymbol{I}_{\boldsymbol{k}}(\sigma^2,\rho,,\rho^*) \end{pmatrix} \quad \text{(C.19)}$$

where

$$\boldsymbol{I}(\delta,\gamma,\lambda_0,\mu_1,\mu_2) = \mathbf{H} + \begin{pmatrix} \mathbf{O}_{3\times 3} & \mathbf{O}_{3\times 2} \\ \mathbf{O}_{2\times 3} & \frac{1}{\sigma^2}\sum\limits_{\boldsymbol{k}\in\Upsilon}\left[e^{-\sum\limits_{i=0}^{2}\lambda_i}\frac{\lambda_1^{k_1}}{k_1!}\frac{\lambda_2^{k_2}}{k_2!}\sum\limits_{i=0}^{\min(k_1,k_2)}\binom{k_1}{i}\binom{k_2}{i}i!\left(\frac{\lambda_0}{\lambda_1\lambda_2}\right)^i\right]\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} \end{pmatrix},$$

and the elements of the matrix $\boldsymbol{I}_{\boldsymbol{k}}(\sigma^2,\rho,\rho^*)$ are defined by (A.10) - (A.15).

C.1.3 Derivation of the Information Matrix for the One Population Model

Suppose now that we have n independent observations from the one population model. The resulting information about the parameter $\boldsymbol{\theta}$ from all the data can be computed as the sum of the information contained in the n independent observations, found using (C.18). Hence, we get

$$\boldsymbol{I}_n(\boldsymbol{\theta}) = n\boldsymbol{I}^*(\boldsymbol{\theta}) + n\sum_{\boldsymbol{k}\in\Upsilon}\left[e^{-\sum\limits_{i=0}^{2}\lambda_i}\frac{\lambda_1^{k_1}}{k_1!}\frac{\lambda_2^{k_2}}{k_2!}\sum_{i=0}^{\min(k_1,k_2)}\binom{k_1}{i}\binom{k_2}{i}i!\left(\frac{\lambda_0}{\lambda_1\lambda_2}\right)^i\right]\boldsymbol{I}(\boldsymbol{\theta}|\boldsymbol{k}). \quad \text{(C.20)}$$

Using the expression in (C.19), it easy to show that $\boldsymbol{I}_n(\boldsymbol{\theta})$ also has a block diagonal form

$$\boldsymbol{I}_n(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{I}_n(\delta,\gamma,\lambda_0,\mu_1,\mu_2) & \mathbf{O}_{5\times 3} \\ \mathbf{O}_{3\times 5} & n\sum\limits_{\boldsymbol{k}\in\Upsilon}\left[e^{-\sum\limits_{i=0}^{2}\lambda_i}\frac{\lambda_1^{k_1}}{k_1!}\frac{\lambda_2^{k_2}}{k_2!}\sum\limits_{i=0}^{\min(k_1,k_2)}\binom{k_1}{i}\binom{k_2}{i}i!\left(\frac{\lambda_0}{\lambda_1\lambda_2}\right)^i\right]\boldsymbol{I}_{\boldsymbol{k}}(\sigma^2,\rho,\rho^*) \end{pmatrix} \quad \text{(C.21)}$$

where

$$\begin{aligned}\boldsymbol{I}_n(\delta,\gamma,\lambda_0,\mu_1,\mu_2) &= n\boldsymbol{I}(\delta,\gamma,\lambda_0,\mu_1,\mu_2)\\ &= n\mathbf{H}+\begin{pmatrix}\mathbf{O}_{3\times3} & \mathbf{O}_{3\times2}\\ \mathbf{O}_{2\times3} & n\frac{1}{\sigma^2}\sum\limits_{\boldsymbol{k}\in\Upsilon}\left[e^{-\sum\limits_{i=0}^{2}\lambda_i}\frac{\lambda_1^{k_1}}{k_1!}\frac{\lambda_2^{k_2}}{k_2!}\sum\limits_{i=0}^{\min(k_1,k_2)}\binom{k_1}{i}\binom{k_2}{i}i!\left(\frac{\lambda_0}{\lambda_1\lambda_2}\right)^i\right]\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1}\end{pmatrix},\end{aligned}$$

and the elements of the matrix $\boldsymbol{I}_{\boldsymbol{k}}(\sigma^2,\rho)$ are defined by (A.10) - (A.15). As before, we denote

$$\Upsilon=\left\{\boldsymbol{k}=(k_1,k_2)|\ k_i=0,1,...\ \text{ for }\forall i=1,2\text{ and }\sum_{i=1}^{2}k_i\geq 1\right\}.$$

C.2 DERIVATION OF THE INFORMATION MATRIX FOR THE MULTIPLE POPULATION MODEL

Let us now consider the multiple population model introduced in Chapter 4, Section 4.2. The parameter vector is $\boldsymbol{\theta}=(\delta,\gamma,\lambda_0,\mu_{11},\mu_{12},...,\mu_{m1},\mu_{m2},\sigma^2,\rho,\rho^*)'$. We denote by $\boldsymbol{I}_{\boldsymbol{i}}(\boldsymbol{\theta})$ the information matrix for $\boldsymbol{\theta}$ contained in the one multivariate random length vector $\boldsymbol{X}_{ij}$ from population i with random lengths $\boldsymbol{K}=(K_{ij1},K_{ij2})$. This vector belongs to a one population model with parameter vector $\boldsymbol{\theta}_{\boldsymbol{i}}=(\delta,\gamma,\lambda_0,\mu_{i1},\mu_{i2},\sigma^2,\rho,\rho^*)'$. We can apply the results from the previous section to find the information matrix about $\boldsymbol{\theta}_{\boldsymbol{i}}$ contributed by this one observation from the multiple population model. The random variables (K_{ij1},K_{ij2}) follow jointly a bivariate Poisson distribution and the joint probability function is given by

$$\begin{aligned}f_i(r,s) &= \mathbf{P}_{\boldsymbol{\theta}}(K_{ij1}=r,K_{ij2}=s)\\ &= e^{-\lambda_0-\sum\limits_{t=1}^{2}\lambda_{it}}\frac{\lambda_{i1}^r}{r!}\frac{\lambda_{i2}^s}{s!}\sum_{i=0}^{\min(r,s)}\binom{r}{i}\binom{s}{i}i!\left(\frac{\lambda_0}{\lambda_{i1}\lambda_{i2}}\right)^i.\end{aligned}$$

