uOttawa

L'Université canadienne
Canada's university

t

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Laura Dumitrescu
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Mathematics)
GRADE / DEGREE

Department of Mathematics and Statistics
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Asymptotic Results for Some Longitudinal Regression Models
TITRE DE LA THÈSE / TITLE OF THESIS

Raluca Balan
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

Ioana Schiopu-Kratina
CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

Rafal Kulik

Mohamedou Ould-Haye

Mary Thompson
University of Waterloo

Mahmoud Zarepour

Gary W. Slater
Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

Asymptotic Results for Some Longitudinal Regression Models

Laura Dumitrescu

Thesis Submitted to the Faculty of Graduate and Postdoctoral Studies
In partial fulfilment of the requirements for the degree of Doctor of Philosophy in Mathematics [1]

Department of Mathematics and Statistics
Faculty of Science
University of Ottawa

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

Your file Votre référence
ISBN: 978-0-494-79737-2
Our file Notre référence
ISBN: 978-0-494-79737-2

Abstract

The scope of this thesis is to examine asymptotic properties for some longitudinal regression models. These models are often used in biostatistics, econometrics and psychology and involve the collection of several measurements, at different occasions, on the same individuals or clusters. First, we examine the classical regression model for longitudinal data and define the asymptotically optimal estimating equation, with strongly consistent solutions. Then, we consider a simple model with error in covariates, adapted to longitudinal data, obtain estimators of the regression parameters and examine their asymptotic behavior.

Acknowledgements

I want to express my gratitude to my supervisor Raluca Balan and co-supervisor Ioana Schiopu-Kratina for guiding me in all these years with patience, encouragement and knowledge.

I am also grateful to all my professors in Romania, particularly to Professor Monica Dumitrescu for the first Statistics lectures and for being a great mentor.

I am especially thankful to my father for his definite intellectual influence, my mother for her unconditional support, my sister and Mihail. I am aware that this work would not have been successful without their help.

Dedication

To my parents, Marin and Ecaterina

Contents

Chapter 1

Introduction

Longitudinal models are often used in psychology, sociology and biostatistics, where individuals or units are followed over time. This type of investigation can be very complex, involving a large population which might be followed over decades. For example, the National Population Health Survey of Canadians is a study which has been conducted by Statistics Canada since 1994/1995, where information is collected every two years for a period of twenty years. The scope of the present work is to evaluate asymptotic properties of estimators in some regression models for longitudinal data, without a subject specific random effect.

When data are collected on the same unit or individual across time, the observations of the variable of interest are usually correlated, and the correlation matrix might be unknown. A popular method of estimation for longitudinal data was proposed in [24]. Their approach is an extension of the generalized linear models, introduced in [28], to a regression setting with correlated observations within subjects. They assume only a functional form for the marginal distribution on each occasion, without specifying the joint distribution of the measurements. and propose solving the derived generalized estimating equation (GEE) by replacing the true correlation matrix with a "working" matrix, $\mathbf{R}(\alpha)$, which is treated as a nuisance. Their main result is that the estimators, which are roots of such GEEs, are consistent, as long as α is estimated consistently, for any correlation. In addition, they conjecture that estimators are most efficient if the true correlation matrix is correctly modeled or if the sample mean of the observed correlations of each individual is used as the "working" correlation matrix. Hence, the downside of their method is that the estimator of the regression parameter obtained is efficient only when the "working" correlation matrix is correctly specified.

Following the ideas of [24] that improvement in efficiency of the estimators of mean parameters can be done by also estimating parameters in the covariance matrices for the response vector, in the case of binary data, the author of [30] introduces a second set of estimating equations for α. The approach is extended by the authors of [31] to the case of a general multivariate response. However, the suggested method requires additional assumption such as the functional form of the third and forth moments.

The author of [9] shows that there are cases when the use of "working" correlation matrix suggested by the authors of [24] leads to a breakdown of the asymptotic properties of the estimators. The article points out that even if an estimator of α, derived from an estimating function, exists and converges, its limit has to depend on the chosen form of the "working" correlation matrix, $\mathbf{R}(\alpha)$ and on the chosen estimating function. An example using particular choices of correlation structures is used to illustrate the problem. The author concludes that the work of [24] may have theoretical problems, generated by the use of a "working" correlation matrix involving a parameter α, which does not have a precise parametric definition. Consequently, to circumvent this issue, the author of [10], suggests using a method which combines the GEE with Gaussian estimation.

Furthermore, when the correlation structure is misspecified, the authors of [35] show that the estimators of the regression parameter derived from the GEE are usually inefficient as compared to those obtained by using the "working" independence estimating equations approach. They argue that the structure of the correlation is quite important and that, especially in the cases involving an exchangeable true correlation structure. the estimator suggested in [24] is less efficient.

In [32], the authors apply the principles of generalized method of moments to the GEE framework. Their approach is to estimate the "working" correlation $\mathbf{R}(\alpha)$ so that the resulting estimator always exists and is optimal within a specific family. Their use of quadratic inference functions leads to estimators which are more efficient than the ones obtained applying the GEE method when the correlation matrix is misspecified.

Therefore, a GEE in which the true correlation matrix is substituted by another matrix is non-optimal. The goal of Chapter 2 is to identify conditions under which a GEE is (asymptotically) optimal. A comprehensive study of optimality of estimating functions and quasi-likelihood is available in [18] and [21].

As in [21]. the principle used for obtaining quasi-score estimating functions is to maximize a matrix value information with respect to the order relation of non-negative definite matrices. Hence, in order to obtain an optimal estimating equation (EE), we define an appropriate class of estimating equations, find the quasi-score in this class and then find sufficient conditions for which a GEE is asymptotically equivalent to it. Optimality and efficiency considerations of transform martingale estimating functions have been considered in [29]. The proposed estimating equation is robust and satisfies C. R. Rao's asymptotic first order efficiency criterion (see [21]) which leads to a minimal asymptotic confidence interval.

The authors of [23] suggest using an iterative estimating equations procedure to achieve weak consistency and asymptotic efficiency of the estimator. They suggest iterating between the estimation of the unknown parameter β and the estimation of the covariance matrix (using the method of moments). It is shown that their procedure converges at an exponential rate and that the estimator obtained by this procedure is asymptotically as efficient as the optimal estimator.

In [37], the authors consider the GEE suggested in [24], which contains the "working" correlation matrix $\mathbf{R}(\alpha)$. They obtain some sufficient conditions for the asymptotic existence, weak and strong consistency and asymptotic normality of an estimator $\widehat{\beta}_n$, which is defined as a solution of the GEE. Their rigorous results hold when either the number of independent observations, or the number of observations within each subject, (or both), go to infinity.

The authors of [2] consider a generalization of the model to the case when the residuals form a martingale difference sequence. Under the "working independence" assumption, they first construct a preliminary sequence of weakly consistent estimators which is used to estimate the average of the correlation matrices. Their "working" correlation matrix is robust and therefore, this approach avoids the problems associated with the nuisance parameter α. As in [37], they obtain conditions for the asymptotic existence, weak consistency and asymptotic normality of an estimator $\widehat{\beta}_n$, which is defined as a root of their GEE.

In Chapter 3, we prove the existence of a strongly consistent estimator $\widehat{\beta}_n$, defined as a root of a GEE which is a martingale. In addition, as in [37] and [2], to simplify assumptions, we examine the asymptotic behavior of the derivative of the estimating function.

In Chapter 4, we introduce a longitudinal error-in-variables linear model, with a single predictor and no subject specific random effects. To motivate our approach, we discuss first some of the contributions in the field, beginning with the univariate case.

