

Modeling Recurrent Gap Times Through Conditional GEE

Hai Yan Liu

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements for the degree of Doctor of Philosophy in
Mathematics ¹

Department of Mathematics and Statistics
Faculty of Science
University of Ottawa

© Hai Yan Liu, Ottawa, Canada, 2018

¹The Ph.D. program is a joint program with Carleton University, administered by the Ottawa-Carleton Institute of Mathematics and Statistics

Abstract

We present a theoretical approach to the statistical analysis of the dependence of the gap time length between consecutive recurrent events, on a set of explanatory random variables and in the presence of right censoring. The dependence is expressed through regression-like and overdispersion parameters, estimated via estimating functions and equations. The mean and variance of the length of each gap time, conditioned on the observed history of prior events and other covariates, are known functions of parameters and covariates, and are part of the estimating functions. Under certain conditions on censoring, we construct normalized estimating functions that are asymptotically unbiased and contain only observed data. We then use modern mathematical techniques to prove the existence, consistency and asymptotic normality of a sequence of estimators of the parameters. Simulations support our theoretical results.

Keywords: Conditional estimating functions; Recurrent events; Censoring; Covariates; Strong consistency of estimators; Asymptotic normality of estimators.

Acknowledgements

I want to express my gratitude to my supervisors Mayer Alvo, Ioana Schiopu-Kratina and Pierre-Jérôme Bergeron for guiding me in all these years with patience, encouragement and knowledge. My sincere thanks also go to my thesis examining committee. I also take this opportunity to express gratitude to the Department of Mathematics and Statistics of the University of Ottawa and Carleton University and the Faculty of Graduate and Postdoctoral Studies, which provided me an excellent academic environment. I also appreciate the financial support from University of Ottawa Admission Scholarship, NSERC and CIHR, for financial support obtained through my supervisors, which funded parts of the research presented in this dissertation.

Last but not the least, I would like to thank my office mates and friends in the Department of Mathematics and Statistics for all times we were working together and for all the fun we have had in the last years.

Dedication

This work is dedicated to my mother, wife and daughter for their unconditional support. I am aware that this work would not have been successful without their help.

Contents

List of Figures	vii
List of Tables	viii
1 Introduction	1
2 The Model and Basic Assumptions	10
2.1 Model Assumptions and Examples	10
2.2 Short Description of Previous Work	16
2.3 Unbiased Estimating Functions	18
2.4 Observed Estimating Functions	23
2.5 Other Estimating Functions	30
2.6 Connection with Other Longitudinal Studies	35
2.7 A Preliminary Result	36
2.8 Examples	41
3 Strong Consistency	49
3.1 General Results	49
3.2 Asymptotic Results for the Derivatives	52
3.2.1 The Derivative of the Non-imputed Terms of $\hat{g}_{n,1}^{obs}$	53
3.2.2 The Derivative of the Imputed Part of $\hat{g}_{n,1}^{obs}$	73
3.2.3 Further Results on the Derivative of $\hat{g}_{n,1}^{obs}$	85

3.2.4	Results for the Derivative of $\hat{g}_{n,2}^{obs}$	89
3.3	The Strong Consistency Theorem	94
3.3.1	The Strong Consistency of $\hat{\theta}_n$	94
3.3.2	The Strong Consistency of $\hat{\sigma}_n^2$	98
4	The Asymptotic Normality of Estimators	105
4.1	The Asymptotic Normality of $\hat{\theta}_n$	106
4.2	The Asymptotic Normality of $\hat{\sigma}_n^2$	111
5	Simulations Results	118
5.1	First Scenario: the Mean and the Variance Functions are Unrelated	121
5.2	Second Scenario: the Variance Function is the Absolute Value of the Mean Function	124
5.3	Discussion	127
A	Extended Simulation Tables	130
A.1	Full Tables for Independent Mean and Variance Function	130
A.2	Full Tables for Simulations where Variance Function is the Ab- solute Mean	134
	Bibliography	139
	List of Symbols	143
	Index	146

List of Figures

5.1 Bias Comparison for Scenario 1	122
5.2 Standard Error Comparison for Scenario 1	122
5.3 Bias Comparison for Scenario 2	125
5.4 Standard Error Comparison for Scenario 2	126

List of Tables

5.1	Simulated data format	120
5.2	Simulated tidy data format	121
5.3	Comparison with C&S, $n = 50$, V_{ij} in (5.0.2), $C_{\max} = 225$	123
5.4	Comparison with C&S, $n = 50$, V_{ij} in (5.0.2), $C_{\max} = 125$	123
5.5	Comparison with C&S, $n = 200$, V_{ij} in (5.0.2), $C_{\max} = 225$	124
5.6	Comparison with C&S, $n = 200$, V_{ij} in (5.0.2), $C_{\max} = 125$	124
5.7	Comparison with C&S, $n = 50$, $V_{ij}(\theta) = \mu_{ij}(\theta) $, $C_{\max} = 225$	126
5.8	Comparison with C&S, $n = 50$, $V_{ij}(\theta) = \mu_{ij}(\theta) $, $C_{\max} = 125$	127
5.9	Comparison with C&S, $n = 200$, $V_{ij}(\theta) = \mu_{ij}(\theta) $, $C_{\max} = 225$	127
5.10	Comparison with C&S, $n = 200$, $V_{ij}(\theta) = \mu_{ij}(\theta) $, $C_{\max} = 125$	127
A.1	Comparison with C&S, $n = 50$, V_{ij} in (5.0.2), $C_{\max} = 225$	131
A.2	Comparison with C&S, $n = 50$, V_{ij} in (5.0.2), $C_{\max} = 125$	132
A.3	Comparison with C&S, $n = 200$, V_{ij} in (5.0.2), $C_{\max} = 225$	133
A.4	Comparison with C&S, $n = 200$, V_{ij} in (5.0.2), $C_{\max} = 125$	134
A.5	Comparison with C&S, $n = 50$, $V_{ij}(\theta) = \mu_{ij}(\theta) $, $C_{\max} = 225$	135
A.6	Comparison with C&S, $n = 50$, $V_{ij}(\theta) = \mu_{ij}(\theta) $, $C_{\max} = 125$	136
A.7	Comparison with C&S, $n = 200$, $V_{ij}(\theta) = \mu_{ij}(\theta) $, $C_{\max} = 225$	137
A.8	Comparison with C&S, $n = 200$, $V_{ij}(\theta) = \mu_{ij}(\theta) $, $C_{\max} = 125$	138

Chapter 1

Introduction

Data collected nowadays from recurrent events constitute a well studied area in statistics, with application ranging from economics to engineering to biomedical. The development of useful regression models for recurrent event data is a problem of significant, practical and methodological interest. When the events are all considered to be of the same type, the waiting times or gap times constitute a natural and informative time scale for analysis. Studies based on gap times are of interest when the events themselves are not of direct interest or when there is a renewal after the occurrence of an event. Examples are in the study of system failures or in cyclical phenomena where it is of interest to characterize the cycle length. An interesting application was considered by Cook and Lawless in [8] on the data collected from a carcinogenicity experiment involving female rats. The subjects were randomized to treatment or control and then exposed to a carcinogen. They were subsequently examined for the development of new tumors whose times of occurrence were recorded. An important question is how to estimate the mean and variance function when the individual subjects are not observed over the same time interval. As well, there is the issue of right censoring if the subject dies.

Recurrent event data is collected on a preferably large sample of independently

selected individuals. We view it as longitudinal data collected on each individual before or at the time each recurrent event occurs. Naturally, data collected on the same individual on different occasions is correlated and therefore can be viewed as a cluster of observations, where an intra-cluster dependence exists. Missing data in the form of censoring, creates, for many subjects, partly observed last gap times.

Even if each individual were observed for a finite number of complete gap times, the drop-out of subjects would create partially observed last gap times. A typical example, which we will often refer to, is given in [18]. The authors study the menstrual pattern of a sample of Lese women from Zaire. They parameterize and analyze the mean length of the menstrual cycle, conditioned on the past cycles and covariates such as age and the BMI (Body Mass Index). This study encapsulates most of the challenges encountered in the analysis of recurrent events: the covariates are time dependent random variables and there is right censoring.

Gap times have been studied by Cook and Lawless in [8] using parametric models based on the hazard function. They also provide an extensive literature review of such methods. In this thesis, we base our approach on the methodology developed in [18] wherein they relax the stringent restrictions imposed by simpler marginal models, while avoiding the need to fully specify how the probability of subsequent recurrence depends on the prior events and covariate history. This approach lends itself to more readily model similar phenomena such as, for example, the pattern of repeated work injuries, or recurrent asthma in children (details provided in [6]).

The theoretical background developed in [17] accommodates the drop-out of individuals from the study. The focus of the study in [18] is the identification of factors that contribute to the menstrual cycle variability.

The approach adopted in [18] and [6]-[7] for the analysis of recurrent events is based on the use of *estimating function(s)* (*e.f.*) to produce estimators of the parameters. More precisely, these authors use a particular case of *generalized estimating equation(s)* (GEE) introduced in [14], which may be succinctly described as follows.

We now use the notation in [2]-[3], [9] and [14]. Let $y_{ij} \in R$ represent the response variables from independent individuals and $x_{ij} \in R^p, p \geq 1$ the nonrandom covariates, for $1 \leq j \leq m_i, i \geq 1$, where j refers to the occasion on which measurements are taken for the i^{th} individual. We are dealing with a regression-type model, with regression parameter $\beta \in R^p$. In these *marginal models*, only the means and variances are specified at each occasion, namely

$$E_{\beta}(y_{ij}) = \mu(x_{ij}^T \beta) = \mu_{ij}(\beta), \quad \text{Var}_{\beta}(y_{ij}) = \mu'(x_{ij}^T \beta) = \phi \sigma_{ij}^2(\beta), \quad (1.0.1)$$

where μ is a canonical link function, μ' its derivative and ϕ is a scaling parameter.

To estimate the true parameter β_0 , Liang and Zeger in [14] obtained a sequence of estimators $\hat{\beta}_n$, which are roots of the GEE

$$\sum_{i=1}^n \left[\frac{\partial \mu_i(\beta)}{\partial \beta^T} \right]^T V_i^{-1}(\beta, \alpha) [y_i - \mu_i(\beta)] = 0. \quad (1.0.2)$$

In (1.0.2), $y_i, \mu_i(\beta)$ are m_i -dimensional vectors, α is a “nuisance” parameter and $V_i(\beta, \alpha)$ is a “working” covariance matrix, which replaces the correct, but unknown intra-cluster covariance. An important particular case occurs under the “working independence” assumption, when the correlation matrix corresponding to V_i in (1.0.2) is the identity matrix (see (2.6.1) and Section 2.6).

It is shown in [14] that the sequence $\{\hat{\beta}_n\}_{n \geq 1}$ is *consistent*, *i.e.*, it converges to β_0 , regardless of what the true covariance is, and this sequence of estimators is asymptotically normally distributed. The penalty for using a “working” covariance in lieu of the true one results in a decrease in efficiency.

Theoretical justification and extensions of the results in [14] were given in [27], [2] as well as in [3]. Further applications and examples of the GEE method can be found in [30] and [31].

We now briefly present the use of the GEE (under the working independence assumption) in the analysis of recurrent events, as is done in [6]. We first point out the difference in notation.

In [6], the response process is denoted Y_{ij} , and represents a measure of the gap time between the $(j - 1)^{th}$ and the j^{th} occurrence of the event for individual i . The main, regression parameter β is denoted θ in [6], whereas $\sigma_{ij}(\beta)$ in (1.0.1) is $V_{ij}(\theta)$ in [6]. Censoring aside, it can be seen that the first expression in (2.7) of [6] generates an *e.f.* that corresponds to (1.0.2), with $Z_{ij}(\theta)$ and $f_{ij}(\theta)$ defined in (2.6) of [6]. Further details on the connection between [6] and the analysis of longitudinal data are presented in [6] and Section 2.6 of this dissertation.

Generalizing the marginal model approach to the case of random covariates requires some form of conditioning in (1.0.1).

The simplest situation occurs when conditioning in (1.0.1) is done on the last observed covariate x_{ij} (marginal models). It is shown in [21] that, when some covariates are random and time-dependent, the expectation of the *e.f.* in (1.0.2) may not be zero (the *e.f.* is *biased*) so the *e.f.* may generate estimators that may not be consistent. The main reason is that (1.0.2) contains all components of the covariates $x_{ij}, j = 1, \dots, m_i, i \geq 1$. This situation does not occur under the working independence assumption. For this reason, it is suggested in [21] that the use of *e.f.* with random covariates in this context be restricted to the working independence case, unless a strong condition on independence among covariates is imposed (see condition (5) of [21]). Note that this condition is not needed in [28].

Within the framework of marginal models, Lai and Small in [13] showed that the unrestricted use of the working independence assumption leads to a loss of efficiency in many situations. They classified the random covariates into three types and solved *estimating equation(s)* (EE) appropriate for each type. The resulting blocks of estimating functions are then put together by applying the generalized method of moments.

Type I and type II covariates are those for which there is no “feed-back” from the response process to the covariate process, whereas for type III covariates there is such feed-back. It is only for type III that Lai and Small in [13] advocate the use of

the working independence assumption. As an example of a type III situation, let us consider a study meant to analyze the dependence of an individual's blood pressure (response process) on the individual's amount of salt intake and daily physical exercise (the covariate processes). Some participants may notice the dependence and its direction, *e.g.*, that a lower blood pressure follows a reduction in salt intake. They may change their habits during the study by further reducing the salt intake to further reduce the blood pressure at the time future measurements are recorded. In this example, a reduction in blood pressure on one occasion induces a change in covariates recorded on future occasions.

A similar problem of potential bias was encountered in [6], while attempting to use the GEE method in connection with the recurrence of similar events at the individual level. This problem was not resolved satisfactorily and, consequently, in this dissertation we confine our research to *e.f.* which reflect an underlying working independence assumption. Returning to marginal models with random covariates, we add that the model assumptions (1.0.1) have been relaxed, in that the marginal, conditional variance is no longer required to be related to the conditional mean.

In the context of recurrent events, conditioning in (1.0.1) is done on the history of each individual available right after the $(j - 1)^{th}$ occurrence of the event, for each $j \geq 1$. This type of model, which we use in the dissertation, is a *partially conditional model*. In a fully conditional model, also called transitional model, all covariates, past, present and future are part of the conditioning σ -field.

Having established a connection between conditional models and marginal models in (1.0.1), one can now attempt to use the techniques for GEE developed in [27] and [2]-[3] to explore, in the first place, the statistical properties of the main, regression estimators, within the context of analyzing data from recurrent events process. More precisely, we aim to establish the existence, strong consistency and the asymptotic normality of estimators of the true regression and overdispersion parameters. Before doing so, one must slightly modify the *e.f.* and the GEE approach, to accommo-

date censoring and replace the last, incompletely observed gap time, with imputed data that fits into an appropriate model. It is known that simply discarding the last, incompletely observed gap time leads to biased *e.f.* (see, for instance, [6], [18]). Therefore, the normalized *e.f.* that contain imputed data must be, at the very least, asymptotically unbiased.

An imputation method usually fits a model presumed to govern the mechanism that generates the missing data. In [18], the last term in the main *e.f.*, which defines estimators of the regression parameter, is replaced by its conditional distribution, given the observed event indicating that censoring has occurred. A similar change is made on the *e.f.* defining the estimators of the overdispersion parameter ϕ in (1.0.1), which is denoted σ^2 in [18]. In order to solve this system of equations, one has to ultimately use actual numbers. The authors of [18] use data generated by an independent random variable, which has the same conditional marginal mean and variance as the incomplete gap time. Then they use the EM algorithm to produce the estimators. From their simulation results, these estimators appear to be *a.s.* consistent. Their theoretical method of imputation is based on observed data, namely the occurrence of censoring. The model assumptions that govern the missing data mechanism are not described in detail and there is no formal proof of either unbiasedness or asymptotic unbiasedness of the *e.f.* used in the simulations.

We adopt the initial “conditional” equations in [18] and provide conditions under which these *e.f.* are unbiased (*e.g.*, our Proposition 2.4.6). Condition (A0) is taken from [6]. Our conditions model the missing data mechanism by expressing the degree of independence between some information on the censoring time and the recurrent events process.

A different approach to the imputation of the incomplete gap time is presented in [6]. The authors build a solid mathematical framework to support their work. The contribution of each individual to the original *e.f.*, which contains the last, partially observed gap time (see (2.7) of [6]) is conditioned on the σ -field containing

the individual's observed data. This produces truly conditional *e.f.* (see (2.8)-(2.9) of [6]), which are unbiased, since the original *e.f.* are unbiased under (A0). As remarked in [6], using only the observed data in the *e.f.* to produce estimators points to a *missing at random* (MAR) type of mechanism that generates the nonresponse. Under MAR, the missing data can be generated solely by a subset of the observed data. More about MAR can be found in [16].

While the conditional *e.f.* corresponding to (2.8)-(2.9) in [6] have good unbiasedness properties, the problem of actually imputing numbers in the conditional expectations of functions of the incompletely observed gap times still needs to be solved. It is more complex here, because the conditioning σ -field in [6] is larger than the conditioning σ -field in [18]. The authors of [6] propose a parametric solution to this problem. They assume that the complete distribution generating the two conditional moments in (2.8)-(2.9) is known.

Our approach to conditioning and imputation relies solely on the observed data. As in [18], we start with *e.f.* that contain the last, unknown gap-time, and condition it on a smaller σ -field than the one generated by the entire individual history. We impose a condition that reduces our *e.f.* to those in [18], and prove their unbiasedness. The last terms in the *e.f.* are still unknown. Under (A0), we impute these terms using the observed data from all individuals. Our simulation results show that the estimators that are produced by solving our EE are close to the value of the parameters and have, overall, good statistical properties.

So far we discussed our methodological approach and compared it to that of [6] and [18]. We now compare the mathematical techniques we use in this dissertation with those in [6], [18] and [27].

Once we set the EE with the imputed terms, we implicitly define our estimators and then prove the strong consistency and asymptotic normality of these estimators. We do this in two steps. We first define $\hat{\theta}_n$, a sequence of estimators of the main, regression parameter, by solving an EE that contains θ , but not σ^2 . We recall that

σ^2 is the overdispersion parameter. Next, we place $\hat{\theta}_n$ in the EE that contains (θ, σ^2) , and prove the existence, strong consistency and asymptotic normality of a sequence of estimators $\hat{\sigma}_n^2$ of σ_0^2 . In Section 4.1, we prove the asymptotic normality of the sequence of regression estimators $\hat{\theta}_n$. Then Theorem 4.2.3 gives the asymptotic normal distribution of $\hat{\sigma}_n^2$, taking into account the asymptotic distribution of $\hat{\theta}_n$. We followed the ideas presented in Section 3.5 of [28] and adapted the technical details from Section 3.4.2 of [20]. We recall that a repeated substitution method, similar to the EM algorithm is used in [18], but the analogy with [18] ends here. The imputed part of the *e.f.* in [18] is different from ours. Furthermore, we use the analytical and stochastic properties of the terms of our *e.f.* to discuss the properties of our estimators.

A more appropriate comparison of our mathematical techniques could be made with [6]. The more difficult result is the existence and strong consistency of the estimators. Clement and Strawderman (see [6]) base their theoretical results on a 1998 paper by Yuan and Jennrich (see [29]), while we apply the more modern results in [27], which were generalized in [2]-[3]. While [29] is quite important pedagogically, some of the conditions required are difficult to prove, *e.g.*, conditions (A2)-(A3) listed in [6]. With parameter $\eta^T = (\theta^T, \sigma^2) \in R^{p+1}$, they consider the *e.f.* $S_n(\eta)$ defined in (2.16) of [6]. Condition (A2) requires that $S_n(\eta)$ converge uniformly in η in probability to $S(\eta)$, while condition (A3) requires the same type of convergence of $\frac{d}{d\eta} S_n(\eta)$ to $\frac{d}{d\eta} S(\eta)$. By contrast, our conditions rely on analytical properties of the derivatives $\mathcal{D}_n(\eta)$ of our *e.f.*, $n \geq 1$, without appealing to the existence of a limiting process and any type of convergence to it (see our Theorem 3.1.1). Furthermore, our conditions (*e.g.* S(ii)) can be expressed in terms of simpler random variables that make-up our *e.f.* (see, for instance, our Section 3.2).

While we base our technical approach on [27] and [2]-[3], some differences are worth mentioning. Since we consider independent individuals as opposed to martingales (see [3]), our results on the asymptotic behaviour of the terms of the derivatives

containing the residuals require much weaker conditions than those in [3].

On the other hand, we are dealing here with a more general model. In our case the conditional variance need not be analytically linked with the conditional mean. Furthermore, the function representing this mean cannot be decomposed to contain the scalar product of the vector of parameters and covariates, so we had to introduce and study an additional function. More importantly, we had to develop the technical approach to deal with censoring and the imputed data. We treated the observed and the imputed parts separately. The advantage is that the treatment of the observed part is valid regardless of the imputation method used. To treat the imputed part, we expanded the method we used for the observed part.

This dissertation is organized as follows. Chapter 2 presents the basic assumptions, and our suggested *e.f.* Chapter 3 is dedicated to the strong consistency of our estimators, while their asymptotic normality is discussed in Chapter 4. In Chapter 5 we present and discuss simulation results.

Chapter 2

The Model and Basic Assumptions

In Section 2.1 of this chapter we introduce the model we use and the basic assumptions needed to prove the main results. We follow the formal set-up in [6], but also adopt ideas from [18]. It allows us to use as covariates values of the previous gap times, which might be strong predictors of current and future gap times. Section 2.2 gives a short description of the related literature. In Section 2.3 we introduce various *e.f.*, which are tools for generating estimators of the parameters associated with the conditional model. In Section 2.4, we present our observed *e.f.* and discuss their unbiasedness. In Section 2.5, we propose other *e.f.*, while in Section 2.6 we outline the connection with other longitudinal models. In Section 2.7 we give a preliminary result, while Section 2.8 is devoted to examples.

2.1 Model Assumptions and Examples

In this section we introduce the set-up and the *e.f.* defined in [6], with a slightly different notation. We state the conditional model used and the conditional independence assumptions governing the censoring times.

Data d_i , collected on each subject i , is generated by a distribution indexed by a

true parameter $\eta_0 = (\theta_0^T, \sigma_0^2)^T \in R^{p+1}, i \geq 1$. Viewed as random vectors, $\{d_i\}_{i \geq 1}$ are independent and identically distributed (*i.i.d.*).

Assume that the time of origin for the analysis is $S_{i0} = 0$, with subsequent events occurring at times $0 < S_{i1} < S_{i2} \cdots < S_{ij} < \cdots$, and terminating at an observed censoring time $C_i > 0, i \geq 1$. The observed, uncensored data for subject i at time j , generate the σ -field

$$\mathcal{F}_{ij} = \sigma(S_{i1}, S_{i2}, \cdots, S_{ij}, x_{i1}, \cdots, x_{ij}, x_{i,j+1}), \quad (2.1.1)$$

where x_{ij} denotes the covariate information (available at time $S_{i,j-1}$), $i, j \geq 1$. Therefore, \mathcal{F}_{ij} may include covariates, measured at or before S_{ik} , and events or summaries of past events up to and including time $S_{ik}, k < j, j \geq 1$. Note that $\{\mathcal{F}_{ij}\}$ is a filtration, meaning that $\mathcal{F}_{ij} \subset \mathcal{F}_{i,j+1}, i, j \geq 1$. Let $h(\cdot)$ be a monotone nondecreasing transformation $h : R^+ \rightarrow R$, and define

$$Y_{ij} = h(S_{ij} - S_{i,j-1}),$$

where $S_{ij} - S_{i,j-1}$ is the j^{th} gap time.

We assume throughout that each Y_{ij} has a finite second moment, $j, i \geq 1$. With $\eta = (\theta^T, \sigma)^T \in R^{p+1}$, the basic modeling assumption is:

$$E_{\theta}[Y_{ij} | \mathcal{F}_{i,j-1}] = \mu_{ij}(\theta) \quad \text{and} \quad \text{Var}_{\eta}[Y_{ij} | \mathcal{F}_{i,j-1}] = \sigma^2 V_{ij}^2(\theta), \quad j \geq 1, \quad (2.1.2)$$

where $\mu_{ij}(\theta) \in R$ and $V_{ij}(\theta) > 0$ are known scalar functions of the parameter vector $\theta \in R^p$, and of covariates, which will be displayed in the detailed examples. The model in (2.1.2) specifies conditional marginal means and variances, given the past of each individual. We think of θ as a regression parameter and of σ^2 as the overdispersion parameter.

Remark 2.1.1 *While (2.1.2) holds for possible values of the parameter η , convergence results, which involve the true parameter η_0 hold when the probability measure is P_{η_0} .*

Examples:

1. $h(x) = x$, $E_\theta[Y_{ij}|\mathcal{F}_{i,j-1}] = \mu_{ij}(\theta)$, $Var_\eta[Y_{ij}|\mathcal{F}_{i,j-1}] = \sigma^2 V_{ij}^2(\theta)$,
2. $h(x) = x$, $E_\theta[Y_{ij}|\mathcal{F}_{i,j-1}] = \mu_{ij}(\theta)$, $Var_\eta[Y_{ij}|\mathcal{F}_{i,j-1}] = \sigma^2 \mu_{ij}^2(\theta)$,
3. $h(x) = \log x$, $E_\theta[Y_{ij}|\mathcal{F}_{i,j-1}] = \mu_{ij}(\theta)$, $Var_\eta[Y_{ij}|\mathcal{F}_{i,j-1}] = \sigma^2 V_{ij}^2(\theta)$,

where $i, j \geq 1$.

When no confusion may arise, we omit writing the subscript of E .

The models in 1 and 3 do not specify a connection between $\mu_{ij}(\theta)$ and $V_{ij}(\theta)$. The models in 2 and 3 generalize the accelerated gap times (AGT) model proposed in [24], which assumes that the gap times of the recurrent event process satisfy $S_{ij} - S_{i,j-1} = R_{ij}\mu(\theta)$, where $\{R_{ij}\}_{i,j \geq 1}$ are *i.i.d.* random variables, with distribution independent of θ . Here $\mu(\theta)$ accelerates or decelerates the baseline gap times based on the values of the time-independent covariates. When $E[R_{ij}] = 1$, model 2 is a direct generalization of this AGT model with $\mu_{ij}(\theta) = \mu(\theta)$, $V_{ij}(\theta) = \mu(\theta)$ and $\sigma^2 = Var[R_{ij}]$, $i, j \geq 1$. An alternative version of this model specifies $\log(S_{ij} - S_{i,j-1}) = \log\mu(\theta) + \log R_{ij}$, and is covered by model 3, with $\mu_{ij}(\theta) = \log\mu(\theta)$, $E[\log R_{ij}] = 0$, $\sigma^2 = Var[\log R_{ij}]$ and $V_{ij}(\theta) = 1$, $i, j \geq 1$.

The following example, taken from [18], will often be invoked to illustrate the conditions under which our results hold. In this example $h(x) = x$.

Example 2.1.2 *As in (3.3) of [6], we assume that $\mu_{ij}(\theta)$ and $V_{ij}(\theta)$ are given by*

$$\mu_{ij}(\theta) = \gamma_0 + \gamma_1 \overline{BMI}_{ij} + \frac{\rho}{\rho(j-1) + 1 - \rho} \left[\sum_{l=1}^{j-1} Y_{il} - \sum_{l=1}^{j-1} (\gamma_0 + \gamma_1 \overline{BMI}_{il}) \right], \quad (2.1.3)$$

$$V_{ij}(\theta) = \left(\left| 1 + \frac{\rho}{\rho(j-1) + 1 - \rho} \right| \right)^{1/2}, \quad \text{or} \quad V_{ij}(\theta) = |\mu_{ij}(\theta)|, \quad (2.1.4)$$

where $\theta^T = (\gamma_0, \gamma_1, \rho)$. In (2.1.3), \overline{BMI}_{ij} represents an average of several measurements of the Body Mass Index taken on individual i at the $(j-1)^{th}$ occurrence of the event. We note that the constant 28 in [6] was incorporated in γ_0 .

The model in Example 2.1.2, where the conditional variance is given by the first formula in (2.1.4), constitutes our main example.

Remark 2.1.3 *Example 2.1.2 illustrates a major difference between the marginal models introduced in [14] and the conditional models studied here. While for marginal models, each marginal mean and variance can be expressed as a function of a linear combination of covariates and parameters, the conditional mean in (2.1.3), though linear in covariates, is not a function of a scalar product of the parameter vector and the vector of covariates.*

The following example of a more general conditional model serves as the basis for our theoretical results.

Example 2.1.4 *In this example, only the conditional mean in (2.1.2) is specified, and it is of the form*

$$\mu_{ij}(\theta) = \mu(c_{ij}^T(\theta)x_{ij}), \quad (2.1.5)$$

where $\mu : R \rightarrow R$, $c_{ij}(\theta) \in R^q$ are nonrandom and $x_{ij}(\omega) \in R^q$ is the random vector of covariates available before the occurrence of the j^{th} event. The function μ , and the components $c_{ij,h}$, $h = 1, 2, \dots, q$ are three times continuously differentiable.

Remark 2.1.5 *The number of covariates generally increases with $j \geq 1$. We assume that there is an upper bound q for this number.*

Example 2.1.6 *This is the generalization of the (AGT) model described earlier. The gap times are modeled as $Y_{ij} = R_{ij} \exp(-\theta^T x_i)$, where x_i is a baseline vector of covariates that is not time dependent and is thus \mathcal{F}_{ij} -measurable for any $j \geq 1, i \geq 1$. The R_{ij} are independent in both i and j and also of the covariates, and $E[R_{ij}] = 1$. Then*

$$E[Y_{ij} | \mathcal{F}_{i,j-1}] = E[R_{ij}] \exp(-\theta^T x_i) = \exp(-\theta^T x_i) := \mu_i(\theta),$$

$$E[Y_{ij}^2 | \mathcal{F}_{i,j-1}] = E[R_{ij}^2] \mu_i^2(\theta),$$

hence

$$\text{Var}[Y_{ij}|\mathcal{F}_{i,j-1}] = (E[R_{ij}^2] - 1)\mu_i^2(\theta) := \sigma^2\mu_i^2(\theta),$$

since by definition $\text{Var}[R_{ij}] = E[R_{ij}^2] - 1 := \sigma^2$. It follows that $V_{ij}(\theta)$ in (2.1.2) equals $\mu_i(\theta)$. We have $\mu(x) = \exp(-\theta^T x)$, $x \in R^q$.

In Example 2.1.7 below, we show that $\mu_{ij}(\theta)$ in Example 2.1.2 is of the form (2.1.5).

Example 2.1.7 Let $\mu(t) = t$, $t \in R$, $\theta^T = (\gamma_0, \gamma_1, \rho)$, and $c_{ij}^T(\theta)$ and x_{ij} are q -dimensional vectors, with components:

$$c_{ij,h}(\theta) = \begin{cases} \gamma_0 - \frac{(j-1)\gamma_0\rho}{\rho(j-1)+1-\rho} & \text{if } h = 1, \\ \frac{-\rho\gamma_1}{\rho(j-1)+1-\rho} & \text{if } 1 < h \leq j, \\ \gamma_1 & \text{if } h = j + 1, \\ \frac{\rho}{\rho(j-1)+1-\rho} & \text{if } j + 1 < h \leq 2j, \\ 0 & 2j < h \leq q. \end{cases}$$

$$x_{ij,h} = \begin{cases} 1 & \text{if } h = 1, \\ \overline{BMI}_{i,h-1} & \text{if } 1 < h \leq j, \\ \overline{BMI}_{ij} & \text{if } h = j + 1, \\ Y_{i,h-(j+1)} & \text{if } j + 1 < h \leq 2j, \\ 0 & 2j < h \leq q. \end{cases}$$

Then (2.1.3) can be expressed as $\mu_{ij}(\theta) = c_{ij}^T(\theta)x_{ij}$. Note that $h = 1$ gives the intercept. Indeed,

$$\begin{aligned} c_{ij}^T(\theta)x_{ij} &= \gamma_0 - \frac{(j-1)\gamma_0\rho}{\rho(j-1)+1-\rho} + \sum_{h=2}^j \frac{-\rho\gamma_1}{\rho(j-1)+1-\rho} \overline{BMI}_{i,h-1} \\ &+ \gamma_1 \overline{BMI}_{ij} + \sum_{h=j+2}^{2j} \frac{\rho}{\rho(j-1)+1-\rho} Y_{i,h-(j+1)} \end{aligned}$$

$$\begin{aligned}
&= \gamma_0 - \frac{\rho(j-1)\gamma_0}{\rho(j-1)+1-\rho} - \frac{\rho}{\rho(j-1)+1-\rho} \sum_{l=1}^{j-1} \gamma_1 \overline{BMI}_{il} \\
&+ \gamma_1 \overline{BMI}_{ij} + \frac{\rho}{\rho(j-1)+1-\rho} \sum_{l=1}^{j-1} Y_{il} \\
&= \gamma_0 + \gamma_1 \overline{BMI}_{ij} + \frac{\rho}{\rho(j-1)+1-\rho} \left[\sum_{l=1}^{j-1} Y_{il} - \sum_{l=1}^{j-1} (\gamma_0 + \gamma_1 \overline{BMI}_{il}) \right].
\end{aligned}$$

We note that the sum on the right hand side above looks like a sum of linear regression residuals, where

$$\varepsilon_{il} = (Y_{il} - \gamma_0 - \gamma_1 \overline{BMI}_{il}), \quad 1 \leq l \leq j-1.$$

As in [6], we adopt the following notation:

$$f_{ij}(\theta) = \frac{\partial \mu_{ij}(\theta)}{\partial \theta} V_{ij}^{-1}(\theta) \quad \text{and} \quad Z_{ij}(\theta) = \frac{Y_{ij} - \mu_{ij}(\theta)}{V_{ij}(\theta)}. \quad (2.1.6)$$

We assume that $E_{\theta} Z_{ij}^2(\theta) < \infty$, for all $j \geq 1$. Note that $f_{ij}(\theta)$ is a p -dimensional vector. From (2.1.2), it follows that $f_{ij}(\theta)$ is $\mathcal{F}_{i,j-1}$ -measurable.

We now define a filtration larger than $\{\mathcal{F}_{ij}\}$, $i, j \geq 1$, which contains some information about censoring times C_i , $i \geq 1$.

Let

$$\mathcal{G}_{ij} = \sigma(\mathcal{F}_{ij}, S_{ik} \leq C_i, \{S_{ik} = C_i\}, k = 1, 2, \dots, j), \quad i, j \geq 1. \quad (2.1.7)$$

The *i.i.d.* vectors d_i can now be written $d_i = \{S_{ij}, x_{ij}, C_i, j \geq 1\}$, where x_{ij} are random covariates. As in [6], we make the following basic assumptions:

$$(A0) \quad E_{\theta}[Y_{ij} | \mathcal{G}_{i,j-1}] = E_{\theta}[Y_{ij} | \mathcal{F}_{i,j-1}] \quad \text{and} \quad \text{Var}_{\eta}[Y_{ij} | \mathcal{G}_{i,j-1}] = \text{Var}_{\eta}[Y_{ij} | \mathcal{F}_{i,j-1}] \quad i, j \geq 1.$$

Remark 2.1.8 *Assumption (A0) holds if, conditional on $\mathcal{F}_{i,j-1}$, the sigma fields generated by $\{S_{ik} \leq C_i\}, \{S_{ik} = C_i\}, 1 \leq k < j$, and Y_{ij} , are independent, for all $i \geq 1$.*

We now give some insight into the meaning of (A0) (see Exercise 34.11 in [4]).

First, we recall that $S_{i0} = 0, i \geq 1$ and note that $C_i > 0$, for all individuals i , which are part of the survey.

Example 2.1.9 Consider Example 2.1.7 and let us examine (A0) for $j = 1, 2$. At $S_{i0} = 0, x_{i1}$ is known, so, by convention, we set $\mathcal{F}_{i0} := \sigma(x_{i1})$. At $j = 1$, the only past information on the position of C_i is $C_i > 0$, which is the entire probability space of events. Therefore, no new information is contained in \mathcal{G}_{i0} , $\mathcal{F}_{i0} = \mathcal{G}_{i0}$ and (A0) holds for $j = 1$. Consider now $j = 2$, and let us examine \mathcal{F}_{i1} and \mathcal{G}_{i1} . We have $\mathcal{F}_{i1} = \sigma(x_{i1}, x_{i2}, S_{i1})$ and $\mathcal{G}_{i1} = \sigma(\mathcal{F}_{i1}, \{S_{i1} \leq C_i\}, \{S_{i1} = C_i\})$. For $j = 2$, (A0) tells us that, in predicting the value of Y_{i2} , past information, given by the covariate and response process, prevails over some information on the position of the censoring times, relative to the time of occurrence of the previous event. In other words, the information provided by $\sigma(\{S_{i1} \leq C_i\}, \{S_{i1} = C_i\})$ is irrelevant in the presence of \mathcal{F}_{i1} . A similar interpretation holds for $j > 2$.

2.2 Short Description of Previous Work

We now give a short description of the estimating methodologies used in [6] and [18].

As in [6], we define

$$N_i := \max\{j \geq 1, S_{ij} \leq C_i\}, \quad i \geq 1,$$

which is the number of events up to and including the censoring time $C_i, i \geq 1$.

We introduce the $(p + 1)$ -dimensional vector (2.7) of [6], which constitutes the following set of e.f.:

$$G_{n,1}(\theta) := \sum_{i=1}^n \sum_{j=1}^{N_i+1} f_{ij}(\theta) Z_{ij}(\theta), \quad G_{n,2}(\eta) := \sum_{i=1}^n \sum_{j=1}^{N_i+1} b_{ij}(\eta) (Z_{ij}^2(\theta) - \sigma^2), \quad (2.2.1)$$

where $b_{ij}(\eta)$ is an $\mathcal{F}_{i,j-1}$ -measurable scalar function, which serves as weight. We assume that all terms in (2.2.1) are P_η -integrable.

In [6], data observed on subject i generate the σ -field:

$$\mathbb{O}_i = \sigma(S_{i1}, S_{i2}, \dots, S_{iN_i}, x_{i1}, x_{i2}, \dots, x_{i,N_i+1}, C_i).$$

The *e.f.* (2.2.1) cannot be used directly for estimating the true parameter η_0 from the observed data, since, for each individual, the last terms in (2.2.1) are unobserved.

As in [6], define the transform:

$$W_{ij}(\eta) = \sigma^{-1} Z_{ij}(\theta), i, j \geq 1,$$

and the normalized *e.f.*

$$S_{n,1}(\eta) = \frac{1}{n} \sum_{i=1}^n \left[\sum_{j=1}^{N_i} f_{ij}(\theta) W_{ij}(\eta) + f_{i,N_i+1}(\theta) K_1(\omega_i(\eta)) \right], \quad (2.2.2)$$

$$S_{n,2}(\eta) = \frac{1}{n} \sum_{i=1}^n \left[\sum_{j=1}^{N_i} b_{ij}(\eta) (W_{ij}^2(\eta) - 1) + b_{i,N_i+1}(\eta) (K_2(\omega_i(\eta)) - 1) \right], \quad (2.2.3)$$

where

$$K_r(\omega_i(\eta)) = E_\eta[W_{i,N_i+1}^r(\eta) I\{W_{i,N_i+1}(\eta) > \omega_i(\eta)\} | \mathbb{O}_i], \quad r = 1, 2. \quad (2.2.4)$$

$$\omega_i(\eta) = \frac{h(C_i - S_{iN_i}) - \mu_{i,N_i+1}(\theta)}{\sigma V_{i,N_i+1}(\theta)}.$$

In (2.2.4), $K_r(\omega) = \int_\omega^\infty \frac{u^r}{1-F_0(\omega-)} dF_0(u)$ $r = 1, 2$, where it is assumed that the conditional distribution of $W_{i,N_i+1}(\eta)$, given some of the observed data, is fully specified by $F_0(\cdot)$.

It is shown in [6] that the *e.f.* in (2.2.1) are *unbiased*, *i.e.*,

$$E_\theta[G_{n,1}(\theta)] = 0, \quad E_\eta[G_{n,2}(\eta)] = 0, \quad n \geq 1.$$

This is also true of (2.2.2)-(2.2.3), which can be viewed as normalized projections of (2.2.1) on the observed data. This parametric approach requires that the distribution F_0 be known.

The approach to this problem is slightly different in [18]. With $h(x) \equiv x$, the *e.f.* used to estimate the main regression parameter θ is :

$$\sum_{i=1}^n \sum_{j=1}^{\infty} f_{ij}(\theta) Z_{ij}(\theta) I\{S_{i,j-1} < C_i\}. \quad (2.2.5)$$

The projection of this *e.f.* on the observed data is given in [18] as

$$\sum_{i=1}^n \sum_{j=1}^{\infty} f_{ij}(\theta) V_{ij}^{-1}(\theta) (E_{\theta}[Y_{ij}|obs] - \mu_{ij}(\theta)) I\{S_{i,j-1} < C_i\}, \quad (2.2.6)$$

where $E_{\theta}[Y_{ij}|obs] = Y_{ij}$, if $\sum_{l=1}^j Y_{il} \leq C_i$ and

$$E_{\theta}[Y_{ij}|obs] = E_{\theta}[Y_{ij}|Y_{ij} > C_i - \sum_{l=1}^{j-1} Y_{il}], \quad \text{otherwise.}$$

In [18], two EE (one of which is generated by (2.2.6) above and the other by (2) in [18]) are solved for θ and σ^2 using an idea from the EM algorithm.

The *e.f.* in (2.2.6) is not the projection of an unbiased *e.f.* on a σ -field. It is only the gap times Y_i that are projected on the σ -field generated by the knowledge that censoring has occurred, an idea that we borrow in defining the *e.f.* we need. The *e.f.* in (2.2.6) is, in general, not unbiased, unless additional conditions are imposed on modeling the data.

2.3 Unbiased Estimating Functions

In this section we introduce the observed *e.f.*, which will be used throughout, and we study their unbiasedness properties.

From here on, we assume that (A0) holds. We often appeal to the strong law of large numbers (SLLN) for the *i.i.d.* case, which is Theorem 22.1 of [4]. We always assume that the hypotheses of Theorem 22.1 hold, *i.e.*, the appropriate random variables have finite expectations.

Recalling (2.2.1) and the conditions on its terms, we consider now the *e.f.*:

$$\begin{aligned} g_{n,1}(\theta) &= \sum_{i=1}^n \sum_{j=1}^{\infty} f_{ij}(\theta) Z_{ij}(\theta) I\{S_{i,j-1} < C_i\}, \\ g_{n,2}(\eta) &= \sum_{i=1}^n \sum_{j=1}^{\infty} b_{ij}(\eta) [Z_{ij}^2(\theta) - \sigma^2] I\{S_{i,j-1} < C_i\}, \end{aligned} \quad (2.3.1)$$

where we assume that the second sums are *a.s.* finite.

The definition of a stopping time is given in (7.18) of [4]. In our set-up, we have

Definition 2.3.1 *Let $\tau_i = \min\{j \geq 1 : S_{ij} \geq C_i\}$, if such j exists, and $\tau_i = \infty$ otherwise.*

We assume throughout that $\tau_i < \infty$ *a.s.*

Proposition 2.3.2 *For every i , τ_i is a stopping time with respect to the filtration $\{\mathcal{G}_{ij}\}_{j \geq 0}$.*

Proof. We have $\{\tau_i = j\} = \{S_{i,j-1} < C_i \leq S_{ij}\} \in \mathcal{G}_{ij}, j \geq 1$. \square

Now (2.3.1) can be written :

$$\begin{aligned} g_{n,1}(\theta) &= \sum_{i=1}^n \sum_{j=1}^{\tau_i} f_{ij}(\theta) Z_{ij}(\theta), \\ g_{n,2}(\eta) &= \sum_{i=1}^n \sum_{j=1}^{\tau_i} b_{ij}(\eta) (Z_{ij}^2(\theta) - \sigma^2). \end{aligned} \quad (2.3.2)$$

Remark 2.3.3 *We have that $\tau_i \leq N_i + 1$. As in [6], our last gap time Y_{ij} could be fully observed. On the other hand, we ignore $f_{i,N_i+1}(\theta) Z_{i,N_i+1}(\theta)$ when $S_{N_i} = C_i$. In other words, we only account for gap times that are at least partially observed. This is also the approach in [18].*

We consider the following conditions:

(T0) $E[\tau_1] < \infty$,

(T1) For each $i \geq 1$, we assume that τ_i is bounded, *i.e.*, there exists a nonrandom integer m_i such that $\tau_i \leq m_i$, *a.s.*

(T2) There exists a non-random integer m , such that $\tau_i \leq m$, for all $i \geq 1$.

Note that (T2) \Rightarrow (T1) \Rightarrow (T0).

Remark 2.3.4 Assume that $P\{\tau_1 > 0\} = 1$. Then condition (T0), which is the weakest of the three conditions above, implies that $0 < \tau_i < \infty$ *a.s.*, $i \geq 1$. Indeed,

$$E[\tau_1] = E[\tau_1 I\{\tau_1 < \infty\}] + E[\tau_1 I\{\tau_1 = \infty\}] < \infty,$$

implies that $P\{\tau_1 = \infty\} = 0$.

Under (T1), our *e.f.* in (2.3.1) become:

$$g_{n,1}(\theta) = \sum_{i=1}^n \sum_{j=1}^{m_i} f_{ij}(\theta) Z_{ij}(\theta) I\{S_{i,j-1} < C_i\}, \quad (2.3.3)$$

$$g_{n,2}(\eta) = \sum_{i=1}^n \sum_{j=1}^{m_i} b_{ij}(\eta) (Z_{ij}^2(\theta) - \sigma^2) I\{S_{i,j-1} < C_i\}, \quad (2.3.4)$$

where the terms with $j > \tau_i$ are zero.

Remark 2.3.5 Condition (T2) is not unreasonable. In practice, there is always a time the entire study ends, say, for lack of funds. Furthermore, for many recurrent events processes, there is a minimum amount of time separating consecutive events. In such cases, the sequence of event times does not accumulate and is bounded, for all individuals.

As in [6], we have the following result.

Proposition 2.3.6 Under assumptions (A0) and (T1), the *e.f.* in (2.3.3)-(2.3.4) are unbiased. If (T2) holds, then

$$n^{-1} g_{n,i}(\eta_0) \rightarrow 0 \quad \text{a.s. in } P_{\eta_0}, \quad i = 1, 2. \quad (2.3.5)$$

Proof.

$$\begin{aligned}
E_\theta[g_{n,1}(\theta)] &= \sum_{i=1}^n \sum_{j=1}^{m_i} E_\theta[f_{ij}(\theta)Z_{ij}(\theta)I\{S_{i,j-1} < C_i\}] \\
&= \sum_{i=1}^n \sum_{j=1}^{m_i} E_\theta[E_\theta[f_{ij}(\theta)Z_{ij}(\theta)I\{S_{i,j-1} < C_i\}|\mathcal{G}_{i,j-1}]] \\
&= \sum_{i=1}^n \sum_{j=1}^{m_i} E_\theta[f_{ij}(\theta)I\{S_{i,j-1} < C_i\}E_\theta[Z_{ij}(\theta)|\mathcal{G}_{i,j-1}]] \\
&= \sum_{i=1}^n \sum_{j=1}^{m_i} E_\theta[f_{ij}(\theta)I\{S_{i,j-1} < C_i\}E_\theta[Z_{ij}(\theta)|\mathcal{F}_{i,j-1}]] \\
&= 0.
\end{aligned}$$

The third equality holds because $f_{ij}(\theta)I\{S_{i,j-1} < C_i\}$ is $\mathcal{G}_{i,j-1}$ -measurable, the fourth by assumption (A0) and the last by (2.1.2).

Similarly,

$$\begin{aligned}
E_\eta[g_{n,2}(\eta)] &= \sum_{i=1}^n \sum_{j=1}^{m_i} E_\eta[b_{ij}(\eta)(Z_{ij}^2(\theta) - \sigma^2)I\{S_{i,j-1} < C_i\}] \\
&= \sum_{i=1}^n \sum_{j=1}^{m_i} E_\eta[E_\eta[b_{ij}(\eta)(Z_{ij}^2(\theta) - \sigma^2)I\{S_{i,j-1} < C_i\}|\mathcal{G}_{i,j-1}]] \\
&= \sum_{i=1}^n \sum_{j=1}^{m_i} E_\eta[b_{ij}(\eta)I\{S_{i,j-1} < C_i\}E_\eta[Z_{ij}^2(\theta) - \sigma^2|\mathcal{G}_{i,j-1}]] \\
&= \sum_{i=1}^n \sum_{j=1}^{m_i} E_\eta[b_{ij}(\eta)I\{S_{i,j-1} < C_i\}E_\eta[Z_{ij}^2(\theta) - \sigma^2|\mathcal{F}_{i,j-1}]] \\
&= 0.
\end{aligned}$$

As before, the third equality holds because $b_{ij}(\eta)I\{S_{i,j-1} < C_i\}$ is $\mathcal{G}_{i,j-1}$ -measurable, the fourth by assumption (A0) and the last by (2.1.2).

To prove (2.3.5), it suffices to show that, for each $j \leq m$, we have, *a.s.* in P_{η_0} ,

$$n^{-1} \sum_{i=1}^n f_{ij}(\theta_0)Z_{ij}(\theta_0)I\{S_{i,j-1} < C_i\} \rightarrow 0, \quad (2.3.6)$$

$$n^{-1} \sum_{i=1}^n b_{ij}(\eta_0)[Z_{ij}^2(\theta_0) - \sigma_0^2]I\{S_{i,j-1} < C_i\} \rightarrow 0. \quad (2.3.7)$$

This follows from the SLLN, since the expectation of each term in (2.3.6)-(2.3.7) is zero, by the proof of the first statement of the proposition. \square

Remark 2.3.7 *In the course of the proof of Proposition 2.3.6, we showed that $E_\theta[Z_{ij}(\theta)|\mathcal{G}_{i,j-1}] = 0$ and $E_\eta[Z_{ij}^2(\theta) - \sigma^2|\mathcal{G}_{i,j-1}] = 0, j, i \geq 1$. Thus, in (2.3.3)-(2.3.4), each sum in j forms a martingale difference.*

We introduce some notation. For $i, j \geq 1$, consider the set $\{S_{i,j-1} < C_i < S_{ij}\} = \{S_{ij} - S_{i,j-1} > C_i - S_{i,j-1} > 0\}$ and note that it belongs to \mathcal{G}_{ij} . The complement of this set is $\{S_{i,j-1} \geq C_i\} \cup \{S_{ij} \leq C_i\} = \{C_i - S_{i,j-1} \leq 0\} \cup \{S_{ij} - S_{i,j-1} \leq C_i - S_{i,j-1}\}$, *i.e.*, it is a union of two disjoint sets. We define the set indicators:

$$I_{ij}^{obs} := I\{S_{ij} \leq C_i\}, \quad I_{ij}^{cen} := I\{S_{i,j-1} < C_i < S_{ij}\}, \quad I_{ij}^{out} := I\{S_{i,j-1} \geq C_i\}.$$

Here “obs” stands for to fully observed gap times, “cen” for censored gap times, and “out” for fully unobserved gap times. Note that I_{ij}^{out} is $\mathcal{G}_{i,j-1}$ -measurable and the other two set indicators are \mathcal{G}_{ij} -measurable.