We denote

$$\delta_{i1} = \lambda_0\left[1-\lambda_0(\tau_i-1)\right] \tag{C.22}$$

$$\delta_{i2} = -(\lambda_{i1}+\lambda_{i2})+\left[\lambda_{i1}^*\lambda_{i2}^*-\lambda_0^2\right](\tau_i-1) \tag{C.23}$$

$$\delta_{i3} = \left[\lambda_{i1}^*\lambda_{i2}^*-\lambda_0^2\right]\left[\tau_1-(\lambda_{i1}+\lambda_{i2})\right] \tag{C.24}$$

$$\tau_i = \sum_{r,s=1}^{\infty}\frac{f_i^2(r-1,s-1)}{f_i(r,s)}, \tag{C.25}$$

where

$$\begin{aligned}\lambda_{i1}^* &= \lambda_{i1} + \lambda_0 \\ \lambda_{i2}^* &= \lambda_{i2} + \lambda_0.\end{aligned}$$

It follows that the information about $\boldsymbol{\theta_i} = (\delta, \gamma, \lambda_0, \mu_{i1}, \mu_{i2}, \sigma^2, \rho, \rho^*)'$ contained in the random lengths for one subject has the expression

$$\boldsymbol{I_i^*}(\boldsymbol{\theta_i}) = \begin{pmatrix} \mathbf{H}_i & \mathbf{O}_{5\times 2} \\ \mathbf{O}_{2\times 5} & \mathbf{O}_{2\times 2} \end{pmatrix},$$

where we define the matrix $\mathbf{H}_i$ as

$$\mathbf{H}_i = \begin{pmatrix} \lambda_{i1} + \lambda_{i2} - 4\delta_{i1} & \sum_{t=1}^{2} \mu_{it}(\lambda_{it} - 2\delta_{i1}) & -\frac{2\delta_{i2}\lambda_0}{\lambda_{i1}\lambda_{i2}} & \gamma(\lambda_{i1} - 2\delta_{i1}) & \gamma(\lambda_{i2} - 2\delta_{i1}) \\ \sum_{t=1}^{2} \mu_{it}(\lambda_{it} - 2\delta_{i1}) & \sum_{t=1}^{2} \mu_{it}^2\lambda_{it} - \delta_{i1}(\sum_{t=1}^{2} \mu_{it})^2 & -\frac{\delta_{i2}\lambda_0 \sum_{t=1}^{2} \mu_{it}}{\lambda_{i1}\lambda_{i2}} & \gamma(\mu_{i1}\lambda_{i1} - \delta_{i1} \sum_{t=1}^{2} \mu_{it}) & \gamma(\mu_{i2}\lambda_{i2} - \delta_{i1} \sum_{t=1}^{2} \mu_{it}) \\ -\frac{2\delta_{i2}\lambda_0}{\lambda_{i1}\lambda_{i2}} & -\frac{\delta_{i2}\lambda_0(\mu_{i1}+\mu_{i2})}{\lambda_{i1}\lambda_{i2}} & \frac{\delta_{i3}}{\lambda_{i1}^2\lambda_{i2}^2} & -\frac{\gamma\delta_{i2}\lambda_0}{\lambda_{i1}\lambda_{i2}} & -\frac{\gamma\delta_{i2}\lambda_0}{\lambda_{i1}\lambda_{i2}} \\ \gamma(\lambda_{i1} - 2\delta_{i1}) & \gamma(\mu_{i1}\lambda_{i1} - \delta_{i1} \sum_{t=1}^{2} \mu_{it}) & -\frac{\gamma\delta_{i2}\lambda_0}{\lambda_{i1}\lambda_{i2}} & \gamma^2(\lambda_{i1} - \delta_{i1}) & -\gamma^2\delta_{i1} \\ \gamma(\lambda_{i2} - 2\delta_{i1}) & \gamma(\mu_{i2}\lambda_{i2} - \delta_{i1} \sum_{t=1}^{2} \mu_{it}) & -\frac{\gamma\delta_{i2}\lambda_0}{\lambda_{i1}\lambda_{i2}} & -\gamma^2\delta_{i1} & \gamma^2(\lambda_{i2} - \delta_{i1}) \end{pmatrix}.$$