Error-in-variables (also known as measurement error) models are regression models where the covariates cannot be measured directly or without error. Thus, the response variable y is assumed to depend on an unobservable variable ξ, called the latent variable. Instead of ξ we observe $\mathbf{x}$. A simple linear error-in-variables model can be written in the following form:

$$\begin{aligned} y &= \alpha + \beta\xi + \varepsilon, \\ x &= \xi + \delta. \end{aligned}$$

If ξ is a random variable, the model is called a structural error-in-variables model. We also mention here that the latent variable ξ is assumed to be independent of the errors ε and δ. Thus, if an n dimensional sample is available, the collected data are $(x_i, y_i)_{1\leq i\leq n}$ and the unknown parameters of the model are α and β.

The error-in-variables model cannot be reformulated as a classical regression model with random design since the regressor x is correlated with the error:

$$y = \alpha + \beta x + (\varepsilon - \beta\delta).$$

It is known that the error-in-variables models are, in general, not identifiable and therefore, it becomes impossible to consistently estimate the parameters from the data. In the univariate case, several additional assumptions that make the model identifiable can be found in the literature (for further details, see [8], Section 1.2.1). Under the assumption that the variance of δ is known, the authors of [8] obtain the modified least squares estimators. However, their asymptotic covariance matrix is a function of unknown parameters.

The author of [27] obtains some central limit theorems (CLTs) for the univariate linear structural models when the explanatory variables are in the domain of attraction of the normal law (DAN), which constitutes a new approach in the study of measurement error models. The concept of DAN is also used in a regression context by other researchers (see [25] for instance). This approach has two advantages. First, the assumption about the finiteness of the variance of the latent variable can be relaxed. Second, Studentized and self-normalized CLTs (that depend only on the data) for distributions which are in DAN, are already available due to [17].

In the multivariate measurement error regression context, [19] assumes the covariance of the error $\boldsymbol{\delta}$ to be known, or estimated to acquire identifiability of the model. The estimation of parameters is suggested to be performed in two steps: first, the reliability matrix (the correspondent of the reliability ratio in the univariate case) needs to be estimated. Second, classical estimation methods, such as the least squares method are used to obtain estimators of the unknown parameters.

The authors of [5] consider a longitudinal linear mixed model with measurement error. Identifiability of their model follows from assuming a restricted model for the covariance matrix of the latent variables. They describe extensions to the case when replicate or validation data is available. For estimation of the regression parameters, they use a regression calibration method, with a substitution for the unknown covariates, which is corrected for estimation of the variance parameters.

Many of the early results regarding error-in-variables problems and applications are summarized in [15], whereas [8] contains updated results and [6] is an exhaustive review of measurement error models.

In our longitudinal error-in-variables model, the true predictor, $\boldsymbol{\xi}$ is assumed to have the same effect for each repeated measurement within the individual. We employ the method of moments, as used in [8] to obtain consistent estimators appropriate to the multivariate structure of the data. They correspond to the modified least squares estimators of a univariate structural measurement error model. The estimators are constructed under the assumption that the diagonal of the covariance matrix of the error $\boldsymbol{\delta}$ is known. This additional assumption ensures the identifiability of our model. We do not discuss here the case when this information is not available. (In this situation, an estimator of the diagonal of the covariance matrix should be provided; for additional details see Section 2 in [19]).

We use the technique developed in [27] to obtain the CLTs for the estimators of the parameters of interest of our model. In our multivariate context, we use the assumption that the vector formed by the latent variables belongs to the generalized domain of attraction of the normal law (GDAN). The concept of GDAN is defined in [20] and it does not constitute a trivial extension of DAN to the multivariate case, as it was previously assumed (see [26] or [34]). By applying the results obtained in [17], this approach enables us to obtain Studentized and self-normalized CLTs for our estimators.

Thus, our results apply to the case of multivariate data where the components of the vectors formed by the explanatory variables are not necessarily independent, allowing for some degree of correlation between the components of the latent vector. As in [27], this approach allows us to obtain CLTs which are data-based and do not involve unknown parameters. Our results can be viewed as a generalization of the results obtained in [27] to include the case of longitudinal data. This generalization is not trivial, the technical difficulty arising from the use of the GDAN concept. For the sake of consistency with the literature and to facilitate a comparison of results, we employ the notation and the structure of proofs laid out in [27].

Chapter 2

The Asymptotically Optimal GEE

A generalized estimating equation in which the true longitudinal correlation matrix is substituted by other matrix is non-optimal. The goal of this chapter is to identify a sequence of generalized estimating equations which are asymptotically optimal in the sense introduced in [21]. We should point out that the principle used in [21] for obtaining an optimal estimating function within a given class is based on the maximization of a "matrix-valued" information function with respect to the order relation of non-negative definite matrices.

In Section 2.1 we establish our general framework: we consider the response variable to be a random variable on a probability space $(\Omega, \mathcal{F}, P_\beta)$, where β is an unknown parameter. Section 2.2 introduces the appropriate class of estimating equations and the concept of optimal equation within this class. In Section 2.3 we identify some sufficient conditions under which a sequence of generalized estimating equations is asymptotically optimal, in the sense that it is asymptotically equivalent to the sequence of optimal equations. Section 2.4 gives an example of a sequence of generalized estimating equations which is, under certain conditions, asymptotically optimal.

2.1 Model Assumptions

Assume that $\mathbf{y}_i = (y_{i1}, \dots, y_{im})^T$, $i \in \{1, \dots, n\}$ are observations made on n individuals. Each y_{ij} denotes the observation made on individual i at time $j \in \{1, \dots, m\}$. We suppose that each $\mathbf{y}_{ij}$ is a random variable on a probability space $(\Omega, \mathcal{F}, P_\beta)$. Here β is a parameter which lies in an open set $\mathcal{T} \subset \mathbb{R}^p$. We assume that the random vectors $\mathbf{y}_1 \dots, \mathbf{y}_n$ are independent. We denote by $\mathrm{E}_\beta(\cdot)$ and $\mathrm{Var}_\beta(\cdot)$ the expectation, respectively the variance with respect to P_β.

The value of y_{ij} is considered to depend on a regressor $\mathbf{x}_{ij} \in \mathbb{R}^p$ through the value $\theta_{ij} = \mathbf{x}_{ij}^T \beta$. Our model assumptions are the following:

$$\begin{aligned}
\mathrm{E}_\beta(y_{ij}) &= \mu(\mathbf{x}_{ij}^T \beta) := \mu_{ij}(\beta), \\
\mathrm{Var}_\beta(y_{ij}) &= \phi \mu'(\mathbf{x}_{ij}^T \beta) := \phi \sigma_{ij}^2(\beta), \ \phi \neq 0,
\end{aligned}$$

where μ is a differentiable function with $\mu' > 0$. (μ is called the **link function**.) Note that these assumptions are satisfied when dealing with distributions in the exponential family.

Some examples of link functions μ are:

1. $\mu(y) = y$ in the linear regression;
2. $\mu(y) = \exp(y)$ in the log regression for count data;
3. $\mu(y) = \exp(y)/[1 + \exp(y)]$ in the logistic regression for binary data;
4. $\mu(y) = \Phi(y)$, where Φ is the standard normal distribution function, in the probit regression for binary data.

Following [37], we set the nuisance parameter ϕ to 1, as we do not need to estimate it here. Let $\mathbf{A}_i(\beta)$ be the diagonal matrix with elements $\sigma_{i1}^2(\beta), \dots, \sigma_{im}^2(\beta)$.