Since $\{S_{i,j-1} < C_i\} = \{S_{ij} \leq C_i\} \cup \{S_{i,j-1} < C_i < S_{ij}\}$, we can express the *e.f.* in (2.3.1) as

$$g_{n,1}(\theta) = \sum_{i=1}^n \sum_{j=1}^{\infty} [f_{ij}(\theta)Z_{ij}(\theta)I_{ij}^{obs} + f_{ij}(\theta)Z_{ij}(\theta)I_{ij}^{cen}], \quad (2.3.8)$$

$$g_{n,2}(\eta) = \sum_{i=1}^n \sum_{j=1}^{\infty} b_{ij}(\eta)(Z_{ij}^2(\theta) - \sigma^2)(I_{ij}^{obs} + I_{ij}^{cen}). \quad (2.3.9)$$

In the expressions above, it is the terms restricted to $I_{ij}^{cen}, i, j \geq 1$ that have to be imputed, using the observed data.

Let

$$g_{n,1}^{imputed}(\theta) := \sum_{i=1}^n \sum_{j=1}^{\infty} f_{ij}(\theta)[Z_{ij}(\theta)I_{ij}^{obs} + Z_{ij}^{imp}(\theta)I_{ij}^{cen}], \quad (2.3.10)$$

$$g_{n,2}^{imputed}(\eta) := \sum_{i=1}^n \sum_{j=1}^{\infty} b_{ij}(\eta)\{Z_{ij}^2(\theta)I_{ij}^{obs} + [Z_{ij}^{imp}(\theta)]^2 I_{ij}^{cen} - \sigma^2(1 - I_{ij}^{out})\}, \quad (2.3.11)$$

where in defining $Z_{ij}^{imp}(\theta)$ we will be using the observed data. This is done in Section 2.4.

2.4 Observed Estimating Functions

In this section we describe a three-step procedure for imputing and estimating the last terms of (2.3.10)-(2.3.11). We also discuss the unbiasedness properties of the resulting *e.f.* We start with (2.3.10).

The first step consists of replacing $Z_{ij}^{imp}(\theta)$ by $\frac{E_\theta[Z_{ij}(\theta)I_{ij}^{cen}]}{E_\theta[I_{ij}^{cen}]}$ in (2.3.10) and showing that, under certain conditions, the resulting $g_{n,1}^{obs}(\theta)$ is unbiased.

The second step consists in showing that

$$E_\theta[Z_{ij}(\theta)I_{ij}^{cen}] = -E_\theta[Z_{ij}(\theta)I_{ij}^{obs}], i, j \geq 1,$$

which justifies the definition of the *e.f.* $\hat{g}_{n,1}^{obs}(\theta)$ in (2.4.2) below.

Finally, we show that $g_{n,1}^{obs}(\theta)$ and $\hat{g}_{n,1}^{obs}(\theta)$ defined below are, in some sense, asymptotically equivalent.

First, we write

$$g_{n,1}^{obs}(\theta) := \sum_{i=1}^n \sum_{j=1}^{m_i} \left[f_{ij}(\theta) Z_{ij}(\theta) I_{ij}^{obs} + f_{ij}(\theta) I_{ij}^{cen} \frac{E_\theta[Z_{ij}(\theta)I_{ij}^{cen}]}{E_\theta[I_{ij}^{cen}]} \right]. \quad (2.4.1)$$

Let us now define an *empirical e.f.*, where we subtract only terms with $E_\theta[I_{ij}^{cen}] > 0$.

$$\hat{g}_{n,1}^{obs}(\theta) := \sum_{i=1}^n \sum_{j=1}^{m_i} \left[f_{ij}(\theta) Z_{ij}(\theta) I_{ij}^{obs} - f_{ij}(\theta) I_{ij}^{cen} \frac{\sum_{k=1}^n Z_{kj}(\theta) I_{kj}^{obs}}{\sum_{k=1}^n I_{kj}^{cen}} \right]. \quad (2.4.2)$$

We note that $\hat{g}_{n,1}^{obs}(\theta)$ is used in our simulations in Chapter 5.

For estimating σ , we consider the following *e.f.*:

$$g_{n,2}^{obs}(\eta) = \sum_{i=1}^n \sum_{j=1}^{m_i} \left\{ b_{ij}(\eta) (Z_{ij}^2(\theta) - \sigma^2) I_{ij}^{obs} + b_{ij}(\eta) I_{ij}^{cen} \frac{E_\eta[(Z_{ij}^2(\theta) - \sigma^2) I_{ij}^{cen}]}{E_\eta I_{ij}^{cen}} \right\}, \quad (2.4.3)$$

$$\hat{g}_{n,2}^{obs}(\eta) = \sum_{i=1}^n \sum_{j=1}^{m_i} b_{ij}(\eta) \left\{ (Z_{ij}^2(\theta) - \sigma^2) I_{ij}^{obs} - I_{ij}^{cen} \frac{\sum_{k=1}^n (Z_{kj}^2(\theta) - \sigma^2) I_{kj}^{obs}}{\sum_{k=1}^n I_{kj}^{cen}} \right\}. \quad (2.4.4)$$

When $b_{ij}(\eta) \equiv 1, i, j \geq 1$, (2.4.4) becomes:

$$\hat{g}_{n,2}^{obs}(\eta) = \sum_{i=1}^n \sum_{j=1}^{m_i} \left\{ (Z_{ij}^2(\theta) - \sigma^2) I_{ij}^{obs} - I_{ij}^{cen} \frac{\sum_{k=1}^n (Z_{kj}^2(\theta) - \sigma^2) I_{kj}^{obs}}{\sum_{k=1}^n I_{kj}^{cen}} \right\}. \quad (2.4.5)$$

Remark 2.4.1 *Since*

$$-\sum_{i=1}^n I_{ij}^{cen} \frac{\sum_{k=1}^n (Z_{kj}^2(\theta) - \sigma^2) I_{kj}^{obs}}{\sum_{k=1}^n I_{kj}^{cen}} = -\sum_{k=1}^n (Z_{kj}^2(\theta) - \sigma^2) I_{kj}^{obs},$$

the e.f. $\hat{g}_{n,2}^{obs}(\eta) = 0$ and hence cannot be used in the estimation of σ^2 when $b_{ij}(\eta) \equiv 1$.

We give in Section 2.7 a direct proof of the consistency of the estimators of σ^2 , when $b_{ij}(\eta) \equiv 1, i, j \geq 1$.

Remark 2.4.2 *When $b_{ij}(\eta) \equiv 1, i, j \geq 1$ for some $\eta \in R^{p+1}$, we can replace the e.f. in (2.4.5) by:*

$$\hat{g}_{n,2}^{sim}(\eta) := \sum_{i=1}^n \sum_{j=1}^{m_i} \left\{ (Z_{ij}^2(\theta) - \sigma^2) I_{ij}^{obs} - I_{ij}^{cen} \frac{\sum_{k=1}^n (Z_{kj}^2(\theta) - \sigma^2) I_{kj}^{obs}}{\sum_{k=1, k \neq i}^n I_{kj}^{cen}} \right\}. \quad (2.4.6)$$

It is seen that $\hat{g}_{n,2}^{sim}(\eta)$ is not zero and it has the same asymptotic properties as $\hat{g}_{n,2}^{obs}(\eta)$. While $\hat{g}_{n,2}^{sim}(\eta)$ may not satisfy all the conditions required for our theoretical results, it gives good results in our simulations.

We now turn our attention to (2.4.1) and its estimator (2.4.2), justify their definition and study their unbiasedness properties.

We proceed with the first step in proving the unbiasedness of (2.4.1). Following an idea in [6], we define a sigma-field, for each $i \geq 1$:

$$\mathbb{O}(f_i(\theta)) := \sigma(f_{ij}(\theta) I_{ij}^{cen}, I_{ij}^{cen}, j \geq 1). \quad (2.4.7)$$

We recall that $I_{ij}^{cen} I_{ij'}^{cen} = 0$ if $j \neq j'$.

We introduce a new condition $(B_{f(\theta)})$:

$$(B_{f(\theta)}) \quad I_{ij}^{cen} E_{\theta}[Z_{ij}(\theta) | \mathbb{O}(f_i(\theta))] = I_{ij}^{cen} E_{\theta}[Z_{ij}(\theta) | I_{ij}^{cen}], \quad i, j \geq 1.$$

We illustrate next condition $(B_{f(\theta)})$ for $j = 1, 2$, and compare it to condition (A0), described in Example 2.1.9. We refer to examples 2.1.2 and 2.1.7. For simplicity, we introduce the notation

$$\overline{BMI}_{ij} := x_{ij,j+1}, \quad \sum_{l=1}^{j-1} Y_{il} := Y_i^{(j-1)}, \quad \sum_{h=2}^j x_{ij,h} := x_i^{(j)},$$

$$f_j(\rho) := \frac{1}{j-2+\rho^{-1}}, \quad \rho \neq 0, 1; i, j \geq 1. \quad (2.4.8)$$

Then (2.1.3) can be written, for $\rho \neq 0, 1$

$$\mu_{ij}(\theta) = \gamma_0 + \gamma_1 x_{ij,j+1} + f_j(\rho)[Y_i^{(j-1)} - (j-1)\gamma_0 - \gamma_1 x_i^{(j)}].$$

Example 2.4.3 Consider examples 2.1.2 and 2.1.7, where $V_{ij}(\theta)$ is given by the first formula in (2.1.4). We first identify the generators of $\sigma(f_{ij}(\theta))$, $i, j \geq 1$. Since V_{ij} is nonrandom in this case, we only need to look at $\frac{\partial \mu_{ij}(\theta)}{\partial \theta}$. We have

$$\begin{aligned} \frac{\partial \mu_{ij}(\theta)}{\partial \gamma_0} &= 1 - (j-1)f_j(\rho), \\ \frac{\partial \mu_{ij}(\theta)}{\partial \gamma_1} &= x_{ij,j+1} - f_j(\rho)x_i^{(j)}, \\ \frac{\partial \mu_{ij}(\theta)}{\partial \rho} &= \frac{df_j(\rho)}{d\rho}[Y_i^{(j-1)} - (j-1)\gamma_0 - \gamma_1 x_i^{(j)}]. \end{aligned}$$

So the generators of $\sigma(\frac{\partial \mu_{ij}(\theta)}{\partial \theta})$ are $x_{ij,j+1} - f_j(\rho)x_i^{(j)}$ and $Y_i^{(j-1)} - \gamma_1 x_i^{(j)}$, which are both $\mathcal{F}_{i,j-1}$ -measurable. Consider the case $j = 1$. Since $x_i^{(1)}$ and $Y_i^{(0)}$ are not defined, we take only $x_{i1,2} = \overline{BM\bar{I}}_{i1}$ as generator of $\sigma(\frac{\partial \mu_{i1}(\theta)}{\partial \theta})$. Condition $(B_{f(\theta)})$ tells us that, for $j = 1$,

$$I_{i1}^{cen} E[Z_{i1}(\theta) | \sigma(x_{i1,2} I_{i1}^{cen}, I_{i1}^{cen})] = I_{i1}^{cen} E[Z_{i1}(\theta) | I_{i1}^{cen}],$$

while (A0) imposes no restriction (see Example 2.1.9). In other words, if censoring happens before S_{i1} , the incomplete value of $S_{i1} - S_{i0}$ would be influenced by the occurrence of censoring, regardless of $x_{i1,2} I_{i1}^{cen}$. For $j = 2$, we recall from Example 2.1.9 that, according to (A0), in the presence of \mathcal{F}_{i1} , information of the type $\{C_i < S_{i1}\}, \{C_i = S_{i1}\}, \{C_i > S_{i1}\}$ can be discarded in predicting the future gap time Z_{i2} . Let us now interpret $(B_{f(\theta)})$ in this case. The generators of $\sigma(\frac{\partial \mu_{i2}(\theta)}{\partial \theta})$ are $x_{i2,3} - f_2(\rho)x_{i2,2} = \overline{BM\bar{I}}_{i2} - f_2(\rho)\overline{BM\bar{I}}_{i1}$ and $Y_{i1} - \gamma_1 x_{i2,2} = Y_{i1} - \gamma_1 \overline{BM\bar{I}}_{i1}$, both $\mathcal{F}_{i,1}$ -measurable. Thus, $\sigma(f_{i2}(\theta)) \subset \mathcal{F}_{i1}$ i.e., $f_{i2}(\theta)$ contains only partial information from the history of the recurrent events process. The event $\{S_{i1} < C_i < S_{i2}\}$, which,

for $j = 2$, does not belong to $\mathcal{G}_{i,j-1}$, tells us that censoring has “just” occurred. Condition $(B_{f(\theta)})$ tells us here that such current information about censoring prevails over the incomplete information on the history of the process provided by $\sigma(f_{i2}(\theta)I_{i2}^{cen})$ in influencing the value of $Z_{i2}(\theta)$.

Remark 2.4.4 *Examples 2.1.9 and 2.4.3 show that conditions (A0) and $(B_{f(\theta)})$ do not contradict each other, rather they complement each other. The following example describes a practical situation.*

Example 2.4.5 *Let us envisage a process where recurrent events consist of work related injuries that occur while operating some complex piece of equipment. The censoring time C_i is the time at which individual i participates in a training session meant to refresh the knowledge of the safety procedures that need to be observed while operating this equipment. An analyst is interested in finding a connection between work related accidents and covariates such as individual training, experience, age, some physical health indicators, and the history of previous work related accidents. This information is available at time C_i , but not after that. It would be expensive to follow-up a large number of individuals until the next work related accident occurs, so the analyst will have to use an incomplete data set. To obtain unbiased estimators, the analyst will have to impute the last, incomplete gap time, in an appropriate manner. One reasonable assumption would be that, the information provided by all covariates, past and present, should override the usefulness of the information that the training session took place, even in the not-so-distant past (condition (A0)). On the other hand, if the individual training session just happened, it will have a stronger influence on the length of the next gap time than some partial information about past and present covariates (condition $(B_{f(\theta)})$).*

Proposition 2.4.6 *Assume that (T1), (A0) and $(B_{f(\theta)})$ hold. Then $g_{n,1}^{obs}(\theta)$ is unbiased.*

Proof. By (2.3.8) and Proposition 2.3.6, it suffices to show that, for $i, j \geq 1$

$$E \left[f_{ij}(\theta) I_{ij}^{cen} \frac{E[Z_{ij}(\theta) I_{ij}^{cen}]}{E[I_{ij}^{cen}]} \right] = E[f_{ij}(\theta) I_{ij}^{cen} Z_{ij}(\theta)]. \quad (2.4.9)$$

Since

$$I_{ij}^{cen} E[Z_{ij}(\theta) | I_{ij}^{cen}] = I_{ij}^{cen} \frac{E[Z_{ij}(\theta) I_{ij}^{cen}]}{E[I_{ij}^{cen}]},$$

the left hand side of (2.4.9) is:

$$\begin{aligned} E[f_{ij}(\theta) I_{ij}^{cen} E[Z_{ij}(\theta) | I_{ij}^{cen}]] &= E[f_{ij}(\theta) I_{ij}^{cen} E[Z_{ij}(\theta) | \mathbb{O}(f_i(\theta))]] \\ &= E[E[f_{ij}(\theta) I_{ij}^{cen} Z_{ij}(\theta) | \mathbb{O}(f_i(\theta))]] \\ &= E[f_{ij}(\theta) I_{ij}^{cen} Z_{ij}(\theta)]. \end{aligned}$$

The first equality above is due to $(B_{f(\theta)})$, and the second to (2.4.7). \square

Remark 2.4.7 Expression (2.4.1) for $g_{n,1}^{obs}(\theta)$ coincides with our (2.2.6), which is used in [18]. Note that we needed $(B_{f(\theta)})$ to prove unbiasedness.

We now proceed with step two, embodied in the following lemma which provides the calculation of the expectation under censoring.

Lemma 2.4.8 Let $g_{ij}(\theta)$ be $\mathcal{G}_{i,j-1}$ -measurable, $i, j \geq 1$. We have, when all integrals exist:

$$E_{\theta}[g_{ij}(\theta) Z_{ij}(\theta) I_{ij}^{cen} | \mathcal{G}_{i,j-1}] = -E_{\theta}[g_{ij}(\theta) Z_{ij}(\theta) I_{ij}^{obs} | \mathcal{G}_{i,j-1}]. \quad (2.4.10)$$

Similarly, with $h_{ij}(\eta)$ $\mathcal{G}_{i,j-1}$ -measurable, we have:

$$E_{\eta}[h_{ij}(\eta) [Z_{ij}^2(\theta) - \sigma^2] I_{ij}^{cen} | \mathcal{G}_{i,j-1}] = -E_{\eta}[h_{ij}(\eta) [Z_{ij}^2(\theta) - \sigma^2] I_{ij}^{obs} | \mathcal{G}_{i,j-1}]. \quad (2.4.11)$$

Consequently:

$$E_{\theta}[Z_{ij}(\theta) I_{ij}^{cen}] = -E_{\theta}[Z_{ij}(\theta) I_{ij}^{obs}], \quad (2.4.12)$$

$$E_{\eta}[[Z_{ij}^2(\theta) - \sigma^2] I_{ij}^{cen}] = -E_{\eta}[[Z_{ij}^2(\theta) - \sigma^2] I_{ij}^{obs}], \quad i, j \geq 1. \quad (2.4.13)$$

Proof. Since (2.4.12)-(2.4.13) follow from (2.4.10)-(2.4.11) with $g_{ij}(\theta) \equiv 1, h_{ij}(\eta) \equiv 1, i, j \geq 1$ by taking expectations, it suffices to prove (2.4.10)-(2.4.11).

Since $1 = I_{ij}^{obs} + I_{ij}^{cen} + I_{ij}^{out}$, it follows from (A0) and (2.1.2) that

$$\begin{aligned} 0 &= E_{\theta}[g_{ij}(\theta)E_{\theta}[Z_{ij}(\theta)|\mathcal{G}_{i,j-1}]] \\ &= E_{\theta}[g_{ij}(\theta)E_{\theta}[Z_{ij}(\theta)I_{ij}^{obs}|\mathcal{G}_{i,j-1}]] + E_{\theta}[g_{ij}(\theta)E_{\theta}[Z_{ij}(\theta)I_{ij}^{cen}|\mathcal{G}_{i,j-1}]] \\ &\quad + E_{\theta}[g_{ij}(\theta)E_{\theta}[Z_{ij}(\theta)I_{ij}^{out}|\mathcal{G}_{i,j-1}]]. \end{aligned}$$

Continuing the argument, we have

$E_{\theta}[Z_{ij}(\theta)I_{ij}^{out}|\mathcal{G}_{i,j-1}] = I\{S_{i,j-1} \geq C_i\}E_{\theta}[Z_{ij}(\theta)|\mathcal{G}_{i,j-1}] = 0$, by (A0) and the model assumptions (2.1.2), which proves (2.4.10).

To prove (2.4.11), we write

$$\begin{aligned} 0 &= E_{\eta}[h_{ij}(\eta)E_{\eta}[(Z_{ij}^2(\theta) - \sigma^2)|\mathcal{G}_{i,j-1}]] \\ &= E_{\eta}[h_{ij}(\eta)E_{\eta}[(Z_{ij}^2(\theta) - \sigma^2)I_{ij}^{obs}|\mathcal{G}_{i,j-1}]] + E_{\eta}[h_{ij}(\eta)E_{\eta}[(Z_{ij}^2(\theta) - \sigma^2)I_{ij}^{cen}|\mathcal{G}_{i,j-1}]] \\ &\quad + E_{\eta}[h_{ij}(\eta)E_{\eta}[(Z_{ij}^2(\theta) - \sigma^2)I_{ij}^{out}|\mathcal{G}_{i,j-1}]], \end{aligned}$$

where the first equality is due to (A0) and (2.1.2). Now $E_{\eta}[(Z_{ij}^2(\theta) - \sigma^2)I_{ij}^{out}|\mathcal{G}_{i,j-1}] = I\{S_{i,j-1} \geq C_i\}E_{\eta}[Z_{ij}^2(\theta) - \sigma^2|\mathcal{G}_{i,j-1}] = 0$, by (A0) and (2.1.2). \square

Expression (2.4.12) justifies the use of $\hat{g}_{n,1}^{obs}$ as “estimator” of $g_{n,1}^{obs}$.

We now continue with step three, which shows the asymptotic equivalence of the normalized *e.f.* $g_{n,1}^{obs}$ and $\hat{g}_{n,1}^{obs}$.

We recall the definition of $f_{ij}(\theta)$ in (2.1.6). Below we will use the Euclidean norm for each vector $v \in R^p$.

Theorem 2.4.9 *Assume that (A0), $(B_{f(\theta)})$ and (T2) hold. Furthermore, assume that, there exists $C > 0$, such that, in P_{θ_0} ,*

$$\sup_{i,j \geq 1} \|f_{ij}(\theta_0)\| \leq C < \infty \quad a.s. \quad (2.4.14)$$

Then

$$n^{-1} \| g_{n,1}^{obs}(\theta_0) - \hat{g}_{n,1}^{obs}(\theta_0) \| \rightarrow 0 \quad a.s. \quad n \rightarrow \infty. \quad (2.4.15)$$

Furthermore,

$$n^{-1} \| g_{n,1}(\theta_0) - g_{n,1}^{obs}(\theta_0) \| \rightarrow 0 \quad a.s. \quad n \rightarrow \infty, \quad (2.4.16)$$

and therefore, by (2.3.5), $n^{-1} \hat{g}_{n,1}^{obs}(\theta_0) \rightarrow 0$ a.s., when $n \rightarrow \infty$.

Proof. From (2.4.1)-(2.4.2), to prove the first assertion, we see that

$$\begin{aligned} & n^{-1} \| g_{n,1}^{obs}(\theta_0) - \hat{g}_{n,1}^{obs}(\theta_0) \| \\ \leq & n^{-1} \sum_{i=1}^n \sum_{j=1}^m \left\| I_{ij}^{cen} f_{ij}(\theta_0) \left\{ \frac{E[Z_{1j}(\theta_0) I_{1j}^{obs}]}{E[I_{1j}^{cen}]} - \frac{\sum_{k=1}^n Z_{kj}(\theta_0) I_{kj}^{obs}}{\sum_{k=1}^n I_{kj}^{cen}} \right\} \right\| \\ \leq & C \sum_{j=1}^m \left| \frac{E[Z_{1j}(\theta_0) I_{1j}^{obs}]}{E[I_{1j}^{cen}]} - \frac{\sum_{k=1}^n Z_{kj}(\theta_0) I_{kj}^{obs}}{\sum_{k=1}^n I_{kj}^{cen}} \right|, \end{aligned}$$

where the second inequality follows from (2.4.14).

Now, for each j we can apply the SLLN to both numerator and denominator of $\frac{n^{-1} \sum_{k=1}^n Z_{kj}(\theta_0) I_{kj}^{obs}}{n^{-1} \sum_{k=1}^n I_{kj}^{cen}}$, and (2.4.15) follows, as m is nonrandom and independent of i .

We now prove (2.4.16). We write

$$n^{-1} [g_{n,1}(\theta_0) - g_{n,1}^{obs}(\theta_0)] = n^{-1} \sum_{i=1}^n \sum_{j=1}^m f_{ij}(\theta_0) I_{ij}^{cen} \left[Z_{ij}(\theta_0) - \frac{E[Z_{1j}(\theta_0) I_{1j}^{cen}]}{E[I_{1j}^{cen}]} \right].$$

By SLLN, the right hand side above converges to

$$\sum_{j=1}^m \left\{ E[f_{1j}(\theta_0) I_{1j}^{cen} Z_{1j}(\theta_0)] - \frac{E[Z_{1j}(\theta_0) I_{1j}^{cen}] E[f_{1j}(\theta_0) I_{1j}^{cen}]}{E[I_{1j}^{cen}]} \right\}.$$

By assumption $(B_f(\theta))$ and (2.4.7),

$$\begin{aligned} E[f_{ij}(\theta_0) I_{ij}^{cen} Z_{ij}(\theta_0) | \mathcal{O}(f_i(\theta_0))] &= f_{ij}(\theta_0) I_{ij}^{cen} E[Z_{ij}(\theta_0) I_{ij}^{cen} | \mathcal{O}(f_i(\theta_0))] \\ &= f_{ij}(\theta_0) I_{ij}^{cen} E[Z_{ij}(\theta_0) I_{ij}^{cen} | I_{ij}^{cen}] \\ &= f_{ij}(\theta_0) I_{ij}^{cen} \frac{E[Z_{1j}(\theta_0) I_{1j}^{cen}]}{E[I_{1j}^{cen}]}. \end{aligned}$$

Taking expectation above and looking at the first and last expression in these equalities, we conclude that, for each $1 \leq j \leq m$,

$$E[f_{1j}(\theta_0)I_{1j}^{cen}Z_{1j}(\theta_0)] = E[f_{1j}(\theta_0)I_{1j}^{cen}] \frac{E[Z_{1j}(\theta_0)I_{1j}^{cen}]}{E[I_{1j}^{cen}]},$$

which proves (2.4.16). \square

Remark 2.4.10 *By (2.4.15)-(2.4.16), the normalized $g_{n,1}(\theta_0)$ and $\hat{g}_{n,1}^{obs}(\theta_0)$ are also asymptotically equivalent.*

Condition $(B_{f(\theta)})$ was needed to ensure the unbiasedness of the *e.f.* $g_{n,1}^{obs}(\theta)$. We now define, for each $i \geq 1$, the σ -field

$$\mathbb{O}(f_i, \frac{\partial^2 \mu_i}{\partial \theta^T \partial \theta} V_i^{-1}) := \sigma(f_{ij}(\theta_0)I_{ij}^{cen}, \frac{\partial^2 \mu_{ij}(\theta_0)}{\partial \theta^T \partial \theta} V_i^{-1}(\theta_0)I_{ij}^{cen}, I_{ij}^{cen}, j \geq 1), \quad (2.4.17)$$

and the corresponding condition

$$(B_{(f, \frac{\partial^2 \mu}{\partial \theta^T \partial \theta} V^{-1})}) \quad I_{ij}^{cen} E_{\theta_0}[Z_{ij}(\theta_0)|\mathbb{O}(f_i, \frac{\partial^2 \mu_i}{\partial \theta^T \partial \theta} V_i^{-1})] = I_{ij}^{cen} E_{\theta_0}[Z_{ij}(\theta_0)|I_{ij}^{cen}], \quad i, j \geq 1.$$

We note that the σ -field in (2.4.17) is larger than $\mathbb{O}(f_i(\theta_0))$ in (2.4.7), and thus ensures the unbiasedness of $g_{n,1}^{obs}(\theta_0)$. Condition $(B_{(f, \frac{\partial^2 \mu}{\partial \theta^T \partial \theta} V^{-1})})$, which will be used in Chapter 3, is stronger than condition $(B_{f(\theta)})$. We further note that $\mathbb{O}(f_i, \frac{\partial^2 \mu_i}{\partial \theta^T \partial \theta} V_i^{-1})$ contains less information than $\mathcal{F}_{i,j-1}, j \geq 1$, and so Remark 2.4.4 also applies here.

2.5 Other Estimating Functions

We start this section by examining the properties of the estimator $\hat{g}_{n,2}^{obs}(\eta)$ defined in (2.4.4), when $b_{ij}(\eta) \neq 1$. As in Section 2.4, the first step is to study the unbiasedness of $g_{n,2}^{obs}(\eta)$ defined by (2.4.3). First, we define, for each η and $i \geq 1$, the σ -field

$$\mathbb{O}(b_i(\eta)) := \sigma(b_{ij}(\eta)I_{ij}^{cen}, I_{ij}^{cen}, j \geq 1). \quad (2.5.1)$$

We introduce the condition:

$$(B_{b(\eta)}) \quad I_{ij}^{cen} E_{\eta}[Z_{ij}^2(\theta)|\mathbb{O}(b_i(\eta))] = I_{ij}^{cen} E_{\eta}[Z_{ij}^2(\theta)|I_{ij}^{cen}] \quad i, j \geq 1,$$

which is similar to $(B_{f(\theta)})$. We note that, when b_{ij} are nonrandom, $(B_{b(\eta)})$ holds.

The next result is similar to Proposition 2.4.6.

Proposition 2.5.1 *Assume that (T1), (A0) and $(B_{b(\eta)})$ hold. Then $g_{n,2}^{obs}(\eta)$ defined in (2.4.3) is unbiased.*

Proof. By (2.3.9) and Proposition 2.3.6, it suffices to show

$$E_\eta \left[b_{1j}(\eta) I_{1j}^{cen} \frac{E_\eta[(Z_{1j}^2(\theta) - \sigma^2) I_{1j}^{cen}]}{E_\eta[I_{1j}^{cen}]} \right] = E_\eta[b_{1j}(\eta) I_{1j}^{cen} (Z_{1j}^2(\theta) - \sigma^2)]. \quad (2.5.2)$$

We start with the left hand side, and observe that it can be written

$$E_\eta[b_{1j}(\eta) I_{1j}^{cen} E_\eta[(Z_{1j}^2(\theta) - \sigma^2) | I_{1j}^{cen}]].$$

By $(B_{b(\eta)})$, this equals $E_\eta[b_{1j}(\eta) I_{1j}^{cen} E_\eta[(Z_{1j}^2(\theta) - \sigma^2) | \mathbb{O}(b_1(\eta))]]$, which is $E_\eta [E_\eta[b_{1j}(\eta) I_{1j}^{cen} (Z_{1j}^2(\theta) - \sigma^2) | \mathbb{O}(b_1(\eta))]]$, the right hand side of (2.5.2). \square

The next result corresponds to Theorem 2.4.9. Let $b_{ij} := b_{ij}(\eta_0)$.

Theorem 2.5.2 *Assume that (A0), $(B_{b(\eta)})$ and (T2) hold. Furthermore, assume that there exists $C > 0$ such that, in P_{η_0} ,*

$$\sup_{i,j \geq 1} |b_{ij}(\eta_0)| \leq C < \infty \quad a.s. \quad (2.5.3)$$

Then

$$n^{-1} |g_{n,2}^{obs}(\eta_0) - \hat{g}_{n,2}^{obs}(\eta_0)| \rightarrow 0 \quad a.s. \quad (2.5.4)$$

Furthermore,

$$n^{-1} |g_{n,2}(\eta_0) - g_{n,2}^{obs}(\eta_0)| \rightarrow 0 \quad a.s., \quad n \rightarrow \infty, \quad (2.5.5)$$

and therefore, by (2.3.5), $n^{-1} \hat{g}_{n,2}^{obs}(\eta_0) \rightarrow 0$ a.s., when $n \rightarrow \infty$.

Proof. By (2.4.3)-(2.4.4),

$$n^{-1} |g_{n,2}^{obs}(\eta_0) - \hat{g}_{n,2}^{obs}(\eta_0)|$$

$$\begin{aligned}
&\leq n^{-1} \sum_{i=1}^n \sum_{j=1}^m |b_{ij}(\eta_0)| \left| \frac{E[(Z_{1j}^2(\theta_0) - \sigma_0^2)I_{1j}^{cen}]}{E[I_{1j}^{cen}]} - \frac{\sum_{k=1}^n (Z_{kj}^2(\theta_0) - \sigma_0^2)I_{kj}^{obs}}{\sum_{k=1}^n I_{kj}^{cen}} \right| \\
&\leq Cn^{-1} \sum_{i=1}^n \sum_{j=1}^m \left| \frac{E[(Z_{1j}^2(\theta_0) - \sigma_0^2)I_{1j}^{cen}]}{E[I_{1j}^{cen}]} - \frac{\sum_{k=1}^n (Z_{kj}^2(\theta_0) - \sigma_0^2)I_{kj}^{obs}}{\sum_{k=1}^n I_{kj}^{cen}} \right|,
\end{aligned}$$

where the second inequality follows from (2.5.3). As before, for each j we can apply SLLN to both numerator and denominator of $\frac{\sum_{k=1}^n (Z_{kj}^2(\theta_0) - \sigma_0^2)I_{kj}^{obs}}{\sum_{k=1}^n I_{kj}^{cen}}$ to obtain (2.5.4).

To prove (2.5.5), we write

$$n^{-1}[g_{n,2}(\eta_0) - g_{n,2}^{obs}(\eta_0)] = n^{-1} \sum_{i=1}^n \sum_{j=1}^m b_{ij}(\eta_0) I_{ij}^{cen} \left[(Z_{ij}^2(\theta_0) - \sigma_0^2) - \frac{E[(Z_{1j}^2(\theta_0) - \sigma_0^2)I_{1j}^{cen}]}{E[I_{1j}^{cen}]} \right].$$

By SLLN, the right hand side above converges to

$$\sum_{j=1}^m \left\{ E[b_{1j}(\eta_0)(Z_{1j}^2(\theta_0) - \sigma_0^2)I_{1j}^{cen}] - \frac{E[b_{1j}(\eta_0)I_{1j}^{cen}]E[(Z_{1j}^2(\theta_0) - \sigma_0^2)I_{1j}^{cen}]}{E[I_{1j}^{cen}]} \right\}.$$

To show that this is 0, we use (2.5.1) and $(B_{b(\eta)})$ to write

$$\begin{aligned}
&E_{\eta_0}[b_{ij}(\eta_0)(Z_{ij}^2(\theta_0) - \sigma_0^2)I_{ij}^{cen} | \mathbb{O}(b_i(\eta_0))] \\
&= b_{ij}(\eta_0)I_{ij}^{cen} E_{\eta_0}[(Z_{ij}^2(\theta_0) - \sigma_0^2) | \mathbb{O}(b_i(\eta_0))] \\
&= b_{ij}(\eta_0)I_{ij}^{cen} E_{\eta_0}[(Z_{ij}^2(\theta_0) - \sigma_0^2) | I_{ij}^{cen}] \\
&= b_{ij}(\eta_0)I_{ij}^{cen} \frac{E_{\eta_0}[(Z_{1j}^2(\theta_0) - \sigma_0^2)I_{1j}^{cen}]}{E_{\eta_0}I_{1j}^{cen}}.
\end{aligned}$$

We now apply E_{η_0} to this string of equalities. Looking at the first and the last expression, we conclude that, for each $1 \leq j \leq m$,

$$E_{\eta_0}[b_{1j}(\eta_0)(Z_{1j}^2(\theta_0) - \sigma_0^2)I_{1j}^{cen}] = E_{\eta_0}[b_{1j}(\eta_0)I_{1j}^{cen}] \frac{E_{\eta_0}[(Z_{1j}^2(\theta_0) - \sigma_0^2)I_{1j}^{cen}]}{E_{\eta_0}[I_{1j}^{cen}]},$$

which concludes the proof of (2.5.5). \square

For the rest of the section, we consider a marginal model in which $\mu_{ij}(\theta)$ and $V_{ij}(\theta)$ in (2.1.2) depend on x_{ij} only, which becomes available at the $(j-1)^{th}$ occurrence, $i, j \geq 1$. We can also think of a conditional model where only the covariate measured

last is influential. We note that in a *bona fide* marginal model x_{ij} is measured at the same time as the response variable (see, for instance [28]).

The *e.f.* $g_{n,1}^{cov}$ that we define next captures the influence of the covariates last observed before censoring. It also captures the presence of censoring. In this sense, it is more general than $\hat{g}_{n,1}^{obs}$. In Example 2.1.7, this would be either \overline{BMI}_{ij} , or $Y_{i,j-1}$.

To define $g_{n,1}^{cov}$ we need further notation. We start by considering scalar covariates with a finite number of values.

Let $a_{ij,k}$ be the possible values of x_{ij} , $1 \leq k \leq k_{ij}$. We define the following partitions of the entire space of events. For $i \geq 1$, let

$$F_{ij,k} := x_{ij}^{-1}(a_{ij,k}) \in \mathcal{F}_{i,j-1}, \quad 1 \leq k \leq k_{ij}, \quad j \geq 1.$$

Let

$$\mathbb{O}(x_i) := \sigma(\chi_{ij,k} I_{ij}^{cen}, I_{ij}^{cen}, 1 \leq k \leq k_{ij}, j \geq 1), \quad (2.5.6)$$

where $\chi_{ij,k}$ is the characteristic function of $F_{ij,k}$. We write, for each $i, j \geq 1$

$$(B_x) \quad I_{ij}^{cen} E_\theta[Z_{ij}(\theta) | \mathbb{O}(x_i)] = I_{ij}^{cen} E_\theta[Z_{ij}(\theta) | \sigma(x_{ij} I_{ij}^{cen}, I_{ij}^{cen})],$$

which is valid in this scenario.

Let

$$Z_{ij}^{imp}(\theta) := \sum_{k=1}^{k_{ij}} \chi_{ij,k} I_{ij}^{cen} (E_\theta[\chi_{ij,k} I_{ij}^{cen}])^{-1} E_\theta[Z_{ij}(\theta) \chi_{ij,k} I_{ij}^{cen}]. \quad (2.5.7)$$

Assuming that (T1) holds, define

$$g_{n,1}^{cov}(\theta) = \sum_{i=1}^n \sum_{j=1}^{m_i} f_{ij}(\theta) \left[Z_{ij}(\theta) I_{ij}^{obs} + I_{ij}^{cen} Z_{ij}^{imp}(\theta) \right]. \quad (2.5.8)$$

The imputed term (2.5.7) is a conditional expectation of $Z_{ij}(\theta)$, given the finite field (2.5.6), and expressed in terms of its atoms. To prove unbiasedness of $g_{n,1}^{cov}$, we proceed as in Section 2.4. We need to show:

$$E[f_{ij}(\theta) I_{ij}^{cen} Z_{ij}^{imp}(\theta)] = E[f_{ij}(\theta) I_{ij}^{cen} Z_{ij}(\theta)]. \quad (2.5.9)$$

We start with the left hand side of (2.5.9), which we write

$$\begin{aligned} E[f_{ij}(\theta)I_{ij}^{cen} E[Z_{ij}(\theta)|\mathbb{O}(x_i)]] &= E[f_{ij}(\theta)I_{ij}^{cen} E[Z_{ij}(\theta)|\sigma(x_{ij}I_{ij}^{cen}, I_{ij}^{cen})]] \\ &= E[E[f_{ij}(\theta)I_{ij}^{cen} Z_{ij}(\theta)|\sigma(x_{ij}I_{ij}^{cen}, I_{ij}^{cen})]] \\ &= E[f_{ij}(\theta)I_{ij}^{cen} Z_{ij}(\theta)]. \end{aligned}$$

The second equality holds because $f_{ij}(\theta)$ is an analytic function of x_{ij} , $i, j \geq 1$. As before, the next step in estimating $Z_{ij}^{imp}(\theta)$ using only collected data is to replace $E[Z_{ij}(\theta)\chi_{ij,k}I_{ij}^{cen}]$ by $-E[Z_{ij}(\theta)\chi_{ij,k}I_{ij}^{obs}]$, which is possible by Lemma 2.4.8, since $\chi_{ij,k}$ are $\mathcal{F}_{i,j-1}$ -measurable. Finally, to estimate the expectations in (2.5.7), we use data from all sampled individuals, as in (2.4.2).

Consider now the more general situation, when x_{ij} is a vector with components that have a continuous range. It is known that such measurable vectors can be approximated, componentwise, by simple functions. For x_{ij} scalar, we can apply (13.6) in Theorem 13.5 of [4] to obtain simple functions $x_{ij}^{(n)}$ such that, for each $i, j \geq 1$

$$x_{ij}^{(n)} \rightarrow x_{ij} \quad \text{a.s.,} \quad n \rightarrow \infty.$$

We then replace $\chi_{ij,k}$ in (B_x) by $\chi_{ij,k}^{(n)}$ for a sufficiently large $n > 1$.

In this case of infinite range for the components, the left hand side of (B_x) is only an approximation of the right hand side, where we stop at a convenient value of n . To go from components to the vector, one can take as partitions associated with the vector the intersection of the partitions corresponding to all components.

In this dissertation we do not pursue the study of $g_{n,1}^{cov}$ and its associated estimator of θ_0 , $n \geq 1$. It is, however worthwhile doing so when the covariate x_{ij} has few discrete values.

When x_{ij} is continuous, or has many possible discrete values (*e.g.*, multinomial, Poisson distributions), obtaining a good estimator for θ_0 from $g_{n,1}^{cov}$ is problematic. The sample of individuals participating in the study would have to be fairly large to obtain good estimators for each of the $k_{ij,n}$ categories, as we did in (2.4.2).

2.6 Connection with Other Longitudinal Studies

We first complete the comparison between our EE and GEE under working independence assumption, as presented in Chapter 1. Under this assumption, we have that the variances in (1.0.2) are the diagonal matrices

$$V_i(\beta) = \text{diag}_{1 \leq j \leq m_i} \{\sigma_{ij}(\beta)\} I_{m_i} \text{diag}_{1 \leq j \leq m_i} \{\sigma_{ij}(\beta)\} = \text{diag}_{1 \leq j \leq m_i} \{\sigma_{ij}^2(\beta)\},$$

where I_{m_i} is the $m_i \times m_i$ identity matrix and the diagonal entries $\sigma_{ij}(\beta)$ are defined in (1.0.1), $i \geq 1$. So the GEE in (1.0.2) becomes:

$$\sum_{i=1}^n \left[\frac{\partial \mu_i(\beta)}{\partial \beta^T} \right]^T \text{diag}_{1 \leq j \leq m_i} \{\sigma_{ij}^{-2}(\beta)\} [y_i - \mu_i(\beta)] = 0. \quad (2.6.1)$$

Returning now to the first EE of (2.3.2) and in the context of recurrent events, we recall (2.1.6) and the definition of $V_{ij}(\theta)$ in (2.1.2) to write this EE as:

$$\sum_{i=1}^n \frac{\partial \mu_i(\theta)}{\partial \theta} \text{diag}_{1 \leq j \leq m_i} \{V_{ij}^{-2}(\theta)\} [y_i - \mu_i(\theta)] = 0. \quad (2.6.2)$$

Since $\frac{\partial \mu_i(\theta)}{\partial \theta} = \left[\frac{\partial \mu_i(\theta)}{\partial \theta^T} \right]^T$, the similarity with (2.6.1) is now apparent, due to the correspondence in notation established in Chapter 1.

Consider a longitudinal study, where each individual i is observed at random times $S_{ij}, 1 \leq j \leq \tau_i, i \geq 1$. The covariates x_{ij} are available at time $S_{i,j-1}$, while the response variable y_{ij} is recorded at time S_{ij} . The response variable satisfies the model assumptions (2.1.2). It could represent, for instance, the result of a blood test, or some characteristic that may require some costly effort to obtain (*e.g.*, a personal interview to obtain some sensitive data). In other words, in this case the recurrent events are the object of the analysis, rather than the gap times. In the study of asthma in children we could be interested in the intensity, duration or type of the asthma episodes, rather than the time gaps between the end of one episode and the beginning of the next. Covariates could contain similar information from previous episodes and on the application of the treatment.

In our set-up, for each individual i , the study ends at a random time C_i , subject to the conditions spelt out in Section 2.1. We note a practical difference in the recording of y_{ij} versus Y_{ij} , where the latter response variable is based on the gap time $S_{ij} - S_{i,j-1}$. In the event that $S_{ij} = C_i$, we can still observe S_{ij} with relative ease and at low cost, and calculate $S_{ij} - S_{i,j-1}$, the just-completed gap time. Recording an important event occurring at S_{ij} may be done routinely by organizations that are not even involved in the study. For instance, health units or hospitals can provide information on events such as work injuries or serious asthma attacks (see the examples in [6]). On the other hand, obtaining y_{ij} might be costly, and we may forgo doing so at the time C_i when the study has just ended. We may consider using our *e.f.* (2.3.1), since all the necessary information is available and no imputation is required. The important thing is to ensure that the *e.f.* used to obtain our estimators are unbiased in all cases. Note that this approach also provides estimators for the overdispersion parameter.

Since our research can cover a large class of longitudinal models, we can include as examples specific conditional distributions of the response variable, *e.g.*, the normal, log-linear for count-type regression and the logistic distributions, for binary responses, which are not necessarily related to the analysis of recurrent events. In these cases, and with our notation, we have the following examples:

$$(4) \quad E_{\theta}[Y_{ij}|\mathcal{F}_{i,j-1}] = \theta^T x_{ij}, \quad Var_{\theta}[Y_{ij}|\mathcal{F}_{i,j-1}] = 1,$$

$$(5) \quad E_{\theta}[Y_{ij}|\mathcal{F}_{i,j-1}] = \exp(\theta^T x_{ij}) = Var_{\theta}[Y_{ij}|\mathcal{F}_{i,j-1}],$$

$$(6) \quad E_{\theta}[Y_{ij}|\mathcal{F}_{i,j-1}] = \frac{\exp(\theta^T x_{ij})}{1 + \exp(\theta^T x_{ij})}, \quad Var_{\theta}[Y_{ij}|\mathcal{F}_{i,j-1}] = \frac{\exp(\theta^T x_{ij})}{(1 + \exp(\theta^T x_{ij}))^2}.$$

2.7 A Preliminary Result

In this section, we give a sequence of consistent estimator of σ^2 when $b_{ij}(\eta) \equiv 1$. In the process of doing this, we obtain asymptotic results that will be used in subsequent proofs.

From here on, we will use the following form of the Cauchy-Schwarz inequality, for vectors $a, b \in R^m$, for some $m \geq 1$.

$$\left| \sum_{i=1}^n a_i^T b_i \right| \leq \left[\sum_{i=1}^n a_i^T a_i \right]^{1/2} \left[\sum_{i=1}^n b_i^T b_i \right]^{1/2}. \quad (2.7.1)$$

Lemma 2.7.1 *Assume that (T0) holds. Then (2.7.2) holds:*

$$n^{-1} \sum_{i=1}^n \sum_{j=1}^{\infty} Z_{ij}^2(\theta_0) I\{S_{i,j-1} < C_i\} \rightarrow \sigma_0^2 E_{\eta_0}(\tau_1) \quad \text{a.s. in } P_{\eta_0}. \quad (2.7.2)$$

In particular, the left hand side above is a.s. asymptotically bounded.

Proof. We first look at

$$\begin{aligned} E_{\eta_0} \left[\sum_{j=1}^{\infty} Z_{ij}^2(\theta_0) I\{S_{i,j-1} < C_i\} \right] &= \sum_{j=1}^{\infty} E_{\eta_0} [Z_{ij}^2(\theta_0) I\{S_{i,j-1} < C_i\}] \\ &= \sum_{j=1}^{\infty} E_{\eta_0} [I\{S_{i,j-1} < C_i\} E_{\eta_0} [Z_{ij}^2(\theta_0) | \mathcal{G}_{i,j-1}]] \\ &= \sigma_0^2 E_{\eta_0} \left[\sum_{j=1}^{\infty} I\{S_{i,j-1} < C_i\} \right] \\ &= \sigma_0^2 E_{\eta_0} [\tau_1] < \infty. \end{aligned} \quad (2.7.3)$$

In the first and the third equality above we used the monotone convergence theorem (Theorem 16.2 of [4]). For the third equality, we used (A0) and $E_{\eta_0} [Z_{ij}^2(\eta_0) | \mathcal{F}_{i,j-1}] = \frac{\text{Var}_{\eta_0} [Y_{ij} | \mathcal{F}_{i,j-1}]}{V_{ij}^2(\theta_0)} = \sigma_0^2$, which follows from (2.1.2) and (2.1.6). By SLLN, since the summands are *i.i.d.* for each i , (2.7.2) holds. \square

Remark 2.7.2 *The first e.f. in (2.3.1) does not depend on σ^2 . Therefore, $g_{n,1}(\theta) = 0$ can be solved first in θ , to obtain an estimator $\hat{\theta}_n$ of θ_0 . Next, $\hat{\theta}_n$ can be placed in the second e.f. to find an estimator $\hat{\sigma}_n^2(\hat{\theta}_n)$ of σ_0^2 .*

In (2.3.1), when $b_{ij}(\eta) \equiv 1, i, j \geq 1, \hat{\sigma}_n^2(\theta)$ can be written explicitly, for any $\theta \in R^p$, as

$$\hat{\sigma}_n^2(\theta) = \frac{\sum_{i=1}^n \sum_{j=1}^{\infty} Z_{ij}^2(\theta) I\{S_{i,j-1} < C_i\}}{\sum_{i=1}^n \tau_i}. \quad (2.7.4)$$

We use the following notation:

$$B_r(\theta) := \{\theta' \in R^p : \|\theta' - \theta\| \leq r\}, \quad (2.7.5)$$

$$\delta_n(r) := \sup_{\theta, \theta' \in B_r(\theta_0)} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} \frac{|\mu_{ij}(\theta) - \mu_{ij}(\theta')|}{V_{ij}(\theta)}, \quad (2.7.6)$$

$$\delta_n^{[1]}(r) := \sup_{\theta, \theta' \in B_r(\theta_0)} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} \frac{|\dot{\mu}_{ij}(\theta) - \dot{\mu}_{ij}(\theta')|}{V_{ij}(\theta)}, \quad (2.7.7)$$

$$\delta_n^{[2]}(r) := \sup_{\theta, \theta' \in B_r(\theta_0)} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} \frac{|\ddot{\mu}_{ij}(\theta) - \ddot{\mu}_{ij}(\theta')|}{V_{ij}(\theta)}, \quad (2.7.8)$$

$$\eta_n(r) := \sup_{\theta, \theta' \in B_r(\theta_0)} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} \left| \frac{V_{ij}^2(\theta) - V_{ij}^2(\theta')}{V_{ij}^2(\theta)} \right|. \quad (2.7.9)$$

We will be using the following conditions:

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} \delta_n(r) = 0 \quad a.s., \quad (2.7.10)$$

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} \eta_n(r) = 0 \quad a.s. \quad (2.7.11)$$

These conditions will be illustrated in Section 2.8.

We have the following result.