To simplify the notations, we partition the matrix $\mathbf{H}_i$ into

$$\mathbf{H}_i = \begin{pmatrix} \mathbf{H}_{11}^{(i)} & \mathbf{H}_{12}^{(i)} \\ \mathbf{H}_{12}^{(i)'} & \mathbf{H}_{22}^{(i)} \end{pmatrix}, \tag{C.26}$$

where $\mathbf{H}_{11}^{(i)}$ is the upper left-corner 3×3 submatrix of $\mathbf{H}_i$. Since there is no information in the random lengths about the other $\mu_{i't}$'s, where $i' \neq i$, the information about $\boldsymbol{\theta}$ contained in the random lengths for one subject in population i has the expression

$$\boldsymbol{I_i^*}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{H}_{11}^{(i)} & \mathbf{O}_{3\times 2(i-1)} & \mathbf{H}_{12}^{(i)} & \mathbf{O}_{3\times 2(m-i)} & \mathbf{O}_{3\times 3} \\ \mathbf{O}_{2(i-1)\times 3} & \mathbf{O}_{2(i-1)\times 2(i-1)} & \mathbf{O}_{2(i-1)\times 2} & \mathbf{O}_{2(i-1)\times 2(m-i)} & \mathbf{O}_{2(i-1)\times 3} \\ \mathbf{H}_{12}^{(i)'} & \mathbf{O}_{2\times 2(i-1)} & \mathbf{H}_{22}^{(i)} & \mathbf{O}_{2\times 2(m-i)} & \mathbf{O}_{2\times 3} \\ \mathbf{O}_{2(m-i)\times 3} & \mathbf{O}_{2(m-i)\times 2(i-1)} & \mathbf{O}_{2(m-i)\times 2} & \mathbf{O}_{2(m-i)\times 2(m-i)} & \mathbf{O}_{2(m-i)\times 3} \\ \mathbf{O}_{3\times 3} & \mathbf{O}_{3\times 2(i-1)} & \mathbf{O}_{3\times 2} & \mathbf{O}_{3\times 2(m-i)} & \mathbf{O}_{3\times 3} \end{pmatrix}.$$

Using (C.17), the information contributed by the vectors of severities given the lengths $\boldsymbol{I_i}(\boldsymbol{\theta_i}|\boldsymbol{k})$ can be computed as

$$\boldsymbol{I_i}(\boldsymbol{\theta_i}|\boldsymbol{k}) = \begin{pmatrix} \mathbf{O}_{3\times 3} & \mathbf{O}_{3\times 2} & \mathbf{O}_{3\times 3} \\ \mathbf{O}_{2\times 3} & \frac{1}{\sigma^2}\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} & \mathbf{O}_{2\times 3} \\ \mathbf{O}_{3\times 3} & \mathbf{O}_{3\times 2} & \boldsymbol{I_k}(\sigma^2, \rho, \rho^*) \end{pmatrix}.$$

Similarly to the above, since $\boldsymbol{\theta_i}$ involves only μ_{it}'s, $t = 1, 2$, we may write the information about $\boldsymbol{\theta}$ contributed by the vectors of severities given the lengths $\boldsymbol{I_i}(\boldsymbol{\theta}|\boldsymbol{k})$ as

$$\boldsymbol{I_i}(\boldsymbol{\theta}|\boldsymbol{k}) = \begin{pmatrix} \mathbf{O}_{3\times 3} & \mathbf{O}_{3\times 2(i-1)} & \mathbf{O}_{3\times 2} & \mathbf{O}_{3\times 2(m-i)} & \mathbf{O}_{3\times 3} \\ \mathbf{O}_{2(i-1)\times 3} & \mathbf{O}_{2(i-1)\times 2(i-1)} & \mathbf{O}_{2(i-1)\times 2} & \mathbf{O}_{2(i-1)\times 2(m-i)} & \mathbf{O}_{2(i-1)\times 3} \\ \mathbf{O}_{3\times 2} & \mathbf{O}_{2\times 2(i-1)} & \frac{1}{\sigma^2}\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} & \mathbf{O}_{2\times 2(m-i)} & \mathbf{O}_{2\times 3} \\ \mathbf{O}_{2(m-i)\times 2} & \mathbf{O}_{2(m-i)\times 2(i-1)} & \mathbf{O}_{2(m-i)\times 2} & \mathbf{O}_{2(m-i)\times 2(m-i)} & \mathbf{O}_{2(m-i)\times 3} \\ \mathbf{O}_{3\times 2} & \mathbf{O}_{3\times 2(i-1)} & \mathbf{O}_{3\times 2} & \mathbf{O}_{3\times 2(m-i)} & \boldsymbol{I_k}(\sigma^2, \rho, , \rho^*) \end{pmatrix}.$$

Finally, adding all the corresponding pieces gives us the information matrix for the multiple population model from Section 4.2

$$\boldsymbol{I}_n(\boldsymbol{\theta}) = \sum_{i=1}^{m} n_i \boldsymbol{I_i^*}(\boldsymbol{\theta}) + \sum_{i=1}^{m} n_i \sum_{\boldsymbol{k}\in\Upsilon} \left[e^{-\lambda_0 - \sum_{t=1}^{2}\lambda_{it}} \frac{\lambda_{i1}^{k_1}}{k_1!}\frac{\lambda_{i2}^{k_2}}{k_2!} \sum_{i=0}^{\min(k_1,k_2)} \binom{k_1}{i}\binom{k_2}{i} i! \left(\frac{\lambda_0}{\lambda_{i1}\lambda_{i2}}\right)^i \right] \boldsymbol{I_i}(\boldsymbol{\theta}|\boldsymbol{k_i}).$$

The summation above is over all the elements of the set

$$\Upsilon = \left\{ \boldsymbol{k} = (k_1, k_2) |\ k_j = 0, 1, ... \text{ for } \forall j = 1, 2 \text{ and } \sum_{t=1}^{2} k_t \geq 1 \right\}.$$