Note that:

$$\mathrm{E}_\beta(y_{ij} - \mu_{ij}(\beta)) = 0, \ \forall \beta \in \mathcal{T}.$$

Let $\Sigma_i(\beta)$ be the covariance matrix of $\mathbf{y}_i$, with respect to P_β, i.e the $m \times m$ matrix whose elements are:

$$\begin{aligned}\sigma_{i,jk}(\beta) : &= \mathrm{Cov}_\beta(y_{ij}, y_{ik}) \\ &= \mathrm{E}_\beta\{[y_{ij} - \mu_{ij}(\beta)][y_{ik} - \mu_{ik}(\beta)]\}.\end{aligned}$$

Recall that by definition,

$$\sigma^2_{ij}(\beta) = \mathrm{Var}_\beta(y_{ij}) = \mathrm{E}_\beta\{[y_{ij} - \mu_{ij}(\beta)]^2\}.$$

Hence the matrix $\Sigma_i(\beta)$ has the elements $\sigma^2_{i1}(\beta), \ldots, \sigma^2_{im}(\beta)$ on the diagonal. Let $\overline{\mathbf{R}}_i(\beta)$ be the <u>*true*</u> correlation matrix of $\mathbf{y}_i$, with respect to P_β, i.e. the $m \times m$ matrix whose elements are:

$$\overline{r}_{i,jk}(\beta) := \mathrm{Corr}_\beta(y_{ij}, y_{ik}) = \frac{\mathrm{Cov}_\beta(y_{ij}, y_{ik})}{\sqrt{\mathrm{Var}_\beta(y_{ij})\mathrm{Var}_\beta(y_{ik})}} = \frac{\sigma_{i,jk}(\beta)}{\sigma_{ij}(\beta)\sigma_{ik}(\beta)}.$$

We assume that $\{\overline{\mathbf{R}}_i(\beta)\}_{i\geq 1}$ is a sequence of positive definite matrices whose entries are continuously differentiable with respect to β.

We denote:

$$\begin{aligned}\mu_i(\beta) &= (\mu_{i1}(\beta), \ldots, \mu_{im}(\beta))^T, \\ \mathbf{y}_i &= (y_{i1}, \ldots, y_{im})^T.\end{aligned}$$

In matrix notation we have:

$$\begin{aligned}\Sigma_i(\beta) &= \mathrm{Cov}_\beta(\mathbf{y}_i) = \mathrm{E}_\beta[(\mathbf{y}_i - \mu_i(\beta))(\mathbf{y}_i - \mu_i(\beta))^T] \\ \overline{\mathbf{R}}_i(\beta) &= \mathbf{A}_i(\beta)^{-1/2}\Sigma_i(\beta)\mathbf{A}_i(\beta)^{-1/2}.\end{aligned}$$

Lemma 2.1.1 *For any $i \leq n$, let $\boldsymbol{X}_i$ be the $m \times p$ matrix with rows $\boldsymbol{x}^T_{ij}$, $j \leq m$. Then, $\dfrac{\partial \mu_i(\beta)}{\partial \beta^T} = \boldsymbol{A}_i(\beta)\boldsymbol{X}_i$, for any $\beta \in \mathcal{T}$ and $i \geq 1$.*

Proof. Let $\beta \in \mathcal{T}$ and $i \in 1, \ldots n$ be arbitrary. It is sufficient to prove that for any $j \leq m$, $\dfrac{\partial \mu_{ij}(\beta)}{\partial \beta^T} = \sigma^2_{ij}(\beta)\mathbf{x}^T_{ij}$. Using the chain rule, we have:

$$\frac{\partial \mu_{ij}(\beta)}{\partial \beta^T} = \frac{\partial \mu(\mathbf{x}^T_{ij}\beta)}{\partial \beta^T} = \mathbf{x}^T_{ij}\mu'(\mathbf{x}^T_{ij}\beta) = \mathbf{x}^T_{ij}\sigma^2_{ij}(\beta),$$

which proves our conclusion. □

2.2 The Quasi-Score

Let $\mathcal{F}_i$ be the σ- field generated by $\mathbf{y}_1, \ldots, \mathbf{y}_i$ for each $i \leq n$. We denote by $\mathrm{E}_\beta(\cdot|\mathcal{F}_{i-1})$ and $\mathrm{Var}_\beta(\cdot|\mathcal{F}_{i-1})$ the conditional expectation, respectively the conditional variance, given $\mathcal{F}_{i-1}$, with respect to P_β.

Throughout the remaining part of this chapter we consider estimating functions of the form:

$$\mathbf{g}_n(\beta) = \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2} \mathcal{R}_{i-1}(\beta)^{-1} \mathbf{A}_i(\beta)^{-1/2} (\mathbf{y}_i - \mu_i(\beta)), \tag{2.2.1}$$

where $\{\mathcal{R}_n(\beta)\}_n$ is a sequence of $m \times m$ positive definite random matrices which satisfy the following assumptions:

(A) $\mathcal{R}_n(\beta)$ is positive definite for all $n \geq 1$ and $\beta \in \mathcal{T}$; the entries of $\mathcal{R}_n(\beta)$ are continuously differentiable with respect to β, for all $n \geq 1$ and $\beta \in \mathcal{T}$,

(B) the entries of $\mathcal{R}_n(\beta)$ are $\mathcal{F}_n$ - measurable for all $n \geq 1$ and $\beta \in \mathcal{T}$.

We want to find a matrix $\mathcal{R}_{i-1}(\beta)$ which is a good approximation of true correlation matrix $\overline{\mathbf{R}}_i(\beta)$. The fact that we consider $\mathcal{R}_{i-1}(\beta)$ instead of $\mathcal{R}_i(\beta)$ allows us to consider a rich class of estimating functions, as it can be seen in the examples below.

For each $n \geq 1$, we consider the following class of estimating functions:

$$\mathcal{H}_n = \{\mathbf{g}_n(\beta) = \sum_{i=1}^{n} \mathbf{C}_i(\beta)[\mathbf{y}_i - \mu_i(\beta)], \beta \in \mathcal{T}\},$$

where $\mathbf{C}_i(\beta)$ is a $p \times m$ random matrix whose entries are $\mathcal{F}_{i-1}$ - measurable and continuously differentiable with respect to β for any $i \leq n$ and for all $\beta \in \mathcal{T}$. We assume that $\mathrm{E}_\beta[\mathbf{g}_n(\beta)\mathbf{g}_n(\beta)^T] < \infty$ and $\mathrm{E}_\beta\left[\dfrac{\partial \mathbf{g}_n(\beta)}{\partial \beta^T}\right] < \infty$ for any $\beta \in \mathcal{T}$.

Example 2.2.1 The generalized estimating equation (GEE) considered in [37] is:

$$\mathbf{g}_n^{\mathrm{XY}}(\beta) = \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2} \mathbf{R}_i(\alpha)^{-1} \mathbf{A}_i(\beta)^{-1/2} (\mathbf{y}_i - \mu_i(\beta)) = 0, \tag{2.2.2}$$

where $\mathbf{R}_i(\alpha)$ is a known positive definite non-random matrix. In this case, $\mathcal{R}_{i-1}(\beta) = \mathbf{R}_i(\alpha)$, for all $\beta \in \mathcal{T}$ and therefore, $\mathcal{R}_n(\beta)$ clearly satisfies (A) and (B). Hence, (2.2.2) is an element of $\mathcal{H}_n$.

Example 2.2.2 The "working independence" GEE:

$$\mathbf{g}_n^{\mathrm{indep}}(\beta) = \sum_{i=1}^{n} \mathbf{X}_i^T(\mathbf{y}_i - \mu_i(\beta)) = 0$$

is of the form (2.2.1), with $\mathcal{R}_n(\beta) = \mathbf{I}$. This GEE is a particular case of (2.2.2) and hence it is an element of $\mathcal{H}_n$.