Theorem 2.7.3 *Let $b_{ij}(\eta) \equiv 1$ in (2.3.1) and assume that (2.7.10)-(2.7.11) and (T1) hold. Assume further that $\hat{\theta}_n \rightarrow \theta_0$ a.s. in P_{η_0} . Then $\hat{\sigma}_n^2(\hat{\theta}_n) \rightarrow \sigma_0^2$ a.s. in P_{η_0} , as $n \rightarrow \infty$, where $\hat{\sigma}_n^2(\hat{\theta}_n)$ is given by (2.7.4) with $\theta = \hat{\theta}_n$.*

Proof. For $\theta \in R^p$ and $g_{n,2}$ in (2.3.1), we solve for σ^2 the equation $g_{n,2}(\theta, \sigma^2) = 0$ and obtain

$$\hat{\sigma}_n^2(\theta) = \frac{\sum_{i=1}^n \sum_{j=1}^{\infty} b_{ij}(\theta, \hat{\sigma}_n^2(\theta)) Z_{ij}^2(\theta) I\{S_{i,j-1} < C_i\}}{\sum_{i=1}^n \sum_{j=1}^{\infty} b_{ij}(\theta, \hat{\sigma}_n^2(\theta)) I\{S_{i,j-1} < C_i\}}. \quad (2.7.12)$$

When $b_{ij}(\eta) \equiv 1, i, j \geq 1$, (2.7.12) becomes (2.7.4), since $\sum_{j=1}^{\infty} I\{S_{i,j-1} < C_i\} = \tau_i, i \geq 1$. We now take $\theta = \theta_0$ in (2.7.4), divide the numerator and the denominator

by n , and study separately their asymptotic behaviour. The denominator is now $\bar{\tau}_n = n^{-1} \sum_{i=1}^n \tau_i$, which converges *a.s.* in P_{η_0} to $E_{\eta_0}[\tau_1] < \infty$, by (T1) and the SLLN for the *i.i.d.* case.

For the numerator of (2.7.4), we have that (2.7.2) holds. Consequently, $\hat{\sigma}_n^2(\theta_0) \rightarrow \sigma_0^2$ *a.s.*, when $n \rightarrow \infty$. To complete the proof, it suffices to show that

$$n^{-1} \sum_{i=1}^n \sum_{j=1}^{\infty} |Z_{ij}^2(\hat{\theta}_n) - Z_{ij}^2(\theta_0)| I\{S_{i,j-1} < C_i\} \rightarrow 0 \quad \textit{a.s.} \quad \text{in} \quad P_{\eta_0}. \quad (2.7.13)$$

We start by writing

$$Z_{ij}^2(\hat{\theta}_n) = \frac{\left(Y_{ij} - \mu_{ij}(\hat{\theta}_n)\right)^2}{V_{ij}^2(\theta_0)} \frac{V_{ij}^2(\theta_0)}{V_{ij}^2(\hat{\theta}_n)}.$$

Since

$$\frac{V_{ij}^2(\theta_0)}{V_{ij}^2(\hat{\theta}_n)} = \frac{V_{ij}^2(\theta_0) - V_{ij}^2(\hat{\theta}_n) + V_{ij}^2(\hat{\theta}_n)}{V_{ij}^2(\hat{\theta}_n)} = \frac{V_{ij}^2(\theta_0) - V_{ij}^2(\hat{\theta}_n)}{V_{ij}^2(\hat{\theta}_n)} + 1,$$

$$\begin{aligned} Z_{ij}^2(\hat{\theta}_n) &:= \frac{\left(Y_{ij} - \mu_{ij}(\hat{\theta}_n)\right)^2}{V_{ij}^2(\theta_0)} \left(1 + \nu_{ij}(\hat{\theta}_n)\right) \\ &= \frac{\left(Y_{ij} - \mu_{ij}(\hat{\theta}_n)\right)^2}{V_{ij}^2(\theta_0)} + \nu_{ij}(\hat{\theta}_n) \frac{\left(Y_{ij} - \mu_{ij}(\hat{\theta}_n)\right)^2}{V_{ij}^2(\theta_0)} \\ &:= a_1^{(n)}(i, j) + \nu_{ij}(\hat{\theta}_n) a_1^{(n)}(i, j) \\ &:= a_1^{(n)}(i, j) + a_2^{(n)}(i, j). \end{aligned} \quad (2.7.14)$$

The first term above can be written:

$$\begin{aligned} V_{ij}^{-2}(\theta_0) \left[Y_{ij} - \mu_{ij}(\hat{\theta}_n)\right]^2 &= V_{ij}^{-2}(\theta_0) \left[Y_{ij} - \mu_{ij}(\theta_0) + \mu_{ij}(\theta_0) - \mu_{ij}(\hat{\theta}_n)\right]^2 \\ &= V_{ij}^{-2}(\theta_0) \left\{ [Y_{ij} - \mu_{ij}(\theta_0)]^2 + 2[Y_{ij} - \mu_{ij}(\theta_0)][\mu_{ij}(\theta_0) - \mu_{ij}(\hat{\theta}_n)] \right. \\ &\quad \left. + [\mu_{ij}(\theta_0) - \mu_{ij}(\hat{\theta}_n)]^2 \right\} \\ &:= a_{11}^{(n)}(i, j) + a_{12}^{(n)}(i, j) + a_{13}^{(n)}(i, j). \end{aligned} \quad (2.7.15)$$

By Lemma 2.7.1,

$$A_{11}^{(n)} := \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{\tau_i} a_{11}^{(n)}(i, j) \quad \text{is } a.s. \text{ asymptotically bounded.}$$

We will show that

$$A_{12}^{(n)} := \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{\tau_i} a_{12}^{(n)}(i, j) \rightarrow 0 \quad a.s. \quad P_{\eta_0}, \quad \text{as } n \rightarrow \infty, \quad (2.7.16)$$

$$A_{13}^{(n)} := \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{\tau_i} a_{13}^{(n)}(i, j) \rightarrow 0 \quad a.s. \quad P_{\eta_0}, \quad \text{as } n \rightarrow \infty, \quad (2.7.17)$$

$$A_2^{(n)} := \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{\tau_i} a_2^{(n)}(i, j) \rightarrow 0 \quad a.s. \quad P_{\eta_0}, \quad \text{as } n \rightarrow \infty. \quad (2.7.18)$$

To obtain (2.7.13), it suffices to prove (2.7.16)-(2.7.18), as (2.7.13) and the decomposition (2.7.15) shows.

By (2.7.1) and the definition of $a_{12}^{(n)}(i, j)$ in (2.7.15), we have

$$|A_{12}^{(n)}| \leq 2(A_{11}^{(n)})^{1/2}(A_{13}^{(n)})^{1/2},$$

where we used (2.7.1) with

$$a_{ij} = Z_{ij}(\theta_0), \quad b_{ij} = V_{ij}^{-1}(\theta_0)[\mu_{ij}(\theta_0) - \mu_{ij}(\hat{\theta}_n)], \quad j = 1, 2, \dots, \tau_i, \quad i = 1, 2, \dots, n.$$

From (2.7.15), (2.7.17) and (2.7.6) $A_{13}^{(n)} \leq \delta_n^2(r) \bar{\tau}_n$, where $\hat{\theta}_n \in B_r(\theta_0)$, which happens for $n \geq n_0(\omega)$, since $\hat{\theta}_n \rightarrow \theta_0$ *a.s.* The first factor converges to 0, as $r \rightarrow 0$, by condition (2.7.10), while $\bar{\tau}_n \rightarrow E_{\eta_0}(\tau_1) < \infty$ *a.s.*, by SLLN, which proves (2.7.17). Combining (2.7.2) with (2.7.17) gives (2.7.16).

It remains to prove (2.7.18). Let us write

$$A_1^{(n)} := A_{11}^{(n)} + A_{12}^{(n)} + A_{13}^{(n)}.$$

By (2.7.2), (2.7.16) and (2.7.17), which we have proved, $A_1^{(n)}$ is *a.s.* asymptotically bounded. Since

$$|A_2^{(n)}| \leq \eta_n(r) |A_1^{(n)}|,$$

when $\hat{\theta}_n \in B_r(\theta_0)$, $A_2^{(n)} \rightarrow 0$ *a.s.* in P_{η_0} , as $n \rightarrow \infty$ and $r \rightarrow 0$, by assumption (2.7.11).

This proves Theorem 2.7.3. \square

2.8 Examples

We first illustrate conditions (2.7.10)-(2.7.11) on Example 2.1.2.

Example 2.8.1 *We start with condition (2.7.11). Assume that $\rho_0 \in (0, 1)$ in Example 2.1.2, with the first formula for $V_{ij}(\rho)$ and ρ_0 the true parameter, and select $r_0 \in R^+$, $0 < r_0 < \min\{\rho_0, 1 - \rho_0\}$. We write*

$$V_{ij}^2(\rho) = 1 + \frac{\rho}{\rho(j-2)+1} = 1 + \frac{1}{j-2+\rho^{-1}}, \quad i, j \geq 1, \rho \in B_{r_0}(\rho_0) \cap (0, 1),$$

and evaluate :

$$\frac{V_{ij}^2(\rho) - V_{ij}^2(\rho')}{V_{ij}^2(\rho)}, \quad \rho, \rho' \in B_{r_0}(\rho_0) \cap (0, 1). \quad (2.8.1)$$

We define the family of functions $\{f_j\}_{j \geq 1}$, where, as in (2.4.8),

$$f_j(\rho) := \frac{1}{j-2+\rho^{-1}}, \quad 0 < \rho < 1.$$

Note that $f_{j'}(\rho) < f_j(\rho)$, $f_j(\rho') < f_j(\rho)$, if $\rho' < \rho, j' > j \geq 1$, i.e., $f_j(\rho)$ is increasing in ρ and decreasing in j . We show first that the denominator in (2.8.1) is equibounded in $i, j \geq 1$, for $\rho \in B_{r_0}(\rho_0)$. Since $0 < \rho_0 - r_0 \leq \rho \leq \rho_0 + r_0$, we have, from the definitions:

$$1 + f_j(\rho_0 - r_0) \leq V_{ij}^2(\rho) \leq 1 + f_j(\rho_0 + r_0). \quad (2.8.2)$$

From (2.8.2) and the properties of $f_j(\rho)$, we obtain:

$$1 < 1 + f_{\tau_i}(\rho_0 - r_0) \leq V_{ij}^2(\rho) \leq 1 + f_1(\rho_0 + r_0), \quad (2.8.3)$$

since the denominator of $f_{\tau_i}(\rho_0 - r_0)$ is $\tau_i - 2 + (\rho_0 - r_0)^{-1} \geq -1 + (\rho_0 - r_0)^{-1} > 0$. The upper and the lower bound in (2.8.3) are independent of $j, i, \rho \in B_{r_0}(\rho_0)$, and so the same holds for the bounds of $V_{ij}^{-2}(\rho)$, $1 \leq j \leq \tau_i, i \geq 1$.

To prove (2.7.11), it suffices to show that

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} \sup_{\rho \in B_r(\rho_0)} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} |V_{ij}^2(\rho) - V_{ij}^2(\rho_0)| = 0. \quad (2.8.4)$$

From (2.8.2) we have, for $i, j \geq 1$:

$$f_j(\rho_0 - r_0) - f_j(\rho_0) \leq V_{ij}^2(\rho) - V_{ij}^2(\rho_0) \leq f_j(\rho_0 + r_0) - f_j(\rho_0). \quad (2.8.5)$$

From the definition and the properties of $f_j(\rho)$,

$$\max\{f_j(\rho_0 - r_0), f_j(\rho_0), f_j(\rho_0 + r_0)\} = f_j(\rho_0 + r_0) \rightarrow 0, \text{ as } j \rightarrow \infty.$$

Thus, for $\varepsilon > 0$, let $j_0(r_0, \varepsilon)$ be the first integer such that $j_0(r_0, \varepsilon) \geq \frac{1}{\varepsilon} - \frac{1}{\rho_0 + r_0} + 2$.

This ensures that $f_j(\rho_0 + r_0) \leq \varepsilon$, for all $j \geq j_0(r_0, \varepsilon)$. Since $r \leq r_0$ implies that $j_0(r, \varepsilon) \leq j_0(r_0, \varepsilon)$, we have that, with $0 < r \leq r_0$

$$\max\{f_j(\rho_0 - r), f_j(\rho_0), f_j(\rho_0 + r)\} \leq \varepsilon \text{ for all } j \geq j_0(r_0, \varepsilon) := j_0. \quad (2.8.6)$$

Since f_j is continuous at ρ_0 for each $j \geq 1$, for $\varepsilon > 0$, we can find $r_0(\varepsilon, j_0) > 0$ such that, for any $\rho \in B_{r_0(\varepsilon, j_0)}(\rho_0)$,

$$\max_{1 \leq j \leq j_0} |V_{ij}^2(\rho) - V_{ij}^2(\rho_0)| \leq \varepsilon. \quad (2.8.7)$$

Let $r_0(\varepsilon) := \min\{r_0, r_0(\varepsilon, j_0)\}$, $\rho \in B_{r_0(\varepsilon)}(\rho_0)$, $\eta_{ij}(\rho) := |V_{ij}^2(\rho) - V_{ij}^2(\rho_0)|$,

$\tau^{(n)} := \max_{1 \leq i \leq n} \tau_i$. Then

$$\begin{aligned} & \max_{i \leq n, j \leq \tau_i} \eta_{ij}(\rho) \leq \max_{1 \leq i \leq n} \max_{1 \leq j \leq \tau^{(n)}} \eta_{ij}(\rho) \\ & \leq \max_{1 \leq i \leq n} \left[\max_{1 \leq j \leq j_0} \eta_{ij}(\rho) I\{\tau^{(n)} \leq j_0\} + \max_{1 \leq i \leq n} \left\{ \max_{1 \leq j \leq j_0} \eta_{ij}(\rho), \max_{j_0 < j \leq \tau^{(n)}} \eta_{ij}(\rho) \right\} I\{\tau^{(n)} > j_0\} \right] \\ & \leq \varepsilon I\{\tau^{(n)} \leq j_0\} + 2\varepsilon I\{\tau^{(n)} > j_0\} \leq 2\varepsilon. \end{aligned} \quad (2.8.8)$$

In the last inequality we used (2.8.5)-(2.8.7). This completes the proof of (2.8.4).

We now discuss conditions under which (2.7.10) holds. Since, by (2.8.3), the denominator of $\delta_n(r)$ is equibounded, we concentrate on the numerator. Let $\theta^T = (\gamma_0, \gamma_1, \rho) \in B_{r_0}(\theta_0)$, $\theta_0^T = (\gamma_{00}, \gamma_{01}, \rho_0)$. From here on in this example we use $x_{ij,h}$ as mentioned in Example 2.1.7, so

$$\mu_{ij}(\theta) = \gamma_0 + f_j(\rho) \sum_{l=2}^j [x_{ij,j+l} - (\gamma_0 + \gamma_1 x_{ij,l})] + \gamma_1 x_{ij,j+1}, \quad j \leq \tau_i, i \geq 1. \quad (2.8.9)$$

With $x_i^{(j-1)} := \sum_{l=1}^{j-1} x_{ij,l+1}$, we have:

$$\begin{aligned} \mu_{ij}(\theta) - \mu_{ij}(\theta_0) &= \gamma_0 - \gamma_{00} - (j-1)[\gamma_0 f_j(\rho) - \gamma_{00} f_j(\rho_0)] \\ &\quad + [f_j(\rho) - f_j(\rho_0)] S_{i,j-1} - [\gamma_1 f_j(\rho) - \gamma_{01} f_j(\rho_0)] x_i^{(j-1)} \\ &\quad + [\gamma_1 - \gamma_{01}] x_{ij,j+1}. \end{aligned} \quad (2.8.10)$$

Since $\gamma_k f_j(\rho) - \gamma_{0k} f_j(\rho_0) = \gamma_k [f_j(\rho) - f_j(\rho_0)] + f_j(\rho_0) [\gamma_k - \gamma_{0k}]$, $k = 0, 1$,

$$|\gamma_k f_j(\rho) - \gamma_{0k} f_j(\rho_0)| \leq r'_0 |f_j(\rho) - f_j(\rho_0)| + f_1(\rho_0) |\gamma_k - \gamma_{0k}|,$$

where we used the fact that $f_j(\rho)$ decreases with $j \geq 1$, with $r'_0 = r_0 + \max_{k=0,1} |\gamma_{0k}|$.

Now

$$\begin{aligned} |\mu_{ij}(\theta) - \mu_{ij}(\theta_0)| &\leq |\gamma_0 - \gamma_{00}| + (\tau_i - 1) [r'_0 |f_j(\rho) - f_j(\rho_0)| + C |\gamma_0 - \gamma_{00}|] \\ &\quad + |f_j(\rho) - f_j(\rho_0)| S_{i,j-1} + [r'_0 |f_j(\rho) - f_j(\rho_0)| \\ &\quad + C |\gamma_1 - \gamma_{01}|] |x_i^{(j-1)}| + |\gamma_1 - \gamma_{01}| |x_{ij,j+1}|. \end{aligned} \quad (2.8.11)$$

Let $0 < r < r_0$ and define $\rho_n(r) := \sup_{|\rho - \rho_0| \leq r} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} |f_j(\rho) - f_j(\rho_0)|$. Recall that we have previously shown that

$$\lim_{r \rightarrow 0} \limsup_n \rho_n(r) = 0. \quad (2.8.12)$$

With $\theta \in B_r(\theta_0)$, (2.8.11) implies that, for $1 \leq j \leq \tau_i$, $i \leq n$ and appropriately chosen constants,

$$\begin{aligned} |\mu_{ij}(\theta) - \mu_{ij}(\theta_0)| &\leq |\gamma_0 - \gamma_{00}| + C_0 \tau^{(n)} \max\{\rho_n(r), |\gamma_0 - \gamma_{00}|\} + \rho_n(r) \max_{1 \leq i \leq n} S_{i,\tau_i-1} \\ &\quad + C_1 \max\{\rho_n(r), |\gamma_1 - \gamma_{01}|\} \max_{1 \leq i \leq n} \sum_{l=1}^{\tau_i} |x_{ij,l+1}|. \end{aligned} \quad (2.8.13)$$

Since $\max\{|\gamma_0 - \gamma_{00}|, |\gamma_1 - \gamma_{01}|\} \leq \|\theta - \theta_0\| \leq r$,

$$\max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} |\mu_{ij}(\theta) - \mu_{ij}(\theta_0)| \leq \|\theta - \theta_0\| + C_0 \tau^{(n)} \max\{\rho_n(r), \|\theta - \theta_0\|\}$$

$$\begin{aligned}
& + \rho_n(r) \max_{1 \leq i \leq n} S_{i, \tau_i - 1} \\
& + C_1 \max\{\rho_n(r), \|\theta - \theta_0\|\} \max_{1 \leq i \leq n} \sum_{l=1}^{\tau_i} |x_{ij, l+1}|.
\end{aligned}$$

Then, a.s.,

$$\begin{aligned}
\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} |\mu_{ij}(\theta) - \mu_{ij}(\theta_0)| & = 0, \quad \text{if} \\
\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} \tau^{(n)} \max\{\rho_n(r), r\} & = 0, \quad (2.8.14)
\end{aligned}$$

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} \rho_n(r) \max_{1 \leq i \leq n} S_{i, \tau_i - 1} = 0, \quad (2.8.15)$$

$$\text{and } \lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} \max\{\rho_n(r), r\} \max_{1 \leq i \leq n} \sum_{l=1}^{\tau_i} |x_{ij, l+1}| = 0. \quad (2.8.16)$$

We now discuss these conditions. If (T2) holds, (2.8.14) holds because (2.8.12) holds. In this example there is a common bound for all $x_{ij, l+1}$, as they estimate the BMI of each individual i at the time of the j th event. Thus, in this case, (T2) also implies (2.8.16). If $\limsup_{i \rightarrow \infty} S_{i, \tau_i - 1} = \infty$, then (2.8.15) controls the rate of this convergence, by (2.8.12). Note that $S_{i, \tau_i - 1} < C_i$, so, if $\sup_{i \geq 1} C_i < \infty$, then (2.8.15) holds. \square

In the following examples, we will be using the exponential inequality:

$$|1 - \exp(x)| \leq |x| \exp(|x|), \quad x \in \mathbb{R}. \quad (2.8.17)$$

In what follows, we take $\theta \in B_r(\theta_0), r > 0$.

Example 2.8.2 This is Example 2.1.6 revisited. We have

$$\begin{aligned}
\frac{|\mu_{ij}(\theta) - \mu_{ij}(\theta_0)|}{V_{ij}(\theta)} & = \frac{|\exp(-\theta^T x_i) - \exp(-\theta_0^T x_i)|}{\exp(-\theta^T x_i)} \\
& = |1 - \exp(-(\theta_0 - \theta)^T x_i)| \\
& \leq |(\theta_0 - \theta)^T x_i| \exp(|(\theta_0 - \theta)^T x_i|) \\
& \leq r \max_{1 \leq i \leq n} \|x_i\| \exp(r \max_{1 \leq i \leq n} \|x_i\|),
\end{aligned}$$

where we used (2.8.17) in the first inequality. If

$$\limsup_{n \rightarrow \infty} \max_{1 \leq i \leq n} \|x_i\| < \infty, \quad (2.8.18)$$

then

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} r \max_{1 \leq i \leq n} \|x_i\| \exp(r \max_{1 \leq i \leq n} \|x_i\|) = 0,$$

which gives (2.7.10).

Next, using (2.8.17), we obtain

$$\begin{aligned} \frac{|V_{ij}^2(\theta) - V_{ij}^2(\theta_0)|}{V_{ij}^2(\theta)} &= \frac{|\exp(-2\theta^T x_i) - \exp(-2\theta_0^T x_i)|}{\exp(-2\theta^T x_i)} \\ &= |1 - \exp(-2(\theta_0 - \theta)^T x_i)| \\ &\leq |-2(\theta_0 - \theta)^T x_i| \exp(|-2(\theta_0 - \theta)^T x_i|) \\ &\leq 2r \max_{1 \leq i \leq n} \|x_i\| \exp(2r \max_{1 \leq i \leq n} \|x_i\|). \end{aligned}$$

Therefore, if (2.8.18) holds, then (2.7.11) holds.

Example 2.8.3 This is the linear model example (example (4), Section 2.6). Clearly, (2.7.11) holds. To prove (2.7.10), we write

$$\begin{aligned} \frac{|\mu_{ij}(\theta) - \mu_{ij}(\theta_0)|}{V_{ij}(\theta)} &= \frac{|\theta^T x_{ij} - \theta_0^T x_{ij}|}{1} \\ &\leq r \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} \|x_{ij}\|. \end{aligned}$$

Let us assume that

$$x_\infty := \limsup_{n \rightarrow \infty} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} \|x_{ij}\| < \infty. \quad (2.8.19)$$

Then

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} r \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} \|x_{ij}\| = 0,$$

which proves (2.7.10).

Example 2.8.4 This is example (5) of Section 2.6. We assume that (2.8.19) holds.

We start with

$$\begin{aligned} \frac{|\mu_{ij}(\theta) - \mu_{ij}(\theta_0)|}{V_{ij}(\theta)} &= \frac{|\exp(\theta^T x_{ij}) - \exp(\theta_0^T x_{ij})|}{\exp(\frac{1}{2}\theta^T x_{ij})} \\ &= \frac{\exp(\theta^T x_{ij}) |1 - \exp((\theta_0 - \theta)^T x_{ij})|}{\exp(\frac{1}{2}\theta^T x_{ij})} \end{aligned}$$

$$\begin{aligned}
&\leq r \|x_{ij}\| \exp(r \|x_{ij}\|) \exp\left(\frac{1}{2} \|\theta\| \|x_{ij}\|\right) \\
&\leq r \|x_{ij}\| \exp\left(\frac{3}{2}r \|x_{ij}\|\right) [\exp(\|x_{ij}\|)]^C, \quad (2.8.20)
\end{aligned}$$

with $C = \frac{1}{2} \|\theta_0\|$.

The first inequality is due to (2.8.17) and the second to $\|\theta\| \leq \|\theta_0\| + r$, since $\theta \in B_r(\theta_0)$. Applying $\limsup_{n \rightarrow \infty} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i}$ on both sides of (2.8.20) and using (2.8.19), we obtain

$$\limsup_{n \rightarrow \infty} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} \frac{|\mu_{ij}(\theta) - \mu_{ij}(\theta_0)|}{V_{ij}(\theta)} \leq r x_\infty \exp\left(\frac{3}{2}r x_\infty\right) [\exp(x_\infty)]^C.$$

Taking now the limit as $r \rightarrow 0$ on both sides of the above inequality gives (2.7.10).

To prove (2.7.11), we evaluate

$$\begin{aligned}
\frac{|V_{ij}^2(\theta) - V_{ij}^2(\theta_0)|}{V_{ij}^2(\theta)} &= \frac{|\exp(\theta^T x_{ij}) - \exp(\theta_0^T x_{ij})|}{\exp(\theta^T x_{ij})} \\
&= |1 - \exp((\theta_0 - \theta)^T x_{ij})| \\
&\leq r \|x_{ij}\| \exp(r \|x_{ij}\|),
\end{aligned}$$

by (2.8.17). Again, if (2.8.19) holds, then so does (2.7.11).

Example 2.8.5 This is example (6) of Section 2.6. As before, we assume (2.8.19).

To prove (2.7.10), we write

$$\begin{aligned}
\frac{|\mu_{ij}(\theta) - \mu_{ij}(\theta_0)|}{V_{ij}(\theta)} &= \frac{|\exp(\theta^T x_{ij}) - \exp(\theta_0^T x_{ij})|}{[1 + \exp(\theta_0^T x_{ij})] \exp\left(\frac{1}{2}\theta^T x_{ij}\right)} \\
&= \frac{\exp\left(\frac{1}{2}\theta^T x_{ij}\right) |1 - \exp[(\theta_0 - \theta)^T x_{ij}]|}{1 + \exp(\theta_0^T x_{ij})} \\
&\leq \frac{\exp\left(\frac{1}{2}\theta^T x_{ij}\right) |1 - \exp[(\theta_0 - \theta)^T x_{ij}]|}{\exp(\theta_0^T x_{ij})} \\
&= \exp\left(\frac{1}{2}(\theta - \theta_0)^T x_{ij}\right) \exp\left(-\frac{1}{2}\theta_0^T x_{ij}\right) |1 - \exp[(\theta_0 - \theta)^T x_{ij}]| \\
&\leq \exp\left(\frac{1}{2}r \|x_{ij}\|\right) [\exp(\|x_{ij}\|)]^C |1 - \exp[(\theta_0 - \theta)^T x_{ij}]|
\end{aligned}$$

$$\begin{aligned}
&\leq r \|x_{ij}\| \exp(r \|x_{ij}\|) \exp\left(\frac{1}{2}r \|x_{ij}\|\right) [\exp(\|x_{ij}\|)]^C \\
&= r \|x_{ij}\| \exp\left(\frac{3}{2}r \|x_{ij}\|\right) [\exp(\|x_{ij}\|)]^C, \tag{2.8.21}
\end{aligned}$$

where C was defined after (2.8.20).

Expression (2.8.21) was obtained by reasoning as we did to obtain (2.8.20). As in

Example 2.8.4, we apply $\limsup_{n \rightarrow \infty} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i}$ on both sides of (2.8.21), and then we

take the limit as $r \rightarrow 0$ to obtain (2.7.10).

For (2.7.11), we write:

$$\begin{aligned}
\frac{V_{ij}^2(\theta) - V_{ij}^2(\theta_0)}{V_{ij}^2(\theta)} &= 1 - \frac{\exp(\theta_0^T x_{ij})}{\exp(\theta^T x_{ij})} \left(\frac{1 + \exp(\theta^T x_{ij})}{1 + \exp(\theta_0^T x_{ij})} \right)^2 \\
&= 1 - \exp((\theta_0 - \theta)^T x_{ij}) \times I^2 \\
&= 1 - \exp((\theta_0 - \theta)^T x_{ij}) \\
&+ \exp((\theta_0 - \theta)^T x_{ij})(1 - I^2), \tag{2.8.22}
\end{aligned}$$

where

$$I := \frac{1 + \exp(\theta^T x_{ij})}{1 + \exp(\theta_0^T x_{ij})}.$$

Now

$$\begin{aligned}
\left| \frac{1 + \exp(\theta^T x_{ij})}{1 + \exp(\theta_0^T x_{ij})} - 1 \right| &= \left| \frac{\exp(\theta_0^T x_{ij}) - \exp(\theta^T x_{ij})}{1 + \exp(\theta_0^T x_{ij})} \right| \\
&= \frac{\exp(\theta_0^T x_{ij}) |1 - \exp((\theta - \theta_0)^T x_{ij})|}{1 + \exp(\theta_0^T x_{ij})} \\
&\leq \frac{\exp(\theta_0^T x_{ij}) |1 - \exp((\theta - \theta_0)^T x_{ij})|}{\exp(\theta_0^T x_{ij})} \\
&= |1 - \exp((\theta - \theta_0)^T x_{ij})| \\
&\leq |(\theta - \theta_0)^T x_{ij}| \exp(|(\theta - \theta_0)^T x_{ij}|) \\
&\leq r \|x_{ij}\| \exp(r \|x_{ij}\|). \tag{2.8.23}
\end{aligned}$$

If (2.8.19) holds, we apply $\limsup_{n \rightarrow \infty} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i}$ on both sides of (2.8.23). Then we take the limit as $r \rightarrow 0$ to conclude that the right hand side of (2.8.23) goes to zero.

Returning to (2.8.22), we have

$$\begin{aligned} \frac{|V_{ij}^2(\theta) - V_{ij}^2(\theta_0)|}{V_{ij}^2(\theta)} &\leq |1 - \exp((\theta_0 - \theta)^T x_{ij})| + \exp((\theta_0 - \theta)^T x_{ij})|1 - I^2| \\ &\leq r \|x_{ij}\| \exp(r \|x_{ij}\|) + \exp(r \|x_{ij}\|)|1 - I^2|, \end{aligned} \quad (2.8.24)$$

where we applied (2.8.17) to the last inequality. Since $1 - I^2 = (1 - I)(1 + I)$, and

$$1 + I = 1 + \frac{1 + \exp(\theta^T x_{ij})}{1 + \exp(\theta_0^T x_{ij})},$$

is bounded (see (2.8.19)), we now apply $\limsup_{n \rightarrow \infty} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i}$ on both sides of (2.8.24), and take the limit as $r \rightarrow 0$ to obtain (2.7.11).

Chapter 3

Strong Consistency

In this chapter we fix η_0 , the true, unknown value of the parameter η , and construct a sequence $\{\hat{\eta}_n\}_{n \geq 1}$ of strongly consistent estimators of η_0 . In Section 3.1 we introduce a general result, which guarantees the existence and *a.s.* convergence of a sequence of estimators of η_0 , defined as roots of EE. The asymptotic behaviour of terms of the derivative of the *e.f.* is treated in Section 3.2. Section 3.3 concludes the proof of strong consistency.

Throughout this chapter, all random variables are defined on a probability space, and almost sure relations are assumed to hold in the probability measure P_{η_0} .

3.1 General Results

As in [27], the spectral radius of a $p \times p$ matrix A is

$$\|A\| := \sup_{\|\lambda\|=1} |\lambda^T A \lambda|, \quad \lambda \in R^p.$$

We denote by $\lambda_{\max}(A)$ the largest eigenvalue of the matrix A . The operator norm $\|A\|$ and the Euclidean norm $\|A\|_E$ are defined in Appendix A of [9], for instance. It is shown there that these three norms are equivalent, and, depending on the situation, one can use the norm that is the most convenient to prove asymptotic results. Let

$q_n(\eta) = \sum_{i=1}^n u_i(\eta)$, where $\eta \in \mathcal{T} \subset R^{p+1}$ is a parameter, $u_i(\eta) \in R^{p+1}$ are random vectors, which are square integrable and continuously differentiable in η . Let $\mathcal{D}_n(\eta) = -\frac{\partial q_n(\eta)}{\partial \eta^T}$ be the $(p+1) \times (p+1)$ matrix of derivatives.

The following theorem gives sufficient conditions for the *a.s.* existence and strong consistency of a sequence of estimators of η_0 .

Theorem 3.1.1 *Assume that the following conditions hold a.s.:*

(LN) $n^{-1}q_n(\eta_0) \rightarrow 0$, *a.s.*, when $n \rightarrow \infty$,

(S) *There exist random variables $C_0 > 0$, $r_1 > 0$, a.s. and a random integer $n_1 \geq 1$, such that, with $B_{r_1}(\eta_0)$ defined as in (2.7.5), for all $\lambda \in R^{p+1}$, $\|\lambda\| = 1$,*

$$(i) \quad \inf_{n \geq n_1} \inf_{\eta \in B_{r_1}(\eta_0)} |\lambda^T \mathcal{D}_n(\eta) \lambda| > 0;$$

$$(ii) \quad \lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} \sup_{\eta \in B_r(\eta_0)} n^{-1} \|\mathcal{D}_n(\eta) - \mathcal{D}_n(\eta_0)\| = 0;$$

$$(iii) \quad \inf_{n \geq n_1} n^{-1} |\lambda^T \mathcal{D}_n(\eta_0) \lambda| \geq C_0.$$

Then, there exists a sequence of random vectors $\{\hat{\eta}_n\} \subset R^{p+1}$, and a random integer n_0 , such that:

$$(a) \quad P(q_n(\hat{\eta}_n) = 0, \text{ for all } n \geq n_0) = 1;$$

$$(b) \quad \hat{\eta}_n \rightarrow \eta_0 \quad \text{a.s.}, \quad n \rightarrow \infty.$$

Proof. The proof is identical to the proof of Theorem 4.2 in [3], once $\alpha_n^{1/2+\delta}$ is replaced by n . Conditions *S(i)* and *S(iii)* ensure nonsingularity and *S(ii)* the equicontinuity of the derivatives at η_0 . \square

The proof of Theorem 3.1.1 presupposes that the GEE's considered have unique roots, which may not be the case. A discussion of this more general situation is given in [23].

Remark 3.1.2 Condition (LN) replaces the unbiasedness of $q_n(\eta)$ at $\eta_0, n \geq 1$. It is satisfied if we can find an e.f. $q_{n,0}(\eta)$, for which (LN) holds at η_0 , and such that:

$$n^{-1}[q_n(\eta_0) - q_{n,0}(\eta_0)] \rightarrow 0 \quad a.s., \quad n \rightarrow \infty. \quad (3.1.1)$$

To apply the theorem, we could take $q_n(\eta) = (\hat{g}_{n,1}^{obs}(\eta), \hat{g}_{n,2}^{obs}(\eta))$ from (2.4.2), (2.4.4) and $q_{n,0}(\eta) = (g_{n,1}^{obs}(\eta), g_{n,2}^{obs}(\eta))$ from (2.4.1), (2.4.3) to obtain (3.1.1) and $\hat{\eta}_n$, such that $q_n(\hat{\eta}_n) = 0$, a.s., $n \geq n_0$ and $\hat{\eta}_n \rightarrow \eta_0$ a.s. To check condition (LN), by (2.4.15)-(2.4.16), (2.5.4)-(2.5.5) we just need to check that

$$n^{-1}g_{n,1}(\eta_0) \rightarrow 0 \quad \text{and} \quad n^{-1}g_{n,2}(\eta_0) \rightarrow 0.$$

both hold, a.s. This was discussed in Proposition 2.3.6.

However, in this dissertation we will be using Remark 3.1.2 differently. We will find a specific sequence of estimators $\hat{\eta}_n$ of η_0 , which is consistent and has a well-defined asymptotic distribution. This sequence is obtained in two steps, by contrast to the one-step procedure described in Remark 3.1.2.

Noting that $\hat{g}_{n,1}^{obs} \in R^p$ is a function of $\theta \in R^p$ alone (see (2.4.2)), we first “solve” the system of p -equations $\hat{g}_{n,1}^{obs}(\theta) = 0$, and obtain $\hat{\theta}_n, \hat{\theta}_n \rightarrow \theta_0$ a.s. We then place $\hat{\theta}_n$ in $\hat{g}_{n,2}^{obs}$ and “solve” $\hat{g}_{n,2}^{obs}(\hat{\theta}_n, \sigma^2) = 0$, to obtain $\hat{\sigma}_n^2 := \sigma_n^2(\hat{\theta}_n) \rightarrow \sigma_0^2$ a.s., when $n \rightarrow \infty$. Put together, $(\hat{\theta}_n, \hat{\sigma}_n^2)$ is a sequence of consistent estimators of η_0 . The advantage of this method is two-fold. Firstly, if the practitioner is interested only in the main regression parameter θ , he/she has to deal with a simpler, self-contained version of Theorem 3.1.1, as σ^2 , along with $\hat{g}_{n,2}^{obs}(\eta)$, can be completely ignored. Next, to obtain $\hat{\sigma}_n^2$, one has to “solve” a scalar equation. Secondly, from a technical point of view, fewer conditions have to be verified. Most conditions involve the partial derivatives of $(\hat{g}_{n,1}^{obs}(\eta), \hat{g}_{n,2}^{obs}(\eta))$ (see condition (S) of Theorem 3.1.1) and $\mathcal{D}_n(\eta)$ in the one-step procedure is a $(p+1) \times (p+1)$, non-sparse matrix. On the other hand, solving the initial problem in two steps requires first a $p \times p$ matrix of derivatives and then a scalar derivative. In all cases, the number of derivatives has to be multiplied by two,

as we have to account for both the non-imputed and the imputed terms of the *e.f.*, which strengthens the case for our two-step procedure.

As in [3], in order to apply Theorem 3.1.1, we provide conditions that are easier to verify. These conditions are imposed on the moduli of continuity of functions related to the derivative and on the matrix of covariates. Thus, properties of the derivatives rely on analytical properties of the functions $\mu, c_{ij}(\theta), V_{ij}(\theta), b_{ij}(\eta), i, j \geq 1$. In [3], only the analytical properties of μ are present, as $V_{ij}(\theta)$ is a function of $\mu_{ij}(\theta)$ (see (2.1.5)), and $c_{ij}(\theta)$ is a linear function there. As for $b_{ij}(\eta)$, it is not present in [3], since the EE for the overdispersion parameter is not considered there. Therefore, in this sense, the results presented here are more general than those presented in [3] and [27].

An additional complication we encountered is the necessity to deal with derivatives of two distinct sums: the sum representing the observed terms (*e.g.*, Section 3.2.1) and that corresponding to the imputed term (*e.g.*, Section 3.2.2). Our strategy was to prove results for the observed term first, which are valid regardless of what imputation method is used, and then adapt them for use with the imputed term.

3.2 Asymptotic Results for the Derivatives

The use of the analytical properties of the derivative in proving the consistency of the sequence of EE was introduced in [27] and expanded on in [2]-[3]. Technically, the main difference between earlier results and ours is that we have to deal with the additional factor $c_{ij}(\theta)$ (see (2.1.5)), which is not necessarily linear in $\theta \in R^p$. Another major difference is the presence of censoring in our work.

This section gives sufficient conditions for $(S)(ii)$ in Theorem 3.1.1 to hold. It also gives preliminary conditions for $(S)(i)$ and $(S)(iii)$ to hold. In Sections 3.2.1-3.2.3 we examine the behaviour of the derivative of $\hat{g}_{n,1}^{obs}$, whereas in Section 3.2.4, we deal with the derivative of $\hat{g}_{n,2}^{obs}$. The nonimputed and the imputed terms are treated

separately.

Following ideas from [27], we decompose the derivative into three types of terms. The first is the leading term in the asymptotics and corresponds to the design matrix (3.2.1) in the linear regression case. The normalized second-type terms converge to 0, and the third contain the residuals. Each of these terms is treated separately, and may be further decomposed into simpler terms. We note that we considerably relaxed the hypotheses required in [3] to prove $(S)(ii)$ for the third type of terms.

To simplify notation, C denotes a generic constant, which may differ from one case to another. We now introduce some notation, in conjunction with Example 2.1.4.

Notation:

With $\mu : R \rightarrow R$ and μ', μ'' its first two derivatives, let $\dot{\mu}_{ij}(\theta) := \mu'(c_{ij}^T(\theta)x_{ij})$, $\ddot{\mu}_{ij}(\theta) := \mu''(c_{ij}^T(\theta)x_{ij})$. Let X_i denote the $\tau_i \times q$ matrix of covariates, with (a, b) -entry $(X_i)_{ab} := x_{ia,b}$, $a = 1, 2, \dots, \tau_i$, $b = 1, 2, \dots, q$, where q is the maximum number of covariates over all individuals, assumed to be finite. Now $X_i^T X_i$ is a $q \times q$ matrix, and we use the following notation, for the design matrix of covariates.

$$D_n := \sum_{i=1}^n X_i^T X_i, \quad n \geq 1. \quad (3.2.1)$$

3.2.1 The Derivative of the Non-imputed Terms of $\hat{g}_{n,1}^{obs}$

In this section we prove several lemmas, which, when put together, provide sufficient conditions for $S(ii)$ to hold for the non-imputed part of $\hat{g}_{n,1}^{obs}$.

We first calculate the derivative of

$$\begin{aligned} \sum_{i=1}^n \sum_{j=1}^{\tau_i} f_{ij}(\theta) Z_{ij}(\theta) I_{ij}^{obs} &= \sum_{i=1}^n \sum_{j=1}^{\tau_i} \frac{\partial \mu_{ij}(\theta)}{\partial \theta} V_{ij}^{-2}(\theta) [Y_{ij} - \mu_{ij}(\theta)] I_{ij}^{obs} \\ &:= g_{n,1}^{[1]}(\theta) + g_{n,1}^{[2]}(\theta), \end{aligned} \quad (3.2.2)$$

where

$$g_{n,1}^{[1]}(\theta) := \sum_{i=1}^n \sum_{j=1}^{\tau_i} \frac{\partial \mu_{ij}(\theta)}{\partial \theta} V_{ij}^{-2}(\theta) [\mu_{ij}(\theta_0) - \mu_{ij}(\theta)] I_{ij}^{obs}, \quad (3.2.3)$$

$$g_{n,1}^{[2]}(\theta) := \sum_{i=1}^n \sum_{j=1}^{\tau_i} \frac{\partial \mu_{ij}(\theta)}{\partial \theta} V_{ij}^{-2}(\theta) [Y_{ij} - \mu_{ij}(\theta_0)] I_{ij}^{obs}. \quad (3.2.4)$$

Taking the derivative in (3.2.3) gives rise to terms corresponding to the first two types mentioned above.

$$-\frac{\partial g_{n,1}^{[1]}(\theta)}{\partial \theta^T} = H_n(\theta) - B_n^{[1]}(\theta) - B_n^{[2]}(\theta) - B_n^{[3]}(\theta), \quad (3.2.5)$$

where the leading term in the asymptotic is

$$H_n(\theta) := \sum_{i=1}^n \sum_{j=1}^{\tau_i} \frac{\partial \mu_{ij}(\theta)}{\partial \theta} V_{ij}^{-2}(\theta) \frac{\partial \mu_{ij}(\theta)}{\partial \theta^T} I_{ij}^{obs}, \quad (3.2.6)$$

$$B_n^{[1]}(\theta) := \sum_{i=1}^n \sum_{j=1}^{\tau_i} V_{ij}^{-2}(\theta) \ddot{\mu}_{ij}(\theta) \frac{\partial c_{ij}^T(\theta)}{\partial \theta} x_{ij} x_{ij}^T \frac{\partial c_{ij}(\theta)}{\partial \theta^T} [\mu_{ij}(\theta_0) - \mu_{ij}(\theta)] I_{ij}^{obs}, \quad (3.2.7)$$

$$B_n^{[2]}(\theta) := \sum_{i=1}^n \sum_{j=1}^{\tau_i} \ddot{\mu}_{ij}(\theta) \frac{\partial^2 [c_{ij}^T(\theta) x_{ij}]}{\partial \theta^T \partial \theta} V_{ij}^{-2}(\theta) [\mu_{ij}(\theta_0) - \mu_{ij}(\theta)] I_{ij}^{obs}, \quad (3.2.8)$$

$$B_n^{[3]}(\theta) := -2 \sum_{i=1}^n \sum_{j=1}^{\tau_i} \frac{\partial \mu_{ij}(\theta)}{\partial \theta} V_{ij}^{-3}(\theta) \frac{\partial V_{ij}(\theta)}{\partial \theta^T} [\mu_{ij}(\theta_0) - \mu_{ij}(\theta)] I_{ij}^{obs}, \quad (3.2.9)$$

$$B_n(\theta) = \sum_{i=1}^3 B_n^{[i]}(\theta). \quad (3.2.10)$$

We note that $B_n^{[k]}(\theta_0) = 0$, $k = 1, 2, 3$.

For the second term in (3.2.2),

$$\frac{\partial g_{n,1}^{[2]}(\theta)}{\partial \theta^T} = \mathcal{E}_n(\theta) := \mathcal{E}_n^{[1]}(\theta) + \mathcal{E}_n^{[2]}(\theta), \quad (3.2.11)$$

where

$$\mathcal{E}_n^{[1]}(\theta) := \sum_{i=1}^n \sum_{j=1}^{\tau_i} \frac{\partial^2 \mu_{ij}(\theta)}{\partial \theta^T \partial \theta} V_{ij}^{-2}(\theta) [Y_{ij} - \mu_{ij}(\theta_0)] I_{ij}^{obs}, \quad (3.2.12)$$

$$\mathcal{E}_n^{[2]}(\theta) := -2 \sum_{i=1}^n \sum_{j=1}^{\tau_i} \frac{\partial \mu_{ij}(\theta)}{\partial \theta} V_{ij}^{-3}(\theta) \frac{\partial V_{ij}(\theta)}{\partial \theta^T} [Y_{ij} - \mu_{ij}(\theta_0)] I_{ij}^{obs}. \quad (3.2.13)$$

We will evaluate the difference $\mathcal{D}_{n,1}^{obs}(\theta) - \mathcal{D}_{n,1}^{obs}(\theta_0)$ by way of three lemmas, where

$$\mathcal{D}_{n,1}^{obs}(\theta) = -\frac{\partial g_{n,1}^{[1]}(\theta)}{\partial \theta^T} - \frac{\partial g_{n,1}^{[2]}(\theta)}{\partial \theta^T}. \quad (3.2.14)$$

Before we proceed, we introduce some notation.