It can easily be shown that the information matrix $\boldsymbol{I}_n(\boldsymbol{\theta})$ about the parameter vector $\boldsymbol{\theta}$, $\boldsymbol{\theta} = (\delta, \gamma, \lambda_0, \mu_{11}, \mu_{12}, ..., \mu_{m1}, \mu_{m2}, \sigma^2, \rho, \rho^*)'$, contained in the $n = \sum_{i=1}^{m} n_i$ independent observations from the multiple population model with $T = 2$ has a block diagonal form

$$\boldsymbol{I}_n(\boldsymbol{\theta}) = \begin{pmatrix} \boldsymbol{I}_n(\delta, \gamma, \lambda_0, \mu_{11}, ..., \mu_{m2}) & \mathbf{O}_{(2m+3)\times 3} \\ \mathbf{O}_{3\times(2m+3)} & \sum_{i=1}^{m} n_i \sum_{\boldsymbol{k}\in\Upsilon} f_i(k_1, k_2)\boldsymbol{I_k}(\sigma^2, \rho, \rho^*) \end{pmatrix}, \tag{C.27}$$

where

$$f_i(k_1, k_2) = e^{-\lambda_0 - \sum_{t=1}^{2}\lambda_{it}} \frac{\lambda_{i1}^{k_1}}{k_1!}\frac{\lambda_{i2}^{k_2}}{k_2!} \sum_{i=0}^{\min(k_1,k_2)} \binom{k_1}{i}\binom{k_2}{i} i! \left(\frac{\lambda_0}{\lambda_{i1}\lambda_{i2}}\right)^i,$$

$$\boldsymbol{I}_n(\delta,\gamma,\lambda_0,\mu_{11},...,\mu_{m2}) \;=\; \sum_{i=1}^{m} n_i \begin{pmatrix} \mathbf{H}_{11}^{(i)} & \mathbf{O}_{3\times 2(i-1)} & \mathbf{H}_{12}^{(i)} & \mathbf{O}_{3\times 2(m-i)} \\ \mathbf{O}_{2(i-1)\times 3} & \mathbf{O}_{2(i-1)\times 2(i-1)} & \mathbf{O}_{2(i-1)\times 2} & \mathbf{O}_{2(i-1)\times 2(m-i)} \\ \mathbf{H}_{12}^{(i)\prime} & \mathbf{O}_{2\times 2(i-1)} & \mathbf{H}_{22}^{(i)} & \mathbf{O}_{2\times 2(m-i)} \\ \mathbf{O}_{2(m-i)\times 3} & \mathbf{O}_{2(m-i)\times 2(i-1)} & \mathbf{O}_{2(m-i)\times 2} & \mathbf{O}_{2(m-i)\times 2(m-i)} \end{pmatrix}$$

$$+ \begin{pmatrix} \mathbf{O}_{3\times 3} & \mathbf{O}_{3\times 2m} \\ \mathbf{O}_{2m\times 3} & \sum_{i=1}^{m} n_i \sum_{\boldsymbol{k}\in\Upsilon} f_i(k_1,k_2)\boldsymbol{\Sigma}_{\boldsymbol{k}}^{-1} \end{pmatrix},$$

the matrices $\mathbf{H}_{kl}^{(i)}$'s $(k,l=1,2)$ are given in (C.26) and the elements of the matrix $\boldsymbol{I}_{\boldsymbol{k}}(\sigma^2,\rho,\rho^*)$ are defined by (A.10) - (A.15).

APPENDIX D

RESULTS OF THE SIMULATIONS WHEN $T = 2$ FOR THE MODELS IN CHAPTER 3 (MODEL 1) AND CHAPTER 4 (MODEL 2)

Table 24: Simulation results for the first choice of parameters, $T = 2, n = 20$

	μ_{11}	μ_{12}	μ_{21}	μ_{22}	δ	γ	σ^2	ρ	ρ^*	λ_0
True(θ)	1	1.1	1.5	1.6	-1.3	2.3	1	0.5	0.2	1
$\lambda = e^{\delta+\gamma\mu}$	2.72	3.42	8.58	10.80						
Model 1										
average	0.996	1.095	1.5	1.6	-2.564	3.22	0.966	0.478	0.172	
bias	-0.004	-0.005	0	0	-1.264	0.92	-0.034	-0.022	-0.028	
sd	0.174	0.145	0.127	0.152	11.829	8.579	0.101	0.057	0.095	
mse$^{1/2}$	0.174	0.145	0.127	0.152	11.896	8.628	0.107	0.061	0.099	
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.168	0.142	0.126	0.15	1.267	0.934	0.106	0.055	0.093	
Model 2										
average	1.005	1.099	1.5	1.595	-1.998	2.831	0.985	0.485	0.185	0.559
bias	0.005	-0.001	0	-0.005	-0.698	0.531	-0.015	-0.015	-0.015	-0.441
sd	0.167	0.142	0.14	0.161	4.064	3.027	0.221	0.057	0.092	4.577
mse$^{1/2}$	0.167	0.142	0.14	0.161	4.124	3.073	0.221	0.059	0.093	4.598
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.162	0.137	0.125	0.149	1.251	0.924	0.102	0.053	0.089	0.923