Example 2.2.3 The following GEE is of the form (2.2.1):

$$\begin{aligned} \widetilde{\mathbf{g}}_n(\beta) &= \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2} \widetilde{\mathbf{R}}_{i-1}^{-1} \mathbf{A}_i(\beta)^{-1/2}(\mathbf{y}_i - \mu_i(\beta)) = 0 \qquad (2.2.3) \\ \mathcal{R}_n(\beta) &:= \widetilde{\mathbf{R}}_n = \frac{1}{n}\sum_{i=1}^{n} \mathbf{A}_i(\widetilde{\beta}_n)^{-1/2}(\mathbf{y}_i - \mu_i(\widetilde{\beta}_n))(\mathbf{y}_i - \mu_i(\widetilde{\beta}_n))^T \mathbf{A}_i(\widetilde{\beta}_n)^{-1/2}. \end{aligned}$$

Here $\{\widetilde{\beta}_n\}_{n\geq 1}$ represents a sequence of weakly consistent estimators of β_0 which are defined as roots of the equation $\mathbf{g}_n^{\mathrm{indep}}(\beta) = 0$; β_0 is interpreted as the "true" value of the parameter. In this case $\mathcal{R}_n(\beta)$ also satisfies (A) and (B). Therefore, (2.2.3) is also an element of $\mathcal{H}_n$. (Note that (2.2.3) is different from the GEE considered in [2], where $\widetilde{\mathbf{R}}_{i-1}^{-1}$ is replaced by $\widetilde{\mathbf{R}}_n^{-1}$.)

Example 2.2.4 Consider the GEE given by (2.2.1), where:

$$\mathcal{R}_n(\beta) = \mathcal{R}_n^*(\beta) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{A}_i(\beta)^{-1/2}(\mathbf{y}_i - \mu_i(\beta))(\mathbf{y}_i - \mu_i(\beta))^T \mathbf{A}_i(\beta)^{-1/2}.$$

The sequence $\{\mathcal{R}_n(\beta)\}_{n\geq 1}$ clearly satisfies (A) and (B). In Section 2.4 we show that, under some conditions, $\mathcal{R}_n(\beta)$ is a good approximation of $\overline{\mathbf{R}}_i(\beta)$, when $\overline{\mathbf{R}}_i(\beta)$ does not depend on i.

As in [21], we introduce:

$$\mathcal{E}[\mathbf{g}_n(\beta)] := \left\{ \mathrm{E}_\beta \left[\frac{\partial \mathbf{g}_n(\beta)}{\partial \beta^T} \right] \right\}^T \left\{ \mathrm{E}_\beta[\mathbf{g}_n(\beta)\mathbf{g}_n^T(\beta)] \right\}^{-1} \left\{ \mathrm{E}_\beta \left[\frac{\partial \mathbf{g}_n(\beta)}{\partial \beta^T} \right] \right\}.$$

For all $\beta \in \mathcal{T}$ and $\mathbf{g}_n(\beta) \in \mathcal{H}_n$, we use the following notation:

$$\begin{aligned} \mathbf{M}_n(\beta) &:= \mathrm{Cov}_\beta[\mathbf{g}_n(\beta)] = \mathrm{E}_\beta[\mathbf{g}_n(\beta)\mathbf{g}_n^T(\beta)]. \\ \mathbf{H}_n(\beta) &:= -\mathrm{E}_\beta \left[\frac{\partial \mathbf{g}_n(\beta)}{\partial \beta} \right]. \end{aligned}$$

It follows that $\mathcal{E}[\mathbf{g}_n(\beta)] = \mathbf{H}_n(\beta)\mathbf{M}_n(\beta)^{-1}\mathbf{H}_n(\beta)$.

The next definition introduces the concept of optimal estimating function within the class $\mathcal{H}_n$ (see Definition 2.1, [21]).

Definition 2.2.5 *An estimating equation $\overline{\boldsymbol{g}}_n \in \mathcal{H}_n$ is* ***optimal*** *(or* ***quasi-score****) within the class $\mathcal{H}_n$ if for any $\boldsymbol{g}_n \in \mathcal{H}_n$ and for any $\beta \in \mathcal{T}$,*

$$\mathcal{E}[\overline{\boldsymbol{g}}_n(\beta)] - \mathcal{E}[\boldsymbol{g}_n(\beta)] \textit{ is non-negative definite.}$$

The following remarks provide some insight into the motivations leading to the above definition.

Remark 2.2.6 The matrix $\mathcal{E}[\mathbf{g}_n(\beta)]$ is a natural generalization of the Fisher information matrix. If there exists a score function $\mathbf{s}_n(\beta)$, then:

$$\begin{aligned} \mathrm{E}_\beta[\mathbf{s}_n(\beta)] &= 0, \\ \mathbf{I}_n(\beta) &:= \mathrm{Cov}_\beta[\mathbf{s}_n(\beta)] = -\mathrm{E}_\beta \left[\frac{\partial \mathbf{s}_n(\beta)}{\partial \beta^T} \right] \quad \text{is the Fisher information matrix.} \end{aligned}$$

It follows that $\mathcal{E}[\mathbf{s}_n(\beta)] = \mathbf{I}_n(\beta)$. The Cramèr - Rao inequality states that the inverse $\mathcal{E}[\mathbf{s}_n(\beta)]^{-1}$ of the Fisher information matrix is a lower bound for the variance of any unbiased estimator of β. Hence a best unbiased estimator of β is one for which $\mathcal{E}[\mathbf{s}_n(\beta)]$ is maximal (w.r.t. the order of nonnegative definite matrices).

Consider the following estimating equation:

$$\overline{\mathbf{g}}_n(\beta) = \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2} \overline{\mathbf{R}}_i(\beta)^{-1} \mathbf{A}_i(\beta)^{-1/2} (\mathbf{y}_i - \mu_i(\beta)), \quad \beta \in \mathcal{T}.$$

Note that $\overline{\mathbf{g}}_n(\beta)$ is an element of the class $\mathcal{H}_n$.

Proposition 2.2.7 *The estimating equation $\overline{\mathbf{g}}_n$ is a quasi-score in $\mathcal{H}_n$.*

Proof. Let $n \geq 1$ be fixed. We apply Theorem 2.1 of [21]. For an arbitrary $\mathbf{g}_n \in \mathcal{H}_n$, we show that:

$$\mathrm{E}_\beta[\mathbf{g}_n(\beta)\overline{\mathbf{g}}_n(\beta)^T] = -\mathrm{E}_\beta\left[\frac{\partial \mathbf{g}_n(\beta)}{\partial \beta^T}\right], \quad \forall \beta \in \mathcal{T}. \tag{2.2.4}$$

Denote $\overline{\mathbf{C}}_i(\beta) = \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2} \overline{\mathbf{R}}_i(\beta)^{-1} \mathbf{A}_i(\beta)^{-1/2} = \mathbf{X}_i^T \mathbf{A}_i(\beta) \Sigma_i(\beta)^{-1}$ and note that $\overline{\mathbf{C}}_i(\beta)$ is non-random. Then:

$$\overline{\mathbf{g}}_n(\beta) = \sum_{i=1}^{n} \overline{\mathbf{C}}_i(\beta)(\mathbf{y}_i - \mu_i(\beta)), \quad \beta \in \mathcal{T}.$$