With $i, j \geq 1$, let

$$\nu_{ij}^T(\theta) := \dot{\mu}(c_{ij}^T(\theta) x_{ij}) \frac{\partial c_{ij}^T(\theta)}{\partial \theta},$$

$$d_{ij}^T(\theta) := \nu_{ij}^T(\theta) - \nu_{ij}^T(\theta_0). \quad (3.2.15)$$

For an arbitrary $r > 0$, let

$$\nu_n(r) := \sup_{\theta, \theta' \in B_r(\theta_0)} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} \| V_{ij}^{-1}(\theta') \nu_{ij}(\theta) \|, \quad (3.2.16)$$

$$d_n(r) := \sup_{\theta \in B_r(\theta_0)} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} \| V_{ij}^{-1}(\theta) d_{ij}(\theta) \|. \quad (3.2.17)$$

Lemma 3.2.1 *Assume that, a.s.*

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} d_n(r) \nu_n(r) \lambda_{\max}(D_n) = 0, \quad (3.2.18)$$

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} \eta_n(r) [\nu_n(r)]^2 \lambda_{\max}(D_n) = 0. \quad (3.2.19)$$

Then

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} \sup_{\theta \in B_r(\theta_0)} \| \| H_n(\theta) - H_n(\theta_0) \| \| = 0, \quad a.s. \quad (3.2.20)$$

Proof. We write

$$H_n(\theta) - H_n(\theta_0) = H_n^{[1]}(\theta) + H_n^{[2]}(\theta) + H_n^{[3]}(\theta), \quad (3.2.21)$$

where

$$H_n^{[1]}(\theta) := \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left(\frac{\partial \mu_{ij}(\theta)}{\partial \theta} - \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} \right) V_{ij}^{-2}(\theta) \frac{\partial \mu_{ij}(\theta)}{\partial \theta^T} I_{ij}^{obs},$$

$$H_n^{[2]}(\theta) := \sum_{i=1}^n \sum_{j=1}^{\tau_i} \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} (V_{ij}^{-2}(\theta) - V_{ij}^{-2}(\theta_0)) \frac{\partial \mu_{ij}(\theta)}{\partial \theta^T} I_{ij}^{obs},$$

$$H_n^{[3]}(\theta) := \sum_{i=1}^n \sum_{j=1}^{\tau_i} \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} V_{ij}^{-2}(\theta_0) \left(\frac{\partial \mu_{ij}(\theta)}{\partial \theta^T} - \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta^T} \right) I_{ij}^{obs}.$$

Let $r > 0$ be arbitrarily chosen. We first examine the spectral radius of $H_n^{[1]}(\theta)$, $\theta \in B_r(\theta_0)$. Note that

$$H_n^{[1]}(\theta) = \sum_{i=1}^n \Delta_i^T(\theta) A_i(\theta), \quad \text{where, with } i \geq 1,$$

$$[\Delta_i^T(\theta)]_{kj} := V_{ij}^{-1}(\theta) \left[\frac{\partial \mu_{ij}(\theta)}{\partial \theta_k} - \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta_k} \right],$$

$$[A_i(\theta)]_{jl} := V_{ij}^{-1}(\theta) \frac{\partial \mu_{ij}(\theta)}{\partial \theta_l} I_{ij}^{obs}, \quad 1 \leq j \leq \tau_i, \quad k, l = 1, 2, \dots, p.$$

Let λ be a $p \times 1$ vector of norm 1. By (2.7.1) with $m = \tau_i$,

$$\begin{aligned} \left| \lambda^T \sum_{i=1}^n \Delta_i^T(\theta) A_i(\theta) \lambda \right| &= \left| \sum_{i=1}^n \lambda^T \Delta_i^T(\theta) A_i(\theta) \lambda \right| \\ &\leq I_1^{1/2} \times I_2^{1/2}, \end{aligned} \quad (3.2.22)$$

$I_1 = \sum_{i=1}^n \lambda^T \Delta_i^T(\theta) \Delta_i(\theta) \lambda$, $I_2 = \sum_{i=1}^n \lambda^T A_i^T(\theta) A_i(\theta) \lambda$. Expanding I_1 , we obtain

$$\begin{aligned} |I_1| &= \left| \sum_{i=1}^n \sum_{j=1}^{\tau_i} V_{ij}^{-2}(\theta) \lambda^T d_{ij}^T(\theta) x_{ij} x_{ij}^T d_{ij}(\theta) \lambda \right| \\ &\leq \sum_{i=1}^n \sum_{j=1}^{\tau_i} V_{ij}^{-2}(\theta) \left| \lambda^T d_{ij}^T(\theta) x_{ij} \right|^2 \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{i=1}^n \sum_{j=1}^{\tau_i} \|x_{ij}\|^2 \|V_{ij}^{-1}(\theta) d_{ij}(\theta)\|^2 \\
&\leq d_n^2(r) \sum_{i=1}^n \sum_{j=1}^{\tau_i} \sum_{k=1}^q x_{ij,k}^2 \\
&\leq d_n^2(r) \sum_{i=1}^n \text{trace}(X_i^T X_i).
\end{aligned}$$

The second last inequality is due to (3.2.17) and the last inequality follows from the fact that

$$\sum_{j=1}^{\tau_i} \|x_{ij}\|^2 = \sum_{j=1}^{\tau_i} \sum_{k=1}^q x_{ij,k}^2 = \text{trace}(X_i^T X_i).$$

Hence

$$\begin{aligned}
|I_1| &\leq d_n^2(r) \text{trace}(D_n) \\
&\leq C^2 d_n^2(r) \lambda_{\max}(D_n),
\end{aligned}$$

where C is a positive constant. The last inequality is due to the fact that $\text{trace}(D_n)$ is also the sum of all eigenvalues of D_n , which are all positive. Consequently,

$$I_1^{1/2} \leq C d_n(r) \lambda_{\max}^{1/2}(D_n). \quad (3.2.23)$$

For

$$I_2 = \sum_{i=1}^n \sum_{j=1}^{\tau_i} \lambda^T \frac{\partial \mu_{ij}(\theta)}{\partial \theta} V_{ij}^{-2}(\theta) \frac{\partial \mu_{ij}(\theta)}{\partial \theta^T} \lambda I_{ij}^{obs},$$

we have the string of inequalities:

$$\begin{aligned}
|I_2| &\leq \sum_{i=1}^n \sum_{j=1}^{\tau_i} \lambda^T \frac{\partial \mu_{ij}(\theta)}{\partial \theta} V_{ij}^{-2}(\theta) \frac{\partial \mu_{ij}(\theta)}{\partial \theta^T} \lambda \\
&\leq \nu_n^2(r) \text{trace}(D_n) \\
&\leq C^2 \nu_n^2(r) \lambda_{\max}(D_n).
\end{aligned}$$

The second inequality is due to (3.2.16) and the third follows by the same reasoning that led to (3.2.23). Therefore, we obtain,

$$I_2^{1/2} \leq C \nu_n(r) \lambda_{\max}^{1/2}(D_n). \quad (3.2.24)$$

Combining (3.2.22)-(3.2.24), we can write the inequality,

$$\|H_n^{[1]}(\theta)\| \leq C d_n(r) \nu_n(r) \lambda_{\max}(D_n). \quad (3.2.25)$$

To obtain a bound for the spectral radius of $H_n^{[2]}(\theta)$ (defined after (3.2.21)), we write:

$$|\lambda^T H_n^{[2]}(\theta) \lambda| \leq \eta_n(r) \sum_{i=1}^n \sum_{j=1}^{\tau_i} \lambda^T \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} V_{ij}^{-2}(\theta_0) \frac{\partial \mu_{ij}(\theta)}{\partial \theta^T} \lambda I_{ij}^{obs},$$

with $\eta_n(r)$ defined by (2.7.9). Next, by (2.7.1)

$$|\lambda^T H_n^{[2]}(\theta) \lambda| \leq \eta_n(r) I_3^{1/2} I_4^{1/2},$$

where

$$I_3 := \sum_{i=1}^n \sum_{j=1}^{\tau_i} \lambda^T \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} V_{ij}^{-2}(\theta_0) \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta^T} \lambda I_{ij}^{obs},$$

$$I_4 := \sum_{i=1}^n \sum_{j=1}^{\tau_i} \lambda^T \frac{\partial \mu_{ij}(\theta)}{\partial \theta} V_{ij}^{-2}(\theta_0) \frac{\partial \mu_{ij}(\theta)}{\partial \theta^T} \lambda I_{ij}^{obs}.$$

Since $I_{ij}^{obs} \leq 1$ a.s., I_3 and I_4 have the same upper bound as I_2 . Therefore, with $\theta \in B_r(\theta_0)$,

$$\|H_n^{[2]}(\theta)\| \leq C \eta_n(r) [\nu_n(r)]^2 \lambda_{\max}(D_n). \quad (3.2.26)$$

We now turn to $\|H_n^{[3]}(\theta)\|$. From the definitions following (3.2.21), the upper bound of $\|H_n^{[1]}(\theta)\|$ is also appropriate here. Therefore

$$\|H_n(\theta) - H_n(\theta_0)\| \leq \nu_n(r) [C_{1,3} d_n(r) + C_2 \eta_n(r) \nu_n(r)] \lambda_{\max}(D_n).$$

The conclusion of the lemma follows now from the hypotheses. \square

We introduce further notation.

$$k_n^{[1]}(r) := \sup_{\theta, \theta' \in B_r(\theta_0)} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} \frac{|\dot{\mu}_{ij}(\theta)|}{V_{ij}(\theta')}, \quad (3.2.27)$$

$$\tilde{k}_n^{[1]}(r) := \max_{i=1,2} [k_n^{[1]}(r)]^i, \quad (3.2.28)$$

$$k_n^{[2]}(r) := \sup_{\theta \in B_r(\theta_0)} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} \frac{|\ddot{\mu}_{ij}(\theta)|}{V_{ij}(\theta)}. \quad (3.2.29)$$

Let $\dot{c}_{ij}(r)$ be the $p \times q$ matrix with (a, b) entry $\sup_{\theta \in B_r(\theta_0)} \left| \frac{\partial c_{ij,b}(\theta)}{\partial \theta_a} \right|$, $1 \leq a \leq p$, $1 \leq b \leq q$, and

$$c_n^{[1]}(r) := \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} \|\dot{c}_{ij}(r)\|, \quad (3.2.30)$$

$$\tilde{c}_n^{[1]}(r) := \max_{i=1,2,3} [c_n^{[1]}(r)]^i. \quad (3.2.31)$$

Similarly

$$c_n^{[2]}(r) := \sup_{\theta \in B_r(\theta_0)} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i, h \leq q} \left\| \frac{\partial^2 c_{ij,h}(\theta)}{\partial \theta^T \partial \theta} \right\|, \quad (3.2.32)$$

where $\frac{\partial^2 c_{ij,h}(\theta)}{\partial \theta^T \partial \theta}$ is a $p \times p$ matrix, for every $1 \leq h \leq q$ and $\left\| \frac{\partial^2 c_{ij,h}(\theta)}{\partial \theta^T \partial \theta} \right\|$ is the h^{th} component of a q -dimensional vector. Furthermore, let

$$v_n^{[1]}(r) := \sup_{\theta \in B_r(\theta_0)} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} \left\| V_{ij}^{-1}(\theta) \frac{\partial V_{ij}(\theta)}{\partial \theta} \right\|. \quad (3.2.33)$$

Remark 3.2.2 With $d_n(r)$ defined in (3.2.17) and $\delta_n^{[1]}(r)$ defined in (2.7.7), we have the inequality

$$d_n(r) \leq \delta_n^{[1]}(r) c_n^{[1]}(r) + k_n^{[1]}(r) c_n^{[2]}(r) r.$$

Lemma 3.2.3 Assume that, a.s.

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} r \tilde{k}_n^{[1]}(r) \tilde{c}_n^{[1]}(r) \{k_n^{[2]}(r) \lambda_{\max}^{1/2}(D_n) + c_n^{[2]}(r) + v_n^{[1]}(r)\} \lambda_{\max}(D_n) = 0, \quad (3.2.34)$$

then

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} \sup_{\theta \in B_r(\theta_0)} \|B_n(\theta)\| = 0 \quad a.s. \quad (3.2.35)$$

Proof. Since, by (3.2.7)

$$B_n^{[1]}(\theta) = \sum_{i=1}^n \sum_{j=1}^{\tau_i} V_{ij}^{-2}(\theta) \ddot{\mu}_{ij}(\theta) \frac{\partial c_{ij}^T(\theta)}{\partial \theta} x_{ij} x_{ij}^T \frac{\partial c_{ij}(\theta)}{\partial \theta^T} [\mu_{ij}(\theta_0) - \mu_{ij}(\theta)] I_{ij}^{obs},$$

$$\begin{aligned} |\lambda^T B_n^{[1]}(\theta) \lambda| &\leq \sum_{i=1}^n \sum_{j=1}^{\tau_i} V_{ij}^{-1}(\theta) |\ddot{\mu}_{ij}(\theta)| \lambda^T \frac{\partial c_{ij}^T(\theta)}{\partial \theta} x_{ij} x_{ij}^T \frac{\partial c_{ij}(\theta)}{\partial \theta^T} \lambda |\mu_{ij}(\theta_0) - \mu_{ij}(\theta)| V_{ij}^{-1}(\theta) \\ &\leq k_n^{[2]}(r) I_5 I_6^{1/2}, \quad \text{where} \end{aligned}$$

$$I_5 := \sum_{i=1}^n \sum_{j=1}^{\tau_i} \lambda^T \frac{\partial c_{ij}^T(\theta)}{\partial \theta} x_{ij} x_{ij}^T \frac{\partial c_{ij}(\theta)}{\partial \theta^T} \lambda,$$

$$I_6 := \sum_{i=1}^n \sum_{j=1}^{\tau_i} |\mu_{ij}(\theta_0) - \mu_{ij}(\theta)|^2 V_{ij}^{-2}(\theta).$$

We have

$$\begin{aligned} I_5 &\leq \sum_{i=1}^n \sum_{j=1}^{\tau_i} \|\lambda\|^2 \left\| \frac{\partial c_{ij}(\theta)}{\partial \theta} \right\|^2 \|x_{ij}\|^2 \\ &\leq C [c_n^{[1]}(r)]^2 \lambda_{\max}(D_n). \end{aligned} \tag{3.2.36}$$

From the mean value theorem, it follows that

$$I_6 = \sum_{i=1}^n \sum_{j=1}^{\tau_i} V_{ij}^{-2}(\theta) [\dot{\mu}_{ij}(\bar{\theta})]^2 \left| [c_{ij}(\theta_0) - c_{ij}(\theta)]^T x_{ij} \right|^2,$$

for some point $\bar{\theta}$, $\|\bar{\theta} - \theta_0\| \leq \|\theta - \theta_0\|$. Next,

$$I_6 \leq [k_n^{[1]}(r)]^2 \sum_{i=1}^n \sum_{j=1}^{\tau_i} \|c_{ij}(\theta_0) - c_{ij}(\theta)\|^2 \|x_{ij}\|^2.$$

$$\begin{aligned} \text{Now } \|c_{ij}(\theta_0) - c_{ij}(\theta)\|^2 &= \sum_{h=1}^q [c_{ij,h}(\theta_0) - c_{ij,h}(\theta)]^2 \\ &= \sum_{h=1}^q \sum_{k=1}^p \left[\frac{\partial c_{ij,h}(\bar{\theta}_{ij,h,k})}{\partial \theta_k} \right]^2 [\theta_k - \theta_{0k}]^2 \\ &\leq \sum_{h=1}^q \sum_{k=1}^p \left[\frac{\partial c_{ij,h}(\bar{\theta}_{ij,h,k})}{\partial \theta_k} \right]^2 \|\theta - \theta_0\|^2 \end{aligned}$$

$$\leq \sum_{h=1}^q \sum_{k=1}^p \left[\sup_{\theta \in B_r(\theta_0)} \left| \frac{\partial c_{ij,h}(\theta)}{\partial \theta_k} \right| \right]^2 r^2, \quad \text{since } \theta \in B_r(\theta_0).$$

Therefore,

$$\begin{aligned} I_6 &\leq Cr^2 [k_n^{[1]}(r)]^2 \sum_{i=1}^n \sum_{j=1}^{\tau_i} \| \dot{c}_{ij}(r) \|^2 \| x_{ij} \|^2 \\ &\leq C [rk_n^{[1]}(r) c_n^{[1]}(r)]^2 \lambda_{\max}(D_n). \end{aligned} \quad (3.2.37)$$

Finally,

$$\sup_{\theta \in B_r(\theta_0)} ||| B_n^{[1]}(\theta) ||| \leq Cr k_n^{[1]}(r) k_n^{[2]}(r) [c_n^{[1]}(r)]^3 (\lambda_{\max}(D_n))^{3/2}, \quad (3.2.38)$$

which gives, using (3.2.34)

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} \sup_{\theta \in B_r(\theta_0)} ||| B_n^{[1]}(\theta) ||| = 0.$$

We now turn to $B_n^{[2]}(\theta)$ defined in (3.2.8). To evaluate the spectral radius, we write, using (2.7.1)

$$\begin{aligned} |\lambda^T B_n^{[2]}(\theta) \lambda| &\leq \left[\sum_{i=1}^n \sum_{j=1}^{\tau_i} [V_{ij}^{-1}(\theta) \dot{\mu}_{ij}(\theta)]^2 [\lambda^T \omega_{ij}(\theta) \lambda]^2 \right]^{1/2} I_6^{1/2} \\ &:= I_7^{1/2} I_6^{1/2}. \end{aligned} \quad (3.2.39)$$

Here $\omega_{ij}(\theta)$ is the $p \times p$ matrix with (a, b) -entry $\sum_{h=1}^q \frac{\partial^2 c_{ij,h}^T(\theta)}{\partial \theta_a \partial \theta_b} x_{ij,h}$, and I_6 was defined previously. Now

$$\begin{aligned} [\lambda^T \omega_{ij}(\theta) \lambda]^2 &= \left[\sum_{h=1}^q \sum_{a,b=1}^p \lambda_a \frac{\partial^2 c_{ij,h}(\theta)}{\partial \theta_a \partial \theta_b} \lambda_b x_{ij,h} \right]^2 \\ &\leq \left[\sum_{h=1}^q \left| \lambda^T \frac{\partial^2 c_{ij,h}(\theta)}{\partial \theta^T \partial \theta} \lambda \right| |x_{ij,h}| \right]^2 \\ &\leq \left[\sum_{h=1}^q \left\| \frac{\partial^2 c_{ij,h}(\theta)}{\partial \theta^T \partial \theta} \right\| |x_{ij,h}| \right]^2 \\ &\leq [c_n^{[2]}(r)]^2 \| x_{ij} \|^2. \end{aligned}$$

In the last inequality, we use the expression (3.2.32). Returning to I_7 , and using expression (3.2.27), we have

$$I_7 \leq C[k_n^{[1]}(r)c_n^{[2]}(r)]^2 \lambda_{\max}(D_n). \quad (3.2.40)$$

Combining (3.2.37), (3.2.39) and (3.2.40), we obtain

$$\sup_{\theta \in B_r(\theta_0)} |||B_n^{[2]}(\theta)||| \leq Cr[k_n^{[1]}(r)]^2 c_n^{[1]}(r)c_n^{[2]}(r) \lambda_{\max}(D_n). \quad (3.2.41)$$

Then, by (3.2.34)

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} \sup_{\theta \in B_r(\theta_0)} |||B_n^{[2]}(\theta)||| = 0.$$

To complete the proof of Lemma 3.2.3, we find a bound for the spectral radius of $B_n^{[3]}(\theta)$ in (3.2.9).

$$\begin{aligned} |\lambda^T B_n^{[3]}(\theta) \lambda| &\leq 2 \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left| \lambda^T \frac{\partial \mu_{ij}(\theta)}{\partial \theta} V_{ij}^{-2}(\theta) \frac{\partial V_{ij}(\theta)}{\partial \theta^T} \lambda \right| \frac{|\mu_{ij}(\theta_0) - \mu_{ij}(\theta)|}{V_{ij}(\theta)} \\ &\leq C \left(\sum_{i=1}^n \sum_{j=1}^{\tau_i} \left[\lambda^T \frac{\partial \mu_{ij}(\theta)}{\partial \theta} V_{ij}^{-2}(\theta) \frac{\partial V_{ij}(\theta)}{\partial \theta^T} \lambda \right]^2 \right)^{1/2} I_6^{1/2}, \end{aligned}$$

where we used (2.7.1) with $j = 1, 2, \dots, \tau_i$,

$$(a_i^T)_j = \lambda^T \frac{\partial \mu_{ij}(\theta)}{\partial \theta} V_{ij}^{-2}(\theta) \frac{\partial V_{ij}(\theta)}{\partial \theta^T} \lambda \quad \text{and} \quad (b_i)_j = \frac{|\mu_{ij}(\theta_0) - \mu_{ij}(\theta)|}{V_{ij}(\theta)}.$$

$$\begin{aligned} \text{Let } I_8 &:= \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left[\lambda^T \frac{\partial \mu_{ij}(\theta)}{\partial \theta} V_{ij}^{-2}(\theta) \frac{\partial V_{ij}(\theta)}{\partial \theta^T} \lambda \right]^2 \\ &\leq \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left\| V_{ij}^{-1}(\theta) \frac{\partial \mu_{ij}(\theta)}{\partial \theta} \right\|^2 \left\| V_{ij}^{-1}(\theta) \frac{\partial V_{ij}(\theta)}{\partial \theta^T} \right\|^2. \end{aligned}$$

Using (3.2.33), the chain rule, (3.2.27) and (3.2.30) and , we have

$$\begin{aligned} I_8 &\leq [v_n^{[1]}(r)]^2 \sum_{i=1}^n \sum_{j=1}^{\tau_i} \|V_{ij}^{-1}(\theta) \dot{\mu}_{ij}(\theta)\|^2 \|\dot{c}_{ij}(\theta) x_{ij}\|^2 \\ &\leq C[k_n^{[1]}(r)v_n^{[1]}(r)c_n^{[1]}(r)]^2 \sum_{i=1}^n \sum_{j=1}^{\tau_i} \|x_{ij}\|^2. \end{aligned}$$

Then

$$I_8^{1/2} \leq C k_n^{[1]}(r) v_n^{[1]}(r) c_n^{[1]}(r) \lambda_{\max}^{1/2}(D_n).$$

Using the upper bound of I_6 in (3.2.37), we obtain

$$\sup_{\theta \in B_r(\theta_0)} |||B_n^{[3]}(\theta)||| \leq Cr [k_n^{[1]}(r)]^2 v_n^{[1]}(r) (c_n^{[1]}(r))^2 \lambda_{\max}(D_n). \quad (3.2.42)$$

By (3.2.34),

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} \sup_{\theta \in B_r(\theta_0)} |||B_n^{[3]}(\theta)||| = 0.$$

This completes the proof of Lemma 3.2.3. \square

To formulate the next result, we need the following definitions:

$$v_n(r) := \sup_{\theta, \theta' \in B_r(\theta_0)} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} \frac{V_{ij}(\theta)}{V_{ij}(\theta')}, \quad (3.2.43)$$

$$\tilde{v}_n(r) := \max_{i=1,2} [v_n(r)]^i, \quad (3.2.44)$$

$$s_n(r) := \sup_{\theta, \theta' \in B_r(\theta_0)} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} \frac{|V_{ij}(\theta) - V_{ij}(\theta')|}{V_{ij}(\theta)}, \quad (3.2.45)$$

$$k_n^{[3]}(r) := \sup_{\theta, \theta' \in B_r(\theta_0)} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} \frac{|\ddot{\mu}_{ij}(\theta)|}{V_{ij}(\theta')}, \quad (3.2.46)$$

$$c_n^{[3]}(r) := \sup_{\theta \in B_r(\theta_0)} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} \max_{a,b,l \leq p; h \leq q} \left| \frac{\partial^3 c_{ij,h}(\theta)}{\partial \theta_a \partial \theta_b \partial \theta_l} \right|, \quad (3.2.47)$$

$$w_n^{[1]}(r) := \sup_{\theta \in B_r(\theta_0)} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} \left\| V_{ij}^{-1}(\theta_0) \left(\frac{\partial V_{ij}(\theta)}{\partial \theta^T} - \frac{\partial V_{ij}(\theta_0)}{\partial \theta^T} \right) \right\|. \quad (3.2.48)$$

Note that $v_n(r) \geq 1$, for all $n \geq 1$.

Remark 3.2.4 *We have*

$$\frac{V_{ij}^2(\theta_0) - V_{ij}^2(\theta)}{V_{ij}^2(\theta)} = \frac{[V_{ij}(\theta_0) - V_{ij}(\theta)] [V_{ij}(\theta_0) + V_{ij}(\theta)]}{V_{ij}(\theta) V_{ij}(\theta)}$$

$$\leq s_n(r)(v_n(r) + 1) < 2s_n(r)v_n(r).$$

Note that this implies $\eta_n(r) \leq 2s_n(r)v_n(r)$, with $\eta_n(r)$ defined in (2.7.9).

$$\begin{aligned} \frac{V_{ij}^3(\theta_0) - V_{ij}^3(\theta)}{V_{ij}^3(\theta)} &= \frac{[V_{ij}(\theta_0) - V_{ij}(\theta)] [V_{ij}^2(\theta_0) + V_{ij}(\theta_0)V_{ij}(\theta) + V_{ij}^2(\theta)]}{V_{ij}(\theta) V_{ij}^2(\theta)} \\ &\leq s_n(r)(v_n^2(r) + v_n(r) + 1) < Cs_n(r)\tilde{v}_n(r). \end{aligned} \quad (3.2.49)$$

The conditions in Lemma 3.2.5 below are a lot weaker than the corresponding conditions of Lemma 4.9 of [3].

Lemma 3.2.5 *Assume that (T1) and (2.7.10)-(2.7.11) hold. In addition, assume that, a.s.*

$$\begin{aligned} \lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1/2} \lambda_{\max}^{1/2}(D_n) \left\{ k_n^{[1]}(r) \tilde{c}_n^{[1]}(r) [w_n^{[1]}(r) + s_n(r) \tilde{v}_n(r) v_n^{[1]}(r)] \right. \\ \left. + \tilde{v}_n(r) v_n^{[1]}(r) d_n(r) \right\} = 0, \end{aligned} \quad (3.2.50)$$

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1/2} v_n(r) c_n^{[2]}(r) \lambda_{\max}^{1/2}(D_n) \left\{ s_n(r) k_n^{[1]}(r) + \delta_n^{[1]}(r) \right\} = 0, \quad (3.2.51)$$

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1/2} \tilde{v}_n(r) \tilde{c}_n^{[1]}(r) \lambda_{\max}(D_n) \left\{ \delta_n^{[2]}(r) + s_n(r) k_n^{[2]}(r) \right\} = 0, \quad (3.2.52)$$

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} r n^{-1/2} \tilde{v}_n(r) \lambda_{\max}^{1/2}(D_n) \left\{ k_n^{[1]}(r) c_n^{[3]}(r) + k_n^{[2]}(r) \tilde{c}_n^{[1]}(r) c_n^{[2]}(r) \lambda_{\max}^{1/2}(D_n) \right\} = 0. \quad (3.2.53)$$

Then

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} \sup_{\theta \in B_r(\theta_0)} \|\mathcal{E}_n(\theta) - \mathcal{E}_n(\theta_0)\| = 0 \quad a.s.$$

Proof. We return to expressions (3.2.11)-(3.2.13) and deal first with (3.2.12). Let us write

$$\lambda^T [\mathcal{E}_n^{[1]}(\theta) - \mathcal{E}_n^{[1]}(\theta_0)] \lambda := U_n^{[1]}(\theta, \lambda) + U_n^{[2]}(\theta, \lambda), \quad (3.2.54)$$

where

$$U_n^{[1]}(\theta, \lambda) := \sum_{i=1}^n \sum_{j=1}^{\tau_i} \lambda^T \left(\frac{\partial^2 \mu_{ij}(\theta)}{\partial \theta^T \partial \theta} - \frac{\partial^2 \mu_{ij}(\theta_0)}{\partial \theta^T \partial \theta} \right) V_{ij}^{-2}(\theta) [Y_{ij} - \mu_{ij}(\theta)] I_{ij}^{obs} \lambda,$$

$$U_n^{[2]}(\theta, \lambda) := \sum_{i=1}^n \sum_{j=1}^{\tau_i} \lambda^T \frac{\partial^2 \mu_{ij}(\theta_0)}{\partial \theta^T \partial \theta} [V_{ij}^{-2}(\theta) - V_{ij}^{-2}(\theta_0)] [Y_{ij} - \mu_{ij}(\theta)] I_{ij}^{obs} \lambda. \quad (3.2.55)$$

The aim here is to show that

$$\lim_{r \rightarrow 0} \lim_{n \rightarrow \infty} n^{-1} \sup_{\theta \in B_r(\theta_0)} \sup_{\|\lambda\|=1} |U_n^{[k]}(\theta, \lambda)| = 0, \quad k = 1, 2. \quad (3.2.56)$$

Since

$$\frac{\partial^2 \mu_{ij}(\theta)}{\partial \theta^T \partial \theta} = \ddot{\mu}_{ij}(\theta) \frac{\partial c_{ij}^T(\theta)}{\partial \theta} x_{ij} x_{ij}^T \frac{\partial c_{ij}(\theta)}{\partial \theta^T} + \dot{\mu}_{ij}(\theta) \frac{\partial^2 c_{ij}^T(\theta)}{\partial \theta^T \partial \theta} x_{ij}, \quad (3.2.57)$$

we decompose $U_n^{[1]}(\theta, \lambda)$ into a sum of five terms:

$$U_1^{[1]} = \sum_{i=1}^n \sum_{j=1}^{\tau_i} \lambda^T (\ddot{\mu}_{ij}(\theta) - \ddot{\mu}_{ij}(\theta_0)) \frac{\partial c_{ij}^T(\theta)}{\partial \theta} x_{ij} x_{ij}^T \frac{\partial c_{ij}(\theta)}{\partial \theta^T} \lambda V_{ij}^{-2}(\theta) V_{ij}(\theta) Z_{ij}(\theta) I_{ij}^{obs},$$

$$U_2^{[1]} = \sum_{i=1}^n \sum_{j=1}^{\tau_i} \lambda^T \ddot{\mu}_{ij}(\theta_0) \left(\frac{\partial c_{ij}^T(\theta)}{\partial \theta} - \frac{\partial c_{ij}^T(\theta_0)}{\partial \theta} \right) x_{ij} x_{ij}^T \frac{\partial c_{ij}(\theta)}{\partial \theta^T} \lambda V_{ij}^{-2}(\theta) V_{ij}(\theta) Z_{ij}(\theta) I_{ij}^{obs},$$

$$U_3^{[1]} = \sum_{i=1}^n \sum_{j=1}^{\tau_i} \lambda^T \ddot{\mu}_{ij}(\theta_0) \frac{\partial c_{ij}^T(\theta_0)}{\partial \theta} x_{ij} x_{ij}^T \left(\frac{\partial c_{ij}(\theta)}{\partial \theta^T} - \frac{\partial c_{ij}(\theta_0)}{\partial \theta^T} \right) \lambda V_{ij}^{-2}(\theta) V_{ij}(\theta) Z_{ij}(\theta) I_{ij}^{obs},$$

$$U_4^{[1]} = \sum_{i=1}^n \sum_{j=1}^{\tau_i} \lambda^T (\dot{\mu}_{ij}(\theta) - \dot{\mu}_{ij}(\theta_0)) \frac{\partial^2 c_{ij}^T(\theta)}{\partial \theta^T \partial \theta} x_{ij} \lambda V_{ij}^{-2}(\theta) V_{ij}(\theta) Z_{ij}(\theta) I_{ij}^{obs},$$

$$U_5^{[1]} = \sum_{i=1}^n \sum_{j=1}^{\tau_i} \lambda^T \dot{\mu}_{ij}(\theta_0) \left(\frac{\partial^2 c_{ij}^T(\theta)}{\partial \theta^T \partial \theta} - \frac{\partial^2 c_{ij}^T(\theta_0)}{\partial \theta^T \partial \theta} \right) x_{ij} \lambda V_{ij}^{-2}(\theta) V_{ij}(\theta) Z_{ij}(\theta) I_{ij}^{obs}.$$

We first find an upper bound for $U_1^{[1]}$:

$$|U_1^{[1]}| \leq \sum_{i=1}^n \sum_{j=1}^{\tau_i} \frac{|\ddot{\mu}_{ij}(\theta) - \ddot{\mu}_{ij}(\theta_0)|}{V_{ij}(\theta)} \left| \frac{V_{ij}(\theta_0)}{V_{ij}(\theta)} \right| \left| \lambda^T \frac{\partial c_{ij}^T(\theta)}{\partial \theta} x_{ij} \right|^2 |Z_{ij}(\theta)| I_{ij}^{obs}.$$

Since

$$I_{ij}^{obs} = I\{S_{ij} \leq C_i\} \leq I\{S_{i,j-1} < C_i\}, \quad (3.2.58)$$

then, using (2.7.8) and (3.2.43)

$$|U_1^{[1]}| \leq \delta_n^{[2]}(r)v_n(r) \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left| \lambda^T \frac{\partial c_{ij}^T(\theta)}{\partial \theta} x_{ij} \right|^2 I_0^{1/2},$$

where

$$I_0 := \sum_{i=1}^n \sum_{j=1}^{\infty} Z_{ij}^2(\theta_0) I\{S_{i,j-1} < C_i\}. \quad (3.2.59)$$

By (2.7.2), $n^{-1/2}I_0^{1/2}$ is asymptotically bounded. We now write:

$$\begin{aligned} |U_1^{[1]}| &\leq \delta_n^{[2]}(r)v_n(r) \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left\| \frac{\partial c_{ij}^T(\theta)}{\partial \theta} \right\|^2 \|x_{ij}\|^2 n^{1/2}(n^{-1}I_0)^{1/2} \\ &= \delta_n^{[2]}(r)v_n(r) \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left\| \frac{\partial c_{ij}^T(\theta)}{\partial \theta} \right\|^2 \|x_{ij}\|^2 n^{1/2} O_P(1) \\ &\leq \delta_n^{[2]}(r)v_n(r) [c_n^{[1]}(r)]^2 \text{trace}(D_n) n^{1/2} O_P(1) \\ &\leq C \delta_n^{[2]}(r)v_n(r) [c_n^{[1]}(r)]^2 \lambda_{\max}(D_n) n^{1/2} O_P(1). \end{aligned}$$

where we used (3.2.30) for the second last inequality.

Therefore, if

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1/2} \delta_n^{[2]}(r)v_n(r) [c_n^{[1]}(r)]^2 \lambda_{\max}(D_n) = 0 \quad a.s. \quad (3.2.60)$$

which follows by (3.2.52),

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} \sup_{\theta \in B_r(\theta_0)} \sup_{\|\lambda\|=1} |U_1^{[1]}| = 0.$$

For the rest of the section we will be consistently using (3.2.59) and (2.7.2).

To find an upper bound for $U_2^{[1]}$, we write, with $\theta \in B_r(\theta_0)$ and I_0 in (3.2.59),

$$|U_2^{[1]}| \leq \sum_{i=1}^n \sum_{j=1}^{\tau_i} \frac{|\ddot{\mu}_{ij}(\theta_0)|}{V_{ij}(\theta)} \left| \lambda^T \left(\frac{\partial c_{ij}^T(\theta)}{\partial \theta} - \frac{\partial c_{ij}^T(\theta_0)}{\partial \theta} \right) x_{ij} x_{ij}^T \frac{\partial c_{ij}(\theta)}{\partial \theta^T} \lambda \right| \frac{V_{ij}(\theta_0)}{V_{ij}(\theta)} |Z_{ij}(\theta_0)| I_{ij}^{obs}$$

$$\begin{aligned}
&\leq k_n^{[2]}(r)v_n(r) \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left| \lambda^T \left(\frac{\partial c_{ij}^T(\theta)}{\partial \theta} - \frac{\partial c_{ij}^T(\theta_0)}{\partial \theta} \right) x_{ij} x_{ij}^T \frac{\partial c_{ij}(\theta)}{\partial \theta^T} \lambda \right| I_0^{1/2} \\
&\leq k_n^{[2]}(r)v_n(r) \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left\| \frac{\partial c_{ij}^T(\theta)}{\partial \theta} - \frac{\partial c_{ij}^T(\theta_0)}{\partial \theta} \right\| \|x_{ij}\|^2 \left\| \frac{\partial c_{ij}^T(\theta)}{\partial \theta^T} \right\| n^{1/2} O_P(1) \\
&\leq C k_n^{[2]}(r)v_n(r)c_n^{[1]}(r)c_n^{[2]}(r)r\lambda_{\max}(D_n)n^{1/2}O_P(1).
\end{aligned}$$

We used (3.2.29) and (3.2.43) for the second inequality, (2.7.2) for the third. The fourth inequality follows from the mean value theorem, (3.2.30) and (3.2.32).

Therefore, if

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1/2} r k_n^{[2]}(r)v_n(r)c_n^{[1]}(r)c_n^{[2]}(r)\lambda_{\max}(D_n) = 0 \quad a.s. \quad (3.2.61)$$

which follows from (3.2.53), then

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} \sup_{\theta \in B_r(\theta_0)} \sup_{\|\lambda\|=1} |U_2^{[1]}| = 0.$$

The upper bound of $U_3^{[1]}$ is the same as the upper bound of $U_2^{[1]}$.

We now turn to $U_4^{[1]}$,

$$\begin{aligned}
|U_4^{[1]}| &\leq \sum_{i=1}^n \sum_{j=1}^{\tau_i} \frac{|\dot{\mu}_{ij}(\theta) - \dot{\mu}_{ij}(\theta_0)|}{V_{ij}(\theta)} \left| \lambda^T \frac{\partial^2 c_{ij}^T(\theta)}{\partial \theta^T \partial \theta} x_{ij} \lambda \right| V_{ij}^{-1}(\theta) V_{ij}(\theta_0) |Z_{ij}(\theta_0)| I_{ij}^{obs} \\
&\leq \delta_n^{[1]}(r)v_n(r) \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left| \lambda^T \frac{\partial^2 c_{ij}^T(\theta)}{\partial \theta^T \partial \theta} x_{ij} \lambda \right| |Z_{ij}(\theta_0)| I\{S_{i,j-1} < C_i\} \\
&\leq \delta_n^{[1]}(r)v_n(r) \left(\sum_{i=1}^n \sum_{j=1}^{\tau_i} \left[\lambda^T \frac{\partial^2 c_{ij}^T(\theta)}{\partial \theta^T \partial \theta} x_{ij} \lambda \right]^2 \right)^{1/2} I_0^{1/2} \\
&\leq \delta_n^{[1]}(r)v_n(r) \left(\sum_{i=1}^n \sum_{j=1}^{\tau_i} \left\| \frac{\partial^2 c_{ij}(\theta)}{\partial \theta^T \partial \theta} \right\|^2 \|x_{ij}\|^2 \right)^{1/2} I_0^{1/2} \\
&\leq C \delta_n^{[1]}(r)v_n(r)c_n^{[2]}(r)\lambda_{\max}^{1/2}(D_n)I_0^{1/2} \\
&= C \delta_n^{[1]}(r)v_n(r)c_n^{[2]}(r)\lambda_{\max}^{1/2}(D_n)n^{1/2}O_P(1).
\end{aligned}$$

We used (2.7.7) and (3.2.43) for the second inequality. The third inequality follows by applying (2.7.1) with $1 \leq i \leq n, 1 \leq j \leq \tau_i$, and

$$a_{ij} = \left| \lambda^T \frac{\partial^2 c_{ij}^T(\theta)}{\partial \theta^T \partial \theta} x_{ij} \lambda \right|, \quad b_{ij} = |Z_{ij}(\theta_0)| I\{S_{i,j-1} < C_i\}.$$

We also used notation (3.2.59). We appealed to (3.2.32) for the fifth inequality, and to (2.7.2) for the last equality.

Therefore, if

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1/2} \delta_n^{[1]}(r) v_n(r) c_n^{[2]}(r) \lambda_{\max}^{1/2}(D_n) = 0 \quad a.s. \quad (3.2.62)$$

which follows from (3.2.51), then

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} \sup_{\theta \in B_r(\theta_0)} \sup_{\|\lambda\|=1} |U_4^{[1]}| = 0.$$

Next, we find an upper bound for $U_5^{[1]}$,

$$\begin{aligned} |U_5^{[1]}| &\leq \sum_{i=1}^n \sum_{j=1}^{\tau_i} \frac{|\mu_{ij}(\theta_0)|}{V_{ij}(\theta)} \left| \lambda^T \left(\frac{\partial^2 c_{ij}^T(\theta)}{\partial \theta^T \partial \theta} - \frac{\partial^2 c_{ij}^T(\theta_0)}{\partial \theta^T \partial \theta} \right) x_{ij} \lambda \right| V_{ij}^{-1}(\theta) V_{ij}(\theta_0) |Z_{ij}(\theta_0)| I_{ij}^{obs} \\ &\leq k_n^{[1]}(r) v_n(r) \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left| \lambda^T \left(\frac{\partial^2 c_{ij}^T(\theta)}{\partial \theta^T \partial \theta} - \frac{\partial^2 c_{ij}^T(\theta_0)}{\partial \theta^T \partial \theta} \right) x_{ij} \lambda \right| |Z_{ij}(\theta_0)| I\{S_{i,j-1} < C_i\} \\ &\leq C k_n^{[1]}(r) v_n(r) c_n^{[3]}(r) r \lambda_{\max}^{1/2}(D_n) n^{1/2} O_P(1). \end{aligned}$$

We used (3.2.27) and (3.2.43) in the second inequality. The last inequality follows by first applying (2.7.1) with $1 \leq i \leq n, 1 \leq j \leq \tau_i$,

$$a_{ij} = \left| \lambda^T \left(\frac{\partial^2 c_{ij}^T(\theta)}{\partial \theta^T \partial \theta} - \frac{\partial^2 c_{ij}^T(\theta_0)}{\partial \theta^T \partial \theta} \right) x_{ij} \lambda \right|, \quad b_{ij} = |Z_{ij}(\theta_0)| I\{S_{i,j-1} < C_i\},$$

and then by applying the mean value theorem, coupled with definition (3.2.47).

More precisely,

$$\begin{aligned} &\sum_{i=1}^n \sum_{j=1}^{\tau_i} \left| \lambda^T \left(\frac{\partial^2 c_{ij}^T(\theta)}{\partial \theta^T \partial \theta} - \frac{\partial^2 c_{ij}^T(\theta_0)}{\partial \theta^T \partial \theta} \right) x_{ij} \lambda \right| |Z_{ij}(\theta_0)| I\{S_{i,j-1} < C_i\} \\ &\leq \left(\sum_{i=1}^n \sum_{j=1}^{\tau_i} \left[\lambda^T \left(\frac{\partial^2 c_{ij}^T(\theta)}{\partial \theta^T \partial \theta} - \frac{\partial^2 c_{ij}^T(\theta_0)}{\partial \theta^T \partial \theta} \right) x_{ij} \lambda \right]^2 \right)^{1/2} I_0^{1/2} \\ &\leq C \left(\sum_{i=1}^n \sum_{j=1}^{\tau_i} \left\| \frac{\partial^2 c_{ij}^T(\theta)}{\partial \theta^T \partial \theta} - \frac{\partial^2 c_{ij}^T(\theta_0)}{\partial \theta^T \partial \theta} \right\|^2 \|x_{ij}\|^2 \right)^{1/2} I_0^{1/2} \end{aligned}$$

$$\begin{aligned}
&\leq Cc_n^{[3]}(r)r \left(\sum_{i=1}^n \sum_{j=1}^{\tau_i} \|x_{ij}\|^2 \right)^{1/2} I_0^{1/2} \\
&\leq Cc_n^{[3]}(r)r \lambda_{\max}^{1/2}(D_n) I_0^{1/2} = Cc_n^{[3]}(r)r \lambda_{\max}^{1/2}(D_n) n^{1/2} O_P(1).
\end{aligned}$$

Therefore, if

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1/2} r k_n^{[1]}(r) v_n(r) c_n^{[3]}(r) \lambda_{\max}^{1/2}(D_n) = 0 \quad a.s. \quad (3.2.63)$$

which follows from (3.2.53), then

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} \sup_{\theta \in B_r(\theta_0)} \sup_{\|\lambda\|=1} |U_5^{[1]}| = 0.$$

Let us turn to $U_n^{[2]}(\theta, \lambda)$, which, by (3.2.55) and (3.2.57) can be written as

$$U_n^{[2]}(\theta, \lambda) = U_1^{[2]} + U_2^{[2]}, \quad \text{with}$$

$$U_1^{[2]} = \sum_{i=1}^n \sum_{j=1}^{\tau_i} \lambda^T \frac{\ddot{\mu}_{ij}(\theta_0)}{V_{ij}(\theta_0)} \frac{\partial c_{ij}^T(\theta_0)}{\partial \theta} x_{ij} x_{ij}^T \frac{\partial c_{ij}(\theta_0)}{\partial \theta^T} \frac{[V_{ij}^2(\theta_0) - V_{ij}^2(\theta)]}{V_{ij}^2(\theta)} Z_{ij}(\theta_0) I_{ij}^{obs} \lambda,$$

$$U_2^{[2]} = \sum_{i=1}^n \sum_{j=1}^{\tau_i} \lambda^T \frac{\dot{\mu}_{ij}(\theta_0)}{V_{ij}(\theta_0)} \frac{\partial^2 c_{ij}^T(\theta_0)}{\partial \theta^T \partial \theta} x_{ij} \frac{[V_{ij}^2(\theta_0) - V_{ij}^2(\theta)]}{V_{ij}^2(\theta)} Z_{ij}(\theta_0) I_{ij}^{obs} \lambda.$$

For $U_1^{[2]}$, we have

$$\begin{aligned}
\left| U_1^{[2]} \right| &\leq \eta_n(r) \sum_{i=1}^n \sum_{j=1}^{\tau_i} \frac{|\ddot{\mu}_{ij}(\theta_0)|}{V_{ij}(\theta_0)} \left| \lambda^T \frac{\partial c_{ij}^T(\theta_0)}{\partial \theta} x_{ij} x_{ij}^T \frac{\partial c_{ij}(\theta_0)}{\partial \theta^T} \lambda \right| |Z_{ij}(\theta_0)| I_{ij}^{obs} \\
&\leq \eta_n(r) k_n^{[2]}(r) \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left| \lambda^T \frac{\partial c_{ij}^T(\theta_0)}{\partial \theta} x_{ij} x_{ij}^T \frac{\partial c_{ij}(\theta_0)}{\partial \theta^T} \lambda \right| I_0^{1/2} \\
&\leq \eta_n(r) k_n^{[2]}(r) [c_n^{[1]}(r)]^2 \lambda_{\max}(D_n) n^{1/2} O_P(1).
\end{aligned}$$

We used (2.7.9) for the first inequality, (3.2.29) for the second, (2.7.2) and (3.2.30) for the third.

Therefore, if

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1/2} \eta_n(r) k_n^{[2]}(r) [c_n^{[1]}(r)]^2 \lambda_{\max}(D_n) = 0 \quad a.s. \quad (3.2.64)$$

which follows by Remark 3.2.4 and (3.2.52), then

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} \sup_{\theta \in B_r(\theta_0)} \sup_{\|\lambda\|=1} \left| U_1^{[2]} \right| = 0.$$

Turning to the term $U_2^{[2]}$,

$$\begin{aligned} |U_2^{[2]}| &\leq \eta_n(r) k_n^{[1]}(r) \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left| \lambda^T \frac{\partial^2 c_{ij}^T(\theta_0)}{\partial \theta^T \partial \theta} x_{ij} \lambda \right| |Z_{ij}(\theta_0)| I_{ij}^{obs} \\ &\leq \eta_n(r) k_n^{[1]}(r) \left(\sum_{i=1}^n \sum_{j=1}^{\tau_i} \left| \lambda^T \frac{\partial^2 c_{ij}^T(\theta_0)}{\partial \theta^T \partial \theta} x_{ij} \lambda \right|^2 \right)^{1/2} I_0^{1/2} \\ &\leq C \eta_n(r) k_n^{[1]}(r) c_n^{[2]}(r) \lambda_{\max}^{1/2}(D_n) n^{1/2} O_P(1). \end{aligned}$$

We used (2.7.9) and (3.2.27) for the first inequality. For the second inequality, we used (2.7.1) with $1 \leq i \leq n$, $1 \leq j \leq \tau_i$

$$a_{ij} = \left| \lambda^T \frac{\partial^2 c_{ij}^T(\theta_0)}{\partial \theta^T \partial \theta} x_{ij} \lambda \right|, \quad b_{ij} = |Z_{ij}(\theta_0)| I\{S_{i,j-1} < C_i\}.$$

We applied (2.7.2) and (3.2.32) for the third inequality.

Therefore, if

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1/2} \eta_n(r) k_n^{[1]}(r) c_n^{[2]}(r) \lambda_{\max}^{1/2}(D_n) = 0 \quad a.s. \quad (3.2.65)$$

which follows from (3.2.51) and Remark 3.2.4, then

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} \sup_{\theta \in B_r(\theta_0)} \sup_{\|\lambda\|=1} \left| U_2^{[2]} \right| = 0.$$

Returning to (3.2.11) and (3.2.13), we write

$$-\frac{1}{2} \lambda^T [\mathcal{E}_n^{[2]}(\theta) - \mathcal{E}_n^{[2]}(\theta_0)] \lambda := U_n^{[3]}(\theta, \lambda) + U_n^{[4]}(\theta, \lambda) + U_n^{[5]}(\theta, \lambda), \quad (3.2.66)$$

where

$$U_n^{[3]}(\theta, \lambda) = \sum_{i=1}^n \sum_{j=1}^{\tau_i} \lambda^T \left(\frac{\partial \mu_{ij}(\theta)}{\partial \theta} - \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} \right) V_{ij}^{-3}(\theta) \frac{\partial V_{ij}(\theta)}{\partial \theta^T} (Y_{ij} - \mu_{ij}(\theta_0)) I_{ij}^{obs} \lambda,$$

$$U_n^{[4]}(\theta, \lambda) = \sum_{i=1}^n \sum_{j=1}^{\tau_i} \lambda^T \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} (V_{ij}^{-3}(\theta) - V_{ij}^{-3}(\theta_0)) \frac{\partial V_{ij}(\theta)}{\partial \theta^T} (Y_{ij} - \mu_{ij}(\theta_0)) I_{ij}^{obs} \lambda,$$

$$U_n^{[5]}(\theta, \lambda) = \sum_{i=1}^n \sum_{j=1}^{\tau_i} \lambda^T \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} V_{ij}^{-3}(\theta_0) \left(\frac{\partial V_{ij}(\theta)}{\partial \theta^T} - \frac{\partial V_{ij}(\theta_0)}{\partial \theta^T} \right) (Y_{ij} - \mu_{ij}(\theta_0)) I_{ij}^{obs} \lambda.$$

We now give conditions for

$$n^{-1} \sup_{\theta \in B_r(\theta_0)} \sup_{\|\lambda\|=1} |U_n^{[k]}(\theta, \lambda)| \rightarrow 0, \quad a.s., \quad k = 3, 4, 5.$$

For $k = 3$, we have:

$$\begin{aligned} |U_n^{[3]}(\theta, \lambda)| &\leq \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left| \lambda^T \left(\frac{\partial \mu_{ij}(\theta)}{\partial \theta} - \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} \right) V_{ij}^{-3}(\theta) \frac{\partial V_{ij}(\theta)}{\partial \theta^T} (Y_{ij} - \mu_{ij}(\theta_0)) I_{ij}^{obs} \lambda \right| \\ &\leq v_n(r) \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left| \lambda^T \left(\frac{\partial \mu_{ij}(\theta)}{\partial \theta} - \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} \right) V_{ij}^{-2}(\theta) \frac{\partial V_{ij}(\theta)}{\partial \theta^T} \lambda \right| |Z_{ij}(\theta_0)| I\{S_{i,j-1} < C_i\} \\ &\leq v_n(r) \left(\sum_{i=1}^n \sum_{j=1}^{\tau_i} \left\| \left(\frac{\partial \mu_{ij}(\theta)}{\partial \theta} - \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} \right) V_{ij}^{-1}(\theta) \right\|^2 \left\| V_{ij}^{-1}(\theta) \frac{\partial V_{ij}(\theta)}{\partial \theta^T} \right\|^2 \right)^{1/2} I_0^{1/2} \\ &\leq v_n(r) v_n^{[1]}(r) \left(\sum_{i=1}^n \sum_{j=1}^{\tau_i} \|x_{ij}\|^2 \|V_{ij}^{-1}(\theta) d_{ij}(\theta)\|^2 \right)^{1/2} n^{1/2} O_P(1) \\ &\leq C v_n(r) v_n^{[1]}(r) d_n(r) \lambda_{\max}^{1/2}(D_n) n^{1/2} O_P(1). \end{aligned}$$

We used (3.2.43) for the second inequality. For the third inequality we applied (2.7.1) for $1 \leq j \leq \tau_i, 1 \leq i \leq n$, and

$$a_{ij} = \left| \lambda^T \left(\frac{\partial \mu_{ij}(\theta)}{\partial \theta} - \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} \right) V_{ij}^{-1}(\theta) V_{ij}^{-1}(\theta) \frac{\partial V_{ij}(\theta)}{\partial \theta^T} \lambda \right|, b_{ij} = |Z_{ij}(\theta_0)| I\{S_{i,j-1} < C_i\}.$$

We also used (2.7.2), (3.2.15) and (3.2.33) for the fourth inequality, while the final inequality followed from (3.2.17).