Table 25: Simulation results for the first choice of parameters, $T = 2, n = 50$

	μ_{11}	μ_{12}	μ_{21}	μ_{22}	δ	γ	σ^2	ρ	ρ^*	λ_0
True(θ)	1	1.1	1.5	1.6	-1.3	2.3	1	0.5	0.2	1
$\lambda = e^{\delta+\gamma\mu}$	2.72	3.42	8.58	10.80						
Model 1										
average	0.997	1.097	1.5	1.599	-1.55	2.486	0.981	0.489	0.188	
bias	-0.003	-0.003	0	-0.001	-0.25	0.186	-0.019	-0.011	-0.012	
sd	0.108	0.093	0.08	0.094	1.161	0.852	0.066	0.035	0.057	
mse$^{1/2}$	0.108	0.093	0.08	0.094	1.187	0.872	0.068	0.037	0.058	
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.106	0.09	0.08	0.095	0.801	0.591	0.067	0.035	0.059	
Model 2										
average	0.993	1.096	1.499	1.599	-1.563	2.5	0.987	0.493	0.191	1.01
bias	-0.007	-0.004	-0.001	-0.001	-0.263	0.2	-0.013	-0.007	-0.009	0.01
sd	0.104	0.087	0.083	0.098	1.227	0.913	0.064	0.033	0.057	0.511
mse$^{1/2}$	0.104	0.087	0.083	0.098	1.255	0.934	0.066	0.034	0.058	0.511
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.102	0.087	0.079	0.094	0.791	0.585	0.065	0.033	0.056	0.497

Table 26: Simulation results for the first choice of parameters, $T = 2, n = 100$

	μ_{11}	μ_{12}	μ_{21}	μ_{22}	δ	γ	σ^2	ρ	ρ^*	λ_0
True(θ)	1	1.1	1.5	1.6	-1.3	2.3	1	0.5	0.2	1
$\lambda = e^{\delta+\gamma\mu}$	2.72	3.42	8.58	10.80						
Model 1										
average	1	1.099	1.498	1.598	-1.428	2.396	0.983	0.492	0.192	
bias	0	-0.001	-0.002	-0.002	-0.128	0.096	-0.017	-0.008	-0.008	
sd	0.077	0.066	0.055	0.065	0.687	0.503	0.046	0.024	0.04	
mse$^{1/2}$	0.077	0.066	0.055	0.065	0.699	0.512	0.049	0.025	0.041	
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.075	0.063	0.056	0.067	0.566	0.418	0.047	0.025	0.042	
Model 2										
average	1.004	1.105	1.502	1.602	-1.445	2.402	0.992	0.496	0.195	0.997
bias	0.004	0.005	0.002	0.002	-0.145	0.102	-0.008	-0.004	-0.005	-0.003
sd	0.075	0.062	0.056	0.068	0.722	0.515	0.045	0.023	0.04	0.375
mse$^{1/2}$	0.075	0.063	0.056	0.068	0.736	0.525	0.046	0.024	0.04	0.375
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.072	0.061	0.056	0.067	0.559	0.413	0.046	0.024	0.040	0.372

Table 27: Simulation results for the second choice of parameters, $T = 2, n = 20$

	μ_{11}	μ_{12}	μ_{21}	μ_{22}	δ	γ	σ^2	ρ	ρ^*	λ_0
True(θ)	1.3	1.2	1.2	1.3	-1.0	2.5	1	0.5	0.2	1
$\lambda = e^{\delta+\gamma\mu}$	4.48	5.475	5.75	4.48						
Model 1										
average	1.303	1.195	1.197	1.298	1.323	0.63	0.981	0.487	0.195	
bias	0.003	-0.005	-0.003	-0.002	2.323	-1.87	-0.019	-0.013	-0.005	
sd	0.13	0.127	0.132	0.13	15.556	12.572	0.099	0.051	0.085	
mse$^{1/2}$	0.13	0.128	0.132	0.13	15.728	12.711	0.101	0.052	0.085	
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.12	0.122	0.122	0.12	4.363	3.48	0.098	0.05	0.085	
Model 2										
average	1.293	1.201	1.2	1.302	1.196	0.728	0.983	0.488	0.194	0.817
bias	-0.007	0.001	0	0.002	2.196	-1.772	-0.017	-0.012	-0.006	-0.183
sd	0.131	0.144	0.142	0.139	10.604	8.499	0.095	0.049	0.084	1.912
mse$^{1/2}$	0.131	0.144	0.142	0.139	10.829	8.682	0.096	0.05	0.084	1.921
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.119	0.122	0.122	0.119	4.315	3.443	0.097	0.049	0.084	1.245

Table 28: Simulation results for the second choice of parameters, $T = 2, n = 50$

	μ_{11}	μ_{12}	μ_{21}	μ_{22}	δ	γ	σ^2	ρ	ρ^*	λ_0
True(θ)	1.3	1.2	1.2	1.3	-1.0	2.5	1	0.5	0.2	1
$\lambda = e^{\delta+\gamma\mu}$	4.48	5.475	5.75	4.48						
Model 1										
average	1.299	1.199	1.199	1.298	0.602	1.203	0.991	0.494	0.197	
bias	-0.001	-0.001	-0.001	-0.002	1.602	-1.297	-0.009	-0.006	-0.003	
sd	0.082	0.081	0.079	0.077	13.972	11.372	0.062	0.033	0.054	
mse$^{1/2}$	0.082	0.081	0.079	0.077	14.064	11.446	0.063	0.033	0.054	
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.076	0.077	0.077	0.076	2.759	2.201	0.062	0.032	0.054	
Model 2										
average	1.301	1.198	1.2	1.303	-0.571	2.141	0.993	0.494	0.2	0.908
bias	0.001	-0.002	0	0.003	0.429	-0.359	-0.007	-0.006	0	-0.092
sd	0.081	0.084	0.082	0.082	10.87	8.779	0.061	0.031	0.052	1.158
mse$^{1/2}$	0.081	0.084	0.082	0.082	10.879	8.786	0.062	0.032	0.052	1.162
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.076	0.077	0.077	0.076	2.729	2.177	0.061	0.031	0.053	1.023

Table 29: Simulation results for the second choice of parameters, $T = 2, n = 100$