Since $\mathbf{y}_i - \mu_i(\beta)$ is independent of $\mathcal{F}_j$ for any $j \neq i$ and $\mathbf{C}_i(\beta)$ is $\mathcal{F}_{i-1}$ measurable, we obtain:

$$\mathrm{E}_\beta[\mathbf{g}_n(\beta)\overline{\mathbf{g}}_n(\beta)^T] = \sum_{i,j=1}^{n} \mathrm{E}_\beta[\mathbf{C}_i(\beta)(\mathbf{y}_i - \mu_i(\beta))(\mathbf{y}_j - \mu_j(\beta))^T \overline{\mathbf{C}}_j(\beta)^T]$$

$$= \sum_{i=1}^{n} \mathrm{E}_\beta\{\mathbf{C}_i(\beta)\mathrm{E}_\beta[(\mathbf{y}_i - \mu_i(\beta))(\mathbf{y}_i - \mu_i(\beta))^T | \mathcal{F}_{i-1}]\overline{\mathbf{C}}_i(\beta)^T\}$$

$$+ \sum_{i=1}^{n} \mathrm{E}_\beta\{\mathbf{C}_i(\beta)\mathrm{E}_\beta[\mathbf{y}_i - \mu_i(\beta)|\mathcal{F}_{i-1}](\mathbf{y}_{i-1} - \mu_{i-1}(\beta))^T \overline{\mathbf{C}}_{i-1}(\beta)^T\}$$

$$+ \sum_{i=1}^{n} \sum_{j \neq i-1, i} \mathrm{E}_\beta\{\mathbf{C}_i(\beta)\mathrm{E}_\beta[(\mathbf{y}_i - \mu_i(\beta))(\mathbf{y}_j - \mu_j(\beta))^T | \mathcal{F}_{i-1}]\overline{\mathbf{C}}_j(\beta)^T\}$$

Since $\mathrm{E}_\beta[\mathbf{y}_i - \mu_i(\beta)|\mathcal{F}_{i-1}] = \mathrm{E}_\beta[\mathbf{y}_i - \mu_i(\beta)] = 0$ and $\mathrm{E}_\beta[(\mathbf{y}_i - \mu_i(\beta))(\mathbf{y}_j - \mu_j(\beta))^T | \mathcal{F}_{i-1}] = \mathrm{E}_\beta[(\mathbf{y}_i - \mu_i(\beta))(\mathbf{y}_j - \mu_j(\beta))^T] = 0$ for any $i \geq 1$ and $j \neq i-1$ and $j \neq i$, the left hand side of (2.2.4) is equal to:

$$\mathrm{E}_\beta[\mathbf{g}_n(\beta)\overline{\mathbf{g}}_n(\beta)^T] \;\; = \;\; \sum_{i=1}^{n} \mathrm{E}_\beta[\mathbf{C}_i(\beta)\Sigma_i(\beta)\overline{\mathbf{C}}_i(\beta)^T]$$

$$= \sum_{i=1}^{n} \mathrm{E}_\beta[\mathbf{C}_i(\beta)\mathbf{A}_i(\beta)\mathbf{X}_i]. \qquad (2.2.5)$$

Denote by $\mathbf{c}_{i1}(\beta), \ldots, \mathbf{c}_{im}(\beta)$ the columns of the matrix $\mathbf{C}_i(\beta)$. Hence, $\mathbf{g}_n(\beta) = \sum_{i=1}^{n}\sum_{j=1}^{m} \mathbf{c}_{ij}(\beta)(y_{ij} - \mu_{ij}(\beta))$. The right hand side of (2.2.4) is equal to:

$$\begin{aligned}
-\mathrm{E}_\beta\left[\frac{\partial \mathbf{g}_n(\beta)}{\partial \beta^T}\right] &= -\sum_{i=1}^{n}\sum_{j=1}^{m}\left\{\mathrm{E}_\beta\left[\frac{\partial \mathbf{c}_{ij}(\beta)}{\partial \beta^T}(y_{ij} - \mu_{ij}(\beta))\right] + \mathrm{E}_\beta\left[\mathbf{c}_{ij}(\beta)\frac{\partial \mu_{ij}(\beta)}{\partial \beta^T}\right]\right\} \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{m}\left\{\mathrm{E}_\beta\left[\frac{\partial \mathbf{c}_{ij}(\beta)}{\partial \beta^T}\mathrm{E}_\beta[y_{ij} - \mu_{ij}(\beta)|\mathcal{F}_{i-1}]\right] + \mathrm{E}_\beta[\mathbf{c}_{ij}(\beta)\sigma_{ij}^2(\beta)\mathbf{x}_{ij}^T]\right\} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m}\mathrm{E}_\beta[\mathbf{c}_{ij}(\beta)\sigma_{ij}^2(\beta)\mathbf{x}_{ij}^T] = \sum_{i=1}^{n}\mathrm{E}_\beta[\mathbf{C}_i(\beta)\mathbf{A}_i(\beta)\mathbf{X}_i]. \qquad (2.2.6)
\end{aligned}$$

Here we have used Lemma 2.1.1, the fact that $\mathbf{c}_{ij}(\beta)$ is $\mathcal{F}_{i-1}$- measurable (and therefore, $\dfrac{\partial \mathbf{c}_{ij}(\beta)}{\partial \beta^T}$ is $\mathcal{F}_{i-1}$- measurable), and $\mathrm{E}_\beta[y_{ij} - \mu_{ij}(\beta)|\mathcal{F}_{i-1}] = \mathrm{E}_\beta[y_{ij} - \mu_{ij}(\beta)] = 0$ for any $i \geq 1$. The conclusion follows from (2.2.5) and (2.2.6) . □

Remark 2.2.8 By taking $\mathbf{g}_n = \overline{\mathbf{g}}_n$ in (2.2.4), we obtain that:

$$\mathrm{Cov}_\beta[\overline{\mathbf{g}}_n(\beta)] = -\mathrm{E}_\beta\left[\frac{\partial \overline{\mathbf{g}}_n(\beta)}{\partial \beta^T}\right], \quad \forall \beta \in \mathcal{T}, \qquad (2.2.7)$$

i.e. $\overline{\mathbf{g}}_n$ has the property of a standard score function.

If we denote by:

$$\begin{aligned}
\overline{\mathbf{M}}_n(\beta) &:= \mathrm{Cov}_\beta[\overline{\mathbf{g}}_n(\beta)], \\
\overline{\mathbf{H}}_n(\beta) &:= -\mathrm{E}_\beta\left[\frac{\partial \overline{\mathbf{g}}_n(\beta)}{\partial \beta}\right],
\end{aligned}$$

from (2.2.7) it follows that:

$$\overline{\mathbf{H}}_n(\beta) = \overline{\mathbf{M}}_n(\beta).$$

Remark 2.2.9 In the case of the "working independence" GEE given in Example 2.2.2, we have:

$$\mathbf{H}_n^{\mathrm{indep}}(\beta) := -\mathrm{E}_\beta\left[\frac{\partial \mathbf{g}_n^{\mathrm{indep}}(\beta)}{\partial \beta^T}\right] = \sum_{i=1}^{n}\mathbf{X}_i^T\mathbf{A}_i(\beta)\mathbf{X}_i.$$

2.3 The Asymptotic Quasi-Score

Even if $\overline{\mathbf{g}}_n$ is the optimal estimating function in the class $\mathcal{H}_n$, it cannot be used in practice, since it depends on the unknown correlation matrix $\overline{\mathbf{R}}_n(\beta)$. In what follows we consider a different estimating function $\mathbf{g}_n^*$ which does not depend on $\overline{\mathbf{R}}_n(\beta)$ and turns out to retain the optimality property of $\overline{\mathbf{g}}_n$, in the asymptotic sense. More precisely, let $\{\mathcal{R}_n^*(\beta)\}_{n\geq 1}$ be a sequence of $m \times m$ random matrices which satisfy (A) and (B) and let $\mathbf{g}_n^*(\beta)$ be the associated estimating function given by (2.2.1). Note that $\mathbf{g}_n^* \in \mathcal{H}_n$, for any $n \geq 1$.