Therefore, if

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1/2} v_n(r) v_n^{[1]}(r) d_n(r) \lambda_{\max}^{1/2}(D_n) = 0 \quad a.s. \quad (3.2.67)$$

which follows from (3.2.50), then

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} \sup_{\theta \in B_r(\theta_0)} \sup_{\|\lambda\|=1} |U_n^{[3]}| = 0.$$

For $k = 4$, we have:

$$\begin{aligned} |U_n^{[4]}(\theta, \lambda)| &\leq \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left| \lambda^T \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} (V_{ij}^{-3}(\theta) - V_{ij}^{-3}(\theta_0)) \frac{\partial V_{ij}(\theta)}{\partial \theta^T} (Y_{ij} - \mu_{ij}(\theta_0)) I_{ij}^{obs} \lambda \right| \\ &\leq \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left| \lambda^T \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} \frac{[V_{ij}^3(\theta_0) - V_{ij}^3(\theta)]}{V_{ij}^3(\theta)} V_{ij}^{-2}(\theta_0) \frac{\partial V_{ij}(\theta)}{\partial \theta^T} \lambda \right| |Z_{ij}(\theta_0)| I\{S_{i,j-1} < C_i\} \\ &\leq C s_n(r) \tilde{v}_n(r) \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left| \lambda^T \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} V_{ij}^{-2}(\theta_0) \frac{\partial V_{ij}(\theta)}{\partial \theta^T} \lambda \right| |Z_{ij}(\theta_0)| I\{S_{i,j-1} < C_i\} \\ &\leq C s_n(r) \tilde{v}_n(r) \left(\sum_{i=1}^n \sum_{j=1}^{\tau_i} \left| \lambda^T \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} V_{ij}^{-2}(\theta_0) \frac{\partial V_{ij}(\theta)}{\partial \theta^T} \lambda \right|^2 \right)^{1/2} I_0^{1/2} \\ &\leq C s_n(r) \tilde{v}_n(r) \left(\sum_{i=1}^n \sum_{j=1}^{\tau_i} \left\| \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} V_{ij}^{-1}(\theta_0) \right\|^2 \left\| V_{ij}^{-1}(\theta_0) \frac{\partial V_{ij}(\theta)}{\partial \theta^T} \right\|^2 \right)^{1/2} I_0^{1/2} \\ &\leq C s_n(r) \tilde{v}_n(r) v_n^{[1]}(r) \left(\sum_{i=1}^n \sum_{j=1}^{\tau_i} \left\| V_{ij}^{-1}(\theta_0) \mu_{ij}(\theta_0) \frac{\partial c_{ij}^T(\theta_0)}{\partial \theta} x_{ij} \right\|^2 \right)^{1/2} n^{1/2} O_P(1) \\ &\leq C s_n(r) \tilde{v}_n(r) v_n^{[1]}(r) k_n^{[1]}(r) c_n^{[1]}(r) \lambda_{\max}^{1/2}(D_n) n^{1/2} O_P(1). \end{aligned}$$

The third inequality follows by using (3.2.49), the fourth inequality by using (2.7.1) for $1 \leq j \leq \tau_i, 1 \leq i \leq n$,

$$a_{ij} = \left| \lambda^T \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} V_{ij}^{-2}(\theta_0) \frac{\partial V_{ij}(\theta)}{\partial \theta^T} \lambda \right|, \quad b_{ij} = |Z_{ij}(\theta_0)| I\{S_{i,j-1} < C_i\},$$

and for the sixth inequality we used (2.7.2) and (3.2.33). For the last inequality, we applied (3.2.27) and (3.2.30).

Therefore, if

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1/2} s_n(r) \tilde{v}_n(r) v_n^{[1]}(r) k_n^{[1]}(r) c_n^{[1]}(r) \lambda_{\max}^{1/2}(D_n) = 0 \quad a.s. \quad (3.2.68)$$

which holds if (3.2.50) holds, then

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} \sup_{\theta \in B_r(\theta_0)} \sup_{\|\lambda\|=1} |U_n^{[4]}| = 0.$$

Now we turn to the final term $U_n^{[5]}$,

$$\begin{aligned}
|U_n^{[5]}(\theta, \lambda)| &\leq \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left| \lambda^T \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} V_{ij}^{-3}(\theta_0) \left(\frac{\partial V_{ij}(\theta)}{\partial \theta^T} - \frac{\partial V_{ij}(\theta_0)}{\partial \theta^T} \right) (Y_{ij} - \mu_{ij}(\theta_0)) I_{ij}^{obs} \lambda \right| \\
&\leq \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left| \lambda^T \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} V_{ij}^{-2}(\theta_0) \left(\frac{\partial V_{ij}(\theta)}{\partial \theta^T} - \frac{\partial V_{ij}(\theta_0)}{\partial \theta^T} \right) \lambda \right| |Z_{ij}(\theta_0)| I\{S_{i,j-1} < C_i\} \\
&\leq \left(\sum_{i=1}^n \sum_{j=1}^{\tau_i} \left| \lambda^T \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} V_{ij}^{-2}(\theta_0) \left(\frac{\partial V_{ij}(\theta)}{\partial \theta^T} - \frac{\partial V_{ij}(\theta_0)}{\partial \theta^T} \right) \lambda \right|^2 \right)^{1/2} I_0^{1/2} \\
&\leq \left(\sum_{i=1}^n \sum_{j=1}^{\tau_i} \left\| \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} V_{ij}^{-1}(\theta_0) \right\|^2 \left\| V_{ij}^{-1}(\theta_0) \left(\frac{\partial V_{ij}(\theta)}{\partial \theta^T} - \frac{\partial V_{ij}(\theta_0)}{\partial \theta^T} \right) \right\|^2 \right)^{1/2} I_0^{1/2} \\
&\leq k_n^{[1]}(r) c_n^{[1]}(r) w_n^{[1]}(r) \lambda_{\max}^{1/2}(D_n) n^{1/2} O_P(1).
\end{aligned}$$

For the third inequality we applied (2.7.1) for $1 \leq j \leq \tau_i, 1 \leq i \leq n$

$$a_{ij} = \left| \lambda^T \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} V_{ij}^{-2}(\theta_0) \left(\frac{\partial V_{ij}(\theta)}{\partial \theta^T} - \frac{\partial V_{ij}(\theta_0)}{\partial \theta^T} \right) \lambda \right|, \quad b_{ij} = |Z_{ij}(\theta_0)| I\{S_{i,j-1} < C_i\}.$$

By using (2.7.2), (3.2.27), (3.2.30) and (3.2.48), we obtained the final inequality.

Therefore, if

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1/2} w_n^{[1]}(r) k_n^{[1]}(r) c_n^{[1]}(r) \lambda_{\max}^{1/2}(D_n) = 0 \quad a.s. \quad (3.2.69)$$

which holds under (3.2.50), then

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} \sup_{\theta \in B_r(\theta_0)} \sup_{\|\lambda\|=1} |U_n^{[5]}| = 0 \quad a.s.$$

This completes the proof of Lemma 3.2.5. \square

3.2.2 The Derivative of the Imputed Part of $\hat{g}_{n,1}^{obs}$

In this section we give the proof of $S(ii)$ for the derivative of the imputed term of $\hat{g}_{n,1}^{obs}(\theta)$ using the corresponding proof for the derivative of the non-imputed term, given in Section 3.2.1.

Let

$$\gamma_{j,n}^{cen} := \frac{n}{\sum_{k=1}^n I_{kj}^{cen}}, \quad j, n \geq 1. \quad (3.2.70)$$

From (3.2.70), we obtain

$$\frac{1}{n} \sum_{i=1}^n \gamma_{j,n}^{cen} I_{ij}^{cen} = 1. \quad (3.2.71)$$

We have the following result.

Lemma 3.2.6 *Assume that (T2) holds. Let $0 < e_m := \min_{1 \leq j \leq m} E_{\eta_0}[I_{1j}^{cen}]$. Then, for $0 < \varepsilon < e_m$, there exists an integer $n_0(\varepsilon)$, such that, for all $n \geq n_0(\varepsilon)$, $1 \leq j \leq m$*

$$0 < e_m - \varepsilon \leq n^{-1} \sum_{k=1}^n I_{kj}^{cen} \leq E_m + \varepsilon, \quad (3.2.72)$$

where $E_m := \max_{1 \leq j \leq m} E_{\eta_0}[I_{1j}^{cen}]$. Furthermore

$$0 < (E_m + \varepsilon)^{-1} \leq \frac{n}{\sum_{k=1}^n I_{kj}^{cen}} \leq (e_m - \varepsilon)^{-1}. \quad (3.2.73)$$

Proof. The first string of inequalities (3.2.72) follows from SLLN applied to a finite number of occurrences, *i.e.*, $j \leq m$, and (3.2.73) follows directly from (3.2.72). \square

The imputed part of (2.4.2) is

$$\hat{g}_{n,1}^{imp}(\theta) := n^{-1} \sum_{i=1}^n \sum_{k=1}^n \sum_{j=1}^{\infty} \gamma_{j,n}^{cen} f_{ij}(\theta) I_{ij}^{cen} Z_{kj}(\theta) I_{kj}^{obs}, \quad (3.2.74)$$

with $\gamma_{j,n}^{cen}$ defined in (3.2.70). We note that, in contrast to the non-imputed part, the imputed part contains n^{-1} as a factor.

Remark 3.2.7 *In (3.2.74) we actually have $i \neq k$ in the first two sums, because $I_{ij}^{cen} I_{kj}^{obs} = 0$ if $i = k$. We could use $\gamma_{ij,n}^{cen} := \frac{n-1}{\sum_{k=1, k \neq i}^n I_{kj}^{cen}}$ in lieu of $\gamma_{j,n}^{cen}$ in (3.2.74). However, the asymptotic results are identical.*

To compute the derivative of (3.2.74), we split it into two terms, as in (3.2.3)-(3.2.4):

$$\hat{g}_{n,1}^{imp}(\theta) := g_{n,1}^{[1]}(\theta)^{imp} + g_{n,1}^{[2]}(\theta)^{imp}, \quad \text{where}$$

$$g_{n,1}^{[1]}(\theta)^{imp} := n^{-1} \sum_{k=1}^n \sum_{i=1}^n \sum_{j=1}^{\tau_i} \gamma_{j,n}^{cen} f_{ij}(\theta) I_{ij}^{cen} V_{kj}^{-1}(\theta) [\mu_{kj}(\theta_0) - \mu_{kj}(\theta)] I_{kj}^{obs}, \quad (3.2.75)$$

$$g_{n,1}^{[2]}(\theta)^{imp} := n^{-1} \sum_{k=1}^n \sum_{i=1}^n \sum_{j=1}^{\tau_i} \gamma_{j,n}^{cen} f_{ij}(\theta) I_{ij}^{cen} V_{kj}^{-1}(\theta) [Y_{kj} - \mu_{kj}(\theta_0)] I_{kj}^{obs}. \quad (3.2.76)$$

Taking the derivative in (3.2.75) gives:

$$-\frac{\partial g_{n,1}^{[1]}(\theta)^{imp}}{\partial \theta^T} := H_n(\theta)^{imp} - B_n^{[1]}(\theta)^{imp} - B_n^{[2]}(\theta)^{imp} - B_n^{[3]}(\theta)^{imp} - B_n^{[4]}(\theta)^{imp}, \quad (3.2.77)$$

where:

$$H_n(\theta)^{imp} := n^{-1} \sum_{k=1}^n \sum_{i=1}^n \sum_{j=1}^{\tau_i} \gamma_{j,n}^{cen} f_{ij}(\theta) I_{ij}^{cen} V_{kj}^{-1}(\theta) \frac{\partial \mu_{kj}(\theta)}{\partial \theta^T} I_{kj}^{obs}. \quad (3.2.78)$$

The following decomposition is similar to (3.2.21):

$$H_n^{imp}(\theta) - H_n^{imp}(\theta_0) = H_n^{[1]}(\theta)^{imp} + H_n^{[2]}(\theta)^{imp} + H_n^{[3]}(\theta)^{imp} + H_n^{[4]}(\theta)^{imp}, \quad (3.2.79)$$

where

$$H_n^{[1]}(\theta)^{imp} := n^{-1} \sum_{i,k=1}^n \sum_{j=1}^{\tau_i} \gamma_{j,n}^{cen} \left(\frac{\partial \mu_{ij}(\theta)}{\partial \theta} - \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} \right) V_{ij}^{-1}(\theta) I_{ij}^{cen} V_{kj}^{-1}(\theta) \frac{\partial \mu_{kj}(\theta)}{\partial \theta^T} I_{kj}^{obs},$$

$$\begin{aligned} H_n^{[2]}(\theta)^{imp} &:= n^{-1} \sum_{i,k=1}^n \sum_{j=1}^{\tau_i} \gamma_{j,n}^{cen} \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} (V_{ij}^{-1}(\theta) - V_{ij}^{-1}(\theta_0)) I_{ij}^{cen} V_{kj}^{-1}(\theta) \frac{\partial \mu_{kj}(\theta)}{\partial \theta^T} I_{kj}^{obs} \\ &= n^{-1} \sum_{i,k=1}^n \sum_{j=1}^{\tau_i} \gamma_{j,n}^{cen} \frac{V_{ij}(\theta_0) - V_{ij}(\theta)}{V_{ij}(\theta)} \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} V_{ij}^{-1}(\theta_0) I_{ij}^{cen} V_{kj}^{-1}(\theta) \frac{\partial \mu_{kj}(\theta)}{\partial \theta^T} I_{kj}^{obs}, \end{aligned}$$

$$H_n^{[3]}(\theta)^{imp} := n^{-1} \sum_{i,k=1}^n \sum_{j=1}^{\tau_i} \gamma_{j,n}^{cen} \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} V_{ij}^{-1}(\theta_0) I_{ij}^{cen} (V_{kj}^{-1}(\theta) - V_{kj}^{-1}(\theta_0)) \frac{\partial \mu_{kj}(\theta)}{\partial \theta^T} I_{kj}^{obs},$$

$$H_n^{[4]}(\theta)^{imp} := n^{-1} \sum_{i,k=1}^n \sum_{j=1}^{\tau_i} \gamma_{j,n}^{cen} \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} V_{ij}^{-1}(\theta_0) I_{ij}^{cen} V_{kj}^{-1}(\theta_0) \left(\frac{\partial \mu_{kj}(\theta)}{\partial \theta^T} - \frac{\partial \mu_{kj}(\theta_0)}{\partial \theta^T} \right) I_{kj}^{obs}.$$

Returning to (3.2.77), we have

$$B_n^{[1]}(\theta)^{imp} := n^{-1} \sum_{i,k=1}^n \sum_{j=1}^{\tau_i} \left\{ \gamma_{j,n}^{cen} \ddot{\mu}_{ij}(\theta) \frac{\partial c_{ij}^T(\theta)}{\partial \theta} x_{ij} x_{ij}^T \frac{\partial c_{ij}(\theta)}{\partial \theta^T} V_{ij}^{-1}(\theta) I_{ij}^{cen} \right. \\ \left. \times V_{kj}^{-1}(\theta) [\mu_{kj}(\theta_0) - \mu_{kj}(\theta)] I_{kj}^{obs} \right\}, \quad (3.2.80)$$

$$B_n^{[2]}(\theta)^{imp} := n^{-1} \sum_{i,k=1}^n \sum_{j=1}^{\tau_i} \gamma_{j,n}^{cen} \dot{\mu}_{ij}(\theta) \frac{\partial^2 [c_{ij}^T(\theta) x_{ij}]}{\partial \theta^T \partial \theta} V_{ij}^{-1}(\theta) I_{ij}^{cen} V_{kj}^{-1}(\theta) [\mu_{kj}(\theta_0) - \mu_{kj}(\theta)] I_{kj}^{obs},$$

$$B_n^{[3]}(\theta)^{imp} := -n^{-1} \sum_{i,k=1}^n \sum_{j=1}^{\tau_i} \gamma_{j,n}^{cen} \frac{\partial \mu_{ij}(\theta)}{\partial \theta} V_{ij}^{-2}(\theta) \frac{\partial V_{ij}(\theta)}{\partial \theta^T} I_{ij}^{cen} V_{kj}^{-1}(\theta) [\mu_{kj}(\theta_0) - \mu_{kj}(\theta)] I_{kj}^{obs},$$

$$B_n^{[4]}(\theta)^{imp} := -n^{-1} \sum_{i,k=1}^n \sum_{j=1}^{\tau_i} \gamma_{j,n}^{cen} f_{ij}(\theta) I_{ij}^{cen} V_{kj}^{-2}(\theta) \frac{\partial V_{kj}(\theta)}{\partial \theta^T} [\mu_{kj}(\theta_0) - \mu_{kj}(\theta)] I_{kj}^{obs},$$

$$B_n(\theta)^{imp} := \sum_{i=1}^4 B_n^{[i]}(\theta)^{imp}. \quad (3.2.81)$$

We note that $B_n^{[k]}(\theta_0)^{imp} = 0, k = 1, 2, 3, 4$.

For the second term in (3.2.76)

$$\frac{\partial g_{n,1}^{[2]}(\theta)^{imp}}{\partial \theta^T} = \mathcal{E}_n(\theta)^{imp} := \mathcal{E}_n^{[1]}(\theta)^{imp} + \mathcal{E}_n^{[2]}(\theta)^{imp} + \mathcal{E}_n^{[3]}(\theta)^{imp}, \quad (3.2.82)$$

where

$$\mathcal{E}_n^{[1]}(\theta)^{imp} := n^{-1} \sum_{i,k=1}^n \sum_{j=1}^{\tau_i} \gamma_{j,n}^{cen} \frac{\partial^2 \mu_{ij}(\theta)}{\partial \theta^T \partial \theta} V_{ij}^{-1}(\theta) I_{ij}^{cen} \frac{V_{kj}(\theta_0)}{V_{kj}(\theta)} Z_{kj}(\theta_0) I_{kj}^{obs}, \quad (3.2.83)$$

$$\mathcal{E}_n^{[2]}(\theta)^{imp} := -n^{-1} \sum_{i,k=1}^n \sum_{j=1}^{\tau_i} \gamma_{j,n}^{cen} \frac{\partial \mu_{ij}(\theta)}{\partial \theta} V_{ij}^{-2}(\theta) \frac{\partial V_{ij}(\theta)}{\partial \theta^T} I_{ij}^{cen} \frac{V_{kj}(\theta_0)}{V_{kj}(\theta)} Z_{kj}(\theta_0) I_{kj}^{obs}, \quad (3.2.84)$$

$$\mathcal{E}_n^{[3]}(\theta)^{imp} := -n^{-1} \sum_{i,k=1}^n \sum_{j=1}^{\tau_i} \gamma_{j,n}^{cen} f_{ij}(\theta) I_{ij}^{cen} V_{kj}^{-2}(\theta) \frac{\partial V_{kj}(\theta)}{\partial \theta^T} V_{kj}(\theta_0) Z_{kj}(\theta_0) I_{kj}^{obs}. \quad (3.2.85)$$

The following decomposition is similar to (3.2.54). Let us write:

$$\lambda^T [\mathcal{E}_n^{[1]}(\theta)^{imp} - \mathcal{E}_n^{[1]}(\theta_0)^{imp}] \lambda := \sum_{k=1}^3 U_n^{[k]}(\theta, \lambda)^{imp}, \quad k = 1, 2, 3. \quad (3.2.86)$$

where

$$U_n^{[1]}(\theta, \lambda)^{imp} := n^{-1} \sum_{i,k=1}^n \sum_{j=1}^{\tau_i} \left\{ \lambda^T \gamma_{j,n}^{cen} \left(\frac{\partial^2 \mu_{ij}(\theta)}{\partial \theta^T \partial \theta} - \frac{\partial^2 \mu_{ij}(\theta_0)}{\partial \theta^T \partial \theta} \right) V_{ij}^{-1}(\theta) I_{ij}^{cen} \right. \\ \left. \times V_{kj}^{-1}(\theta) V_{kj}(\theta_0) Z_{kj}(\theta_0) I_{kj}^{obs} \lambda \right\},$$

$$U_n^{[2]}(\theta, \lambda)^{imp} := n^{-1} \sum_{i,k=1}^n \sum_{j=1}^{\tau_i} \left\{ \lambda^T \gamma_{j,n}^{cen} \frac{\partial^2 \mu_{ij}(\theta_0)}{\partial \theta^T \partial \theta} \left(V_{ij}^{-1}(\theta) - V_{ij}^{-1}(\theta_0) \right) I_{ij}^{cen} \right. \\ \left. \times V_{kj}^{-1}(\theta) V_{kj}(\theta_0) Z_{kj}(\theta_0) I_{kj}^{obs} \lambda \right\},$$

$$U_n^{[3]}(\theta, \lambda)^{imp} := n^{-1} \sum_{i,k=1}^n \sum_{j=1}^{\tau_i} \lambda^T \gamma_{j,n}^{cen} \frac{\partial^2 \mu_{ij}(\theta_0)}{\partial \theta^T \partial \theta} V_{ij}^{-1}(\theta_0) I_{ij}^{cen} \frac{V_{kj}(\theta_0) - V_{kj}(\theta)}{V_{kj}(\theta)} Z_{kj}(\theta_0) I_{kj}^{obs} \lambda.$$

Let

$$\mathcal{E}_n^{[l]}(\theta)^{imp} := -n^{-1} \sum_{j=1}^m \sum_{i,k=1}^n \left\{ \gamma_{j,n}^{cen} \frac{\partial \mu_{ij}(\theta)}{\partial \theta} V_{ij}^{-1}(\theta) I_{ij}^{cen} V_{lj}^{-1}(\theta) \frac{\partial V_{lj}(\theta)}{\partial \theta^T} \right. \\ \left. \times V_{kj}^{-1}(\theta) V_{kj}(\theta_0) Z_{kj}(\theta_0) I_{kj}^{obs} \right\}.$$

Note that, when $l = i$ we obtain (3.2.84), whereas for $l = k$ we obtain (3.2.85).

The following decomposition is similar to (3.2.66):

$$-\lambda^T [\mathcal{E}_n^{[l]}(\theta)^{imp} - \mathcal{E}_n^{[l]}(\theta_0)^{imp}] \lambda := \sum_{c=4}^7 U_n^{[c,l]}(\theta, \lambda)^{imp}, \quad (3.2.87)$$

where

$$U_n^{[4,l]}(\theta, \lambda)^{imp} := n^{-1} \sum_{j=1}^m \sum_{i,k=1}^n \left\{ \gamma_{j,n}^{cen} I_{ij}^{cen} \lambda^T \left(\frac{\partial \mu_{ij}(\theta)}{\partial \theta} - \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} \right) V_{ij}^{-1}(\theta) \frac{\partial V_{lj}(\theta)}{\partial \theta^T} V_{lj}^{-1}(\theta) \right. \\ \left. \times V_{kj}^{-1}(\theta) V_{kj}(\theta_0) Z_{kj}(\theta_0) I_{kj}^{obs} \lambda \right\},$$

$$\begin{aligned}
U_n^{[5,l]}(\theta, \lambda)^{imp} &:= n^{-1} \sum_{j=1}^m \sum_{i,k=1}^n \left\{ \gamma_{j,n}^{cen} I_{ij}^{cen} \lambda^T \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} [V_{ij}^{-1}(\theta) V_{lj}^{-1}(\theta) - V_{ij}^{-1}(\theta_0) V_{lj}^{-1}(\theta_0)] \right. \\
&\quad \left. \times \frac{\partial V_{lj}(\theta)}{\partial \theta^T} \lambda V_{kj}^{-1}(\theta) V_{kj}(\theta) Z_{kj}(\theta) I_{kj}^{obs} \right\}, \\
U_n^{[6,l]}(\theta, \lambda)^{imp} &:= n^{-1} \sum_{j=1}^m \sum_{i,k=1}^n \left\{ \gamma_{j,n}^{cen} I_{ij}^{cen} \lambda^T \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} V_{ij}^{-1}(\theta_0) V_{lj}^{-1}(\theta_0) \right. \\
&\quad \left. \times \left(\frac{\partial V_{lj}(\theta)}{\partial \theta^T} - \frac{\partial V_{lj}(\theta_0)}{\partial \theta^T} \right) \lambda V_{kj}^{-1}(\theta) V_{kj}(\theta) Z_{kj}(\theta) I_{kj}^{obs} \right\}, \\
U_n^{[7,l]}(\theta, \lambda)^{imp} &:= n^{-1} \sum_{j=1}^m \sum_{i,k=1}^n \left\{ \gamma_{j,n}^{cen} I_{ij}^{cen} \lambda^T \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} V_{ij}^{-1}(\theta_0) V_{lj}^{-1}(\theta_0) \frac{\partial V_{lj}(\theta_0)}{\partial \theta^T} \lambda \right. \\
&\quad \left. \times (V_{kj}^{-1}(\theta) - V_{kj}^{-1}(\theta_0)) V_{kj}(\theta) Z_{kj}(\theta) I_{kj}^{obs} \right\}.
\end{aligned}$$

We use the following extension of the Cauchy-Schwarz inequality.

Lemma 3.2.8 *Let $a_{ij}, b_{kj}, 1 \leq i, k \leq n, 1 \leq j \leq m$ be real numbers, and let*

$$I_a(j) := \sum_{i=1}^n a_{ij}^2 \quad I_b(j) := \sum_{k=1}^n b_{kj}^2.$$

Then

$$n^{-1} \sum_{j=1}^m \sum_{i,k=1}^n a_{ij} b_{kj} \leq \left(\sum_{j=1}^m I_a(j) \right)^{1/2} \left(\sum_{j=1}^m I_b(j) \right)^{1/2}. \quad (3.2.88)$$

Furthermore, if $a_{ij} \geq 0$, and $I_c := \sum_{j=1}^m \sum_{i=1}^n a_{ij}$,

then

$$n^{-1} \sum_{j=1}^m \sum_{i,k=1}^n a_{ij} b_{kj} \leq n^{-1/2} I_c \left(\sum_{j=1}^m I_b(j) \right)^{1/2} \leq I_c \left(\sum_{j=1}^m I_b(j) \right)^{1/2}. \quad (3.2.89)$$

Proof. Since, by (2.7.1)

$$\sum_{i=1}^n a_{ij} \leq n^{1/2} \left(\sum_{i=1}^n a_{ij}^2 \right)^{1/2}, \quad \sum_{k=1}^n b_{kj} \leq n^{1/2} \left(\sum_{k=1}^n b_{kj}^2 \right)^{1/2}, \quad (3.2.90)$$

the left hand side of (3.2.88) is bounded above by $\sum_{j=1}^m I_a^{1/2}(j) I_b^{1/2}(j)$. Using (2.7.1) again, we have that the left hand side of (3.2.88) is smaller than the right hand side

of (3.2.88). To prove (3.2.89), we use (3.2.90) only for b_{kj} , then use the fact that $I_c(j)$ is non negative. \square

We now look for upper bounds for the derivatives above by comparing each term with the corresponding term of the non-imputed derivatives.

Note that, when (T2) holds, by (3.2.73), the factor $\gamma_{j,n}^{cen}, 1 \leq j \leq m$ can be ignored in finding these upper bounds, since this factor is bounded from above and below. We also note the similarities between $H_n^{[1]}(\theta)^{imp}$ and $H_n^{[4]}(\theta)^{imp}$, or $H_n^{[2]}(\theta)^{imp}$ and $H_n^{[3]}(\theta)^{imp}$.

Lemma 3.2.9 *Assume that (T2) holds, along with the conditions of Lemma 3.2.1.*

Then

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} \frac{1}{n} \sup_{\theta \in B_r(\theta_0)} |||H_n^{imp}(\theta) - H_n^{imp}(\theta_0)||| = 0 \quad a.s. \quad (3.2.91)$$

Proof. We start with $H_n^{[1]}(\theta)^{imp}$, where $\tau_i \leq m, i \geq 1$, by (T2). With $\lambda \in R^p, \|\lambda\| = 1$, we take in Lemma 3.2.8, with $d_{ij}^T(\theta)$ defined in (3.2.15):

$$I_a(j) := \sum_{i=1}^n V_{ij}^{-2}(\theta) \lambda^T d_{ij}^T(\theta) x_{ij} x_{ij}^T d_{ij}(\theta) \lambda I_{ij}^{cen},$$

$$I_b(j) := \sum_{k=1}^n \lambda^T \frac{\partial \mu_{kj}(\theta)}{\partial \theta^T} V_{kj}^{-2}(\theta) \frac{\partial \mu_{kj}(\theta)}{\partial \theta} \lambda I_{kj}^{obs},$$

and conclude that the upper bound in (3.2.25) is also an upper bound for $|||H_n^{[i]}(\theta)^{imp}|||$, $i = 1, 4$, with a modified constant.

Similarly, to find an upper bound for $|||H_n^{[2]}(\theta)^{imp}|||$, after extracting the factor $\eta_n(r)$ defined in (2.7.9), we take

$$I_a(j) := \sum_{i=1}^n \lambda^T \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta^T} V_{ij}^{-2}(\theta_0) \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} \lambda I_{ij}^{cen},$$

$$I_b(j) := \sum_{k=1}^n \lambda^T \frac{\partial \mu_{kj}(\theta)}{\partial \theta^T} V_{kj}^{-2}(\theta) \frac{\partial \mu_{kj}(\theta)}{\partial \theta} \lambda I_{kj}^{obs}.$$

We conclude that the upper bound in (3.2.26) is also an upper bound for $|||H_n^{[i]}(\theta)^{imp}|||$, $i = 2, 3$, with a modified constant. This concludes the proof of Lemma 3.2.9. \square

Lemma 3.2.10 *Assume that (T2) holds, along with the assumptions of Lemma 3.2.3.*

Then

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} \sup_{\theta \in B_r(\theta_0)} |||B_n^{imp}(\theta)||| = 0 \quad a.s. \quad (3.2.92)$$

Proof. For $B_n^{[1]}(\theta)^{imp}$ defined in (3.2.80), after extracting the factor $k_n^{[2]}(r)$ defined in (3.2.29), we take in Lemma 3.2.8

$$I_c := \sum_{j=1}^m \sum_{i=1}^n \lambda^T \frac{\partial c_{ij}^T(\theta)}{\partial \theta} x_{ij} x_{ij}^T \frac{\partial c_{ij}(\theta)}{\partial \theta^T} \lambda I_{ij}^{cen} \leq I_5,$$

$$I_b(j) := \sum_{k=1}^n |\mu_{kj}(\theta_0) - \mu_{kj}(\theta)|^2 V_{kj}^{-2}(\theta) I_{kj}^{obs} \leq I_6, \quad (3.2.93)$$

where I_5, I_6 are defined in Lemma 3.2.3, above (3.2.36). We conclude that the upper bound in (3.2.38) is also an upper bound for $|||B_n^{[1]}(\theta)^{imp}|||$, with a modified constant.

Similarly, to find an upper bound for $|||B_n^{[2]}(\theta)^{imp}|||$, we take

$$I_a(j) := \sum_{i=1}^n [V_{ij}^{-1}(\theta) \dot{\mu}_{ij}(\theta)]^2 \left[\lambda^T \frac{\partial^2 [c_{ij}^T(\theta) x_{ij}]}{\partial \theta^T \partial \theta} \lambda \right]^2 I_{ij}^{cen} \leq I_7,$$

with I_7 in (3.2.39) and $I_b(j)$ as in (3.2.93), to conclude that the upper bound in (3.2.41) is also an upper bound for $|||B_n^{[2]}(\theta)^{imp}|||$, with a modified constant.

Turning to $|||B_n^{[3]}(\theta)^{imp}|||$, we take

$$I_a(j) := \sum_{i=1}^n \left[\lambda^T \frac{\partial \mu_{ij}(\theta)}{\partial \theta} V_{ij}^{-2}(\theta) \frac{\partial V_{ij}(\theta)}{\partial \theta^T} \lambda \right]^2 I_{ij}^{cen} \leq I_8,$$

with I_8 from Lemma 3.2.3, above (3.2.42), and $I_b(j)$ as in (3.2.93), to conclude that the upper bound in (3.2.42) is also an upper bound for $|||B_n^{[3]}(\theta)^{imp}|||$, with a modified constant.

For $B_n^{[4]}(\theta)^{imp}$, we take in Lemma 3.2.8 $I_a(j) \leq I_8$ as above,

$$I_b(j) := \sum_{k=1}^n \lambda^T \frac{\partial V_{kj}(\theta)}{\partial \theta} V_{kj}^{-2}(\theta) \frac{\partial V_{kj}(\theta)}{\partial \theta^T} \lambda \left[\frac{\mu_{kj}(\theta_0) - \mu_{kj}(\theta)}{V_{kj}(\theta)} \right]^2 I_{kj}^{obs}.$$

To evaluate the spectral radius, we write

$$|\lambda^T B_n^{[4]}(\theta)^{imp} \lambda| \leq I_8^{1/2} \left(\sum_{j=1}^m I_b(j) \right)^{1/2}. \quad (3.2.94)$$

Recalling $v_n^{[1]}(r)$ defined in (3.2.33) and I_6 as in (3.2.93) above, we obtain

$$\sum_{j=1}^m I_b(j) \leq \left(v_n^{[1]}(r) \right)^2 I_6. \quad (3.2.95)$$

Combining (3.2.94) and (3.2.95), and using the evaluation of I_8 in Lemma 3.2.3, we can conclude that (3.2.42) also hold for $\|B_n^{[4]}(\theta)^{imp}\|$. This concludes the proof of Lemma 3.2.10. \square

Lemma 3.2.11 *Assume (T2) holds, along with the conditions of Lemma 3.2.5. Then*

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} \sup_{\theta \in B_r(\theta_0)} \|\mathcal{E}_n(\theta)^{imp} - \mathcal{E}_n(\theta_0)^{imp}\| = 0 \quad a.s.$$

Proof. We will use the decomposition (3.2.86)-(3.2.87) and prove that

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} \sup_{\theta \in B_r(\theta_0)} n^{-1} U_n^{[k]}(\theta, \lambda)^{imp} = 0, \quad k = 1, 2, 3. \quad (3.2.96)$$

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} \sup_{\theta \in B_r(\theta_0)} n^{-1} U_n^{[k,l]}(\theta, \lambda)^{imp} = 0, \quad k = 4, 5, 6, 7. \quad (3.2.97)$$

As justified earlier, we can and will ignore the factor $\gamma_{j,n}^{cen}$, $j, n \geq 1$. The idea of the proof is to use upper bounds already obtained in Lemma 3.2.5. In that lemma, as well as in this proof, expression (3.2.57) is often used. We will also extensively use Lemma 3.2.8.

We first deal with the decomposition (3.2.86) and start with $U_n^{[1]}(\theta, \lambda)^{imp}$. Based on (3.2.57), we decompose $U_n^{[1]}(\theta, \lambda)^{imp}$ into five terms, as we did with $U_n^{[1]}(\theta, \lambda)$ in Lemma 3.2.5.

$$U_n^{[1]}(\theta, \lambda)^{imp} = \sum_{l=1}^5 U_l^{[1]}(\theta, \lambda)^{imp}, \quad \text{where}$$

$$U_l^{[1]}(\theta, \lambda)^{imp} = n^{-1} \sum_{j=1}^m \sum_{i,k=1}^n a_{ij,l}^{[1]} b_{kj,l}^{[1]}, \quad l = 1, 2, \dots, 5.$$

$$\begin{aligned}
\left| a_{ij,1}^{[1]} \right| &= \frac{|\ddot{\mu}_{ij}(\theta) - \ddot{\mu}_{ij}(\theta_0)|}{V_{ij}(\theta)} \left| \lambda^T \frac{\partial c_{ij}^T(\theta)}{\partial \theta} x_{ij} \right|^2 I_{ij}^{cen}, \\
\left| a_{ij,2}^{[1]} \right| &= \frac{|\ddot{\mu}_{ij}(\theta_0)|}{V_{ij}(\theta)} \left| \lambda^T \left(\frac{\partial c_{ij}^T(\theta)}{\partial \theta} - \frac{\partial c_{ij}^T(\theta_0)}{\partial \theta} \right) x_{ij} x_{ij}^T \frac{\partial c_{ij}(\theta)}{\partial \theta^T} \lambda \right| I_{ij}^{cen}, \\
\left| a_{ij,3}^{[1]} \right| &= \frac{|\ddot{\mu}_{ij}(\theta_0)|}{V_{ij}(\theta)} \left| \lambda^T \frac{\partial c_{ij}^T(\theta_0)}{\partial \theta} x_{ij} x_{ij}^T \left(\frac{\partial c_{ij}(\theta)}{\partial \theta^T} - \frac{\partial c_{ij}(\theta_0)}{\partial \theta^T} \right) \lambda \right| I_{ij}^{cen}, \\
\left| a_{ij,4}^{[1]} \right| &= \frac{|\dot{\mu}_{ij}(\theta) - \dot{\mu}_{ij}(\theta_0)|}{V_{ij}(\theta)} \left| \lambda^T \frac{\partial^2 c_{ij}^T(\theta)}{\partial \theta^T \partial \theta} x_{ij} \lambda \right| I_{ij}^{cen}, \\
\left| a_{ij,5}^{[1]} \right| &= \frac{|\dot{\mu}_{ij}(\theta_0)|}{V_{ij}(\theta)} \left| \lambda^T \left(\frac{\partial^2 c_{ij}^T(\theta)}{\partial \theta^T \partial \theta} - \frac{\partial^2 c_{ij}^T(\theta_0)}{\partial \theta^T \partial \theta} \right) x_{ij} \lambda \right| I_{ij}^{cen}, \\
\left| b_{kj,l}^{[1]} \right| &= \left| b_{kj}^{[1]} \right| = \frac{V_{kj}(\theta_0)}{V_{kj}(\theta)} |Z_{kj}(\theta_0)| I_{kj}^{obs}, \quad l = 1, \dots, 5, \\
&\leq v_n(r) |Z_{kj}(\theta_0)| I_{kj}^{obs},
\end{aligned}$$

with $v_n(r)$ defined in (3.2.43).

We now apply (3.2.89) to the sequences $(|a_{ij,l}^{[1]}|, b_{kj,l}^{[1]})_{i,k,j}$ with $l = 1, 2, 3$, and (3.2.88) with $l = 4, 5$, and we obtain the same upper bounds as those of $U_l^{[1]}$, $l = 1, 2, \dots, 5$ in Lemma 3.2.5. Therefore (3.2.96) holds with $U_l^{[1]}(\theta, \lambda)^{imp}$ in lieu of $U_n^{[k]}(\theta, \lambda)^{imp}$, $l = 1, 2, \dots, 5$.

We now look at $U_n^{[2]}(\theta, \lambda)^{imp}$ and compare it to $U_n^{[2]}(\theta, \lambda)$ in (3.2.55), which can be similarly split into two terms. We have here, based on (3.2.57)

$$U_n^{[2]}(\theta, \lambda)^{imp} = U_1^{[2]}(\theta, \lambda)^{imp} + U_2^{[2]}(\theta, \lambda)^{imp},$$

where

$$U_l^{[2]}(\theta, \lambda)^{imp} = n^{-1} \sum_{j=1}^m \sum_{i,k=1}^n a_{ij,l}^{[2]} b_{kj,l}^{[2]}, \quad l = 1, 2.$$

$$\begin{aligned}
|a_{ij,1}^{[2]}| &= |\ddot{\mu}_{ij}(\theta_0)| \left[\lambda^T \frac{\partial c_{ij}^T(\theta_0)}{\partial \theta} x_{ij} \right]^2 |V_{ij}^{-1}(\theta) - V_{ij}^{-1}(\theta_0)| I_{ij}^{cen}, \\
|a_{ij,2}^{[2]}| &= |\ddot{\mu}_{ij}(\theta_0)| \left| \lambda^T \frac{\partial^2 c_{ij}^T(\theta_0)}{\partial \theta^T \partial \theta} x_{ij} \lambda \right| |V_{ij}^{-1}(\theta) - V_{ij}^{-1}(\theta_0)| I_{ij}^{cen},
\end{aligned}$$

$$|b_{kj,l}^{[2]}| = |b_{kj}^{[1]}| \leq v_n(r) |Z_{kj}(\theta_0)| I_{kj}^{obs}, \quad l = 1, 2.$$

We have

$$|a_{ij,1}^{[2]}| \leq s_n(r) \frac{|\ddot{\mu}_{ij}(\theta_0)|}{V_{ij}(\theta_0)} \left| \lambda^T \frac{\partial c_{ij}^T(\theta_0)}{\partial \theta} x_{ij} \right|^2,$$

with $s_n(r)$ in (3.2.45). We now apply (3.2.89) with $(|a_{ij,1}^{[2]}|, b_{kj}^{[1]})_{i,k,j}$ and find that, via Remark 3.2.4, (3.2.52) implies (3.2.96) for $U_1^{[2]}(\theta, \lambda)^{imp}$.

Now

$$|a_{ij,2}^{[2]}| \leq s_n(r) \frac{|\dot{\mu}_{ij}(\theta_0)|}{V_{ij}(\theta_0)} \left| \lambda^T \frac{\partial^2 c_{ij}^T(\theta_0)}{\partial \theta^T \partial \theta} x_{ij} \lambda \right|.$$

This time we apply (3.2.88) with $(a_{ij,2}^{[2]}, b_{kj}^{[1]})_{i,k,j}$. By (3.2.51) and via Remark 3.2.4, we obtain (3.2.96) for $U_2^{[2]}(\theta, \lambda)^{imp}$.

We now deal with $U_n^{[3]}(\theta, \lambda)^{imp}$ and, using (3.2.57), we split into two terms, as before, with

$$|a_{ij,1}^{[3]}| = \frac{|\ddot{\mu}_{ij}(\theta_0)|}{V_{ij}(\theta_0)} \left[\lambda^T \frac{\partial c_{ij}^T(\theta_0)}{\partial \theta} x_{ij} \right]^2 I_{ij}^{cen},$$

$$a_{ij,2}^{[3]} = \frac{\dot{\mu}_{ij}(\theta_0)}{V_{ij}(\theta_0)} \left[\lambda^T \frac{\partial^2 c_{ij}^T(\theta_0)}{\partial \theta^T \partial \theta} x_{ij} \lambda \right] I_{ij}^{cen},$$

$$\begin{aligned} b_{kj,1}^{[3]} &= b_{kj,2}^{[3]} = b_{kj}^{[3]} = [V_{kj}^{-1}(\theta) - V_{kj}^{-1}(\theta_0)] V_{kj}(\theta_0) Z_{kj}(\theta_0) I_{kj}^{obs}, \\ &\leq s_n(r) v_n(r) |Z_{kj}(\theta_0)| I_{kj}^{obs}, \quad \text{as } v_n(r) \geq 1. \end{aligned}$$

For $U_1^{[3]}(\theta, \lambda)^{imp}$ we use (3.2.89), condition (3.2.64) and Remark 3.2.4 to obtain (3.2.96).

Now, turning to $U_2^{[3]}(\theta, \lambda)^{imp}$, we use (3.2.88). By (3.2.65), which follows from (3.2.51) and Remark 3.2.4, we have (3.2.96) for $U_2^{[3]}(\theta, \lambda)^{imp}$.

We now turn to

$$U_n^{[4,l]}(\theta, \lambda)^{imp} = n^{-1} \sum_{j=1}^m \sum_{i,k=1}^n a_{ij}^{[4,l]} b_{kj}^{[4]},$$

the first term in (3.2.87), where

$$a_{ij}^{[4,l]} = \lambda^T \left(\frac{\partial \mu_{ij}(\theta)}{\partial \theta} - \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} \right) V_{ij}^{-1}(\theta) \frac{\partial V_{lj}(\theta)}{\partial \theta^T} V_{lj}^{-1}(\theta),$$

$$|b_{kj}^{[4]}| = |V_{kj}^{-1}(\theta) V_{kj}(\theta_0) Z_{kj}(\theta_0) I_{kj}^{obs}| \leq v_n(r) |Z_{kj}(\theta_0)| I_{kj}^{obs}.$$

We apply (3.2.88). From condition (3.2.67), with $d_n(r)$ defined in (3.2.17), (3.2.97) follows for $U_n^{[4,l]}(\theta, \lambda)^{imp}$.

We write

$$U_n^{[5,l]}(\theta, \lambda)^{imp} = n^{-1} \sum_{j=1}^m \sum_{i,k=1}^n a_{ij}^{[5,l]} b_{kj}^{[5]}, \quad \text{where}$$

$$a_{ij}^{[5,l]} = \lambda^T \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} [V_{ij}^{-1}(\theta) V_{lj}^{-1}(\theta) - V_{ij}^{-1}(\theta_0) V_{lj}^{-1}(\theta_0)] \frac{\partial V_{lj}(\theta)}{\partial \theta^T} \lambda,$$

$$|b_{kj}^{[5]}| = |V_{kj}^{-1}(\theta) V_{kj}(\theta_0) Z_{kj}(\theta_0) I_{kj}^{obs}| \leq v_n(r) |Z_{kj}(\theta_0)| I_{kj}^{obs}.$$

An elementary calculation shows that

$$|V_{ij}^{-1}(\theta) V_{lj}^{-1}(\theta) - V_{ij}^{-1}(\theta_0) V_{lj}^{-1}(\theta_0)| \leq |V_{ij}^{-1}(\theta) + V_{ij}^{-1}(\theta_0)| s_n(r) V_{lj}^{-1}(\theta_0),$$

where $s_n(r)$ is defined in (3.2.45). We now put this together with the other factors of $a_{ij}^{[5,l]}$ and $b_{kj}^{[5]}$, and appeal to (3.2.68), which follows from (3.2.50), to conclude that (3.2.97) holds for $U_n^{[5,l]}(\theta, \lambda)^{imp}$.

We now consider $U_n^{[6,l]}(\theta, \lambda)^{imp}$, and use (3.2.88) with

$$a_{ij}^{[6,l]} = \lambda^T \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} V_{ij}^{-1}(\theta_0) V_{lj}^{-1}(\theta_0) \left[\frac{\partial V_{lj}(\theta)}{\partial \theta^T} - \frac{\partial V_{lj}(\theta_0)}{\partial \theta^T} \right] \lambda I_{ij}^{cen},$$

$$|b_{kj}^{[6]}| = |b_{kj}^{[1]}| \leq v_n(r) |Z_{kj}(\theta_0)| I_{kj}^{obs}.$$

By (3.2.69), which follows from hypothesis (3.2.50), (3.2.97) holds for $U_n^{[6,l]}(\theta, \lambda)^{imp}$.

The last term is $U_n^{[7,l]}(\theta, \lambda)^{imp}$, we apply (3.2.88) with

$$a_{ij}^{[7,l]} = \lambda^T \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} V_{ij}^{-1}(\theta_0) V_{lj}^{-1}(\theta_0) \frac{\partial V_{lj}(\theta_0)}{\partial \theta^T} \lambda I_{ij}^{cen},$$

$$b_{kj}^{[7]} = b_{kj}^{[3]} = [V_{kj}^{-1}(\theta) - V_{kj}^{-1}(\theta_0)] V_{kj}(\theta_0) Z_{kj}(\theta_0) I_{kj}^{obs}.$$

Following the search for an upper bound for $U_n^{[5]}(\theta, \lambda)$ in Lemma 3.2.5, and since (3.2.69) holds by hypotheses (3.2.50), we conclude that (3.2.97) holds for $U_n^{[7,l]}(\theta, \lambda)^{imp}$.

This concludes the proof of Lemma 3.2.11. \square

3.2.3 Further Results on the Derivative of $\hat{g}_{n,1}^{obs}$

We first put together the results of sections 3.2.1-3.2.2. Next, we use some results from these sections to prove Proposition 3.2.13, which is further used to prove $S(i)$ and $S(iii)$ for the derivative of $\hat{g}_{n,1}^{obs}$.

Proposition 3.2.12 *Assume that (T2) and the assumptions of lemmas 3.2.1, 3.2.3 and 3.2.5 hold. Then $S(ii)$ in Theorem 3.1.1 holds, with $\mathcal{D}_{n,1}(\theta) := -\frac{\partial}{\partial\theta^T}\hat{g}_{n,1}^{obs}(\theta)$.*

Proof. We first note that the hypotheses imply that the conclusions of lemmas 3.2.9-3.2.11 hold. Let us now recall the notation (3.2.78)-(3.2.85) and, define

$$H_n^{emp}(\theta) := H_n(\theta) - H_n^{imp}(\theta), \quad B_n^{emp}(\theta) := B_n(\theta) - B_n^{imp}(\theta), \quad \mathcal{E}_n^{emp}(\theta) := \mathcal{E}_n(\theta) - \mathcal{E}_n^{imp}(\theta).$$

We have, by recalling (3.2.2)-(3.2.14)

$$\mathcal{D}_{n,1}(\theta) = -\frac{\partial}{\partial\theta^T}\hat{g}_{n,1}^{obs}(\theta) = H_n^{emp}(\theta) - B_n^{emp}(\theta) - \mathcal{E}_n^{emp}(\theta).$$

The conclusion of the proposition follows now from lemmas 3.2.1, 3.2.3, 3.2.5 and 3.2.9-3.2.11. \square

The following results will be used to prove $S(i)$ and $S(iii)$ in Section 3.3.1.

For convenience, we recall definition (3.2.27).

$$k_n^{[1]} := \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} \frac{|\dot{\mu}_{ij}(\theta_0)|}{V_{ij}(\theta_0)}.$$

In a similar vein and using (3.2.30) and (3.2.33), we recall $c_n^{[1]}, v_n^{[1]}$. We also recall the definitions of $\mathcal{E}_n(\theta_0), \mathcal{E}_n^{[k]}(\theta_0), k = 1, 2$ in (3.2.11)-(3.2.13), $\mathcal{E}_n^{imp}(\theta_0), \mathcal{E}_n^{[k]}(\theta_0)^{imp}, k = 1, 2, 3$ in (3.2.82)-(3.2.85) and $\gamma_{j,n}^{cen}$ in (3.2.70) to prove the following result.