	μ_{11}	μ_{12}	μ_{21}	μ_{22}	δ	γ	σ^2	ρ	ρ^*	λ_0
True(θ)	1.3	1.2	1.2	1.3	-1.0	2.5	1	0.5	0.2	1
$\lambda = e^{\delta+\gamma\mu}$	4.48	5.475	5.75	4.48						
Model 1										
average	1.299	1.201	1.201	1.299	-1.272	2.715	0.996	0.498	0.198	
bias	-0.001	0.001	0.001	-0.001	-0.272	0.215	-0.004	-0.002	-0.002	
sd	0.054	0.057	0.056	0.054	10.427	8.33	0.044	0.022	0.039	
mse$^{1/2}$	0.054	0.057	0.056	0.054	10.431	8.333	0.045	0.022	0.039	
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.054	0.055	0.055	0.054	1.951	1.556	0.044	0.022	0.038	
Model 2										
average	1.298	1.199	1.199	1.297	-1.695	3.059	0.996	0.498	0.199	0.936
bias	-0.002	-0.001	-0.001	-0.003	-0.695	0.559	-0.004	-0.002	-0.001	-0.064
sd	0.056	0.056	0.056	0.056	8.124	6.514	0.043	0.022	0.038	0.763
mse$^{1/2}$	0.056	0.056	0.056	0.056	8.154	6.538	0.044	0.022	0.038	0.766
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.053	0.054	0.054	0.053	1.930	1.540	0.043	0.022	0.037	0.742

Table 30: Simulation results for the Third choice of parameters, $T = 2, n = 20$

	μ_{11}	μ_{12}	μ_{21}	μ_{22}	δ	γ	σ^2	ρ	ρ^*	λ_0
True(θ)	1	2	1.5	2.5	2.5	0.01	1	0.5	0.2	1
$\lambda = e^{\delta+\gamma\mu}$	12.30	12.43	12.37	12.49						
Model 1										
average	0.996	1.996	1.503	2.498	2.502	0.008	0.97	0.481	0.192	
bias	-0.004	-0.004	0.003	-0.002	0.002	-0.002	-0.03	-0.019	-0.008	
sd	0.171	0.168	0.168	0.163	0.108	0.059	0.091	0.048	0.08	
mse$^{1/2}$	0.171	0.168	0.168	0.163	0.108	0.059	0.096	0.052	0.08	
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.165	0.165	0.165	0.165	0.105	0.057	0.094	0.048	0.081	
Model 2										
average	0.994	2.004	1.51	2.506	2.462	0.018	0.97	0.48	0.19	0.986
bias	-0.006	0.004	0.01	0.006	-0.038	0.008	-0.03	-0.02	-0.01	-0.014
sd	0.17	0.163	0.165	0.163	0.223	0.065	0.088	0.047	0.08	2.312
mse$^{1/2}$	0.171	0.163	0.166	0.164	0.226	0.065	0.093	0.051	0.08	2.312
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.164	0.164	0.164	0.164	0.107	0.058	0.089	0.047	0.081	2.297

Table 31: Simulation results for the third choice of parameters, $T = 2, n = 50$

	μ_{11}	μ_{12}	μ_{21}	μ_{22}	δ	γ	σ^2	ρ	ρ^*	λ_0
True(θ)	1	2	1.5	2.5	2.5	0.01	1	0.5	0.2	1
$\lambda = e^{\delta+\gamma\mu}$	12.30	12.43	12.37	12.49						
Model 1										
average	1.003	2	1.499	2.504	2.501	0.009	0.985	0.492	0.194	
bias	0.003	0	-0.001	0.004	0.001	-0.001	-0.015	-0.008	-0.006	
sd	0.103	0.106	0.105	0.107	0.07	0.038	0.056	0.029	0.051	
mse$^{1/2}$	0.104	0.106	0.105	0.107	0.07	0.038	0.057	0.03	0.051	
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.104	0.104	0.104	0.104	0.066	0.036	0.059	0.03	0.051	
Model 2										
average	0.998	1.997	1.493	2.491	2.494	0.008	0.989	0.493	0.197	1.053
bias	-0.002	-0.003	-0.007	-0.009	-0.006	-0.002	-0.011	-0.007	-0.003	0.053
sd	0.102	0.105	0.102	0.105	0.133	0.039	0.061	0.031	0.05	1.326
smse	0.102	0.105	0.103	0.106	0.133	0.039	0.062	0.032	0.05	1.327
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.104	0.104	0.104	0.104	0.067	0.036	0.059	0.030	0.051	1.287

Table 32: Simulation results for the third choice of parameters, $T = 2, n = 100$

	μ_{11}	μ_{12}	μ_{21}	μ_{22}	δ	γ	σ^2	ρ	ρ^*	λ_0
True(θ)	1	2	1.5	2.5	2.5	0.01	1	0.5	0.2	1
$\lambda = e^{\delta+\gamma\mu}$	12.30	12.43	12.37	12.49						
Model 1										
average	0.998	1.999	1.498	2.499	2.503	0.008	0.994	0.496	0.198	
bias	-0.002	-0.001	-0.002	-0.001	0.003	-0.002	-0.006	-0.004	-0.002	
sd	0.074	0.072	0.071	0.073	0.048	0.026	0.043	0.022	0.036	
mse$^{1/2}$	0.074	0.072	0.071	0.073	0.049	0.026	0.043	0.022	0.036	
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.074	0.074	0.074	0.074	0.047	0.025	0.042	0.021	0.036	
Model 2										
average	1.005	2.001	1.501	2.5	2.494	0.012	0.994	0.496	0.199	1.006
bias	0.005	0.001	0.001	0	-0.006	0.002	-0.006	-0.004	-0.001	0.006
sd	0.071	0.074	0.075	0.076	0.091	0.026	0.043	0.022	0.038	0.935
mse$^{1/2}$	0.071	0.074	0.075	0.076	0.092	0.026	0.043	0.022	0.038	0.935
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.074	0.073	0.073	0.073	0.048	0.026	0.042	0.021	0.036	0.932