Denote:

$$\begin{aligned} \mathbf{M}_n^*(\beta) &:= \mathrm{Cov}_\beta\left[\mathbf{g}_n^*(\beta)\right], \\ \mathbf{H}_n^*(\beta) &:= -\mathrm{E}_\beta\left[\frac{\partial \mathbf{g}_n^*(\beta)}{\partial \beta^T}\right]. \end{aligned}$$

Lemma 2.3.1 We have the following formulas:

$$\overline{\mathbf{M}}_n(\beta) = \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2} \overline{\mathbf{R}}_i(\beta)^{-1} \mathbf{A}_i(\beta)^{1/2} \mathbf{X}_i \tag{2.3.1}$$

$$\mathbf{M}_n^*(\beta) = \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2} \overline{\mathbf{E}}_{i-1}^*(\beta) \mathbf{A}_i(\beta)^{1/2} \mathbf{X}_i \tag{2.3.2}$$

$$\mathbf{H}_n^*(\beta) = \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2} \mathbf{E}_{i-1}^*(\beta) \mathbf{A}_i(\beta)^{1/2} \mathbf{X}_i, \tag{2.3.3}$$

where $\overline{\mathbf{E}}_{i-1}^*(\beta) = \mathrm{E}_\beta[\mathcal{R}_{i-1}^*(\beta)^{-1}\overline{\mathbf{R}}_i(\beta)\mathcal{R}_{i-1}^*(\beta)^{-1}]$ and $\mathbf{E}_{i-1}^*(\beta) = \mathrm{E}_\beta[\mathcal{R}_{i-1}^*(\beta)^{-1}]$.

Proof. Using an argument similar to the one used for the proof of (2.2.5), we have:

$$\mathrm{E}_\beta[\mathbf{g}_n(\beta)\mathbf{g}_n(\beta)^T] = \sum_{i=1}^{n} \mathrm{E}_\beta[\mathbf{C}_i(\beta)\Sigma_i(\beta)\mathbf{C}_i(\beta)^T], \quad \forall \mathbf{g}_n \in \mathcal{H}_n.$$

We first prove (2.3.1).

$$\begin{aligned} &\overline{\mathbf{M}}_n(\beta) = \mathrm{Cov}_\beta[\overline{\mathbf{g}}_n(\beta)] = \mathrm{E}_\beta[\overline{\mathbf{g}}_n(\beta)\overline{\mathbf{g}}_n(\beta)^T] \\ &= \sum_{i=1}^{n} \mathrm{E}_\beta\{\mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2} \overline{\mathbf{R}}_i(\beta)^{-1} \mathbf{A}_i(\beta)^{-1/2} \mathbf{A}_i(\beta)^{1/2} \overline{\mathbf{R}}_i(\beta) \mathbf{A}_i(\beta)^{1/2} \cdot \end{aligned}$$

$$\cdot \mathbf{A}_i(\beta)^{-1/2}\overline{\mathbf{R}}_i(\beta)^{-1}\mathbf{A}_i(\beta)^{1/2}\mathbf{X}_i\}$$
$$= \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2}\overline{\mathbf{R}}_i(\beta)^{-1}\mathbf{A}_i(\beta)^{1/2}\mathbf{X}_i.$$

Using the fact that the entries of $\mathcal{R}_{i-1}^*(\beta)$ are $\mathcal{F}_{i-1}$ - measurable $\forall \beta \in \mathcal{T}$ we obtain:

$$\mathbf{M}_n^*(\beta) = \mathrm{Cov}_\beta[\mathbf{g}_n^*(\beta)] = \mathrm{E}_\beta[\mathbf{g}_n^*(\beta)\mathbf{g}_n^*(\beta)^T]$$
$$= \sum_{i=1}^{n} \mathrm{E}_\beta\{\mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2}\mathcal{R}_{i-1}^*(\beta)^{-1}\mathbf{A}_i(\beta)^{-1/2}\mathbf{A}_i(\beta)^{1/2}\overline{\mathbf{R}}_i(\beta)\mathbf{A}_i(\beta)^{1/2} \cdot$$
$$\cdot \mathbf{A}_i(\beta)^{-1/2}\mathcal{R}_{i-1}^*(\beta)^{-1}\mathbf{A}_i(\beta)^{1/2}\mathbf{X}_i\}$$
$$= \sum_{i=1}^{n} \mathrm{E}_\beta\left[\mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2}\mathcal{R}_{i-1}^*(\beta)^{-1}\overline{\mathbf{R}}_i(\beta)\mathcal{R}_{i-1}^*(\beta)^{-1}\mathbf{A}_i(\beta)^{1/2}\mathbf{X}_i\right]$$
$$= \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2}\mathrm{E}_\beta[\mathcal{R}_{i-1}^*(\beta)^{-1}\overline{\mathbf{R}}_i(\beta)\mathcal{R}_{i-1}^*(\beta)^{-1}]\mathbf{A}_i(\beta)^{1/2}\mathbf{X}_i$$
$$= \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2}\overline{\mathbf{E}}_{i-1}^*(\beta)\mathbf{A}_i(\beta)^{1/2}\mathbf{X}_i.$$

This proves (2.3.2).

To prove (2.3.3), let $\mathbf{C}_i^*(\beta) = \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2}\mathcal{R}_{i-1}^*(\beta)^{-1}\mathbf{A}_i(\beta)^{-1/2}$. Using (2.2.6) we have:

$$\begin{aligned}
\mathbf{H}_n^*(\beta) &= -\mathrm{E}_\beta\left[\frac{\partial \mathbf{g}_n^*(\beta)}{\partial \beta^T}\right] = \sum_{i=1}^{n} \mathrm{E}_\beta[\mathbf{C}_i^*(\beta)\mathbf{A}_i(\beta)\mathbf{X}_i] \\
&= \sum_{i=1}^{n} \mathrm{E}_\beta[\mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2}\mathcal{R}_{i-1}^*(\beta)^{-1}\mathbf{A}_i(\beta)^{1/2}\mathbf{X}_i] \\
&= \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2}\mathrm{E}_\beta[\mathcal{R}_{i-1}^*(\beta)^{-1}]\mathbf{A}_i(\beta)^{1/2}\mathbf{X}_i \\
&= \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2}\mathbf{E}_{i-1}^*(\beta)\mathbf{A}_i(\beta)^{1/2}\mathbf{X}_i. \ \square
\end{aligned}$$

As in [21], in order to define the concept of asymptotic optimality, for any estimating equation $\mathbf{g}_n \in \mathcal{H}_n$, we introduce the following normalization:

$$\mathbf{g}_n^{(\mathrm{norm})}(\beta) := \left[\mathrm{E}_\beta\left(\frac{\partial \mathbf{g}_n(\beta)}{\partial \beta^T}\right)\right]^{-1} \cdot \mathbf{g}_n(\beta) = -\mathbf{H}_n(\beta)^{-1}\mathbf{g}_n(\beta).$$

We denote $\mathcal{I}[\mathbf{g}_n(\beta)] := \mathrm{Cov}_\beta[\mathbf{g}_n^{(\mathrm{norm})}(\beta)]$.