Proposition 3.2.13 *Assume that (T2) and $(B_{(f, \frac{\partial^2 \mu}{\partial\theta^T \partial\theta} V^{-1})})$ with (2.4.17) hold. Assume further that the E_{θ_0} -expectations of $\frac{\partial^2 \mu_{1j}(\theta_0)}{\partial\theta^T \partial\theta} V_{1j}^{-1}(\theta_0) Z_{1j}(\theta_0)$ and $\frac{\partial^2 \mu_{1j}(\theta_0)}{\partial\theta^T \partial\theta} V_{1j}^{-1}(\theta_0)$ exist. Then*

$$\lim_{n \rightarrow \infty} n^{-1} \|\mathcal{E}_n^{[1]}(\theta_0) - \mathcal{E}_n^{[1]}(\theta_0)^{imp}\| = 0 \quad a.s. \quad (3.2.98)$$

If, in addition, *a.s.*

$$\lim_{n \rightarrow \infty} n^{-1/2} k_n^{[1]} c_n^{[1]} v_n^{[1]} \lambda_{\max}^{1/2}(D_n) = 0, \quad (3.2.99)$$

then, *a.s.*

$$\lim_{n \rightarrow \infty} n^{-1} \|\mathcal{E}_n^{[2]}(\theta_0)\| = 0, \quad (3.2.100)$$

$$\lim_{n \rightarrow \infty} n^{-1} \|\mathcal{E}_n^{[k]}(\theta_0)^{imp}\| = 0, \quad k = 2, 3. \quad (3.2.101)$$

Therefore, *a.s.*

$$\lim_{n \rightarrow \infty} n^{-1} \|\mathcal{E}_n(\theta_0) - \mathcal{E}_n(\theta_0)^{imp}\| = 0. \quad (3.2.102)$$

Proof. To prove (3.2.98), by SLLN, we have that

$$n^{-1} \lambda^T \mathcal{E}_n^{[1]}(\theta_0) \lambda = n^{-1} \sum_{i=1}^n \sum_{j=1}^{\tau_i} \lambda^T \frac{\partial^2 \mu_{ij}(\theta_0)}{\partial \theta^T \partial \theta} \lambda V_{ij}^{-1}(\theta_0) Z_{ij}(\theta_0) I_{ij}^{obs},$$

converges *a.s.* to

$$\sum_{j=1}^m \lambda^T E_{\theta_0} \left[\frac{\partial^2 \mu_{1j}(\theta_0)}{\partial \theta^T \partial \theta} V_{1j}^{-1}(\theta_0) Z_{1j}(\theta_0) I_{1j}^{obs} \right] \lambda.$$

Similarly, and applying SLLN three times in (3.2.83), with $\theta = \theta_0$, we have that

$$\begin{aligned} n^{-1} \lambda^T \mathcal{E}_n^{[1]}(\theta_0)^{imp} \lambda &\rightarrow \\ &\sum_{j=1}^m \lambda^T E_{\theta_0}^{-1} [I_{1j}^{cen}] E_{\theta_0} \left[\frac{\partial^2 \mu_{1j}(\theta_0)}{\partial \theta^T \partial \theta} V_{1j}^{-1}(\theta_0) I_{1j}^{cen} \right] E_{\theta_0} [Z_{1j}(\theta_0) I_{1j}^{obs}] \lambda. \end{aligned}$$

To prove (3.2.98), it suffices to show that, for every $j \geq 1$,

$$\begin{aligned} &\lambda^T E_{\theta_0} \left[\frac{\partial^2 \mu_{1j}(\theta_0)}{\partial \theta^T \partial \theta} V_{1j}^{-1}(\theta_0) Z_{1j}(\theta_0) I_{1j}^{obs} \right] \lambda \\ &- \lambda^T E_{\theta_0}^{-1} [I_{1j}^{cen}] E_{\theta_0} \left[\frac{\partial^2 \mu_{1j}(\theta_0)}{\partial \theta^T \partial \theta} V_{1j}^{-1}(\theta_0) I_{1j}^{cen} \right] E_{\theta_0} [Z_{1j}(\theta_0) I_{1j}^{obs}] \lambda = 0. \end{aligned} \quad (3.2.103)$$

To show this, note first that, by (2.4.12)

$$E_{\theta_0} [Z_{ij}(\theta_0) I_{ij}^{cen}] = -E_{\theta_0} [Z_{ij}(\theta_0) I_{ij}^{obs}].$$

Now (3.2.103) becomes

$$\begin{aligned} & \lambda^T E_{\theta_0} \left[\frac{\partial^2 \mu_{1j}(\theta_0)}{\partial \theta^T \partial \theta} V_{1j}^{-1}(\theta_0) Z_{1j}(\theta_0) I_{1j}^{obs} \right] \lambda \\ & + \lambda^T E_{\theta_0}^{-1} [I_{1j}^{cen}] E_{\theta_0} \left[\frac{\partial^2 \mu_{1j}(\theta_0)}{\partial \theta^T \partial \theta} V_{1j}^{-1}(\theta_0) I_{1j}^{cen} \right] E_{\theta_0} [Z_{1j}(\theta_0) I_{1j}^{cen}] \lambda = 0. \end{aligned} \quad (3.2.104)$$

Note that the second term in (3.2.104) is

$$\begin{aligned} & \lambda^T E_{\theta_0} \left\{ \frac{\partial^2 \mu_{1j}(\theta_0)}{\partial \theta^T \partial \theta} V_{1j}^{-1}(\theta_0) I_{1j}^{cen} E_{\theta_0}^{-1} [I_{1j}^{cen}] E_{\theta_0} [Z_{1j}(\theta_0) I_{1j}^{cen}] \right\} \lambda \\ & = \lambda^T E_{\theta_0} \left\{ \frac{\partial^2 \mu_{1j}(\theta_0)}{\partial \theta^T \partial \theta} V_{1j}^{-1}(\theta_0) I_{1j}^{cen} E_{\theta_0} [Z_{1j}(\theta_0) | I_{1j}^{cen}] \right\} \lambda \\ & = \lambda^T E_{\theta_0} \left\{ \frac{\partial^2 \mu_{1j}(\theta_0)}{\partial \theta^T \partial \theta} V_{1j}^{-1}(\theta_0) I_{1j}^{cen} E_{\theta_0} [Z_{1j}(\theta_0) \mid \mathbb{O}(f_1, \frac{\partial^2 \mu_1}{\partial \theta^T \partial \theta} V_1^{-1})] \right\} \lambda \\ & = \lambda^T E_{\theta_0} \left\{ E_{\theta_0} \left[\frac{\partial^2 \mu_{1j}(\theta_0)}{\partial \theta^T \partial \theta} V_{1j}^{-1}(\theta_0) I_{1j}^{cen} Z_{1j}(\theta_0) \mid \mathbb{O}(f_1, \frac{\partial^2 \mu_1}{\partial \theta^T \partial \theta} V_1^{-1}) \right] \right\} \lambda. \end{aligned}$$

The first equality in the chain above follows from the definition of the conditional expectation. The second equality is due to $(B_{(f, \frac{\partial^2 \mu}{\partial \theta^T \partial \theta} V^{-1})})$ and the last to the fact that $\frac{\partial^2 \mu_{1j}(\theta_0)}{\partial \theta^T \partial \theta} V_{1j}^{-1}(\theta_0) I_{1j}^{cen}$ is $\mathbb{O}(f_1, \frac{\partial^2 \mu_1}{\partial \theta^T \partial \theta} V_1^{-1})$ -measurable (see (2.4.17)). Thus (3.2.104) reduces to

$$\lambda^T E_{\theta_0} \left[\frac{\partial^2 \mu_{1j}(\theta_0)}{\partial \theta^T \partial \theta} V_{1j}^{-1}(\theta_0) Z_{1j}(\theta_0) (I_{1j}^{obs} + I_{1j}^{cen}) \right] \lambda = 0,$$

or

$$\lambda^T E_{\theta_0} \left[\frac{\partial^2 \mu_{1j}(\theta_0)}{\partial \theta^T \partial \theta} V_{1j}^{-1}(\theta_0) Z_{1j}(\theta_0) I \{S_{1,j-1} < C_1\} \right] \lambda = 0, \quad (3.2.105)$$

since $I_{1j}^{obs} + I_{1j}^{cen} = I \{S_{1,j-1} < C_1\} \in \mathcal{G}_{1,j-1}$.

To prove (3.2.105), we note that $\frac{\partial^2 \mu_{1j}(\theta_0)}{\partial \theta^T \partial \theta} V_{1j}^{-1}(\theta_0)$ is $\mathcal{F}_{1,j-1}$ -measurable, therefore,

$$\begin{aligned} & \lambda^T E_{\theta_0} \left[\frac{\partial^2 \mu_{1j}(\theta_0)}{\partial \theta^T \partial \theta} V_{1j}^{-1}(\theta_0) Z_{1j}(\theta_0) I \{S_{1,j-1} < C_1\} \right] \lambda \\ & = \lambda^T E_{\theta_0} \left\{ E_{\theta_0} \left[\frac{\partial^2 \mu_{1j}(\theta_0)}{\partial \theta^T \partial \theta} V_{1j}^{-1}(\theta_0) Z_{1j}(\theta_0) I \{S_{1,j-1} < C_1\} \mid \mathcal{G}_{1,j-1} \right] \right\} \lambda \\ & = \lambda^T E_{\theta_0} \left\{ \frac{\partial^2 \mu_{1j}(\theta_0)}{\partial \theta^T \partial \theta} V_{1j}^{-1}(\theta_0) I \{S_{1,j-1} < C_1\} E_{\theta_0} [Z_{1j}(\theta_0) | \mathcal{G}_{1,j-1}] \right\} \lambda = 0. \end{aligned}$$

The last equality is due to (A0) and the model assumptions (2.1.2). This proves (3.2.104) for all $j \geq 1$ and thus (3.2.98).

The following proof uses some steps of the proof of Lemma 3.2.5, to which we refer here. Under condition (3.2.99) we now prove (3.2.100).

We write

$$\begin{aligned} \left| -\frac{1}{2}\lambda^T \mathcal{E}_n^{[2]}(\theta_0)\lambda \right| &= \left| \sum_{i=1}^n \sum_{j=1}^{\tau_i} \lambda^T \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} V_{ij}^{-2}(\theta_0) \frac{\partial V_{ij}(\theta_0)}{\partial \theta^T} Z_{ij}(\theta_0) I_{ij}^{obs} \lambda \right| \\ &\leq k_n^{[1]} c_n^{[1]} v_n^{[1]} \lambda_{\max}^{1/2}(D_n) n^{1/2} O_P(1). \end{aligned} \quad (3.2.106)$$

To obtain (3.2.106), we used the evaluation of the bound of $U_n^{[5]}$, defined after (3.2.66) and evaluated after (3.2.68), with $w_n^{[1]}(r)$ replaced by $v_n^{[1]}$. This proves (3.2.100).

Similarly, to prove (3.2.101) when $k = 2$, we first examine (3.2.84). Let

$$\begin{aligned} I_a(j) &:= \sum_{i=1}^n \left| \lambda^T \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} V_{ij}^{-2}(\theta_0) \frac{\partial V_{ij}(\theta_0)}{\partial \theta} \lambda \right|^2 I_{ij}^{cen}, \\ I_b(j) &:= \sum_{k=1}^n Z_{kj}^2(\theta_0) I_{kj}^{obs}. \end{aligned}$$

By (3.2.88) in Lemma 3.2.8, we conclude that the upper bound in (3.2.106) is also an upper bound for $\|\mathcal{E}_n^{[2]}(\theta_0)^{imp}\|$ with a modified constant. Using (3.2.99), this completes the proof of (3.2.101) when $k = 2$.

To prove (3.2.101) when $k = 3$ (see (3.2.85)), let

$$\begin{aligned} I_a(j) &:= \sum_{i=1}^n \lambda^T \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} V_{ij}^{-2}(\theta_0) \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} \lambda I_{ij}^{cen}, \\ I_b(j) &:= \sum_{k=1}^n \lambda^T \frac{\partial V_{kj}(\theta_0)}{\partial \theta} V_{kj}^{-2}(\theta_0) \frac{\partial V_{kj}(\theta_0)}{\partial \theta^T} \lambda Z_{kj}^2(\theta_0) I_{kj}^{obs}. \end{aligned}$$

By (3.2.88) in Lemma 3.2.8, we conclude that the upper bound in (3.2.106) is also an upper bound for $\|\mathcal{E}_n^{[3]}(\theta_0)^{imp}\|$ with a modified constant. Using (3.2.99) completes the proof of (3.2.101) when $k = 3$ and thus (3.2.102). This concludes the proof of Proposition 3.2.13. \square

3.2.4 Results for the Derivative of $\hat{g}_{n,2}^{obs}$

In this section we obtain results for the derivative of $\hat{g}_{n,2}^{obs}$, which parallel the results in sections 3.2.1-3.2.3 for the derivative of $\hat{g}_{n,1}^{obs}$. We write $\hat{g}_{n,2}^{obs}(\eta)$ as:

$$\hat{g}_{n,2}^{obs}(\eta) = \sum_{i=1}^n \sum_{j=1}^{\tau_i} b_{ij}(\eta)(Z_{ij}^2(\theta) - \sigma^2)I_{ij}^{obs} - \hat{g}_{n,2}^{imp}(\eta) = g_{n,2}^{[1]}(\eta) - \hat{g}_{n,2}^{imp}(\eta),$$

where

$$g_{n,2}^{[1]}(\eta) := \sum_{i=1}^n \sum_{j=1}^{\tau_i} b_{ij}(\eta)(Z_{ij}^2(\theta) - \sigma^2)I_{ij}^{obs}, \quad (3.2.107)$$

$$\hat{g}_{n,2}^{imp}(\eta) := \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{\tau_i} \gamma_{j,n}^{cen} b_{ij}(\eta) I_{ij}^{cen} \sum_{k=1}^n (Z_{kj}^2(\theta) - \sigma^2) I_{kj}^{obs}. \quad (3.2.108)$$

Taking the partial derivative in (3.2.107) gives:

$$\frac{\partial g_{n,2}^{[1]}(\eta)}{\partial \sigma^2} = \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left\{ \frac{\partial b_{ij}(\eta)}{\partial \sigma^2} (Z_{ij}^2(\theta) - \sigma^2) I_{ij}^{obs} - b_{ij}(\eta) I_{ij}^{obs} \right\}. \quad (3.2.109)$$

We introduce more notation and conditions.

$$b_n(r) := \sup_{\eta, \eta' \in B_r(\eta_0)} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} |b_{ij}(\eta) - b_{ij}(\eta')|, \quad (3.2.110)$$

$$b_n^{[1]}(r) := \sup_{\eta, \eta' \in B_r(\eta_0)} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} \left| \frac{\partial b_{ij}(\eta)}{\partial \sigma^2} - \frac{\partial b_{ij}(\eta')}{\partial \sigma^2} \right|. \quad (3.2.111)$$

Assume that

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} b_n(r) = 0 \quad a.s. \quad (3.2.112)$$

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} b_n^{[1]}(r) = 0 \quad a.s. \quad (3.2.113)$$

Assume further that, $r > 0$, almost surely,

$$\limsup_{n \rightarrow \infty} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} |b_{ij}(\theta_0, \sigma_0^2)| \leq C, \quad (3.2.114)$$

$$\limsup_{n \rightarrow \infty} \sup_{\eta \in B_r(\eta_0)} \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} \left| \frac{\partial b_{ij}(\eta)}{\partial \sigma^2} \right| \leq C, \quad (3.2.115)$$

where C is some constant.

We first note that, if $\hat{\theta}_n \rightarrow \theta_0$ a.s., then, given $r > 0$, there exists $n(\omega)$ such that $\|\hat{\theta}_n - \theta_0\| < r/\sqrt{2}$, for all $n \geq n(\omega)$. If we also take $\sigma^2 \in B_{\frac{r}{\sqrt{2}}}(\sigma_0^2)$, then $(\hat{\theta}_n, \sigma^2) \in B_r(\eta_0)$. Therefore, to prove $S(ii)$ for $\frac{\partial g_{n,2}^{[1]}(\hat{\theta}_n, \sigma^2)}{\partial \sigma^2}$, it suffices to prove Proposition 3.2.14 below.

Proposition 3.2.14 *Assume $\hat{\theta}_n \rightarrow \theta_0$, a.s., (T1), (2.7.10)-(2.7.11) and (3.2.112)-(3.2.115) hold. Then*

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} \sup_{(\hat{\theta}_n, \sigma^2) \in B_r(\eta_0)} \left| \frac{\partial g_{n,2}^{[1]}(\hat{\theta}_n, \sigma^2)}{\partial \sigma^2} - \frac{\partial g_{n,2}^{[1]}(\hat{\theta}_n, \sigma_0^2)}{\partial \sigma^2} \right| = 0 \quad a.s. \quad (3.2.116)$$

Proof. Consider

$$\frac{\partial g_{n,2}^{[1]}(\hat{\theta}_n, \sigma^2)}{\partial \sigma^2} - \frac{\partial g_{n,2}^{[1]}(\hat{\theta}_n, \sigma_0^2)}{\partial \sigma^2} = J_n^{[1]} + J_n^{[2]}, \quad (3.2.117)$$

where

$$J_n^{[1]} := \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left\{ \frac{\partial b_{ij}(\hat{\theta}_n, \sigma^2)}{\partial \sigma^2} (Z_{ij}^2(\hat{\theta}_n) - \sigma^2) I_{ij}^{obs} - \frac{\partial b_{ij}(\hat{\theta}_n, \sigma_0^2)}{\partial \sigma^2} (Z_{ij}^2(\hat{\theta}_n) - \sigma_0^2) I_{ij}^{obs} \right\},$$

$$J_n^{[2]} := - \sum_{i=1}^n \sum_{j=1}^{\tau_i} (b_{ij}(\hat{\theta}_n, \sigma^2) - b_{ij}(\hat{\theta}_n, \sigma_0^2)) I_{ij}^{obs}. \quad (3.2.118)$$

We can split $J_n^{[1]}$ into four terms, as follows

$$\begin{aligned} J_n^{[1]} &= \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left(\frac{\partial b_{ij}(\hat{\theta}_n, \sigma^2)}{\partial \sigma^2} - \frac{\partial b_{ij}(\hat{\theta}_n, \sigma_0^2)}{\partial \sigma^2} \right) (Z_{ij}^2(\hat{\theta}_n) - \sigma^2) I_{ij}^{obs} \\ &\quad - \sum_{i=1}^n \sum_{j=1}^{\tau_i} \frac{\partial b_{ij}(\hat{\theta}_n, \sigma_0^2)}{\partial \sigma^2} (\sigma^2 - \sigma_0^2) I_{ij}^{obs} \\ &= \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left\{ \left(\frac{\partial b_{ij}(\hat{\theta}_n, \sigma^2)}{\partial \sigma^2} - \frac{\partial b_{ij}(\hat{\theta}_n, \sigma_0^2)}{\partial \sigma^2} \right) (Z_{ij}^2(\hat{\theta}_n) - Z_{ij}^2(\theta_0)) I_{ij}^{obs} \right. \\ &\quad \left. + \left(\frac{\partial b_{ij}(\hat{\theta}_n, \sigma^2)}{\partial \sigma^2} - \frac{\partial b_{ij}(\hat{\theta}_n, \sigma_0^2)}{\partial \sigma^2} \right) (Z_{ij}^2(\theta_0) - \sigma_0^2) I_{ij}^{obs} \right\} \end{aligned}$$

$$\begin{aligned}
& - \left(\frac{\partial b_{ij}(\hat{\theta}_n, \sigma^2)}{\partial \sigma^2} - \frac{\partial b_{ij}(\hat{\theta}_n, \sigma_0^2)}{\partial \sigma^2} \right) (\sigma^2 - \sigma_0^2) I_{ij}^{obs} \Big\} \\
& - \sum_{i=1}^n \sum_{j=1}^{\tau_i} \frac{\partial b_{ij}(\hat{\theta}_n, \sigma_0^2)}{\partial \sigma^2} (\sigma^2 - \sigma_0^2) I_{ij}^{obs} \\
& := J_1^{[1]} + J_2^{[1]} - J_3^{[1]} - J_4^{[1]}. \tag{3.2.119}
\end{aligned}$$

We now find an upper bound for $n^{-1} J_1^{[1]}$:

$$\begin{aligned}
\left| n^{-1} J_1^{[1]} \right| & \leq \left| \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left(\frac{\partial b_{ij}(\hat{\theta}_n, \sigma^2)}{\partial \sigma^2} - \frac{\partial b_{ij}(\hat{\theta}_n, \sigma_0^2)}{\partial \sigma^2} \right) (Z_{ij}^2(\hat{\theta}_n) - Z_{ij}^2(\theta_0)) I_{ij}^{obs} \right| \\
& \leq b_n^{[1]}(r) \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{\tau_i} |Z_{ij}^2(\hat{\theta}_n) - Z_{ij}^2(\theta_0)| I_{ij}^{obs} \\
& = b_n^{[1]}(r) o_P(1), \tag{3.2.120}
\end{aligned}$$

with $b_n^{[1]}(r)$ defined in (3.2.111). For the last equality, we used (2.7.13), since $\hat{\theta}_n \rightarrow \theta_0$ *a.s.* By (3.2.113), we conclude that the left hand side of (3.2.120) converges to zero, when $n \rightarrow \infty$ and then $r \rightarrow 0$.

To find the upper bound of $n^{-1} J_2^{[1]}$, we write

$$\begin{aligned}
|n^{-1} J_2^{[1]}| & = \left| n^{-1} \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left(\frac{\partial b_{ij}(\hat{\theta}_n, \sigma^2)}{\partial \sigma^2} - \frac{\partial b_{ij}(\hat{\theta}_n, \sigma_0^2)}{\partial \sigma^2} \right) (Z_{ij}^2(\theta_0) - \sigma_0^2) I_{ij}^{obs} \right| \\
& \leq \left| n^{-1} \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left(\frac{\partial b_{ij}(\hat{\theta}_n, \sigma^2)}{\partial \sigma^2} - \frac{\partial b_{ij}(\hat{\theta}_n, \sigma_0^2)}{\partial \sigma^2} \right) Z_{ij}^2(\theta_0) I_{ij}^{obs} \right| \\
& \quad + \left| n^{-1} \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left(\frac{\partial b_{ij}(\hat{\theta}_n, \sigma^2)}{\partial \sigma^2} - \frac{\partial b_{ij}(\hat{\theta}_n, \sigma_0^2)}{\partial \sigma^2} \right) \sigma_0^2 I_{ij}^{obs} \right| \\
& \leq b_n^{[1]}(r) n^{-1} \sum_{i=1}^n \sum_{j=1}^{\tau_i} Z_{ij}^2(\theta_0) I_{ij}^{obs} + b_n^{[1]}(r) \sigma_0^2 n^{-1} \sum_{i=1}^n \sum_{j=1}^{\tau_i} I_{ij}^{obs} \\
& \leq b_n^{[1]}(r) O_P(1) + b_n^{[1]}(r) \sigma_0^2 \bar{\tau}_n. \tag{3.2.121}
\end{aligned}$$

In the last equality, for the first term we use the asymptotic boundedness given by (2.7.2). Since (T0) holds, $\bar{\tau}_n \rightarrow E(\tau_1) < \infty$, by SLLN. Using (3.2.113), we conclude

that the expression on the left of (3.2.121) converges *a.s.* to zero, as $n \rightarrow \infty$, and then $r \rightarrow 0$.

For the normalized of $J_3^{[1]}$, an upper bound is

$$\begin{aligned} \left| n^{-1} J_3^{[1]} \right| &\leq \left| \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left(\frac{\partial b_{ij}(\hat{\theta}_n, \sigma^2)}{\partial \sigma^2} - \frac{\partial b_{ij}(\hat{\theta}_n, \sigma_0^2)}{\partial \sigma^2} \right) (\sigma^2 - \sigma_0^2) I_{ij}^{obs} \right| \\ &\leq b_n^{[1]}(r) |\sigma^2 - \sigma_0^2| \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{\tau_i} I_{ij}^{obs} \\ &\leq b_n^{[1]}(r) |\sigma^2 - \sigma_0^2| \bar{\tau}_n \leq b_n^{[1]}(r) r \bar{\tau}_n, \end{aligned} \quad (3.2.122)$$

with $b_n^{[1]}(r)$ defined in (3.2.111). Since (T0) holds, $\bar{\tau}_n \rightarrow E(\tau_1) < \infty$, by SLLN. Using (3.2.113), we conclude that the expression on the left of the last inequality converges *a.s.* to zero, as $n \rightarrow \infty$, and then $r \rightarrow 0$.

To evaluate the normalized $J_4^{[1]}$, we write

$$\begin{aligned} \left| n^{-1} J_4^{[1]} \right| &= \left| \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{\tau_i} \frac{\partial b_{ij}(\hat{\theta}_n, \sigma_0^2)}{\partial \sigma^2} (\sigma^2 - \sigma_0^2) I_{ij}^{obs} \right| \\ &\leq C \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{\tau_i} |\sigma^2 - \sigma_0^2| I_{ij}^{obs} \\ &\leq C |\sigma^2 - \sigma_0^2| \bar{\tau}_n \leq Cr \bar{\tau}_n, \end{aligned} \quad (3.2.123)$$

where we used (3.2.115) in the second inequality. Since (T0) holds, $\bar{\tau}_n \rightarrow E(\tau_1) < \infty$, by SLLN. As before, the expression on the right of the last inequality converges *a.s.* to zero, as $n \rightarrow \infty$, and then $r \rightarrow 0$.

For the normalized $J_n^{[2]}$ in (3.2.118), we have following inequalities:

$$\begin{aligned} \left| n^{-1} J_n^{[2]} \right| &= \left| \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{\tau_i} \left(b_{ij}(\hat{\theta}_n, \sigma^2) - b_{ij}(\hat{\theta}_n, \sigma_0^2) \right) I_{ij}^{obs} \right| \\ &\leq b_n(r) \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^{\tau_i} I_{ij}^{obs} \\ &\leq b_n(r) \bar{\tau}_n, \end{aligned} \quad (3.2.124)$$

with $b_n(r)$ defined in (3.2.110). Since (T0) holds, $\bar{\tau}_n \rightarrow E(\tau_1) < \infty$, by SLLN. Combining this with (3.2.112), we conclude that the expression on the right of (3.2.124)

converges *a.s.* to zero, as $n \rightarrow \infty$, and then $r \rightarrow 0$. \square

Corollary 3.2.15 *Assume that the conditions of Proposition 3.2.14 hold. Then*

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} \sup_{\hat{\theta}_n \in B_r(\theta_0)} \left| \frac{\partial g_{n,2}^{[1]}(\hat{\theta}_n, \sigma_0^2)}{\partial \sigma^2} - \frac{\partial g_{n,2}^{[1]}(\theta_0, \sigma_0^2)}{\partial \sigma^2} \right| = 0 \quad a.s.$$

and so

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} \sup_{(\hat{\theta}_n, \sigma^2) \in B_r(\eta_0)} \left| \frac{\partial g_{n,2}^{[1]}(\hat{\theta}_n, \sigma^2)}{\partial \sigma^2} - \frac{\partial g_{n,2}^{[1]}(\theta_0, \sigma_0^2)}{\partial \sigma^2} \right| = 0 \quad a.s.$$

Proof. The proof of the first assertion is a simpler version of the proof of Proposition 3.2.14. The second assertion follows from the first and from (3.2.116). \square

Proposition 3.2.16 *Assume that (T2) and the conditions of Proposition 3.2.14 hold. Then*

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} \sup_{(\hat{\theta}_n, \sigma^2) \in B_r(\eta_0)} \left| \frac{\partial \hat{g}_{n,2}^{imp}(\hat{\theta}_n, \sigma^2)}{\partial \sigma^2} - \frac{\partial \hat{g}_{n,2}^{imp}(\hat{\theta}_n, \sigma_0^2)}{\partial \sigma^2} \right| = 0 \quad a.s.$$

Proof. The decomposition (3.2.119) is also valid for $\frac{\partial \hat{g}_{n,2}^{imp}(\eta)}{\partial \sigma^2} - \frac{\partial \hat{g}_{n,2}^{imp}(\eta_0)}{\partial \sigma^2}$. We use the notations $J_k^{[1]}(\eta)^{imp}$, $k = 1, 2, 3, 4$, $J_n^{[2]}(\eta)^{imp}$ for the corresponding term defined before. Let us focus on the first term $J_1^{[1]}(\eta)^{imp}$, which is

$$J_1^{[1]}(\eta)^{imp} = n^{-1} \sum_{i=1}^n \sum_{j=1}^{\tau_i} \gamma_{j,n}^{cen} \left(\frac{\partial b_{ij}(\hat{\theta}_n, \sigma^2)}{\partial \sigma^2} - \frac{\partial b_{ij}(\hat{\theta}_n, \sigma_0^2)}{\partial \sigma^2} \right) I_{ij}^{cen} \sum_{k=1}^n (Z_{kj}^2(\hat{\theta}_n) - Z_{kj}^2(\theta_0)) I_{kj}^{obs}.$$

Therefore

$$\begin{aligned} & \left| n^{-1} J_1^{[1]}(\eta)^{imp} \right| \\ &= \left| n^{-2} \sum_{i=1}^n \sum_{j=1}^{\tau_i} \gamma_{j,n}^{cen} \left(\frac{\partial b_{ij}(\hat{\theta}_n, \sigma^2)}{\partial \sigma^2} - \frac{\partial b_{ij}(\hat{\theta}_n, \sigma_0^2)}{\partial \sigma^2} \right) I_{ij}^{cen} \sum_{k=1}^n (Z_{kj}^2(\hat{\theta}_n) - Z_{kj}^2(\theta_0)) I_{kj}^{obs} \right| \\ &\leq n^{-2} \sum_{i=1}^n \sum_{j=1}^{\tau_i} \gamma_{j,n}^{cen} \left| \frac{\partial b_{ij}(\hat{\theta}_n, \sigma^2)}{\partial \sigma^2} - \frac{\partial b_{ij}(\hat{\theta}_n, \sigma_0^2)}{\partial \sigma^2} \right| I_{ij}^{cen} \sum_{k=1}^n |Z_{kj}^2(\hat{\theta}_n) - Z_{kj}^2(\theta_0)| I_{kj}^{obs} \end{aligned}$$

$$\begin{aligned}
&\leq b_n^{[1]}(r) \frac{1}{n} \sum_{j=1}^m \left(\frac{1}{n} \sum_{i=1}^n \gamma_{j,n}^{cen} I_{ij}^{cen} \right) \sum_{k=1}^n |Z_{kj}^2(\hat{\theta}_n) - Z_{ij}^2(\theta_0)| I_{kj}^{obs} \\
&= b_n^{[1]}(r) \frac{1}{n} \sum_{k=1}^n \sum_{j=1}^m |Z_{kj}^2(\hat{\theta}_n) - Z_{ij}^2(\theta_0)| I_{kj}^{obs},
\end{aligned}$$

by (3.2.71). Hence, both $n^{-1}J_1^{[1]}(\eta)^{imp}$ and $n^{-1}J_1^{[1]}(\eta)$ have the same upper bound (3.2.120).

Similarly, the upper bound of $n^{-1}J_k^{[1]}$, $k = 2, 3, 4$, in (3.2.121)-(3.2.123) are also upper bounds for $n^{-1}J_k^{[1]}(\eta)^{imp}$. Moreover, the normalized upper bound of $J_n^{[2]}$ in (3.2.124) can also serve as the normalized upper bound for $J_n^{[2]}(\eta)^{imp}$. This concludes the proof of Proposition 3.2.16. \square

The proof of the next result is very similar to the proof of Corollary 3.2.15 and will be omitted.

Corollary 3.2.17 *Assume that the conditions of Proposition 3.2.16 hold. Then*

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} \sup_{\hat{\theta}_n \in B_r(\theta_0)} \left| \frac{\partial \hat{g}_{n,2}^{imp}(\hat{\theta}_n, \sigma_0^2)}{\partial \sigma^2} - \frac{\partial \hat{g}_{n,2}^{imp}(\theta_0, \sigma_0^2)}{\partial \sigma^2} \right| = 0 \quad a.s.$$

and so

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} \sup_{(\hat{\theta}_n, \sigma^2) \in B_r(\eta_0)} \left| \frac{\partial \hat{g}_{n,2}^{imp}(\hat{\theta}_n, \sigma^2)}{\partial \sigma^2} - \frac{\partial \hat{g}_{n,2}^{imp}(\theta_0, \sigma_0^2)}{\partial \sigma^2} \right| = 0 \quad a.s. \quad \square$$

3.3 The Strong Consistency Theorem

3.3.1 The Strong Consistency of $\hat{\theta}_n$

In this section we use Theorem 3.1.1 to show the existence and strong consistency of a sequence of estimators $\hat{\theta}_n$, which are roots of the EE associated with $\hat{g}_{n,1}^{obs}(\theta)$. While condition (LN) of Theorem 3.1.1 is discussed in Theorem 2.4.9 and $S(ii)$ in sections 3.2.1-3.2.2, it remains to give sufficient conditions for $S(i)$ and $S(iii)$ to hold. To this

end, we introduce condition (S') , which is similar to condition (ii) of Theorem 4.13 in [3], or (S'_δ) in [9] with normalizing factor $n, n \geq 1$.

For the rest of this section $\theta = \theta_0$ and so, for the sake of simplicity, we omit writing it.

(S') *There exists a random integer $N \geq 1$ and a constant $C_0 > 0$, such that, for all $n \geq N$*

$$\lambda_{\min}[H_n - H_n^{imp}] > C_0 n.$$

In accordance with (2.1.6), we define here

$$f_{ij} := \frac{\partial \mu_{ij}(\theta_0)}{\partial \theta} V_{ij}^{-1}(\theta_0), \quad i, j \geq 1, \quad \lambda \in R^p, \quad \|\lambda\| = 1.$$

The next result gives a condition for (S') to hold.

Proposition 3.3.1 *Assume that (T2) holds, and that there exists $C_0 > 0$, such that, for all $\lambda \in R^p$*

$$\lambda^T \left[\sum_{j=1}^m \{E[f_{1j} f_{1j}^T I_{1j}^{obs}] - E^{-1}[I_{1j}^{cen}] E[f_{1j} I_{1j}^{cen}] E[f_{1j}^T I_{1j}^{obs}]\} \right] \lambda > C_0, \quad (3.3.1)$$

then

$$\lim_{n \rightarrow \infty} n^{-1} \lambda^T [H_n - H_n^{imp}] \lambda > C_0. \quad (3.3.2)$$

In (3.3.1), we assumed that all expectations exist.

Proof. We first write:

$$n^{-1} H_n^{imp} = n^{-2} \sum_{i=1}^n \sum_{j=1}^m f_{ij} I_{ij}^{cen} \gamma_{j,n}^{cen} \sum_{k=1}^n f_{kj}^T I_{kj}^{obs}, \quad (3.3.3)$$

and note that when $E[I_{1j}^{cen}] > 0$ by the definition of *e.f.*, $\gamma_{j,n}^{cen} \rightarrow E^{-1}[I_{1j}^{cen}]$ by the SLLN. Applying SLLN three times in (3.3.3), we obtain that, *a.s.* and as $n \rightarrow \infty$

$$n^{-1} \lambda^T H_n^{imp} \lambda \rightarrow \sum_{j=1}^m E^{-1}[I_{1j}^{cen}] \lambda^T E[f_{1j} I_{1j}^{cen}] E[f_{1j}^T I_{1j}^{obs}] \lambda. \quad (3.3.4)$$

Now

$$n^{-1}H_n = n^{-1} \sum_{i=1}^n \sum_{j=1}^m f_{ij} f_{ij}^T I_{ij}^{obs}, \quad \text{and, by SLLN,}$$

$$n^{-1} \lambda^T H_n \lambda \rightarrow \sum_{j=1}^m \lambda^T E[f_{1j} f_{1j}^T I_{1j}^{obs}] \lambda. \quad (3.3.5)$$

Using (3.3.1) with (3.3.4)-(3.3.5) gives (3.3.2). \square

Next, we give an example to illustrate condition (3.3.1). For simplicity, we take $p = 1$ and assume that, for some $0 < m < M$,

$$m \leq |f_{ij}| \leq M, \quad i, j \geq 1. \quad (3.3.6)$$

Condition (3.3.6) is satisfied if the derivatives are continuous and the covariates are equibounded, which is often the case in medical studies.

Example 3.3.2 *First, we intend to set, before the study starts, a censoring time C_i , such that j_0 recurrent events can be included in the study, for each individual $i \geq 1$. Assume that we have a good idea of what C_i should be, either from previous, similar studies, or because, as in [18], we know the expected length of time between two consecutive events, and the variability of this length is not large. Still, when we actually carry out the study, we cannot expect $C_i = S_{ij_0}$, rather that*

$$S_{i,j_0-1} < C_i < S_{i,j_0+1},$$

which means that we can have $I_{ij_0}^{obs} > 0, I_{ij_0}^{cen} > 0$ on disjoint sets, each with nonzero probability. Under these conditions, (3.3.1) becomes

$$\sum_{j=1}^{j_0-1} E[f_{1j}^2] + E[f_{1j_0}^2 I_{1j_0}^{obs}] - E^{-1}[I_{1j_0}^{cen}] E[f_{1j_0} I_{1j_0}^{cen}] E[f_{1j_0} I_{1j_0}^{obs}] > C_0, \quad (3.3.7)$$

for some $C_0 > 0$. This is so because $I_{ij}^{obs} \equiv 1, I_{ij}^{cen} \equiv 0$ for $j \leq j_0 - 1$.

By Proposition 3.3.1, in order to carry out the analysis for this study, we need to pick j_0 so that (3.3.7) holds. By (3.3.6)

$$\sum_{j=1}^{j_0-1} E[f_{1j}^2] + E[f_{1j_0}^2 I_{1j_0}^{obs}] \geq (j_0 - 1)m^2.$$

We also have

$$E^{-1}[I_{1j_0}^{cen}]E[f_{1j_0}I_{1j_0}^{cen}]E[f_{1j_0}I_{1j_0}^{obs}] \leq M^2 E[I_{1j_0}^{obs}] \leq M^2.$$

It follows that, if we take

$$j_0 - 1 > \frac{M^2}{m^2}, \quad (3.3.8)$$

then (3.3.7) holds with $C_0 = (j_0 - 1)m^2 - M^2 > 0$.

This example shows that, in order to impute the incomplete gap time using observed data, we must observe a sufficiently large number of gap times (inequality (3.3.8)).

Theorem 3.3.3 *Assume that the conclusions of Theorem 2.4.9 and the hypotheses of Proposition 3.2.12 hold. Furthermore, assume that (S') and the hypotheses of Proposition 3.2.13 hold. Then there exists a sequence $\hat{\theta}_n$ of estimators of the main parameter $\theta_0 \in R^p$, and a random integer $n_0(\omega)$, such that*

$$(a) P(\hat{g}_{n,1}^{obs}(\hat{\theta}_n) = 0, \text{ for all } n \geq n_0) = 1;$$

$$(b) \hat{\theta}_n \rightarrow \theta_0 \text{ a.s., } n \rightarrow \infty.$$

Proof. We apply Theorem 3.1.1 to $q_n = \hat{g}_{n,1}^{obs}$ and restrict it to the set of parameters $\theta \in R^p$. Assumption (LN) holds by the last conclusion of Theorem 2.4.9. By Proposition 3.2.12, condition S(ii) holds for $\mathcal{D}_{n,1}(\theta)$. It remains to prove S(i) and S(iii). Now (3.2.102) from Proposition 3.2.13 is

$$\lim_{n \rightarrow \infty} n^{-1} \|\mathcal{E}_n^{emp}(\theta_0)\| = 0.$$

With

$$\mathcal{D}_{n,1}(\theta) = H_n^{emp}(\theta) - B_n^{emp}(\theta) - \mathcal{E}_n^{emp}(\theta),$$

we write, for a large enough $n \geq N$ from (S'), and a small enough $r_1 > 0$,

$$\begin{aligned} & | \lambda^T \mathcal{D}_{n,1}(\theta) \lambda | \geq | \lambda^T H_n^{emp}(\theta_0) \lambda | \\ & - | \lambda^T [\mathcal{D}_{n,1}(\theta) - \mathcal{D}_{n,1}(\theta_0) - \mathcal{E}_n^{emp}(\theta_0)] \lambda | \end{aligned}$$

$$> C_0 n - \frac{2C_0 n}{3} = \frac{C_0 n}{3} > 0, \quad \text{for all } \theta \in B_{r_1}(\theta_0).$$

We used above the results of propositions 3.2.12-3.2.13. Then $S(i)$ and $S(iii)$ hold and the conclusions of Theorem 3.3.3 follow by Theorem 3.1.1. \square

3.3.2 The Strong Consistency of $\hat{\sigma}_n^2$

We first give sufficient conditions for $\hat{g}_{n,2}^{obs}(\hat{\theta}_n, \sigma_0^2)$ to satisfy (LN) in Theorem 3.1.1. We use Remark 3.1.2 with $q_{n,0}(\sigma^2) := \hat{g}_{n,2}^{obs}(\theta_0, \sigma^2)$, then combine the last conclusion of Theorem 2.5.2 with the result below.

Lemma 3.3.4 *Assume that (T2), (2.7.10)-(2.7.11) and (3.2.112), (3.2.114) hold. Assume further that $\hat{\theta}_n$ is a sequence of a.s. consistent estimators of θ_0 i.e., $\hat{\theta}_n \rightarrow \theta_0$ a.s. Then*

$$n^{-1}[\hat{g}_{n,2}^{obs}(\hat{\theta}_n, \sigma_0^2) - \hat{g}_{n,2}^{obs}(\theta_0, \sigma_0^2)] \rightarrow 0 \quad \text{a.s., } n \rightarrow \infty.$$

Proof. We evaluate the difference $n^{-1}[\hat{g}_{n,2}^{obs}(\hat{\theta}_n, \sigma_0^2) - \hat{g}_{n,2}^{obs}(\theta_0, \sigma_0^2)]$ and show that it converges to 0 a.s. From (2.4.4), this difference can be broken into the sum of two terms, (3.3.9) and (3.3.10):

$$\begin{aligned} & n^{-1} \sum_{i=1}^n \sum_{j=1}^{\tau_i} [b_{ij}(\hat{\theta}_n, \sigma_0^2) Z_{ij}^2(\hat{\theta}_n) - b_{ij}(\theta_0, \sigma_0^2) Z_{ij}^2(\theta_0)] I_{ij}^{obs} \\ & - n^{-1} \sum_{i=1}^n \sum_{j=1}^{\tau_i} [b_{ij}(\hat{\theta}_n, \sigma_0^2) - b_{ij}(\theta_0, \sigma_0^2)] \sigma_0^2 I_{ij}^{obs}, \end{aligned} \quad (3.3.9)$$

$$\begin{aligned} & n^{-2} \sum_{i=1}^n \sum_{j=1}^{\tau_i} \gamma_{j,n}^{cen} [b_{ij}(\hat{\theta}_n, \sigma_0^2) - b_{ij}(\theta_0, \sigma_0^2)] I_{ij}^{cen} \sum_{k=1}^n [Z_{kj}^2(\hat{\theta}_n) - \sigma_0^2] I_{kj}^{obs} \\ & + n^{-2} \sum_{i=1}^n \sum_{j=1}^{\tau_i} \gamma_{j,n}^{cen} b_{ij}(\theta_0, \sigma_0^2) I_{ij}^{cen} \sum_{k=1}^n [Z_{kj}^2(\hat{\theta}_n) - Z_{kj}^2(\theta_0)] I_{kj}^{obs} := L_1 + L_2. \end{aligned} \quad (3.3.10)$$

We start with the expression (3.3.9), the first double summation. A bound for the summands can be found as follows, with $b_n(r)$ defined in (3.2.110).

$$|b_{ij}(\hat{\theta}_n, \sigma_0^2) Z_{ij}^2(\hat{\theta}_n) - b_{ij}(\theta_0) Z_{ij}^2(\theta_0)|$$

$$\begin{aligned}
&\leq |b_{ij}(\hat{\theta}_n, \sigma_0^2) - b_{ij}(\eta_0)| |Z_{ij}^2(\hat{\theta}_n)| + |b_{ij}(\eta_0)| |Z_{ij}^2(\hat{\theta}_n) - Z_{ij}^2(\theta_0)| \\
&\leq b_n(r) |Z_{ij}^2(\hat{\theta}_n) - Z_{ij}^2(\theta_0)| + b_n(r) Z_{ij}^2(\theta_0) + \max_{1 \leq i \leq n, 1 \leq j \leq \tau_i} |b_{ij}(\eta_0)| |Z_{ij}^2(\hat{\theta}_n) - Z_{ij}^2(\theta_0)|.
\end{aligned}$$

Using the hypotheses (3.2.114), an upper bound for the first double sum of (3.3.9) is

$$Cn^{-1} \sum_{i=1}^n \sum_{j=1}^{\infty} |Z_{ij}^2(\hat{\theta}_n) - Z_{ij}^2(\theta_0)| I_{ij}^{obs} + b_n(r) n^{-1} \sum_{i=1}^n \sum_{j=1}^{\infty} Z_{ij}^2(\theta_0) I_{ij}^{obs}. \quad (3.3.11)$$

Since $I_{ij}^{obs} = I\{S_{ij} \leq C_i\} \leq I\{S_{i,j-1} < C_i\}$, the first term above converges *a.s.* to zero, as in (2.7.13). Notice that for this term to converge to zero, the weaker assumption (T1) is used in lieu of (T2). By Lemma 2.7.1 and (3.2.112), the second term converges *a.s.* to zero.

Now

$$\begin{aligned}
&n^{-1} \sum_{i=1}^n \sum_{j=1}^{\infty} |b_{ij}(\hat{\theta}_n, \sigma_0^2) - b_{ij}(\theta_0, \sigma_0^2)| I_{ij}^{obs} \\
&\leq n^{-1} b_n(r) \sum_{i=1}^n \sum_{j=1}^{\infty} I\{S_{i,j-1} < C_i\} = b_n(r) \bar{\tau}_n,
\end{aligned} \quad (3.3.12)$$

with $b_n(r)$ defined in (3.2.110). Since $\bar{\tau}_n \rightarrow E[\tau_1] < \infty$, the last expression converges to zero *a.s.*, by (3.2.112). This completes the proof of the convergence of (3.3.9).

We now show how the proof of convergence of the terms in (3.3.10) reduces to the result proven above. For this purpose, let us look at

$$\begin{aligned}
|L_1| &\leq n^{-2} \sum_{i=1}^n \sum_{j=1}^{\tau_i} \gamma_{j,n}^{cen} |b_{ij}(\hat{\theta}_n, \sigma_0^2) - b_{ij}(\theta_0, \sigma_0^2)| I_{ij}^{cen} \sum_{k=1}^n |Z_{kj}^2(\hat{\theta}_n) - \sigma_0^2| I_{kj}^{obs} \\
&\leq n^{-1} b_n(r) \sum_{k=1}^n \sum_{j=1}^{\tau_i} |Z_{kj}^2(\hat{\theta}_n) - \sigma_0^2| I_{kj}^{obs} \\
&\leq n^{-1} b_n(r) \sum_{k=1}^n \sum_{j=1}^{\tau_i} |Z_{kj}^2(\hat{\theta}_n) - Z_{kj}^2(\theta_0)| I_{kj}^{obs} + n^{-1} b_n(r) \sum_{k=1}^n \sum_{j=1}^{\tau_i} Z_{kj}^2(\theta_0) I_{kj}^{obs} \\
&+ \sigma_0^2 n^{-1} b_n(r) \sum_{k=1}^n \sum_{j=1}^{\tau_i} I_{kj}^{obs}.
\end{aligned} \quad (3.3.13)$$

In the second inequality, we made use of (3.2.71). Since the first and second terms of (3.3.13) are less than (3.3.11), which converges to zero, so they also converge to zero. Using (3.3.12), the third term of (3.3.13) converges to zero too.

We also use (3.2.71) and hypothesis (3.2.114) to obtain

$$|L_2| \leq Cn^{-1} \sum_{k=1}^n \sum_{j=1}^{\tau_i} |Z_{kj}^2(\hat{\theta}_n) - Z_{kj}^2(\theta_0)| I_{kj}^{obs}. \quad (3.3.14)$$

Because the right hand side of (3.3.14) is the first part of (3.3.11), it converges to zero too. This completes the proof of Lemma 3.3.4. \square

We dealt with $S(ii)$ in Section 3.2.4. To deal with $S(i)$ and $S(iii)$, recall (3.2.107)-(3.2.109) and rewrite them as:

$$-\frac{\partial g_{n,2}^{[1]}(\eta)}{\partial \sigma^2} = H_{n,2}(\eta) - \mathcal{E}_{n,2}(\eta),$$

where

$$H_{n,2}(\eta) := \sum_{i=1}^n \sum_{j=1}^{\tau_i} b_{ij}(\eta) I_{ij}^{obs},$$

$$\mathcal{E}_{n,2}(\eta) := \sum_{i=1}^n \sum_{j=1}^{\tau_i} \frac{\partial b_{ij}(\eta)}{\partial \sigma^2} (Z_{ij}^2(\theta) - \sigma^2) I_{ij}^{obs},$$

and

$$-\frac{\partial g_{n,2}^{imp}(\eta)}{\partial \sigma^2} = H_{n,2}^{imp}(\eta) - \mathcal{E}_{n,2}^{imp}(\eta),$$

$$H_{n,2}^{imp}(\eta) := n^{-1} \sum_{i=1}^n \sum_{j=1}^{\tau_i} \gamma_{j,n}^{cen} b_{ij}(\eta) I_{ij}^{cen} \sum_{k=1}^n I_{kj}^{obs},$$

$$\mathcal{E}_{n,2}^{imp}(\eta) := n^{-1} \sum_{i=1}^n \sum_{j=1}^{\tau_i} \gamma_{j,n}^{cen} \frac{\partial b_{ij}(\eta)}{\partial \sigma^2} I_{ij}^{cen} \sum_{k=1}^n (Z_{kj}^2(\theta) - \sigma^2) I_{kj}^{obs}, \quad (3.3.15)$$

$$H_{n,2}^{emp}(\eta) := H_{n,2}(\eta) - H_{n,2}^{imp}(\eta), \quad \mathcal{E}_{n,2}^{emp}(\eta) := \mathcal{E}_{n,2}(\eta) - \mathcal{E}_{n,2}^{imp}(\eta).$$

The next result is similar to Proposition 3.3.1.

Proposition 3.3.5 *Assume that (T2) holds, and $\max_{1 \leq j \leq m} E|b_{1j}(\eta_0)| < \infty$.*

Furthermore, there exists $C_0 > 0$, such that

$$\sum_{j=1}^m \{E[b_{1j}(\eta_0)I_{1j}^{obs}] - E^{-1}[I_{1j}^{cen}]E[b_{1j}(\eta_0)I_{1j}^{cen}]E[I_{1j}^{obs}]\} > C_0. \quad (3.3.16)$$

Then

$$\lim_{n \rightarrow \infty} n^{-1}[H_{n,2}(\eta_0) - H_{n,2}(\eta_0)^{imp}] > C_0. \quad (3.3.17)$$

In (3.3.16), we assumed that expectations exist.