Table 33: Simulation results for the fourth choice of parameters, $T = 2, n = 20$

	μ_{11}	μ_{12}	μ_{21}	μ_{22}	δ	γ	σ^2	ρ	ρ^*	λ_0
True(θ)	4	3	3	4	1.5	0.05	1	0.5	0.2	1
$\lambda = e^{\delta+\gamma\mu}$	5.47	5.21	5.21	5.47						
Model 1										
average	4.002	3.005	3.011	4.002	1.489	0.053	0.97	0.48	0.187	
bias	0.002	0.005	0.011	0.002	-0.011	0.003	-0.03	-0.02	-0.013	
sd	0.172	0.179	0.183	0.177	0.37	0.103	0.102	0.055	0.092	
mse$^{1/2}$	0.173	0.179	0.183	0.177	0.37	0.103	0.106	0.059	0.093	
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.174	0.175	0.175	0.174	0.344	0.097	0.106	0.056	0.092	
Model 2										
average	3.995	2.993	2.995	3.993	1.474	0.054	0.975	0.481	0.19	0.723
bias	-0.005	-0.007	-0.005	-0.007	-0.026	0.004	-0.025	-0.019	-0.01	-0.277
sd	0.181	0.17	0.177	0.175	0.455	0.121	0.19	0.056	0.09	5.876
mse$^{1/2}$	0.182	0.17	0.178	0.176	0.456	0.122	0.192	0.059	0.091	5.883
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.172	0.172	0.172	0.172	0.346	0.097	0.103	0.055	0.088	1.937

Table 34: Simulation results for the fourth choice of parameters, $T = 2, n = 50$

	μ_{11}	μ_{12}	μ_{21}	μ_{22}	δ	γ	σ^2	ρ	ρ^*	λ_0
True(θ)	4	3	3	4	1.5	0.05	1	0.5	0.2	1
$\lambda = e^{\delta+\gamma\mu}$	5.47	5.21	5.21	5.47						
Model 1										
average	3.996	2.997	3.001	4.004	1.493	0.052	0.989	0.491	0.196	
bias	-0.004	-0.003	0.001	0.004	-0.007	0.002	-0.011	-0.009	-0.004	
sd	0.109	0.116	0.112	0.109	0.221	0.062	0.067	0.036	0.059	
mse$^{1/2}$	0.109	0.116	0.112	0.109	0.221	0.062	0.068	0.037	0.059	
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.11	0.111	0.111	0.11	0.218	0.061	0.067	0.035	0.058	
Model 2										
average	4.005	3.004	2.997	3.994	1.485	0.05	0.99	0.493	0.197	1.035
bias	0.005	0.004	-0.003	-0.006	-0.015	0	-0.01	-0.007	-0.003	0.035
sd	0.107	0.109	0.108	0.111	0.274	0.064	0.067	0.036	0.056	0.645
mse$^{1/2}$	0.107	0.109	0.108	0.111	0.274	0.064	0.068	0.036	0.056	0.646
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.108	0.109	0.109	0.108	0.219	0.061	0.065	0.034	0.056	0.639

Table 35: Simulation results for the fourth choice of parameters, $T = 2, n = 100$

	μ_{11}	μ_{12}	μ_{21}	μ_{22}	δ	γ	σ^2	ρ	ρ^*	λ_0
True(θ)	4	3	3	4	1.5	0.05	1	0.5	0.2	
$\lambda = e^{\delta+\gamma\mu}$	5.47	5.21	5.21	5.47						
Model 1										
average	4.004	3.001	3	4.004	1.499	0.05	0.993	0.495	0.199	
bias	0.004	0.001	0	0.004	-0.001	0	-0.007	-0.005	-0.001	
sd	0.081	0.082	0.08	0.079	0.155	0.043	0.047	0.025	0.041	
mse$^{1/2}$	0.081	0.082	0.08	0.079	0.155	0.043	0.048	0.026	0.041	
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.078	0.078	0.078	0.078	0.154	0.043	0.047	0.025	0.041	
Model 2										
average	4.029	3.006	3.014	3.999	1.508	0.047	1.003	0.5	0.213	1.025
bias	0.029	0.006	0.014	-0.001	0.008	-0.003	0.003	0	0.013	0.025
sd	0.091	0.072	0.101	0.087	0.175	0.047	0.038	0.015	0.031	0.326
mse$^{1/2}$	0.095	0.072	0.101	0.087	0.176	0.047	0.038	0.015	0.034	0.327
$\boldsymbol{I}_n^{-1/2}(\theta)$	0.077	0.077	0.077	0.077	0.155	0.043	0.046	0.024	0.040	

BIBLIOGRAPHY

[1] Allen, A. S. (2001). Multivariate random length and missing data. *Ph.D. Dissertation*, Emory University.

[2] Allen, A. S. and Barnhart, H. X. (2005). General marginal regression models for the joint modeling of event frequency and correlated severities with applications to clinical trials. *Journal of Data Science* **3**, 199-219.

[3] Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*, Second edition. New York: John Wiley & Sons, Inc.

[4] Barnhart, H. X. (1992). Models for multivariate random length data with applications in clinical trials. *Ph.D. Dissertation*, University of Pittsburgh.

[5] Barnhart, H. X. (1993). Models for multivariate random length data with applications in clinical trials. *Drug Information Journal* **27**, 1147-1157.