Remark 2.3.2 For any $\mathbf{g}_n \in \mathcal{H}_n$, we have:

$$\begin{aligned}
\mathcal{J}[\mathbf{g}_n(\beta)] &= \mathrm{E}_\beta\left[\mathbf{g}_n^{(\mathrm{norm})}(\beta)\mathbf{g}_n^{(\mathrm{norm})}(\beta)^T\right] \\
&= \mathbf{H}_n(\beta)^{-1}\mathrm{E}_\beta\left[\mathbf{g}_n(\beta)\mathbf{g}_n(\beta)^T\right]\mathbf{H}_n(\beta)^{-1} \\
&= \mathbf{H}_n(\beta)^{-1}\mathbf{M}_n(\beta)\mathbf{H}_n(\beta)^{-1} \\
&= \{\mathcal{E}[\mathbf{g}_n(\beta)]\}^{-1}.
\end{aligned}$$

Intuitively, we say that $\mathbf{g}_n^* \in \mathcal{H}_n$ is asymptotically optimal if, for n large enough, the matrix $\mathcal{J}[\mathbf{g}_n^*(\beta)]$ is minimal. The following definition introduces this concept rigorously; it corresponds to Definition 5.1 in [21].

Definition 2.3.3 *A sequence $\{\boldsymbol{g}_n^*\}_{n\geq 1}$ of estimating equations with $\boldsymbol{g}_n^* \in \mathcal{H}_n$ for any n is **asymptotically optimal** (or **asymptotic quasi-score**) within $\{\mathcal{H}_n\}_{n\geq 1}$ if for any sequence $\{\boldsymbol{g}_n\}_{n\geq 1}$ with $\boldsymbol{g}_n \in \mathcal{H}_n$ for any $n \geq 1$, and for any $\beta \in \mathcal{T}$ the sequence*

$$\{\mathcal{J}[\boldsymbol{g}_n^*(\beta)]^{-1/2}\mathcal{J}[\boldsymbol{g}_n(\beta)]\mathcal{J}[\boldsymbol{g}_n^*(\beta)]^{-T/2} - \boldsymbol{I}\}_{n\geq 1}$$

is asymptotically non-negative definite, i.e. for any $p \times 1$ vector $\boldsymbol{x}$ of norm 1,

$$\liminf_{n\to\infty} \boldsymbol{x}^T\{\mathcal{J}[\boldsymbol{g}_n^*(\beta)]^{-1/2}\mathcal{J}[\boldsymbol{g}_n(\beta)]\mathcal{J}[\boldsymbol{g}_n^*(\beta)]^{-T/2} - \boldsymbol{I}\}\boldsymbol{x} \geq 0.$$

Here, we used the notation $\mathcal{J}[\mathbf{g}_n^*(\beta)]^{-T/2} = \{\mathcal{J}[\mathbf{g}_n^*(\beta)]^{-1/2}\}^T$.

We introduce the following assumptions: (note that the first one is similar to assumption (H') in [2])

(H) for any $\beta \in \mathcal{T}$, there exists a constant $C_\beta > 0$ such that $\lambda_{\min}[\overline{\mathbf{R}}_n(\beta)] \geq C_\beta$, for any $n \geq 1$

(D) $\lambda_{\min}[\mathbf{H}_n^{\mathrm{indep}}(\beta)] \to \infty$, for any $\beta \in \mathcal{T}$.

Here, $\lambda_{\min}(\cdot)$ denotes the smallest eigenvalue of the corresponding matrix (see A.1.1).

The next theorem states the conditions under which a generalized estimating equation of the form (2.2.1) is an asymptotic quasi-score.

Theorem 2.3.4 *Assume that (H) and (D) hold.*

Let $\{\mathcal{R}_n^*(\beta)\}_{n\geq 1}$ be a sequence which satisfies conditions (A) and (B) as well as:

(C) $\mathcal{R}_{n-1}^*(\beta) - \overline{\boldsymbol{R}}_n(\beta) \xrightarrow{P_\beta} 0$ *element-wise, for all* $\beta \in \mathcal{T}$

(R) there exists $K(\beta) > 0$ *such that* $\lambda_{\min}[\mathcal{R}_n^*(\beta)] \geq K(\beta)$ P_β *a.s. for all* $\beta \in \mathcal{T}$ *and* $n \geq 1$.

Then the sequence $\{\boldsymbol{g}_n^*\}_{n\geq 1}$ *is an asymptotic quasi-score within the collection* $\mathcal{H}_n$.

Proof. By Remark 5.2 of [21], as $\overline{\mathbf{g}}_n$ is a quasi-score in $\mathcal{H}_n$, it is therefore an asymptotic quasi-score in $\mathcal{H}_n$. In what follows we show that $\overline{\mathbf{g}}_n$ and $\mathbf{g}_n^*$ are asymptotically equivalent. More precisely, by Proposition 5.5 of [21], it is enough to show that:

$$\frac{\det \mathrm{E}_\beta[\overline{\mathbf{g}}_n^{(norm)}(\beta)\overline{\mathbf{g}}_n^{(norm)}(\beta)^T]}{\det \mathrm{E}_\beta[\mathbf{g}_n^{*(norm)}(\beta)\mathbf{g}_n^{*(norm)}(\beta)^T]} \to 1, \quad \forall \beta \in \mathcal{T}. \tag{2.3.4}$$

By Remark 2 3 2, it follows that:

$$\begin{aligned}
\mathrm{E}_\beta[\overline{\mathbf{g}}_n^{(norm)}(\beta)\overline{\mathbf{g}}_n^{(norm)}(\beta)^T] &= \overline{\mathbf{M}}_n(\beta)^{-1}, \\
\mathrm{E}_\beta[\mathbf{g}_n^{*(norm)}(\beta)\mathbf{g}_n^{*(norm)}(\beta)^T] &= \mathbf{H}_n^*(\beta)^{-1}\mathbf{M}_n^*(\beta)\mathbf{H}_n^*(\beta)^{-1}.
\end{aligned}$$

Therefore, relation (2.3.4) becomes:

$$\frac{\det \overline{\mathbf{M}}_n(\beta)^{-1}}{\det[\mathbf{H}_n^*(\beta)^{-1}\mathbf{M}_n^*(\beta)\mathbf{H}_n^*(\beta)^{-1}]} \to 1, \quad \forall \beta \in \mathcal{T},$$

which can be written as:

$$\frac{\det \mathbf{H}_n^*(\beta)^2}{\det[\overline{\mathbf{M}}_n(\beta)\mathbf{M}_n^*(\beta)]} \to 1. \quad \forall \beta \in \mathcal{T}. \tag{2.3.5}$$

since $\det(\mathbf{AB}) = \det(\mathbf{A})\cdot\det(\mathbf{B})$ for any matrices $\mathbf{A}$ and $\mathbf{B}$ and $\det(\mathbf{A}^{-1}) = (\det \mathbf{A})^{-1}$ for any non-singular matrix $\mathbf{A}$.

Relation (2.3.5) is proved in two steps. Namely, we show that:

$$\frac{\det \mathbf{H}_n^*(\beta)}{\det \overline{\mathbf{M}}_n(\beta)} \to 1, \quad \forall \beta \in \mathcal{T}. \tag{2.3.6}$$

$$\frac{\det \mathbf{M}_n^*(\beta)}{\det \overline{\mathbf{M}}_n(\beta)} \to 1, \quad \forall \beta \in \mathcal{T}. \tag{2.3.7}$$

For this, we first show that:

$$\overline{\mathbf{R}}_n(\beta)\mathbf{E}^*_{n-1}(\beta) \to \mathbf{I} \text{ element-wise}, \ \forall \beta \in \mathcal{T}. \tag{2.3.8}$$

$$\overline{\mathbf{R}}_n(\beta)\overline{\mathbf{E}}^*_{n-1}(\beta) \to \mathbf{I} \text{ element-wise}, \ \forall \beta \in \mathcal{T}. \tag{2.3.9}$$

Let β be arbitrary, but fixed. We begin by proving (2.3.8). From (C), it follows that there exists a subsequence $(n_k)_{k \geq 1}$, such that:

$$\mathcal{R}^*_{n_k-1}(\beta) - \overline{\mathbf{R}}_{n_k}(\beta) \overset{P_\beta \ a.s}{\longrightarrow} 0, \quad \text{element-wise.}$$