Proof. We first write

$$n^{-1}H_{n,2}(\eta_0)^{imp} := n^{-2} \sum_{i=1}^n \sum_{j=1}^{\tau_i} \gamma_{j,n}^{cen} b_{ij}(\eta_0) I_{ij}^{cen} \sum_{k=1}^n I_{kj}^{obs}. \quad (3.3.18)$$

We note that, when $E[I_{1j}^{cen}] > 0$, $\gamma_{j,n}^{cen} \rightarrow E^{-1}[I_{1j}^{cen}]$ by the SLLN. Applying the SLLN three times in (3.3.18), we obtain that, *a.s.* and when $n \rightarrow \infty$

$$n^{-1}H_{n,2}(\eta_0)^{imp} \rightarrow \sum_{j=1}^m E^{-1}[I_{1j}^{cen}]E[b_{1j}(\eta_0)I_{1j}^{cen}]E[I_{1j}^{obs}]. \quad (3.3.19)$$

Now

$$n^{-1}H_{n,2}(\eta_0) = n^{-1} \sum_{i=1}^n \sum_{j=1}^m b_{ij}(\eta_0) I_{ij}^{obs}, \quad \text{and, by the SLLN,}$$

$$n^{-1}H_{n,2}(\eta_0) \rightarrow \sum_{j=1}^m E[b_{1j}(\eta_0)I_{1j}^{obs}]. \quad (3.3.20)$$

Using (3.3.16) with (3.3.19)-(3.3.20) gives (3.3.17), which concludes our proof. \square

Proposition 3.3.6 *Assume that (T2) holds and that $\frac{\partial b_{ij}(\eta_0)}{\partial \sigma^2} I_{ij}^{cen}$ is $\mathcal{O}(b_i(\eta_0))$ measurable. Furthermore, assume that*

$$E_{\eta_0} \left| \frac{\partial b_{1j}(\eta_0)}{\partial \sigma^2} \right| Z_{1j}^2(\theta_0) < \infty \quad \text{and} \quad E_{\eta_0} \left| \frac{\partial b_{1j}(\eta_0)}{\partial \sigma^2} \right| < \infty.$$

Then

$$n^{-1}[\mathcal{E}_{n,2}(\eta_0) - \mathcal{E}_{n,2}^{imp}(\eta_0)] \rightarrow 0, \quad \text{a.s.,} \quad n \rightarrow \infty. \quad (3.3.21)$$

Proof. By SLLN, we have that

$$n^{-1}\mathcal{E}_{n,2}(\eta_0) = n^{-1} \sum_{j=1}^m \sum_{i=1}^n \frac{\partial b_{ij}(\eta_0)}{\partial \sigma^2} (Z_{ij}^2(\theta_0) - \sigma_0^2) I_{ij}^{obs},$$

converges *a.s.* to

$$\sum_{j=1}^m E_{\eta_0} \left[\frac{\partial b_{1j}(\eta_0)}{\partial \sigma^2} (Z_{1j}^2(\theta_0) - \sigma_0^2) I_{1j}^{obs} \right]. \quad (3.3.22)$$

Similarly, and applying the SLLN three times in (3.3.15), we have that

$$n^{-1}\mathcal{E}_{n,2}^{imp}(\eta_0) \rightarrow \sum_{j=1}^m E_{\eta_0}^{-1}[I_{1j}^{cen}] E_{\eta_0} \left[\frac{\partial b_{1j}(\eta_0)}{\partial \sigma^2} I_{1j}^{cen} \right] E_{\eta_0} [(Z_{1j}^2(\theta_0) - \sigma_0^2) I_{1j}^{obs}]. \quad (3.3.23)$$

To prove the proposition, it suffices to show that, for every $j \geq 1$,

$$E_{\eta_0} \left[\frac{\partial b_{1j}(\eta_0)}{\partial \sigma^2} (Z_{1j}^2(\theta_0) - \sigma_0^2) I_{1j}^{obs} \right] - E_{\eta_0}^{-1}[I_{1j}^{cen}] E_{\eta_0} \left[\frac{\partial b_{1j}(\eta_0)}{\partial \sigma^2} I_{1j}^{cen} \right] E_{\eta_0} [(Z_{1j}^2(\theta_0) - \sigma_0^2) I_{1j}^{obs}] = 0. \quad (3.3.24)$$

To show this, note first that, by (2.4.13)

$$E_{\eta_0} [(Z_{1j}^2(\theta_0) - \sigma_0^2) I_{1j}^{obs}] = -E_{\eta_0} [(Z_{1j}^2(\theta_0) - \sigma_0^2) I_{1j}^{cen}].$$

Now (3.3.24) becomes

$$E_{\eta_0} \left[\frac{\partial b_{1j}(\eta_0)}{\partial \sigma^2} (Z_{1j}^2(\theta_0) - \sigma_0^2) I_{1j}^{obs} \right] + E_{\eta_0}^{-1}[I_{1j}^{cen}] E_{\eta_0} \left[\frac{\partial b_{1j}(\eta_0)}{\partial \sigma^2} I_{1j}^{cen} \right] E_{\eta_0} [(Z_{1j}^2(\theta_0) - \sigma_0^2) I_{1j}^{cen}] = 0. \quad (3.3.25)$$

Note that the second term in (3.3.25) is

$$\begin{aligned} & E_{\eta_0} \left\{ \frac{\partial b_{1j}(\eta_0)}{\partial \sigma^2} I_{1j}^{cen} E_{\eta_0}^{-1}[I_{1j}^{cen}] E_{\eta_0} [(Z_{1j}^2(\theta_0) - \sigma_0^2) I_{1j}^{cen}] \right\} \\ &= E_{\eta_0} \left\{ \frac{\partial b_{1j}(\eta_0)}{\partial \sigma^2} I_{1j}^{cen} E_{\eta_0} [(Z_{1j}^2(\theta_0) - \sigma_0^2) | I_{1j}^{cen}] \right\} \\ &= E_{\eta_0} \left\{ \frac{\partial b_{1j}(\eta_0)}{\partial \sigma^2} I_{1j}^{cen} E_{\eta_0} [(Z_{1j}^2(\theta_0) - \sigma_0^2) | \mathbb{O}(b_1(\eta_0))] \right\} \\ &= E_{\eta_0} \left\{ E_{\eta_0} \left[\frac{\partial b_{1j}(\eta_0)}{\partial \sigma^2} (Z_{1j}^2(\theta_0) - \sigma_0^2) I_{1j}^{cen} | \mathbb{O}(b_1(\eta_0)) \right] \right\}. \end{aligned}$$

The second equality in the chain above is due to $(B_{b(\eta)})$ and the last to the fact that $\frac{\partial b_{1j}(\eta_0)}{\partial \sigma^2} I_{1j}^{cen}$ is $\mathbb{O}(b_1(\eta_0))$ -measurable. Thus (3.3.25) reduces to

$$E_{\eta_0} \left[\frac{\partial b_{1j}(\eta_0)}{\partial \sigma^2} (Z_{1j}^2(\theta_0) - \sigma_0^2) (I_{1j}^{obs} + I_{1j}^{cen}) \right] = 0,$$

or

$$E_{\eta_0} \left[\frac{\partial b_{1j}(\eta_0)}{\partial \sigma^2} (Z_{1j}^2(\theta_0) - \sigma_0^2) I\{S_{1,j-1} < C_1\} \right] = 0, \quad (3.3.26)$$

since $I_{1j}^{obs} + I_{1j}^{cen} = I\{S_{1,j-1} < C_1\} \in \mathcal{G}_{1,j-1}$.

To prove (3.3.26), we note that $\frac{\partial b_{1j}(\eta_0)}{\partial \sigma^2}$ is $\mathcal{F}_{1,j-1}$ -measurable, as is $b_{1j}(\eta_0)$. Therefore

$$\begin{aligned} & E_{\eta_0} \left[\frac{\partial b_{1j}(\eta_0)}{\partial \sigma^2} (Z_{1j}^2(\theta_0) - \sigma_0^2) I\{S_{1,j-1} < C_1\} \right] \\ &= E_{\eta_0} \left\{ E_{\eta_0} \left[\frac{\partial b_{1j}(\eta_0)}{\partial \sigma^2} (Z_{1j}^2(\theta_0) - \sigma_0^2) I\{S_{1,j-1} < C_1\} \middle| \mathcal{G}_{1,j-1} \right] \right\} \\ &= E_{\eta_0} \left\{ \frac{\partial b_{1j}(\eta_0)}{\partial \sigma^2} I\{S_{1,j-1} < C_1\} E_{\eta_0} [(Z_{1j}^2(\theta_0) - \sigma_0^2) | \mathcal{G}_{1,j-1}] \right\} = 0. \end{aligned}$$

The last equality is due to (A0) and the model assumptions (2.1.2). This proves (3.3.25) for all $j \geq 1$ and thus (3.3.21). \square

To summarize the above results, we state the following theorem.

Theorem 3.3.7 *We adopt the hypotheses and notation of Theorem 2.5.2 and Theorem 3.3.3. We also assume the hypotheses of propositions 3.2.14 and 3.2.16, Lemma 3.3.4 and propositions 3.3.5-3.3.6. Then there exists a sequence $\hat{\eta}_n^T = (\hat{\theta}_n^T, \hat{\sigma}_n^2)$ of estimators of the parameter $\eta_0 \in R^{p+1}$, and a random integer $n_0(\omega)$, such that*

- (a) $P(\hat{g}_{n,2}^{obs}(\hat{\theta}_n, \hat{\sigma}_n^2) = 0, \text{ for all } n \geq n_0) = 1;$
- (b) $\hat{\eta}_n \rightarrow \eta_0 \text{ a.s., } n \rightarrow \infty.$

Proof. We apply Theorem 3.1.1 to $q_n(\sigma^2) := \hat{g}_{n,2}^{obs}(\hat{\theta}_n, \sigma^2)$. Assumption (LN) holds by Lemma 3.3.4 and the last conclusion of Theorem 2.5.2. By propositions 3.2.14 and 3.2.16, condition S(ii) holds for $\mathcal{D}_{n,2}(\hat{\theta}_n, \sigma^2)$. It remains to prove S(i) and S(iii).

From the results of Proposition 3.3.6, we have

$$\lim_{n \rightarrow \infty} n^{-1} \mathcal{E}_{n,2}^{emp}(\eta_0) = 0.$$

With $\mathcal{D}_{n,2}(\eta) := -\frac{\partial \hat{g}_{n,2}^{obs}(\eta)}{\partial \sigma^2} = H_{n,2}^{emp}(\eta) - \mathcal{E}_{n,2}^{emp}(\eta)$, we write, for a large enough $n \geq n_1$ and a small enough $r_1 > 0$,

$$\begin{aligned} & | \mathcal{D}_{n,2}(\hat{\theta}_n, \sigma^2) | \geq | H_{n,2}^{emp}(\eta_0) | \\ & - | \mathcal{D}_{n,2}(\hat{\theta}_n, \sigma^2) - \mathcal{D}_{n,2}(\eta_0) - \mathcal{E}_{n,2}^{emp}(\eta_0) | \\ & \geq C_0 n - \frac{2C_0 n}{3} = \frac{C_0 n}{3} > 0, \quad \text{for all } \eta \in B_{r_1}(\eta_0). \end{aligned}$$

We used above the results of corollaries 3.2.15 and 3.2.17 and propositions 3.3.5-3.3.6.

Then $S(i)$ and $S(iii)$ hold. \square

Chapter 4

The Asymptotic Normality of Estimators

In this chapter we present the asymptotic normality of sequences of consistent estimators $\hat{\theta}_n$ and $\hat{\sigma}_n^2$, of the parameters θ_0 and σ_0^2 , which are defined in Theorems 3.3.3 and 3.3.7, respectively.

In order to simplify the exposition, we further assume that the conditions of Theorems 3.3.3 and 3.3.7 actually hold.

It is well known that, under certain conditions, estimators $\hat{\beta}_n$ that are implicit roots of unbiased estimating equations are asymptotically normal with mean 0 and covariance matrix commonly referred to as “sandwich”. The proofs given in the literature are usually sketchy and proceed in two steps. First, it is shown that a CLT holds for the original *e.f.*’s evaluated at a true parameter, say β_0 . The limiting distribution is normal, with mean 0 and covariance Σ , say, which is nonsingular. Secondly, using the mean value theorem, a normal asymptotic distribution for $n^{1/2}(\hat{\beta}_n - \beta_0)$ is obtained, with mean zero and variance

$$\mathcal{D}^{-1}(\beta_0)\Sigma[\mathcal{D}^{-1}(\beta_0)]^T,$$

where $\mathcal{D}(\beta_0)$ is nonsingular and equal to the limit of $n^{-1}\mathcal{D}_n(\beta_0)$, while $-\mathcal{D}_n(\beta_0)$ is the

derivative of the *e.f.* with respect to the parameter β , evaluated at β_0 .

In our case, the situation is more complicated and an additional step is required. This is so because the initial *e.f.*'s also contain imputed terms, so they no longer consist of simple sums of random vectors. A further complication arises in determining the asymptotic distribution of $\hat{\sigma}_n^2$. It is due to the fact that the *e.f.*'s, which define $\hat{\sigma}_n^2$, also contain $\hat{\theta}_n$, which has a distribution of its own that must be taken into account. In what follows, we give complete proofs of our results, following the ideas and proofs in [20], Section 3.4.

4.1 The Asymptotic Normality of $\hat{\theta}_n$

In this section we present three results. The first demonstrates that $n^{-1/2}g_{n,1}^{obs}(\theta_0)$, is asymptotically normal with mean zero and covariance Σ . The second result shows that the asymptotic distribution of $n^{-1/2}\hat{g}_{n,1}^{obs}(\theta_0)$ is also $N(0, \Sigma)$. Finally using the mean value theorem, we show that $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is asymptotically normal with mean 0 and a covariance matrix of a sandwich form,

$$\mathcal{D}_1^{-1}(\theta_0)\Sigma[\mathcal{D}_1^{-1}(\theta_0)]^T,$$

where

$$\mathcal{D}_1(\theta_0) = \lim_{n \rightarrow \infty} n^{-1}\mathcal{D}_{n,1}(\theta_0), \quad \text{and} \quad \mathcal{D}_{n,1}(\theta_0) = -\frac{\partial \hat{g}_{n,1}^{obs}(\theta_0)}{\partial \theta^T}, \quad n \geq 1, \quad (4.1.1)$$

with $\mathcal{D}_1(\theta_0)$ nonrandom and nonsingular.

Let us write

$$g_{n,1}^{obs}(\theta_0) := \sum_{i=1}^n u_{i,1}, \quad u_{i,1} := \sum_{j=1}^m [a_{ij}^{(1)} I_{ij}^{obs} - b_{ij}^{(1)} I_{ij}^{cen}],$$

where

$$a_{ij}^{(1)} := f_{ij}(\theta_0)Z_{ij}(\theta_0), \quad b_{ij}^{(1)} := f_{ij}(\theta_0) \frac{E_{\theta_0}[Z_{1j}(\theta_0)I_{1j}^{obs}]}{E_{\theta_0}[I_{1j}^{cen}]}. \quad (4.1.2)$$

In the course of proving the following result, we calculate all entries of the variance matrix.

Theorem 4.1.1 *Assume that (T2) holds, and that*

$$\max_{1 \leq j \leq m} E_{\theta_0} [\| f_{1j}(\theta_0) \|^2 Z_{1j}^2(\theta_0)] < \infty, \quad (4.1.3)$$

$$\max_{1 \leq j \leq m} E_{\theta_0} \| f_{1j}(\theta_0) \|^2 < \infty. \quad (4.1.4)$$

Then

$$n^{-1/2} g_{n,1}^{obs}(\theta_0) \xrightarrow{\mathcal{L}} N(0, \Sigma), \quad (4.1.5)$$

where $N(0, \Sigma)$ is a p -dimensional random vector normally distributed with zero mean and covariance matrix Σ . The (kl) -entry of Σ is:

$$\begin{aligned} & \sum_{j,j'=1}^m E_{\theta_0} [a_{1j,k}^{(1)} a_{1j',l}^{(1)} I_{\max\{j,j'\}}^{obs}] - \sum_{j=1, j'>j}^m E_{\theta_0} [a_{1j,k}^{(1)} I_{1j'}^{cen} b_{1j',l}^{(1)}] \\ & - \sum_{j'=1, j>j'}^m E_{\theta_0} [a_{1j',l}^{(1)} I_{1j}^{cen} b_{1j,k}^{(1)}] + \sum_{j=1}^m E_{\theta_0} [b_{1j,k}^{(1)} b_{1j,l}^{(1)} I_{1j}^{cen}]. \end{aligned} \quad (4.1.6)$$

Proof. To simplify the proof, we write $u_i = u_{i,1}$, with $u_i^T = (u_{ik})_{k=1,2,\dots,p}$, $i \geq 1$. Since $E_{\theta_0}[u_1] = 0$ by Proposition 2.4.6, the proof of Theorem 4.1.1 follows from Theorem 29.5 of [4], once we show that

$$\max_{1 \leq k \leq p} E_{\theta_0}[u_{1k}^2] < \infty. \quad (4.1.7)$$

To this end, we first calculate the entries of Σ . We have, for a fixed i , and $1 \leq k, l \leq p$.

$$\begin{aligned} u_{ik} u_{il} &= \sum_{j,j'=1}^m a_{ij,k}^{(1)} a_{ij',l}^{(1)} I_{ij}^{obs} I_{ij'}^{obs} - \sum_{j,j'=1}^m a_{ij,k}^{(1)} b_{ij',l}^{(1)} I_{ij}^{obs} I_{ij'}^{cen} \\ & - \sum_{j,j'=1}^m a_{ij',l}^{(1)} b_{ij,k}^{(1)} I_{ij'}^{obs} I_{ij}^{cen} + \sum_{j,j'=1}^m b_{ij,k}^{(1)} b_{ij',l}^{(1)} I_{ij}^{cen} I_{ij'}^{cen}. \end{aligned} \quad (4.1.8)$$

Since $I_{ij}^{obs} I_{ij'}^{obs} = I_{i \max\{j,j'\}}^{obs}$, the expectation of the first double sum in (4.1.8) is given by the first double sum in (4.1.6). Because $I_{ij}^{obs} I_{ij'}^{cen} = I_{ij'}^{cen}$, if $j' > j$, and is zero otherwise, the expectation of the second double sum in (4.1.8) equals the second double sum in (4.1.6) and the same reasoning applies to the third double sum. Now

the fourth double sum in (4.1.8) reduces to $\sum_{j=1}^m b_{ij,k}^{(1)} b_{ij,l}^{(1)} I_{ij}^{cen}$, because $I_{ij}^{cen} I_{ij'}^{cen} = I_{ij}^{cen}$ if $j' = j$, and is zero otherwise.

We now show that the expectations of all terms in (4.1.8) exist, which implies (4.1.7). Starting with the first, we have

$$\begin{aligned} \left| \sum_{j,j'=1}^m a_{ij,k}^{(1)} a_{ij',l}^{(1)} I_{i \max\{j,j'\}}^{obs} \right| &\leq \left(\sum_{j=1}^m \|a_{ij}^{(1)}\| \right)^2 \leq C \sum_{j=1}^m \|a_{ij}^{(1)}\|^2 \\ &\leq C \sum_{j=1}^m \|f_{ij}(\theta_0)\|^2 Z_{ij}^2(\theta_0). \end{aligned} \quad (4.1.9)$$

For the next two terms, we have

$$\begin{aligned} \left| \sum_{j,j'=1}^m a_{ij,k}^{(1)} b_{ij',l}^{(1)} I_{ij}^{obs} I_{ij'}^{cen} \right| &\leq \sum_{j,j'=1}^m \|a_{ij}^{(1)}\| \|b_{ij'}^{(1)}\| \\ &\leq C \sum_{j,j'=1}^m \|f_{ij}(\theta_0)\| |Z_{ij}(\theta_0)| \|f_{ij'}(\theta_0)\|. \end{aligned}$$

Taking expectations in the last inequality above gives

$$\begin{aligned} &\sum_{j,j'=1}^m E_{\theta_0} \|f_{1j}(\theta_0)\| |Z_{1j}(\theta_0)| \|f_{1j'}(\theta_0)\| \\ &\leq \sum_{j,j'=1}^m \left(E_{\theta_0} [\|f_{1j}(\theta_0)\|^2 Z_{1j}^2(\theta_0)] \right)^{1/2} \left(E_{\theta_0} \|f_{1j'}(\theta_0)\|^2 \right)^{1/2} \\ &\leq \max_{1 \leq j \leq m} \left(E_{\theta_0} [\|f_{1j}(\theta_0)\|^2 Z_{1j}^2(\theta_0)] \right)^{1/2} \max_{1 \leq j' \leq m} \left(E_{\theta_0} \|f_{1j'}(\theta_0)\|^2 \right)^{1/2}. \end{aligned} \quad (4.1.10)$$

Finally,

$$\left| \sum_{j=1}^m b_{ij,k}^{(1)} b_{ij,l}^{(1)} I_{ij}^{cen} \right| \leq \sum_{j=1}^m \|b_{ij}^{(1)}\|^2 \leq C \sum_{j=1}^m \|f_{ij}(\theta_0)\|^2. \quad (4.1.11)$$

Note that C contains as factor the square of

$$\max_{1 \leq j \leq m} \frac{|E_{\theta_0}[Z_{1j}(\theta_0) I_{1j}^{obs}]|}{E_{\theta_0}[I_{1j}^{cen}]}.$$

By (4.1.3)-(4.1.4) and (4.1.9)-(4.1.11), (4.1.7) holds. This completes the proof of the theorem. \square

Proposition 4.1.2 *Under the conditions of Theorem 4.1.1, $n^{-1/2}\hat{g}_{n,1}^{obs}(\theta_0)$ and $n^{-1/2}g_{n,1}^{obs}(\theta_0)$ are asymptotically equivalent and have the asymptotic distribution $N(0, \Sigma)$.*

Proof. We write

$$n^{-1/2}\hat{g}_{n,1}^{obs}(\theta_0) = n^{-1/2}g_{n,1}^{obs}(\theta_0) + n^{-1/2}[\hat{g}_{n,1}^{obs}(\theta_0) - g_{n,1}^{obs}(\theta_0)]. \quad (4.1.12)$$

We prove first that

$$n^{-1/2}[\hat{g}_{n,1}^{obs}(\theta_0) - g_{n,1}^{obs}(\theta_0)] \rightarrow 0 \quad \text{in probability.} \quad (4.1.13)$$

Since condition (T2) holds, we have

$$\begin{aligned} g_{n,1}^{obs}(\theta_0) - \hat{g}_{n,1}^{obs}(\theta_0) &= \sum_{i=1}^n \sum_{j=1}^m I_{ij}^{cen} f_{ij}(\theta_0) \left\{ -\frac{E_{\theta_0}[Z_{1j}(\theta_0)I_{1j}^{obs}]}{E_{\theta_0}[I_{1j}^{cen}]} + \frac{\sum_{k=1}^n Z_{kj}(\theta_0)I_{kj}^{obs}}{\sum_{k=1}^n I_{kj}^{cen}} \right\} \\ &:= \sum_{i=1}^n \sum_{j=1}^m I_{ij}^{cen} f_{ij}(\theta_0) \delta_{n,j}^{cen}. \end{aligned}$$

Thus

$$n^{-1/2}[g_{n,1}^{obs}(\theta_0) - \hat{g}_{n,1}^{obs}(\theta_0)] = \sum_{j=1}^m \left(n^{-1/2} \sum_{i=1}^n f_{ij}(\theta_0) I_{ij}^{cen} \right) \delta_{n,j}^{cen}, \quad (4.1.14)$$

where

$$\delta_{n,j}^{cen} = -\frac{E_{\theta_0}[Z_{1j}(\theta_0)I_{1j}^{obs}]}{E_{\theta_0}[I_{1j}^{cen}]} + \frac{\sum_{k=1}^n Z_{kj}(\theta_0)I_{kj}^{obs}}{\sum_{k=1}^n I_{kj}^{cen}}.$$

With condition (4.1.4), by Theorem 29.5 of [4] for *i.i.d.* random vectors, we have, for each $1 \leq j \leq m$

$$n^{-1/2} \sum_{i=1}^n f_{ij}(\theta_0) I_{ij}^{cen} \xrightarrow{\mathcal{L}} N(E_j^{cen}(\theta_0), \Sigma_j^{cen}(\theta_0)),$$

where the k^{th} component of $E_j^{cen}(\theta_0)$ is $E_{\theta_0}[f_{1j,k}(\theta_0)I_{1j}^{cen}]$ and the kl -entry of $\Sigma_j^{cen}(\theta_0)$ is, for $1 \leq k, l \leq p$

$$E_{\theta_0}[(f_{1j,k}(\theta_0)I_{1j}^{cen} - E_{j,k}^{cen}(\theta_0))(f_{1j,l}(\theta_0)I_{1j}^{cen} - E_{j,l}^{cen}(\theta_0))].$$

For each j , $\delta_{n,j}^{cen} \rightarrow 0$ a.s., when $n \rightarrow \infty$, by SLLN applied to both numerators and denominators. Therefore, by Slutsky's Theorem 7.7.1 of [1], each term in (4.1.14) converges to zero in distribution, hence in probability, therefore the finite sum in (4.1.14) converges to zero in probability. Now the conclusion of Proposition 4.1.2 follows from (4.1.12) by Slutsky's Theorem. \square

Let us recall $\mathcal{D}_{n,1}$ defined in (4.1.1).

Theorem 4.1.3 *Assume that the conditions of Theorems 3.3.3 and 4.1.1 hold. Assume further that $n^{-1}\mathcal{D}_{n,1}(\theta_0) \rightarrow \mathcal{D}_1(\theta_0)$, and that $\mathcal{D}_1(\theta_0)$ is nonrandom and invertible. Then $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converges to a normal distribution with mean zero and covariance matrix $\mathcal{D}_1^{-1}(\theta_0)\Sigma[\mathcal{D}_1^{-1}(\theta_0)]^T$.*

Proof. We use the mean value theorem to write

$$\hat{g}_{n,1}^{obs}(\hat{\theta}_n) - \hat{g}_{n,1}^{obs}(\theta_0) = -\mathcal{D}_{n,1}(\bar{\theta}_n)(\hat{\theta}_n - \theta_0),$$

where $\bar{\theta}_n$ is defined by Theorem 3.3.3 and $\|\bar{\theta}_n - \theta_0\| \leq \|\hat{\theta}_n - \theta_0\|$. Then

$$\begin{aligned} \hat{g}_{n,1}^{obs}(\theta_0) &= \mathcal{D}_{n,1}(\bar{\theta}_n)(\hat{\theta}_n - \theta_0) \\ &= n \left\{ n^{-1}[\mathcal{D}_{n,1}(\bar{\theta}_n) - \mathcal{D}_{n,1}(\theta_0)] + n^{-1}\mathcal{D}_{n,1}(\theta_0) \right\} (\hat{\theta}_n - \theta_0) \\ &= n \left\{ o_P(1) + n^{-1}\mathcal{D}_{n,1}(\theta_0) - \mathcal{D}_1(\theta_0) + \mathcal{D}_1(\theta_0) \right\} (\hat{\theta}_n - \theta_0), \end{aligned}$$

where we used condition $S(ii)$ of Theorem 3.3.3 and the consistency of $\hat{\theta}_n$ for the last equality. By hypothesis, we obtain

$$\hat{g}_{n,1}^{obs}(\theta_0) = n \left\{ o_P(1) + \mathcal{D}_1(\theta_0) \right\} (\hat{\theta}_n - \theta_0).$$

Multiplying both sides of the equality by $n^{-1/2}$, we have

$$n^{-1/2}\hat{g}_{n,1}^{obs}(\theta_0) = \left\{ o_P(1) + \mathcal{D}_1(\theta_0) \right\} n^{1/2}(\hat{\theta}_n - \theta_0).$$

Now

$$n^{1/2}(\hat{\theta}_n - \theta_0) = \left\{ o_P(1) + \mathcal{D}_1(\theta_0) \right\}^{-1} n^{-1/2}\hat{g}_{n,1}^{obs}(\theta_0), \quad (4.1.15)$$

and by Proposition 4.1.2, $n^{-1/2}\hat{g}_{n,1}^{obs}(\theta_0)$ is asymptotically normal with mean zero and covariance matrix Σ . Then by Theorem 3.2.1 of [25], $n^{1/2}(\hat{\theta}_n - \theta_0)$ is asymptotically normally distributed with mean zero and covariance $\mathcal{D}_1^{-1}(\theta_0)\Sigma[\mathcal{D}_1^{-1}(\theta_0)]^T$. \square

Remark 4.1.4 *The convergence of the normalized derivative in the hypotheses of Theorem 4.1.3 justifies the use of condition (S)(iii) in Theorem 3.1.1.*

4.2 The Asymptotic Normality of $\hat{\sigma}_n^2$

The presentation in this section is similar to that of the previous section. The second result, Proposition 4.2.2, gives the asymptotic distribution of $n^{-1/2}\hat{g}_{n,2}^{obs}(\eta_0)$, which is the same as that of $n^{-1/2}g_{n,2}^{obs}(\eta_0)$, obtained in Theorem 4.2.1. The final result, Theorem 4.2.3, gives the asymptotic distribution of $\sqrt{n}(\hat{\sigma}_n^2 - \sigma_0^2)$, taking into account the asymptotic behaviour of $\hat{\theta}_n$.

Let us write

$$g_{n,2}^{obs}(\eta_0) := \sum_{i=1}^n u_{i,2}, \quad u_{i,2} := \sum_{j=1}^m [a_{ij}^{(2)} I_{ij}^{obs} - b_{ij}^{(2)} I_{ij}^{cen}],$$

where

$$a_{ij}^{(2)} := b_{ij}(\eta_0)(Z_{1j}^2(\theta_0) - \sigma_0^2), \quad b_{ij}^{(2)} := b_{ij}(\eta_0) \frac{E_{\eta_0}[(Z_{1j}^2(\theta_0) - \sigma_0^2)I_{1j}^{obs}]}{E_{\eta_0}[I_{1j}^{cen}]}. \quad (4.2.1)$$

We have the following result.

Theorem 4.2.1 *Assume that (T2) holds, and that*

$$\max_{1 \leq j \leq m} E_{\eta_0}[b_{1j}^2(\eta_0) \max\{1, Z_{1j}^2(\theta_0), Z_{1j}^4(\theta_0)\}] < \infty. \quad (4.2.2)$$

Then

$$n^{-1/2}g_{n,2}^{obs}(\eta_0) \xrightarrow{\mathcal{L}} N(0, \Omega), \quad (4.2.3)$$

where Ω is:

$$\begin{aligned} \Omega &= \sum_{j,j'=1}^m E_{\eta_0}[a_{1j}^{(2)} a_{1j'}^{(2)} I_{\max\{j,j'\}}^{obs}] - \sum_{j=1, j'>j}^m 2E_{\eta_0}[a_{1j}^{(2)} I_{1j'}^{cen} b_{1j'}^{(2)}] \\ &+ \sum_{j=1}^m E_{\eta_0}[(b_{1j}^{(2)})^2 I_{1j}^{cen}]. \end{aligned} \quad (4.2.4)$$

Proof. Since $E_{\theta_0}[u_{1,2}] = 0$ by Proposition 2.5.1, the proof of Theorem 4.2.1 follows from Theorem 29.5 of [4], once we show that

$$E_{\eta_0}[u_{1,2}^2] < \infty. \quad (4.2.5)$$

To this end, we first calculate Ω . We have, for a fixed i

$$\begin{aligned} u_{i,2}^2 &= \sum_{j,j'=1}^m a_{ij}^{(2)} a_{ij'}^{(2)} I_{ij}^{obs} I_{ij'}^{obs} - \sum_{j,j'=1}^m a_{ij}^{(2)} I_{ij}^{obs} b_{ij'}^{(2)} I_{ij'}^{cen} - \sum_{j,j'=1}^m a_{ij'}^{(2)} I_{ij'}^{obs} b_{ij}^{(2)} I_{ij}^{cen} \\ &+ \sum_{j,j'=1}^m b_{ij}^{(2)} b_{ij'}^{(2)} I_{ij}^{cen} I_{ij'}^{cen}. \end{aligned} \quad (4.2.6)$$

Since $I_{ij}^{obs} I_{ij'}^{obs} = I_{i \max\{j,j'\}}^{obs}$, the expectation of the first double sum in (4.2.6) is given by the first double sum in (4.2.4). Because $I_{ij}^{obs} I_{ij'}^{cen} = I_{ij'}^{cen}$, if $j' > j$, and is zero otherwise, the expectation of the second plus the third double sum in (4.2.6) equals the second double sum in (4.2.4). Now the fourth double sum in (4.2.6) reduces to $\sum_{j=1}^m b_{ij}^{(2)} b_{ij}^{(2)} I_{ij}^{cen}$, because $I_{ij}^{cen} I_{ij}^{cen} = I_{ij}^{cen}$ if $j' = j$, and is zero otherwise.

We now show that the expectations of all terms in (4.2.6) exist. Starting with the first, we have

$$\begin{aligned} \left| \sum_{j,j'=1}^m a_{ij}^{(2)} a_{ij'}^{(2)} I_{i \max\{j,j'\}}^{obs} \right| &\leq \left(\sum_{j=1}^m |a_{ij}^{(2)}| \right)^2 \leq C \sum_{j=1}^m [a_{ij}^{(2)}]^2 \\ &\leq C \sum_{j=1}^m b_{ij}^2(\eta_0) (Z_{ij}^2(\theta_0) - \sigma_0^2)^2. \end{aligned} \quad (4.2.7)$$

For the next two terms, we have

$$\left| \sum_{j,j'=1}^m a_{ij}^{(2)} b_{ij'}^{(2)} I_{ij}^{obs} I_{ij'}^{cen} \right| \leq \sum_{j,j'=1}^m |a_{ij}^{(2)}| |b_{ij'}^{(2)}|$$

$$\leq C \sum_{j,j'=1}^m |b_{ij}(\eta_0)(Z_{ij}^2(\theta_0) - \sigma_0^2)| |b_{ij'}(\eta_0)|.$$

Taking expectations in the last inequality above gives

$$\begin{aligned} & C \sum_{j,j'=1}^m E_{\eta_0} [|b_{1j}(\eta_0)(Z_{1j}^2(\theta_0) - \sigma_0^2)| |b_{1j'}(\eta_0)|] \\ & \leq C \sum_{j,j'=1}^m \left(E_{\eta_0} [b_{1j}^2(\eta_0)(Z_{1j}^2(\theta_0) - \sigma_0^2)^2] \right)^{1/2} \left(E_{\eta_0} b_{1j'}^2(\eta_0) \right)^{1/2} \\ & \leq C \max_{1 \leq j \leq m} \left(E_{\eta_0} [b_{1j}^2(\eta_0)(Z_{1j}^2(\theta_0) - \sigma_0^2)^2] \right)^{1/2} \max_{1 \leq j' \leq m} \left(E_{\eta_0} b_{1j'}^2(\eta_0) \right)^{1/2}. \end{aligned} \quad (4.2.8)$$

Finally,

$$\left| \sum_{j=1}^m b_{ij}^{(2)} b_{ij}^{(2)} I_{ij}^{cen} \right| \leq \sum_{j=1}^m [b_{ij}^{(2)}]^2 \leq C \sum_{j=1}^m b_{ij}^2(\eta_0). \quad (4.2.9)$$

Note that C contains as factor the square of

$$\max_{1 \leq j \leq m} \frac{|E_{\eta_0} [(Z_{1j}^2(\theta_0) - \sigma_0^2) I_{1j}^{obs}]|}{E_{\eta_0} [I_{1j}^{cen}]}.$$

By (4.2.2) and (4.2.7)-(4.2.9), (4.2.5) holds. This completes the proof of Theorem 4.2.1. \square

Proposition 4.2.2 *Under the conditions of Theorem 4.2.1, $n^{-1/2} \hat{g}_{n,2}^{obs}(\eta_0)$ and $n^{-1/2} g_{n,2}^{obs}(\eta_0)$ have the same asymptotic distribution $N(0, \Omega)$.*

Proof. We write

$$n^{-1/2} \hat{g}_{n,2}^{obs}(\eta_0) = n^{-1/2} g_{n,2}^{obs}(\eta_0) + n^{-1/2} [\hat{g}_{n,2}^{obs}(\eta_0) - g_{n,2}^{obs}(\eta_0)]. \quad (4.2.10)$$

We prove first that

$$n^{-1/2} [\hat{g}_{n,2}^{obs}(\eta_0) - g_{n,2}^{obs}(\eta_0)] \rightarrow 0, \quad \text{in probability.} \quad (4.2.11)$$

Then, since $n^{-1/2} g_{n,2}^{obs}(\eta_0) \xrightarrow{\mathcal{L}} N(0, \Omega)$, by Theorem 4.2.1 and Theorem 25.4 of [4], the right hand side of (4.2.10) converges to $N(0, \Omega)$. We now return to the proof of

(4.2.11). Since condition (T2) holds, we have

$$\begin{aligned} g_{n,2}^{obs}(\eta_0) - \hat{g}_{n,2}^{obs}(\eta_0) &= \sum_{i=1}^n \sum_{j=1}^m b_{ij}(\eta_0) I_{ij}^{cen} \left\{ - \frac{E_{\eta}[(Z_{1j}^2(\theta_0) - \sigma_0^2) I_{1j}^{obs}]}{E_{\eta}[I_{1j}^{cen}]} \right. \\ &\quad \left. + \frac{\sum_{k=1}^n [Z_{kj}^2(\theta_0) - \sigma_0^2] I_{kj}^{obs}}{\sum_{k=1}^n I_{kj}^{cen}} \right\} \\ &:= \sum_{i=1}^n \sum_{j=1}^m b_{ij}(\eta_0) I_{ij}^{cen} \delta_{n,j}^{cen}, \end{aligned}$$

$$n^{-1/2} [g_{n,2}^{obs}(\eta_0) - \hat{g}_{n,2}^{obs}(\eta_0)] = \sum_{j=1}^m \left(n^{-1/2} \sum_{i=1}^n b_{ij}(\eta_0) I_{ij}^{cen} \right) \delta_{n,j}^{cen}. \quad (4.2.12)$$

Note that $b_{ij}(\eta_0) I_{ij}^{cen}$ are independent in $i \geq 1$. With condition (4.2.2), by Theorem 29.5 of [4] for *i.i.d.* random variable, we have, for $1 < j \leq m$,

$$n^{-1/2} \sum_{i=1}^n b_{ij}(\eta_0) I_{ij}^{cen} \xrightarrow{\mathcal{L}} N(E[b_{1j}(\eta_0) I_{1j}^{cen}], Var[b_{1j}(\eta_0) I_{1j}^{cen}]).$$

Now each term $\delta_{n,j}^{cen} \rightarrow 0$ *a.s.*, when $n \rightarrow \infty$, by the SLLN applied to both numerators and denominators. Therefore, by Slutsky's Theorem 7.7.1 of [1], each term in (4.2.12) converges to zero in distribution, hence in probability. Therefore, the right hand side of (4.2.12) converges to zero in probability. \square

The following result gives the asymptotic distribution of a sequence of estimators of the parameter σ_0^2 , when a sequence of consistent estimators of θ_0 is available as in Theorem 3.3.3. We follow an idea from [28] and the approach in [20], Section 3.4.

We write our *e.f.* in (2.4.2) and (2.4.4) as

$$\hat{g}_{n,1}^{obs}(\theta) := \sum_{i=1}^n \hat{u}_{i,1}(\theta), \quad \text{and} \quad \hat{g}_{n,2}^{obs}(\theta, \sigma^2) := \sum_{i=1}^n \hat{u}_{i,2}(\theta, \sigma^2). \quad (4.2.13)$$

Let

$$\mathcal{D}_{n,2}(\theta, \sigma^2) := - \frac{\partial \hat{g}_{n,2}^{obs}(\theta, \sigma^2)}{\partial \sigma^2}, \quad \mathcal{D}_{n,3}(\theta, \sigma^2) := - \frac{\partial \hat{g}_{n,2}^{obs}(\theta, \sigma^2)}{\partial \theta^T},$$

where $\mathcal{D}_{n,2}(\theta, \sigma^2)$ is a scalar and $\mathcal{D}_{n,3}(\theta, \sigma^2)$ is of dimension $1 \times p$. We recall the definitions of $u_{i,1}$ in (4.1.2) and $u_{i,2}$ in (4.2.1), $i \geq 1$.

Theorem 4.2.3 Let $(\hat{\theta}_n, \hat{\sigma}_n^2)$, $n \geq 1$ be such that the hypotheses of Theorems 3.3.3, 3.3.7 and 4.1.3 hold for these sequences of estimators of $(\theta_0^T, \sigma_0^2)^T$. Assume further that

$$\lim_{r \rightarrow 0} \limsup_{n \rightarrow \infty} n^{-1} \sup_{\eta \in B_r(\eta_0)} \left| \frac{\partial \hat{g}_{n,2}^{obs}(\theta, \sigma^2)}{\partial \theta^T} - \frac{\partial \hat{g}_{n,2}^{obs}(\theta_0, \sigma_0^2)}{\partial \theta^T} \right| = 0, \quad (4.2.14)$$

and

$$n^{-1} \mathcal{D}_{n,2}(\eta_0) \rightarrow \mathcal{D}_2(\eta_0) \neq 0, \quad n^{-1} \mathcal{D}_{n,3}(\eta_0) \rightarrow \mathcal{D}_3(\eta_0), \quad (4.2.15)$$

where the convergence above is a.s., with $\mathcal{D}_2(\eta_0)$ and $\mathcal{D}_3(\eta_0)$ nonrandom. Then

$$\sqrt{n}(\hat{\sigma}_n^2 - \sigma_0^2) \rightarrow N(0, \mathcal{D}_2^{-2}(\eta_0)\Omega_1),$$

where

$$\Omega_1 = E[Q_1^2(\theta_0, \sigma_0^2)], \quad Q_1(\theta_0, \sigma_0^2) := u_{1,2} - \mathcal{D}_3(\eta_0)\mathcal{D}_1^{-1}(\theta_0)u_{1,1},$$

with $\mathcal{D}_1(\theta_0)$ defined in (4.1.1).

Proof. Part of the proof is similar to that of Theorem 4.1.3. We apply the mean value theorem to $\hat{g}_{n,2}^{obs}(\theta, \sigma^2)$

$$\begin{aligned} \hat{g}_{n,2}^{obs}(\hat{\theta}_n, \hat{\sigma}_n^2) - \hat{g}_{n,2}^{obs}(\theta_0, \sigma_0^2) &= \hat{g}_{n,2}^{obs}(\hat{\theta}_n, \hat{\sigma}_n^2) - \hat{g}_{n,2}^{obs}(\theta_0, \hat{\sigma}_n^2) + \hat{g}_{n,2}^{obs}(\theta_0, \hat{\sigma}_n^2) - \hat{g}_{n,2}^{obs}(\theta_0, \sigma_0^2) \\ &= \frac{\partial \hat{g}_{n,2}^{obs}(\bar{\theta}_n, \hat{\sigma}_n^2)}{\partial \theta^T}(\hat{\theta}_n - \theta_0) + \frac{\partial \hat{g}_{n,2}^{obs}(\theta_0, \bar{\sigma}_n^2)}{\partial \sigma^2}(\hat{\sigma}_n^2 - \sigma_0^2), \end{aligned}$$

where $\|\bar{\theta}_n - \theta_0\| \leq \|\hat{\theta}_n - \theta_0\|$, $\|\bar{\sigma}_n^2 - \sigma_0^2\| \leq \|\hat{\sigma}_n^2 - \sigma_0^2\|$.

Since $\hat{g}_{n,2}^{obs}(\hat{\theta}_n, \hat{\sigma}_n^2) = 0$, this leads to the following relation

$$-\hat{g}_{n,2}^{obs}(\theta_0, \sigma_0^2) = \frac{\partial \hat{g}_{n,2}^{obs}(\bar{\theta}_n, \hat{\sigma}_n^2)}{\partial \theta^T}(\hat{\theta}_n - \theta_0) + \frac{\partial \hat{g}_{n,2}^{obs}(\theta_0, \bar{\sigma}_n^2)}{\partial \sigma^2}(\hat{\sigma}_n^2 - \sigma_0^2). \quad (4.2.16)$$

Relation (4.2.16) can be written as follows:

$$\begin{aligned} -\hat{g}_{n,2}^{obs}(\theta_0, \sigma_0^2) &= n \frac{1}{n} \frac{\partial \hat{g}_{n,2}^{obs}(\bar{\theta}_n, \hat{\sigma}_n^2)}{\partial \theta^T}(\hat{\theta}_n - \theta_0) + n \frac{1}{n} \frac{\partial \hat{g}_{n,2}^{obs}(\theta_0, \bar{\sigma}_n^2)}{\partial \sigma^2}(\hat{\sigma}_n^2 - \sigma_0^2) \\ &= n \left[\frac{1}{n} \frac{\partial \hat{g}_{n,2}^{obs}(\bar{\theta}_n, \hat{\sigma}_n^2)}{\partial \theta^T} - \frac{1}{n} \frac{\partial \hat{g}_{n,2}^{obs}(\theta_0, \sigma_0^2)}{\partial \theta^T} + \frac{1}{n} \frac{\partial \hat{g}_{n,2}^{obs}(\theta_0, \sigma_0^2)}{\partial \theta^T} \right] (\hat{\theta}_n - \theta_0) \end{aligned}$$

$$\begin{aligned}
& + n \left[\frac{1}{n} \frac{\partial \hat{g}_{n,2}^{obs}(\theta_0, \bar{\sigma}_n^2)}{\partial \sigma^2} - \frac{1}{n} \frac{\partial \hat{g}_{n,2}^{obs}(\theta_0, \sigma_0^2)}{\partial \sigma^2} + \frac{1}{n} \frac{\partial \hat{g}_{n,2}^{obs}(\theta_0, \sigma_0^2)}{\partial \sigma^2} \right] (\hat{\sigma}_n^2 - \sigma_0^2) \\
& = n \left[o_P(1) + \frac{1}{n} \frac{\partial \hat{g}_{n,2}^{obs}(\theta_0, \sigma_0^2)}{\partial \theta^T} \right] (\hat{\theta}_n - \theta_0) \\
& + n \left[o_P(1) + \frac{1}{n} \frac{\partial \hat{g}_{n,2}^{obs}(\theta_0, \sigma_0^2)}{\partial \sigma^2} \right] (\hat{\sigma}_n^2 - \sigma_0^2) \\
& = n[o_P(1) - \mathcal{D}_3(\eta_0)](\hat{\theta}_n - \theta_0) + n[o_P(1) - \mathcal{D}_2(\eta_0)](\hat{\sigma}_n^2 - \sigma_0^2). \quad (4.2.17)
\end{aligned}$$

The third equality uses (4.2.14), S(ii) of Theorem 3.3.7(see (3.2.116)) and the consistency of the sequences $\hat{\theta}_n$ and $\hat{\sigma}_n$, which follows from the hypotheses. The last equality uses (4.2.15).

Rearranging terms of (4.2.17) results in

$$\begin{aligned}
& (\mathcal{D}_2(\eta_0) - o_P(1))n(\hat{\sigma}_n^2 - \sigma_0^2) \\
& = \hat{g}_{n,2}^{obs}(\theta_0, \sigma_0^2) - (\mathcal{D}_3(\eta_0) - o_P(1))\sqrt{n}\sqrt{n}(\hat{\theta}_n - \theta_0) \\
& = \hat{g}_{n,2}^{obs}(\theta_0, \sigma_0^2) - (\mathcal{D}_3(\eta_0) - o_P(1))\sqrt{n}\{\mathcal{D}_1(\theta_0) + o_P(1)\}^{-1}n^{-1/2}\hat{g}_{n,1}^{obs}(\theta_0) \\
& = \hat{g}_{n,2}^{obs}(\theta_0, \sigma_0^2) - (\mathcal{D}_3(\eta_0) - o_P(1))\{\mathcal{D}_1(\theta_0) + o_P(1)\}^{-1}\hat{g}_{n,1}^{obs}(\theta_0), \quad (4.2.18)
\end{aligned}$$

where we used (4.1.15) from the proof of Theorem 4.1.3 for the second equality.

Using notation (4.2.13), we obtain from (4.2.18)

$$(\mathcal{D}_2(\eta_0) - o_P(1))\sqrt{n}(\hat{\sigma}_n^2 - \sigma_0^2) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [\hat{u}_{i,2}(\eta_0) - \mathcal{D}_3(\eta_0)\mathcal{D}_1^{-1}(\theta_0)\hat{u}_{i,1}(\theta_0)] + o_P(1).$$

We also used Theorem 4.1.3, which implies that $n^{-1/2}\hat{g}_{n,1}^{obs}(\theta_0)$ is $O_P(1)$. By theorems 4.1.1 and 4.2.1 and propositions 4.1.2 and 4.2.2, we can replace $\hat{u}_{i,1}(\theta_0)$ and $\hat{u}_{i,2}(\eta_0)$ above by $u_{i,1}$ and $u_{i,2}$ which are defined in (4.1.2) and (4.2.1), respectively. Now

$$\sqrt{n}(\hat{\sigma}_n^2 - \sigma_0^2) = \mathcal{D}_2^{-1}(\eta_0) \frac{1}{\sqrt{n}} \sum_{i=1}^n Q_i(\theta_0, \sigma_0^2) + o_P(1), \quad \text{almost surely.}$$

Since $\{Q_i(\theta_0, \sigma_0^2)\}_{i \geq 1}$ are *i.i.d.* of mean 0 and variance Ω_1 , by the central limit theorem,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n Q_i(\theta_0, \sigma_0^2) \rightarrow N(0, \Omega_1).$$

Therefore, we have

$$\mathcal{D}_2^{-1}(\eta_0) \frac{1}{\sqrt{n}} \sum_{i=1}^n Q_i(\theta_0, \sigma_0^2) \rightarrow N(0, \mathcal{D}_2^{-2}(\eta_0)\Omega_1).$$

This concludes our proof. \square

Remark 4.2.4 *Condition (4.2.14) differs from (3.2.116), used in the proof of consistency of our $\hat{\sigma}_n^2$. Condition (4.2.14) is needed in Theorem 4.2.3 because we deal with the analytical properties of $\mathcal{D}_{n,3}(\theta, \sigma^2)$, the partial derivative of $\hat{g}_{n,2}^{obs}(\theta, \sigma^2)$ with respect to θ , which is not required in the proof of S(ii) of Theorem 3.3.7, with our definition of $\hat{\sigma}_n^2$.*

Chapter 5

Simulations Results

In this chapter, we investigate whether our proposed estimators in (2.4.2) and (2.4.6) can be implemented and compare their performance to the conditional GEE estimators proposed by Clement and Strawderman in [6] (available through the R package `condGEE`) in a simulated setting, reproducing their simulation approach. In both sections of this chapter, the mean function is the same. However, the sections differ in how the variance function is defined. In Section 5.1, the variance function is given by (5.0.2), where variance is independently defined from the mean. In section 5.2, the variance function is the absolute value of the mean function. This makes finite sample convergence of the estimates more challenging.

The simulation is ultimately motivated by the work of Murphy *et al.* in [18], which modeled the menstrual cycle patterns among Lease women of the Ituri Forest, Zaire, based on multiple covariates. Insight into the relationship between cycle length and covariates such as location, Body Mass Index(BMI), and age is given in [6], henceforth abbreviated C&S in the tables.

To examine the performance of our proposed estimators, we consider an expanded version of these simulations. We consider a *factorial design with 4 parameters*: two sample sizes ($n \in \{50, 200\}$), two censoring schemes ($C_{\max} \in \{125, 225\}$), or

$C_{\max} \sim N(C, 0)$ for $C = 125$ and $C = 225$, to follow the notation in [6]), four standardized distributions of error terms (normal, exponential, uniform and log-normal), and, finally two variance functions $V_{ij}(\theta)$ described below, which we will call our two scenarios. For each combination of the design parameters, 1000 simulated studies are done to obtain parameter estimates, which are then compared in terms of bias and estimated standard error.

The simulated gap times (in days) are generated according to the model $Y_{ij} = \max\{Y_{ij}^*, 1\}$, $j \geq 1$, $i = 1, 2, \dots, n$, where $Y_{ij}^* = \mu_{ij}(\theta) + \sigma V_{ij}(\theta) \varepsilon_{ij}$ and ε_{ij} are independent, identically distributed observations from a density with mean zero and variance one. This translates to each of the following schemes: $N(0, 1)$, shifted exponential with mean 0 and rate one, uniform on $[-\sqrt{3}, \sqrt{3}]$, and lognormal $\exp(X) - \exp(0.4812119)$, where $X \sim N(0, \sqrt{0.4812119})$.