[6] Barnhart, H. X. and Sampson, A. R. (1995). Multiple population models for multivariate random length data without covariates. *Biometrics* **51**, 195-204.

[7] Barnhart, H. X.(1998). A probit model for multivariate random length ordinal data. *Communications in Statistics* **27**, 1693-1713.

[8] Barnhart, H. X., Kosinski, A. S. and Sampson, A. R. (1999). A regression model for multivariate random length data. *Statistics in Medicine* **18**, 199-211.

[9] Birmaher, B., Dahl, R. E., Williamson, D. E., Perel, J. M., Brent, D. A., Axelson, D. A., Kaufman, J. , Dorn, L. D. , Stull, S., Rao, U. and Ryan, N. D. (2000). Growth hormone secretion in children and adolescents at high risk for major depressive disorder. *Archives of General Psychiatry* **57(9)**, 867-872.

[10] Brown, G. W. and Harris, T. O. (1989). *Life Events and Illness.* New York, Guilford Press.

[11] Dahl, R. E., Birmaher, B., Williamson, D. E., Dorn, L., Perel, J., Kaufman, J., Brent, D. A., Axelson, D. A. and Ryan, N. D. (2000). Low growth hormone response to growth hormone-releasing hormone in child depression. *Biological Psychiatry* **48(10)**, 981-988.

[12] Dunson, D. B., Chen, Z. and Harry, J. (2003). A Bayesian approach for joint modeling of cluster size and subunit-specific outcomes. *Biometrics* **59**, 521-530.

[13] Follmanm, D., Proschan, M. and Leifer, E. (2003). Multiple outputations: inference for complex clustered data by averaging analyses from independent data. *Biometrics* **59**, 420-429.

[14] Hillegers, M. H., Burger, H., Wals, M., Reichart, C. G., Verhulst, F. C., Nolen, W. A. and Ormel, J. (2004). Impact of stressful life events, familial loading and their interaction on the onset of mood disorders. *British Journal of Psychiatry* **185**, 97-101.

[15] Hoffman, E. B. (1999). Within-cluster resampling. *Ph.D. Dissertation*, University of North Carolina at Chapel Hill.

[16] Hoffman, E. B., Sen, P. K. and Weinberg, C. R. (2001). Within-cluster resampling. *Biometrika* **88**, 1121-1134.

[17] Holgate, P. (1964). Estimation for the bivariate Poisson distribution. *Biometrika* **51**, 241-245.

[18] Holmes, T. H. and Rahe, R. H. (1967). The social readjustment rating scale. *Journal of Psychosomatic Research* **11**, 213-218.

[19] Johnson, N. , Kotz, S. and Balakrishnan, N. (1997). *Discrete Multivariate Distributions*, Wiley, New York.

[20] Karlis, D. (2003). An EM algorithm for multivariate Poisson distribution and related models. *Journal of Applied Statistics* **30(1)**, 63-77.

[21] Karlis, D. and Ntzoufras, I. (2005). Bivariate Poisson and Diagonal Inflated Bivariate Poisson Regression Models in R. *Journal of Statistical Software* **14(10)**, 63-77.

[22] Karlis, D. and Meligkotsidou, L. (2005). Multivariate Poisson regression with covariance structure. *Statistics and Computing* **15**, 255-265.

[23] Karlis, D. and Meligkotsidou, L. (2007). Finite mixtures of multivariate Poisson distributions with applications. *Journal of Statistical Planning and Inference* **137**, 1942-1960.

[24] Kim-Cohen J. , Caspi, A., Moffitt, T. E., Harrington, H. , Milne B. J. and Poulton, R. (2003). Prior juvenile diagnoses in adults with mental disorder: developmental follow-back of a prospective-longitudinal cohort. *Archives of General Psychiatry.* **60(7)**, 709-717.

[25] Kocherlakota, S. and Kocherlakota, K. (1992). *Bivariate Discrete Distributions*, Marcel Dekker, Inc. , New York.

[26] Lange, K. (1999). *Numerical Analysis for Statisticians*, New York : Springer-Verlag.

[27] Lange, K. (2004). *Optimization*, New York : Springer-Verlag.

[28] Lehman, E. L. and Romano, J. P. (2005). *Testing Statistical Hypothesis*, Third Edition. New York : Springer-Verlag.

[29] Liang, K. -Y. and Zeger, S. L. (1986). Longitudinal analysis using generalized linear models. *Biometrika* **73**, 13-22.

[30] McCulloch, C. E. and Searle, S. R. (2001). *Generalized, Linear and Mixed Models.* New York: John Wiley & Sons, Inc.

[31] Monck E. and Dobbs, R. (1985). Measuring life events in an adolescent population: Methodological issues and related findings. *Psychological Medicine* **15(4)**, 841-850.

[32] Rao, C. R. (1973). *Linear Statistical Inference and its Applications*, Second edition. New York: John Wiley & Sons, Inc.

[33] Sen, P. K. and Singer, J. M. (1993). *Large Sample Methods in Statistics. An Introduction with Applications.* New York & London: Chapman & Hall, Inc.

[34] Surtees, P. G. and Ingham, J. G. (1980). Life stress and depressive outcome: Application of a dissipation model to life events. *Social Psychiatry* **15**, 21-31.

[35] Williamson, J. M., Datta, S. and Satten, G. A. (2003). Marginal analyses of clustered data when cluster size is informative. *Biometrics* **59**, 36-42.

[36] Zeger, S. L. and Liang, K. -Y. (1986). Longitudinal analysis for discrete and continuous outcomes. *Biometrics* **42**, 121-130.