We assume the following conditional mean and variance functions specifications (our Example 2.1.2 is a slightly simplified version of it):

$$\begin{aligned} \mu_{ij}(\theta) &:= 28 + \gamma_0 + \gamma_1 \overline{BMI}_{ij} \\ &+ \frac{\rho}{\rho(j-1) + 1 - \rho} \left[\sum_{l=1}^{j-1} Y_{il} - \sum_{l=1}^{j-1} (28 + \gamma_0 + \gamma_1 \overline{BMI}_{il}) \right], \end{aligned} \quad (5.0.1)$$

$$V_{ij}(\theta) := \left(\left| 1 + \frac{\rho}{\rho(j-1) + 1 - \rho} \right| \right)^{1/2}. \quad (5.0.2)$$

where $\gamma_0 = 0.6$, $\gamma_1 = -0.4$ and $\rho = 0.03$. A single time-varying covariate is used: $\overline{BMI}_{ij} = BMI_{ij} - 21$, where BMI_{ij} is assumed to decrease linearly from 22 kg/m^2 on day 1 to 20 kg/m^2 on day 195, increase linearly to 21 kg/m^2 on day 225, and then remain constant thereafter. We consider the specification (5.0.2) for $V_{ij}(\theta)$ in conjunction with $\sigma^2 = 11$, which we cover in the next section, and finally, in Section 5.2, $V_{ij}(\theta) = |\mu_{ij}(\theta)|$.

With simulated observation periods of 125 and 225 days (corresponding roughly to 4 and 7.5 months), the average number of events per subject under an observation period is approximately 3.9 and 7.4, with little noise between subjects as one would expect from a study of menstrual cycles. All simulations were run using $b_{ij}(\eta) = 1$.

Ultimately, the data structure is of the following form, to match the requirements of the `condGEE` package:

Table 5.1: Simulated data format

Subject ID	Gap Time	Event Indicator	\overline{BMI}
1	$Y_{1,1}$	1	$\overline{BMI}_{1,1}$
1	$Y_{1,2}$	1	$\overline{BMI}_{1,2}$
\vdots	\vdots	\vdots	\vdots
1	Y_{1,m_1}	0	\overline{BMI}_{1,m_1}
2	$Y_{2,1}$	1	$\overline{BMI}_{2,1}$
\vdots	\vdots	\vdots	\vdots
n	Y_{n,m_n}	0	\overline{BMI}_{n,m_n}

While this is the “long” format for recurrent events (as opposed to the wide), it does not quite conform to the idea of “tidy data” introduced by Wickham in [26] to make analyses more straight-forward, reproducible and extensible. While every necessary quantity can be rederived from this format, to make things more compatible with other modern packages in R, particularly for the `survival` package, this would be the format of choice:

Table 5.2: Simulated tidy data format

Subject ID	Start Time	End Time	Gap Time	Event Indicator	Event Number	\overline{BMI}
1	0	$S_{1,1}$	$Y_{1,1}$	1	1	$\overline{BMI}_{1,1}$
1	$S_{1,1}$	$S_{1,2}$	$Y_{1,2}$	1	2	$\overline{BMI}_{1,2}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
1	S_{1,m_1-1}	C_1	Y_{1,m_1}	0	m_1	\overline{BMI}_{1,m_1}
2	0	$S_{2,1}$	$Y_{2,1}$	1	1	$\overline{BMI}_{2,1}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
n	S_{n,m_n-1}	C_n	Y_{n,m_n}	0	m_n	\overline{BMI}_{n,m_n}

Only the results for normal errors are presented in this chapter, comparing our method to a correctly specified conditional model from [6]. The full tables with the other error distributions are in Appendix A, to assess and compare behavior of our methods when the F_0 in [6] is misspecified.

5.1 First Scenario: the Mean and the Variance Functions are Unrelated

We first simulated with the variance function (5.0.2), which depends only on the correlation parameter ρ and the event number j within an individual.

Figures 5.1 and 5.2 summarize scenario 1 for bias and empirical standard error. We note that our proposed method is quite biased in estimating ρ , but seems to perform somewhat better than in [6] for γ_0 , and γ_1 , especially on a shorter observation period with a small sample. On a large sample with a long observation period, the method from [6] when correctly specified is hard to beat. It would appear that the standard errors are similar for both methods, except on a small sample with a short observation period, where, again, our proposed method performs rather well. Note

that the range of the y axis depends on the value of C_{\max} in the graphs.

Figure 5.1: Bias Comparison for Scenario 1

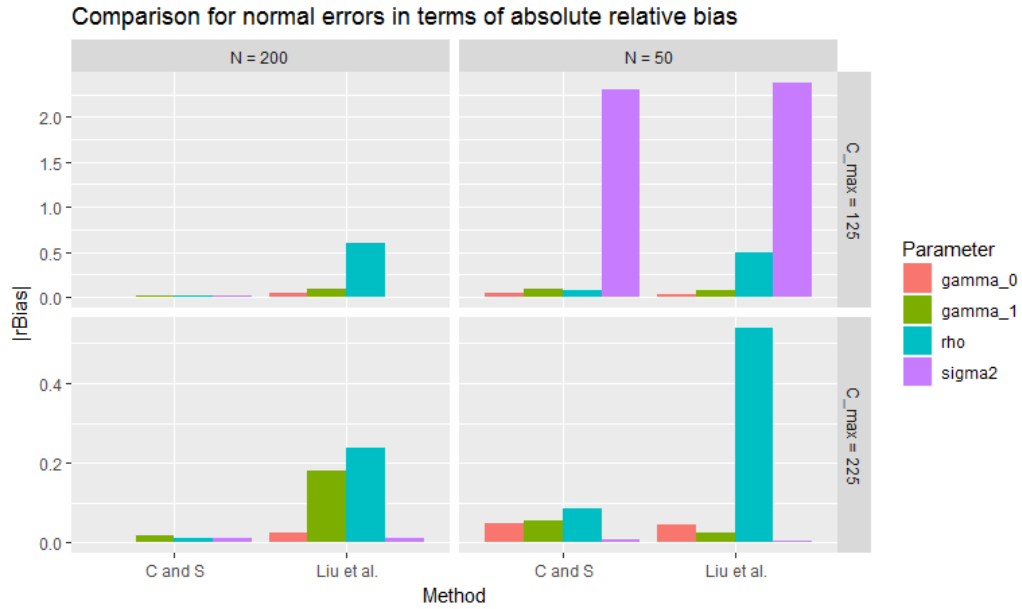
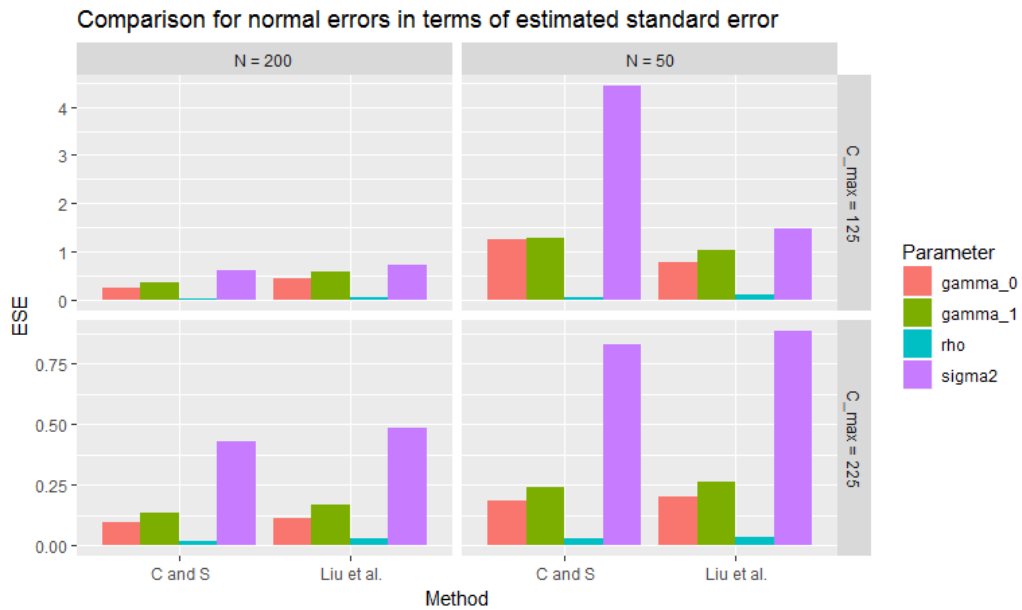


Figure 5.2: Standard Error Comparison for Scenario 1



Tables 5.3 through 5.6 summarize the results for the four combinations of sample sizes n and observation periods in the design. Note that in all tables, we use the following abbreviations:

- ENOES: Expected number of events
- $|rBias|$: Absolute relative bias: $\left| \frac{O-E}{E} \right|$ where O , E stand for “Observed” and “Expected” respectively
- ESE: Empirical standard error
- ASE: Asymptotic standard error (when available)

First, we consider a small sample with a long observation period.

Table 5.3: Comparison with C&S, $n = 50$, V_{ij} in (5.0.2), $C_{\max} = 225$

<i>Estimator</i>	ENOES=7.4	Our method			C&S $F_0 = \text{Normal}$		
	Parameter	$ rBias $	ESE	ASE	$ rBias $	ESE	ASE
Normal	γ_0	0.045	0.202	0.227	0.050	0.185	0.192
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.024	0.263	0.353	0.054	0.242	0.270
$V_{ij}(\theta) := (5.0.2)$	ρ	0.537	0.036	0.037	0.086	0.028	0.033
	σ^2	0.003	0.882	0.900	0.008	0.828	0.815

Next, we have a small sample and a short observation period.

Table 5.4: Comparison with C&S, $n = 50$, V_{ij} in (5.0.2), $C_{\max} = 125$

<i>Estimator</i>	ENOES=3.9	Our method			C&S $F_0 = \text{Normal}$		
	Parameter	$ rBias $	ESE	Bias	$ rBias $	ESE	Bias
Normal	γ_0	0.027	0.772	-0.016	0.051	1.246	-0.030
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.074	1.027	0.029	0.095	1.289	0.038
$V_{ij}(\theta) := (5.0.2)$	ρ	0.500	0.122	0.015	0.084	0.067	0.003
	σ^2	2.376	1.471	7.843	2.305	4.431	7.606

Next, we show the results for a large sample size with a long study period. This takes quite a bit more time to run due to fact that the data generation has to be serial and cannot be done completely in parallel.

Table 5.5: Comparison with C&S, $n = 200$, V_{ij} in (5.0.2), $C_{\max} = 225$

<i>Estimator</i>	ENOES=7.4	Our method			<i>C&S</i> $F_0 = \text{Normal}$		
	Parameter	$ rBias $	ESE	Bias	$ rBias $	ESE	Bias
Normal	γ_0	0.025	0.110	-0.015	0.002	0.097	-0.001
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.180	0.169	0.072	0.019	0.137	0.008
$V_{ij}(\theta) := (5.0.2)$	ρ	0.239	0.028	0.007	0.010	0.017	-0.000
	σ^2	0.011	0.485	0.125	0.012	0.427	0.141

Finally, we present the results for large sample size, short observation period.

Table 5.6: Comparison with C&S, $n = 200$, V_{ij} in (5.0.2), $C_{\max} = 125$

<i>Estimator</i>	ENOES=3.9	Our method			<i>C&S</i> $F_0 = \text{Normal}$		
	Parameter	$ rBias $	ESE	Bias	$ rBias $	ESE	Bias
Normal	γ_0	0.049	0.438	0.029	0.008	0.249	0.005
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.095	0.583	-0.038	0.021	0.359	-0.008
$V_{ij}(\theta) := (5.0.2)$	ρ	0.606	0.070	0.018	0.023	0.033	-0.001
	σ^2	0.008	0.730	0.089	0.012	0.604	0.134

Overall, as expected, increasing sample size improves estimation for both methods, as does the increased observation period, but our method is relatively better suited for small sample size and shorter studies.

5.2 Second Scenario: the Variance Function is the Absolute Value of the Mean Function

The second simulation scenario uses the variance function $V_{ij}(\theta) = |\mu_{ij}(\theta)|$. The numerical solving of this set of GEEs being less stable, bootstrap methods are used to obtain standard error estimates. All conditions in this scenario are otherwise essentially the same as in Section 5.1. However, as $V_{ij}(\theta) = |\mu_{ij}(\theta)|$, the corresponding σ^2

is $1/72.2$, or approximately 0.014 .

Again, we first visualize the comparison for bias and standard error in Figures 5.3 and 5.4. Note that the absolute relative bias of $\hat{\sigma}^2$ is omitted in the first figure due to its size, which dwarfs that of the other estimates. As in Scenario 1, our proposed estimators have trouble estimating ρ , but perform relatively better for small sample size and observation period.

Figure 5.3: Bias Comparison for Scenario 2

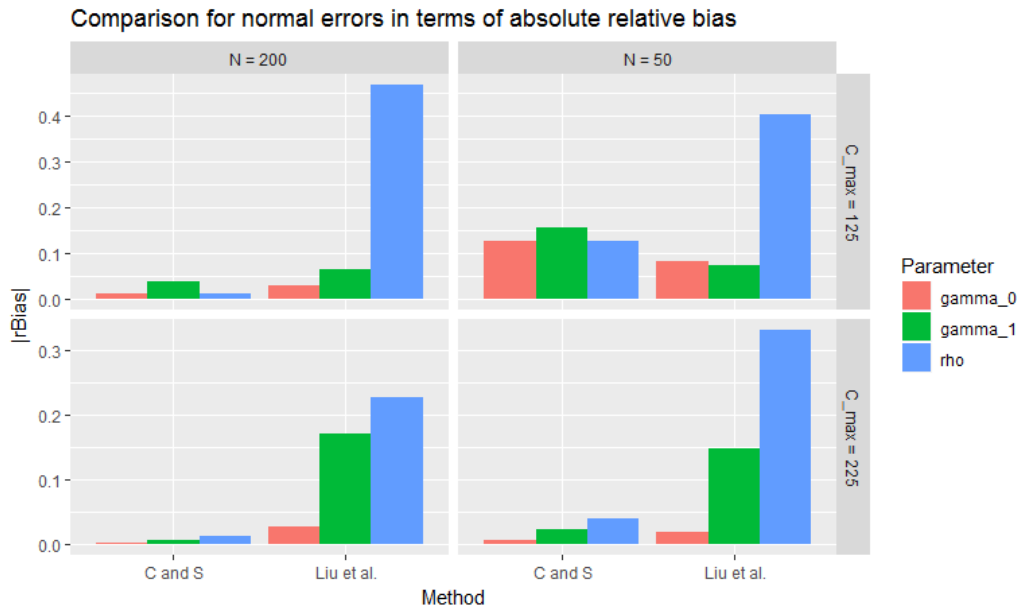
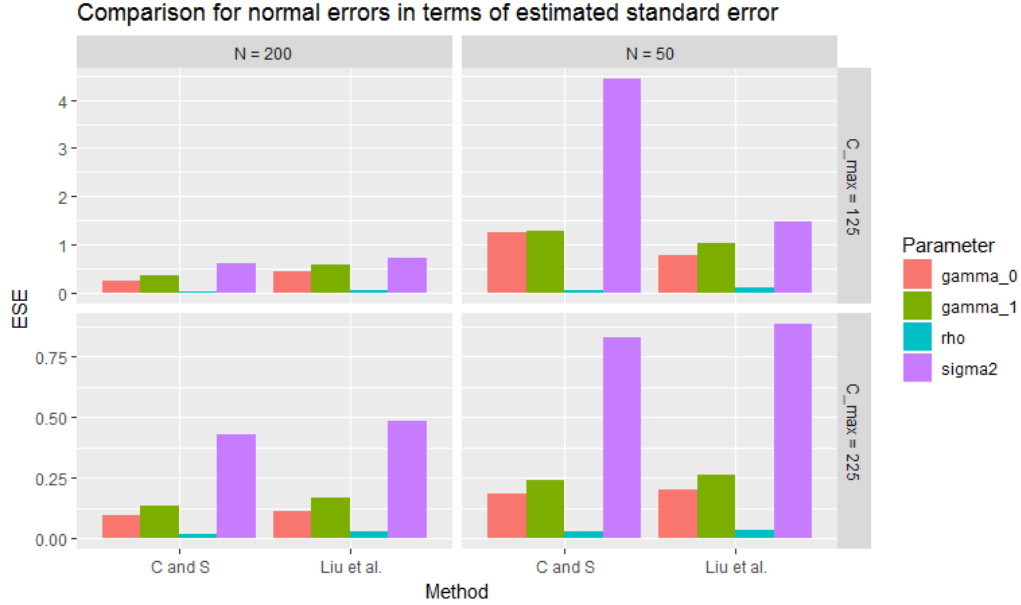


Figure 5.4: Standard Error Comparison for Scenario 2



Tables 5.7 to 5.10, which, as before, correspond to the four design parameters, give more details, including the bias for σ^2 , which is quite bad for both methods. This is in part due to the true value of σ^2 being rather small and thus inflating the relative bias.

First, we consider small sample and long observation period.

Table 5.7: Comparison with C&S, $n = 50$, $V_{ij}(\theta) = |\mu_{ij}(\theta)|$, $C_{\max} = 225$

<i>Estimator</i>	ENOES=7.4		Our method		<i>C&S</i> $F_0 = \text{Normal}$		
	Parameter	$ rBias $	ESE	Bias	$ rBias $	ESE	Bias
Normal	γ_0	0.019	0.206	-0.011	0.006	0.192	-0.004
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.149	0.307	0.060	0.024	0.271	0.010
$V_{ij}(\theta) := \mu_{ij}(\theta) $	ρ	0.331	0.040	0.010	0.040	0.033	-0.001
	σ^2	27.285	0.031	0.378	27.210	0.030	0.377

Table 5.8: Comparison with C&S, $n = 50$, $V_{ij}(\theta) = |\mu_{ij}(\theta)|$, $C_{\max} = 125$

<i>Estimator</i>	ENOES=3.9	Our method			C&S $F_0 = \text{Normal}$		
	Parameter	$ rBias $	ESE	Bias	$ rBias $	ESE	Bias
Normal	γ_0	0.082	0.745	-0.049	0.126	0.456	-0.076
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.073	0.983	0.029	0.155	0.666	0.062
$V_{ij}(\theta) := \mu_{ij}(\theta) $	ρ	0.401	0.126	0.012	0.126	0.062	-0.004
	σ^2	27.045	0.062	0.375	26.533	0.042	0.367

Table 5.9: Comparison with C&S, $n = 200$, $V_{ij}(\theta) = |\mu_{ij}(\theta)|$, $C_{\max} = 225$

<i>Estimator</i>	ENOES=7.4	Our method			C&S $F_0 = \text{Normal}$		
	Parameter	$ rBias $	ESE	Bias	$ rBias $	ESE	Bias
Normal	γ_0	0.027	0.106	-0.016	0.002	0.095	-0.001
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.172	0.166	0.069	0.006	0.137	0.002
$V_{ij}(\theta) := \mu_{ij}(\theta) $	ρ	0.227	0.026	0.007	0.014	0.017	-0.000
	σ^2	27.524	0.016	0.381	27.535	0.015	0.381

Table 5.10: Comparison with C&S, $n = 200$, $V_{ij}(\theta) = |\mu_{ij}(\theta)|$, $C_{\max} = 125$

<i>Estimator</i>	ENOES=3.9	Our method			C&S $F_0 = \text{Normal}$		
	Parameter	$ rBias $	ESE	Bias	$ rBias $	ESE	Bias
Normal	γ_0	0.028	0.452	0.017	0.013	0.236	0.008
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.066	0.596	-0.026	0.037	0.359	-0.014936
$V_{ij}(\theta) := \mu_{ij}(\theta) $	ρ	0.467	0.070	0.014	0.014	0.029	-0.000
	σ^2	27.275	0.023	0.378	27.250	0.019	0.377

5.3 Discussion

Overall, the simulation scenarios, including the extended ones presented in Appendix A, shed some light on the behaviour of our proposed estimators that make no assumptions on the error distribution. In Appendix A, one can see that while the method in

[6] is robust to some misspecifications, it is less so for others, while our method is more consistent in its bias, regardless of the error distribution. The relative robustness of [6] also is quite dependent on sample size and length of observation period, as can be seen by comparing the log-normal cases of tables A.1 and A.2. This shows that there may be reason to prefer our method on samples with fewer events and asymmetric errors. It should be noted that for numerical stability, the implementation of equations (2.4.2) and (2.4.6) require some minor adjustments in “edge cases”, that is, some individual terms may return values of NA (“not available”, that is, a missing value) or NaN (“not a number”), which can propagate to the solution and prevent convergence. This can occur if a denominator is zero or if an entry is missing. Those problematic terms are mapped to zero. This effectively perturbs the equation but helps prevent failures of convergence.

One of the main limitations of these simulations is that only one set of values of η is used, thus we are not assessing behaviour under the null hypothesis (no effect of the covariates) or within a range of weak to strong effects. Additionally, having only one covariate with the exact same deterministic behaviour for all subjects is quite a strong assumption of lack of noise compared to what would be a realistic setting (a varied distribution of BMI, with random fluctuation in time). This plays in favor of any estimation procedure. However, modeling recurrent events with a “full history” as in (5.0.1) (as opposed to, say, a Markov model based only on the previous event) is quite complex even with a single covariate, as the number of entities to input into the formula grows linearly with the number of events. This is quite tricky to implement, hard to parallelize and difficult for subject matter experts to interpret. In our simulations, individuals with longer past gap times will have the mean log-gap tend to be longer as well, and similarly, those with shorter gap times will see their next event tend to have shorter mean log-gap. The effect size of this is harder to explain in simple terms to a data user, unlike, say, the effect size from a proportional hazard model. It is not surprising that the default setup of the `condGEE` package

actually models independent gap times within individuals, even if as a model it is rather uninteresting. Some level of programming expertise is necessary to code a more elaborate model, and adding multiple covariates requires some skills in software engineering. Thus, while it would be worthwhile to try our methods on a real data set, it will remain a future goal for the time being.

Appendix A

Extended Simulation Tables

Models were fitted with four scenarios of errors (all with mean zero and variance one), and comparison was always with [6] assuming normal errors. The four scenarios are standard normal errors, shifted exponential with rate one, uniform $[-a, a]$ where $a = \sqrt{3}$, and log-normal distribution.

A.1 Full Tables for Independent Mean and Variance Function

As explained in Chapter 5, $\mu_{ij}(\theta)$ is given by equation (5.0.1) and the variance function $V_{ij}(\theta)$ is defined by equation (5.0.2). Tables A.1 to A.4 present the results for all combinations of $n \in \{50, 200\}$ and $C_{\max} \in \{125, 225\}$.

Table A.1: Comparison with C&S, $n = 50$, V_{ij} in (5.0.2), $C_{\max} = 225$

<i>Estimator</i>	ENOES=7.4	Our method			C&S $F_0 = \text{Normal}$		
	Parameter	$ rBias $	ESE	ASE	$ rBias $	ESE	ASE
Normal errors	γ_0	0.045	0.202	0.227	0.050	0.185	0.192
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.024	0.263	0.353	0.054	0.242	0.270
$V_{ij}(\theta) := (5.0.2)$	ρ	0.537	0.036	0.037	0.086	0.028	0.033
	σ^2	0.003	0.882	0.900	0.008	0.828	0.815
Exponential errors	γ_0	0.050	0.403	0.240	0.091	0.444	0.199
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.071	0.777	0.371	0.316	0.949	0.289
$V_{ij}(\theta) := (5.0.2)$	ρ	0.321	0.180	0.036	0.236	0.068	0.031
	σ^2	0.068	4.718	1.049	0.074	5.255	0.938
Uniform errors	γ_0	0.018	0.206	0.221	0.006	0.192	0.189
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.137	0.308	0.342	0.004	0.292	0.273
$V_{ij}(\theta) := (5.0.2)$	ρ	0.330	0.040	0.035	0.054	0.034	0.032
	σ^2	0.004	0.558	0.591	0.010	1.365	0.542
Log-normal errors	γ_0	0.032	0.240	0.254	0.010	0.191	0.190
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.184	0.527	0.441	0.007	0.276	0.253
$V_{ij}(\theta) := (5.0.2)$	ρ	0.350	0.112	0.049	0.015	0.037	0.031
	σ^2	0.013	3.302	2.667	0.043	2.295	1.931

$|rBias|$: Absolute relative bias; ESE: Estimated Standard Error; ASE: Asymptotic Standard Error.

Table A.2: Comparison with C&S, $n = 50$, V_{ij} in (5.0.2), $C_{\max} = 125$

<i>Estimator</i>	ENOES=3.9	Our method			<i>C&S</i> $F_0 = \text{Normal}$		
	Parameter	$ rBias $	ESE	Bias	$ rBias $	ESE	Bias
Normal errors	γ_0	0.027	0.772	-0.016	0.051	1.246	-0.030
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.074	1.027	0.029	0.095	1.289	0.038
$V_{ij}(\theta) := (5.0.2)$	ρ	0.500	0.122	0.015	0.084	0.067	0.003
	σ^2	2.376	1.471	7.843	2.305	4.431	7.606
Exponential errors	γ_0	0.027	0.763	-0.016	0.182	0.709	-0.109
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.078	1.015	0.031	0.348	0.770	0.139
$V_{ij}(\theta) := (5.0.2)$	ρ	0.988	0.151	0.030	0.096	0.076	-0.003
	σ^2	2.328	2.482	7.681	2.154	2.081	7.110
Uniform errors	γ_0	0.017	0.766	-0.010	0.012	0.455	0.007
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.076	1.033	0.030	0.022	0.694	0.009
$V_{ij}(\theta) := (5.0.2)$	ρ	0.311	0.119	0.009	0.092	0.059	-0.003
	σ^2	2.381	1.078	7.856	2.357	0.812	7.778
Log-normal errors	γ_0	0.069	0.751	-0.041	1.116	16.019	-0.669
$\mu_{ij}(\theta) = (5.0.1)$	γ_1	0.205	0.985	0.082	1.749	15.348	0.700
$V_{ij}(\theta) = (5.0.2)$	ρ	0.997	0.153	0.030	0.121	0.122	0.004
	σ^2	2.325	4.211	7.672	2.098	3.999	6.903

$|rBias|$: Absolute relative bias; ESE: Estimated Standard Error.

Table A.3: Comparison with C&S, $n = 200$, V_{ij} in (5.0.2), $C_{\max} = 225$

<i>Estimator</i>	ENOES=7.4	Our method			C&S $F_0 = \text{Normal}$		
	Parameter	$ rBias $	ESE	Bias	$ rBias $	ESE	Bias
Normal errors	γ_0	0.025	0.110	-0.015	0.002	0.097	-0.001
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.180	0.169	0.072	0.019	0.137	0.008
$V_{ij}(\theta) := (5.0.2)$	ρ	0.239	0.028	0.007	0.010	0.017	-0.000
	σ^2	0.011	0.485	0.125	0.013	0.428	0.141
Exponential errors	γ_0	0.032	0.109	-0.019	0.016	0.131	0.010
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.203	0.172	0.081	0.011	0.167	-0.004
$V_{ij}(\theta) := (5.0.2)$	ρ	0.273	0.034	0.008	0.250	0.128	-0.008
	σ^2	0.015	0.912	0.170	0.006	1.672	-0.064
Uniform errors	γ_0	0.024	0.108	-0.014	0.003	0.094	-0.002
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.144	0.170	0.058	0.012	0.139	-0.005
$V_{ij}(\theta) := (5.0.2)$	ρ	0.257	0.027	0.008	0.014	0.017	-0.000
	σ^2	0.010	0.308	0.110	0.011	0.265	0.122
Log-normal errors	γ_0	0.046	0.109	-0.028	0.003	0.124	-0.002
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.208	0.166	0.083	0.000	0.161	0.000
$V_{ij}(\theta) := (5.0.2)$	ρ	0.232	0.040	0.007	0.151	0.084	-0.005
	σ^2	0.007	1.631	0.078	0.030	1.896	-0.326

$|rBias|$: Absolute relative bias; ESE: Estimated Standard Error.

Table A.4: Comparison with C&S, $n = 200$, V_{ij} in (5.0.2), $C_{\max} = 125$

<i>Estimator</i>	ENOES=3.9	Our method			C&S $F_0 = \text{Normal}$		
	Parameter	$ rBias $	ESE	Bias	$ rBias $	ESE	Bias
Normal errors	γ_0	0.049	0.438	0.029	0.008	0.249	0.005
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.095	0.583	-0.038	0.021	0.359	-0.008
$V_{ij}(\theta) := (5.0.2)$	ρ	0.606	0.070	0.018	0.023	0.033	-0.001
	σ^2	0.008	0.730	0.089	0.012	0.604	0.134
Exponential errors	γ_0	0.062	0.451	0.037	0.045	0.232	-0.027
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.112	0.591	-0.045	0.086	0.347	0.034
$V_{ij}(\theta) := (5.0.2)$	ρ	1.069	0.092	0.032	0.123	0.031	-0.004
	σ^2	0.008	1.302	0.084	0.024	1.081	-0.264
Uniform errors	γ_0	0.064	0.440	0.039	0.035	0.233	0.021
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.130	0.591	-0.052	0.075	0.355	-0.030
$V_{ij}(\theta) := (5.0.2)$	ρ	0.502	0.067	0.015	0.038	0.029	-0.001
	σ^2	0.007	0.525	0.079	0.011	0.381	0.124
Log-normal errors	γ_0	0.068	0.465	0.041	0.085	0.226	-0.051
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.106	0.631	-0.042	0.157	0.345	0.063
$V_{ij}(\theta) := (5.0.2)$	ρ	0.765	0.358	0.023	0.033	0.031	0.001
	σ^2	0.016	2.693	-0.181	0.064	1.300	-0.703

$|rBias|$: Absolute relative bias; ESE: Estimated Standard Error.

A.2 Full Tables for Simulations where Variance Function is the Absolute Mean

Here, while $\mu_{ij}(\theta)$ is still given by equation (5.0.1), the variance function $V_{ij}(\theta)$ is defined by $|\mu_{ij}(\theta)|$. Tables A.5 to A.8 present the results for all combinations of $n \in \{50, 200\}$ and $C_{\max} \in \{125, 225\}$.

Table A.5: Comparison with C&S, $n = 50$, $V_{ij}(\theta) = |\mu_{ij}(\theta)|$, $C_{\max} = 225$

<i>Estimator</i>	ENOES=7.4	Our method			C&S $F_0 = \text{Normal}$		
	Parameter	$ rBias $	ESE	Bias	$ rBias $	ESE	Bias
Normal errors	γ_0	0.0191	0.206	-0.011	0.006	0.192	-0.004
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.149	0.307	0.060	0.024	0.271	0.010
$V_{ij}(\theta) := \mu_{ij}(\theta) $	ρ	0.330	0.040	0.010	0.040	0.033	-0.001
	σ^2	27.285	0.031	0.378	27.210	0.030	0.377
Exponential errors	γ_0	0.011	0.228	-0.001	0.003	0.207	-0.002
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.110	0.334	0.044	0.003	0.291	-0.001
$V_{ij}(\theta) := \mu_{ij}(\theta) $	ρ	0.654	0.104	0.020	0.008	0.038	-0.000
	σ^2	27.253	0.068	0.377	26.204	0.058	0.363
Uniform errors	γ_0	0.011	0.208	-0.007	0.001	0.199	-0.000
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.113	0.295	0.045	0.013	0.259	-0.005
$V_{ij}(\theta) := \mu_{ij}(\theta) $	ρ	0.305	0.039	0.009	0.038	0.033	-0.001
	σ^2	27.429	0.030	0.380	27.737	0.018	0.379
Log-normal errors	γ_0	0.028	0.242	-0.017	0.009	0.207	-0.005
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.166	0.335	0.067	0.045	0.286	0.018
$V_{ij}(\theta) := \mu_{ij}(\theta) $	ρ	0.163	0.635	-0.005	0.027	0.038	0.001
	σ^2	27.160	0.145	0.376	25.976	0.081	0.360

$|rBias|$: Absolute relative bias; ESE: Estimated Standard Error.

Table A.6: Comparison with C&S, $n = 50$, $V_{ij}(\theta) = |\mu_{ij}(\theta)|$, $C_{\max} = 125$

<i>Estimator</i>	ENOES=3.9	Our method			C&S $F_0 = \text{Normal}$		
	Parameter	$ rBias $	ESE	Bias	$ rBias $	ESE	Bias
Normal errors	γ_0	0.082	0.745	-0.049	0.126	0.456	-0.076
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.073	0.983	0.029	0.155	0.666	0.062
$V_{ij}(\theta) := \mu_{ij}(\theta) $	ρ	0.401	0.126	0.012	0.126	0.062	-0.004
	σ^2	27.045	0.062	0.375	26.533	0.042	0.367
Exponential errors	γ_0	0.079	0.760	-0.047	0.156	0.428	-0.094
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.211	1.016	0.084	0.346	0.650	0.138
$V_{ij}(\theta) := \mu_{ij}(\theta) $	ρ	0.922	0.145	0.028	0.047	0.065	0.001
	σ^2	27.359	0.092	0.379	25.635	0.073	0.355
Uniform errors	γ_0	0.044	0.767	-0.026	0.053	0.472	-0.032
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.088	1.016	0.035	0.099	0.721	0.039
$V_{ij}(\theta) := \mu_{ij}(\theta) $	ρ	0.121	0.121	-0.004	0.491	0.068	-0.015
	σ^2	27.246	0.032	0.377	26.661	0.043	0.369
Log-normal errors	γ_0	0.018	0.757	-0.011	0.137	0.430	-0.082
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.058	0.990	0.023	0.286	0.637	0.114
$V_{ij}(\theta) := \mu_{ij}(\theta) $	ρ	0.051	0.813	0.002	0.046	0.066	-0.001
	σ^2	26.620	0.126	0.369	24.572	0.088	0.340

$|rBias|$: Absolute relative bias; ESE: Estimated Standard Error.

Table A.7: Comparison with C&S, $n = 200$, $V_{ij}(\theta) = |\mu_{ij}(\theta)|$, $C_{\max} = 225$

<i>Estimator</i>	ENOES=7.4	Our method			C&S $F_0 = \text{Normal}$		
	Parameter	$ rBias $	ESE	Bias	$ rBias $	ESE	Bias
Normal errors	γ_0	0.027	0.106	-0.016	0.002	0.095	-0.001
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.172	0.166	0.069	0.006	0.138	0.002
$V_{ij}(\theta) := \mu_{ij}(\theta) $	ρ	0.227	0.026	0.007	0.014	0.017	-0.000
	σ^2	27.525	0.016	0.381	27.535	0.015	0.381
Exponential errors	γ_0	0.042	0.109	-0.025	0.001	0.130	-0.000
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.209	0.169	0.083	0.003	0.160	-0.001
$V_{ij}(\theta) := \mu_{ij}(\theta) $	ρ	0.246	0.034	0.007	0.267	0.140	-0.008
	σ^2	27.302	0.031	0.378	26.444	0.071	0.366
Uniform errors	γ_0	0.023	0.111	-0.014	0.000	0.095	0.000
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.186	0.172	0.075	0.024	0.135	0.010
$V_{ij}(\theta) := \mu_{ij}(\theta) $	ρ	0.243	0.027	0.007	0.018	0.017	-0.001
	σ^2	27.491	0.010	0.381	27.478	0.009	0.381
Log-normal errors	γ_0	0.018	0.757	-0.011	0.137	0.430	-0.082
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.058	0.990	0.023	0.286	0.637	0.114
$V_{ij}(\theta) := \mu_{ij}(\theta) $	ρ	0.051	0.813	0.002	0.046	0.066	-0.001
	σ^2	26.620	0.126	0.369	24.572	0.088	0.340

$|rBias|$: Absolute relative bias; ESE: Estimated Standard Error.

Table A.8: Comparison with C&S, $n = 200$, $V_{ij}(\theta) = |\mu_{ij}(\theta)|$, $C_{\max} = 125$

<i>Estimator</i>	ENOES=3.9	Our method			C&S $F_0 = \text{Normal}$		
	Parameter	$ rBias $	ESE	Bias	$ rBias $	ESE	Bias
Normal errors	γ_0	0.028	0.452	0.017	0.013	0.236	0.008
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.066	0.596	-0.026	0.037	0.359	-0.015
$V_{ij}(\theta) := \mu_{ij}(\theta) $	ρ	0.467	0.070	0.014	0.014	0.029	-0.000
	σ^2	27.275	0.023	0.378	27.25	0.019	0.377
Exponential errors	γ_0	0.052	0.777	-0.031	0.118	0.435	-0.071
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.107	1.039	0.043	0.223	0.653	0.089
$V_{ij}(\theta) := \mu_{ij}(\theta) $	ρ	0.902	0.149	0.027	0.034	0.066	0.001
	σ^2	27.229	0.087	0.377	25.518	0.068	0.353
Uniform errors	γ_0	0.058	0.436	0.035	0.035	0.233	0.021
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.095	0.568	-0.038	0.057	0.354	-0.023
$V_{ij}(\theta) := \mu_{ij}(\theta) $	ρ	0.405	0.068	0.012	0.109	0.031	-0.003
	σ^2	27.299	0.015	0.378	0.27.268	0.013	0.378
Log-normal errors	γ_0	0.036	0.264	0.297	0.210	0.185	0.191
$\mu_{ij}(\theta) := (5.0.1)$	γ_1	0.201	0.543	0.494	0.037	0.268	0.259
$V_{ij}(\theta) := \mu_{ij}(\theta) $	ρ	0.338	0.115	0.057	0.060	0.033	0.031
	σ^2	84.755	4.345	2.511	80.008	3.529	1.781

$|rBias|$: Absolute relative bias; ESE: Estimated Standard Error.

Bibliography

- [1] Ash, R.B. and Doléans-Dade, C.A.(2000). *Probability & Measure Theory*. Second Edition, Academic Press, San Diego, California.
- [2] Balan, R.M. and Schiopu-Kratina, I.(2005). Asymptotic results with generalized estimating equations for longitudinal data. *Annals of Statistics*, 55, pp. 522-541.
- [3] Balan, R.M., Dumitrescu L. and Schiopu-Kratina, I.(2010). Asymptotically optimal estimating equation with strongly consistent solutions for longitudinal data. *Mathematical Methods of Statistics*, 19, pp. 93-120.
- [4] Billingsley, P.(1995). *Probability and Measure*. Third edition, John Wiley, New York.
- [5] Chen, K., Hu, I. and Ying, Z.(1999). Strong consistency of maximum quasi-likelihood estimators in generalized linear models with fixed and adaptive designs. *Annals of Statistics*, 27, pp. 1155-1163.
- [6] Clement, D.Y. and Strawderman, R.L.(2009). Conditional GEE for recurrent event gap times. *Biostatistics*, 10, pp. 451-467.
- [7] Clement, D.Y. and Strawderman, R.L.(2009). Supplemental materials for conditional GEE for recurrent event gap times.
<http://www.biostatistics.oxfordjournals.org>.

- [8] Cook, R.J. and Lawless, J.F.(2007). *The Statistical Analysis of Recurrent Events*. Springer-Verlag, New York.
- [9] Dumitrescu, L.(2011). *Asymptotic results for some longitudinal regression models* (Doctoral dissertation). University of Ottawa, Ottawa, Canada.
- [10] Elashoff, M. and Ryan, L.(2004). An EM algorithm for estimating equations. *Journal of Computational and Graphical Association*, 13, pp. 48-65.
- [11] Heyde, C.C. (1997). *Quasi-Likelihood and its Application. An Optimal Approach to Parameter Estimation*. Springer-Verlag, New York.
- [12] Laird, N.M.(1988). Missing data in longitudinal studies. *Statistics in Medicine*, 7, pp. 305-315.
- [13] Lai, T.L. and Small, D.(2007). Marginal regression analysis of longitudinal data with time-dependent covariates: a generalized method-of-moments approach. *J.R. Statist, Soc, Series B*. 69, pp. 79-99.
- [14] Liang, K.Y. and Zeger, S.L.(1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73, pp. 13-22.
- [15] Lin, Y.X. (1994). On the strong law of large numbers of multivariate martingales with random norming. *Stochastic Process, Appl*, 54, pp. 355-360.
- [16] Little, R. J. A. and Rubin, D. B.(2002). *Statistical Analysis with Missing Data*. Second edition, Wiley Series in Probability and Statistics. New York: Wiley-Interscience.
- [17] Murphy, S.A. and Li, B.(1995). Projected partial likelihood and its application to longitudinal data. *Biometrika*, 82, pp. 399-406.

- [18] Murphy, S.A., Bentley, G.R. and O'Hanesian, M.A.(1995). An analysis for menstrual data with time-varying covariates. *Statistics in Medicine*, 14, pp. 1843-1857.
- [19] Murphy, S.A.(1995). A central limit theorem for local martingales with applications to the analysis of longitudinal data. *The Scandinavian Journal of Statistics*, 22, pp. 279-294.
- [20] Park, J.(2015) *Longitudinal data analysis using generalized linear model with missing responses* (Master's thesis). University of Ottawa, Ottawa, Canada.
- [21] Pepe, M. S. and Anderson, G. L.(1994). A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Communications in Statistics-Simulation and Computation*, 23, pp. 939-951.
- [22] Schott, J.R.(2005). *Matrix Analysis for Statistics*. Second edition. John Wiley, New York.
- [23] Small, C.G., Wang, J. and Yang, Z.(2000). Eliminating multiple root problems in estimation. *Stat. Science*, 15, pp. 313-341.
- [24] Strawderman, R.L.(2005). The accelerated gap times model. *Biometrika*, 92, pp. 647-666.
- [25] Tong, Y.L.(1999). *The Multivariate Normal Distribution*. Springer-Verlag, New York.
- [26] Wickham, H.(2014). Tidy data. *Journal of Statistical Software*, 59, pp. 1-23.
- [27] Xie, M. and Yang, Y.(2003). Asymptotics for generalized estimating equations with large cluster sizes. *Annals of Statistics*, 31, pp. 310-347

-
- [28] Yi, G.Y., Ma, Y.Y and Carroll, R.J.(2012). A functional generalized method of moments approach for longitudinal studies with missing responses and covariate measurement error. *Biometrika*, 99, pp. 151-165.
- [29] Yuan, K.H. and Jennrich, R.I.(1998). Asymptotics of estimating equations under natural conditions. *Journal of Multivariate Analysis*, 65, pp. 245-260.
- [30] Zeger, S.L and Liang, K.Y.(1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, 42, pp. 121-130.
- [31] Zeger, S.L and Liang, K.Y.(1992). An overview of methods for the analysis of longitudinal data. *Statistics in Medicine*, 11, pp. 1825-1839.

List of Symbols

$(Ti), i = 0, 1, 2,$	20	$H_n^{[3]}(\theta),$	56
$B_r(\theta),$	38	$H_n^{[3]}(\theta)^{imp},$	75
$B_n^{[1]}(\theta),$	54	$H_n^{[4]}(\theta)^{imp},$	76
$B_n^{[1]}(\theta)^{imp},$	76	$H_n^{emp}(\theta),$	85
$B_n^{[2]}(\theta),$	54	$I_c,$	78
$B_n^{[2]}(\theta)^{imp},$	76	$I_0,$	66
$B_n^{[3]}(\theta),$	54	$I_a(j),$	78
$B_n^{[3]}(\theta)^{imp},$	76	$I_b(j),$	78
$B_n^{[4]}(\theta)^{imp},$	76	$I_{ij}^{cen},$	22
$B_n^{emp}(\theta),$	85	$I_{ij}^{obs},$	22
$C_i,$	11	$I_{ij}^{out},$	22
$D_n,$	53	$N_i,$	16
$H_{n,2}(\eta),$	100	$S_{ij},$	11
$H_{n,2}^{imp}(\eta),$	100	$S_{n,i}(\eta), i = 1, 2,$	17
$H_n(\theta),$	54	$V_{ij}^2(\theta),$	11
$H_n(\theta)^{imp},$	75	$X_i,$	53
$H_n^{[1]}(\theta),$	56	$Y_{ij},$	11
$H_n^{[1]}(\theta)^{imp},$	75	$Z_{ij}(\theta),$	15
$H_n^{[2]}(\theta),$	56	$Z_{ij}^{imp}(\theta),$	33
$H_n^{[2]}(\theta)^{imp},$	75	$\ddot{\mu}_{ij}(\theta),$	53

$\delta_n(r)$, 38	$\mathcal{E}_n(\theta)$, 55
$\delta_n^{[1]}(r)$, 38	$\mathcal{E}_n(\theta)^{imp}$, 76
$\delta_n^{[2]}(r)$, 38	$\mathcal{E}_n^{[1]}(\theta)$, 55
$\dot{\mu}_{ij}(\theta)$, 53	$\mathcal{E}_n^{[1]}(\theta)^{imp}$, 76
η , 11	$\mathcal{E}_n^{[2]}(\theta)$, 55
η_0 , 11	$\mathcal{E}_n^{[2]}(\theta)^{imp}$, 76
$\eta_n(r)$, 38	$\mathcal{E}_n^{[3]}(\theta)^{imp}$, 76
$\gamma_{j,n}^{cen}$, 74	\mathcal{F}_{ij} , 11
$\hat{g}_{n,1}^{imp}(\theta)$, 74	\mathcal{G}_{ij} , 15
$\hat{g}_{n,1}^{obs}(\theta)$, 23	$\mu_{ij}(\theta)$, 11
$\hat{g}_{n,2}^{obs}(\eta)$, 23	$\nu_{ij}^T(\theta)$, 55
$\hat{g}_{n,2}^{sim}(\eta)$, 24	$\nu_n(r)$, 55
$\lambda_{\max}(A)$, 49	\overline{BMI}_{ij} , 12
$\mathbb{O}(b_i(\eta))$, 30	$\ \cdot\ $, 49
$\mathbb{O}(f_i(\theta))$, 24	$\ \cdot\ _E$, 49
$\mathbb{O}(f_i(\theta), \frac{\partial^2 \mu_i}{\partial \theta^T \partial \theta} V_i^{-1})$, 30	τ_i , 19
$\mathbb{O}(x_i)$, 33	$\tilde{k}_n^{[1]}(r)$, 59
$\mathcal{D}_1(\theta_0)$, 106, 110	$\tilde{v}_n(r)$, 63
$\mathcal{D}_2(\eta_0)$, 115	$a_{ij}^{(1)}$, 106
$\mathcal{D}_3(\eta_0)$, 115	$a_{ij}^{(2)}$, 111
$\mathcal{D}_{n,1}(\theta)$, 85, 97, 106	$b_{ij}(\eta)$, 16
$\mathcal{D}_{n,1}^{obs}(\theta)$, 55	$b_{ij}^{(1)}$, 106
$\mathcal{D}_{n,2}(\eta)$, 104, 114	$b_{ij}^{(2)}$, 111
$\mathcal{D}_{n,3}(\theta, \sigma^2)$, 114	$b_n(r)$, 89
$\mathcal{D}_n(\eta)$, 50	$b_n^{[1]}(r)$, 89
$\mathcal{E}^{emp}(\theta)$, 85	$c_{ij,h}$, 13
$\mathcal{E}_{n,2}(\eta)$, 100	$c_{ij}(\theta)$, 13
$\mathcal{E}_{n,2}^{imp}(\eta)$, 100	$c_n^{[1]}(r)$, 59

$c_n^{[2]}(r)$, 59	$k_n^{[1]}(r)$, 59
$c_n^{[3]}(r)$, 63	$k_n^{[2]}(r)$, 59
$d_{ij}^T(\theta)$, 55	$k_n^{[3]}(r)$, 63
$d_n(r)$, 55	$s_n(r)$, 63
$f_{ij}(\theta)$, 15	$u_{i,1}$, 106
$g_{n,1}(\theta)$, 19	$u_{i,2}$, 111
$g_{n,1}^{cov}(\theta)$, 33	$v_n(r)$, 63
$g_{n,1}^{imputed}(\theta)$, 22	$v_n^{[1]}(r)$, 59
$g_{n,1}^{obs}(\theta)$, 23	$w_n^{[1]}(r)$, 63
$g_{n,2}(\eta)$, 19	x_{ij} , 11
$g_{n,2}^{imputed}(\eta)$, 22	
$g_{n,2}^{obs}(\eta)$, 23	\mathbb{O}_i , 17

Index

- σ -fields, 11, 17, 24, 30
- e.f.*, 2
- Assumption (A0), 15
- AGT, 12
- ASE, 123
- Asymptotic equivalency, 109, 113
- Asymptotic normality of $n^{-1/2}g_{n,1}^{obs}(\theta_0)$,
107
- Asymptotic normality of $\sqrt{n}(\hat{\sigma}_n^2 - \sigma_0^2)$,
115
- Asymptotic normality of $\sqrt{n}(\hat{\theta}_n - \theta_0)$,
110
- Asymptotic normality of $n^{-1/2}g_{n,2}^{obs}(\eta_0)$,
111
- Basic modeling assumption, 11
- BMI, 2, 12, 118, 119
- Cauchy-Schwarz inequality, 37
- Censored event inequality, 74
- condGEE (R package), 118
- Condition $(B_{(f, \frac{\partial^2 \mu}{\partial \theta^T \partial \theta} V^{-1})})$, 30
- Condition $(B(x_i))$, 33
- Condition $(B_{b(\eta)})$, 31
- Condition $(B_{f(\theta)})$, 24
- Condition (S') , 95
- Condition (LN), 50
- Conditional mean, 11, 13
- Conditional variance, 11
- Conditions $(T0), (T1), (T2)$, 20
- Conditions S(i),S(ii),S(iii), 50
- Consistency theorem for $\hat{\sigma}_n^2(\hat{\theta}_n)$, 38
- Covariate information, 11
- Design matrix of covariates, 53
- EE, 4
- EM algorithm, 6
- Empirical *e.f.*, 23
- ENOES, 123
- Equicontinuity of derivatives, 50, 55,

- 64, 79, 81, 90, 93
- ESE, 123
- Euclidean norm, 49
- Existence and strong consistency
theorems, 50, 97, 103
- Expectation under censoring, 27
- Exponential inequality, 44
- Extension of the Cauchy-Schwarz
inequality, 78
- Gap time, 4, 11
- GEE, 2
- Link function, 3, 13
- Mean value theorem, 60, 67, 68, 110,
115
- Observed censoring time C_i , 11
- Observed estimating functions, 23
- Occurring event at time S_{ij} , 11
- Operator norm, 49
- Roots of the GEE, 3
- Set indicators, 22
- SLLN, 18, 40, 74, 86, 91, 92, 95, 101,
102, 110, 114
- Spectral radius, 49
- Stopping time, 19
- Trace, 57
- Unbiased estimating functions, 17, 19,
20, 22, 26, 31