By assumptions. (A) and (H), the matrices $\mathcal{R}^*_{n_k-1}(\beta)$ and $\overline{\mathbf{R}}_{n_k}(\beta)$ are positive definite and therefore invertible. Hence, it follows by Lemma A.2.1(i) that:

$$\mathcal{R}^*_{n_k-1}(\beta)^{-1} - \overline{\mathbf{R}}_{n_k}(\beta)^{-1} \overset{P_\beta \ a.s}{\longrightarrow} 0, \quad \text{element-wise.} \tag{2.3.10}$$

Also, each of the absolute values of the elements of the matrix $\mathcal{R}^*_n(\beta)^{-1}$ is bounded above by $\lambda_{\max}[\mathcal{R}^*_n(\beta)^{-1}]$, by Lemma A.1.12. Since $\lambda_{\max}[\mathcal{R}^*_n(\beta)^{-1}] = \lambda^{-1}_{\min}[\mathcal{R}^*_n(\beta)] \leq K(\beta)^{-1}$, $\forall n \geq 1$, P_β a.s., by (R), we conclude that for any $k \geq 1$:

$$\text{each of the elements of } \mathcal{R}^*_{n_k-1}(\beta)^{-1} \text{ is bounded above by } K(\beta)^{-1}, \ P_\beta \text{ a.s.} \tag{2.3.11}$$

Similarly, by (H),

$$\text{each of the elements of } \overline{\mathbf{R}}_{n_k}(\beta)^{-1} \text{ is bounded above by } C_\beta^{-1}. \tag{2.3.12}$$

Therefore, the convergence in (2.3.10) holds in $L^1(P_\beta)$ as well, and so:

$$\mathrm{E}_\beta[\mathcal{R}^*_{n_k-1}(\beta)^{-1}] - \mathrm{E}_\beta[\overline{\mathbf{R}}_{n_k}(\beta)^{-1}] \to 0, \quad \text{element-wise}$$

or, equivalently,

$$\mathbf{E}^*_{n_k-1}(\beta) - \overline{\mathbf{R}}_{n_k}(\beta)^{-1} \to 0, \quad \text{element-wise (or in norm).} \tag{2.3.13}$$

Using Lemma A.1.12 and the fact that the absolute value of the elements of $\overline{\mathbf{R}}_{n_k}(\beta)$ are bounded above by 1, we conclude that $\lambda_{\max}[\overline{\mathbf{R}}_{n_k}(\beta)] \leq m$, so $\lambda_{\min}[\overline{\mathbf{R}}_{n_k}(\beta)^{-1}] =$

$\{\lambda_{\max}[\overline{\mathbf{R}}_{n_k}(\beta)]\}^{-1} \geq m^{-1}$. By Lemma A.1.14(i), it follows that $\overline{\mathbf{R}}_{n_k}(\beta)^{-1} \geq m^{-1}\mathbf{I}$ for any $k \geq 1$and so $\overline{\mathbf{R}}_{n_k}(\beta) \leq m\mathbf{I}$ for any $k \geq 1$, i.e. (by Lemma A.1.14(ii)),

$$\|\overline{\mathbf{R}}_{n_k}(\beta)\| = \lambda_{\max}[\overline{\mathbf{R}}_{n_k}(\beta)] \leq m. \tag{2.3.14}$$

Since:

$$\begin{aligned}
\|\overline{\mathbf{R}}_{n_k}(\beta)\mathbf{E}^*_{n_k-1}(\beta) - \mathbf{I}\| &= \|\overline{\mathbf{R}}_{n_k}(\beta) \cdot [\mathbf{E}^*_{n_k-1}(\beta) - \overline{\mathbf{R}}_{n_k}(\beta)^{-1}]\| \\
&\leq \|\overline{\mathbf{R}}_{n_k}(\beta)\| \cdot \|\mathbf{E}^*_{n_k-1}(\beta) - \overline{\mathbf{R}}_{n_k}(\beta)^{-1}\|,
\end{aligned}$$

using relations (2.3.13) and (2.3.14) we conclude that:

$$\|\overline{\mathbf{R}}_{n_k}(\beta)\mathbf{E}^*_{n_k-1}(\beta) - \mathbf{I}\| \to 0, \text{ as } k \to \infty. \tag{2.3.15}$$

We now claim that (2.3.8) holds. Suppose, by contradiction, that there exists $\varepsilon_0 > 0$ and a subsequence $(n_l)_{l\geq 1}$ such that:

$$\|\overline{\mathbf{R}}_{n_l}(\beta)\mathbf{E}^*_{n_l-1}(\beta) - \mathbf{I}\| > \varepsilon_0. \quad \forall l \geq 1. \tag{2.3.16}$$

Repeating the previous argument with $\mathcal{R}^*_{n_l}(\beta)$ in lieu of $\mathcal{R}^*_n(\beta)$, and $\overline{\mathbf{R}}_{n_l}(\beta)$ in lieu of $\overline{\mathbf{R}}_n(\beta)$, similarly to (2.3.15), we conclude that there exists a subsequence $(n_{l_k})_{k\geq 1}$ such that:

$$\|\overline{\mathbf{R}}_{n_{l_k}}(\beta)\mathbf{E}^*_{n_{l_k}-1}(\beta) - \mathbf{I}\| \to 0, \text{ as } k \to \infty. \tag{2.3.17}$$

From (2.3.16) and (2.3.17) we obtain a contradiction.

We now turn to the proof of (2.3.9). From (2.3.10), (2.3.11) and the fact that the elements of $\overline{\mathbf{R}}_n(\beta)$ are bounded by 1, we also have:

$$\mathcal{R}^*_{n_k-1}(\beta)^{-1}\overline{\mathbf{R}}_{n_k}(\beta)[\mathcal{R}^*_{n_k-1}(\beta)^{-1} - \overline{\mathbf{R}}_{n_k}(\beta)^{-1}] \stackrel{P_\beta \text{ a.s.}}{\longrightarrow} 0, \quad \text{element-wise},$$

and hence, by using (2.3.10) again we obtain the following element-wise convergence:

$$\begin{aligned}
&[\mathcal{R}^*_{n_k-1}(\beta)^{-1}\overline{\mathbf{R}}_{n_k}(\beta)\mathcal{R}^*_{n_k-1}(\beta)^{-1} - \mathcal{R}^*_{n_k-1}(\beta)^{-1}] \\
&\qquad + [\mathcal{R}^*_{n_k-1}(\beta)^{-1} - \overline{\mathbf{R}}_{n_k}(\beta)^{-1}] \stackrel{P_\beta \text{ a.s.}}{\longrightarrow} 0.
\end{aligned} \tag{2.3.18}$$

The convergence in (2.3.18) holds in $L^1(P_\beta)$ as well, by the bounded convergence theorem and by invoking again (2.3.11), (2.3.12) and the fact that the elements of $\overline{\mathbf{R}}_n(\beta)$ are bounded by 1. Therefore, applying Lemma A.2.2:

$$\mathrm{E}_\beta[\mathcal{R}^*_{n_k-1}(\beta)^{-1}\overline{\mathbf{R}}_{n_k}(\beta)\mathcal{R}^*_{n_k-1}(\beta)^{-1}] - \mathrm{E}_\beta[\overline{\mathbf{R}}_{n_k}(\beta)^{-1}] \to 0, \text{ element-wise, i.e.}$$

$$\overline{\mathbf{E}}^*_{n_k-1}(\beta) - \overline{\mathbf{R}}_{n_k}(\beta)^{-1} \to 0, \text{ element-wise.} \tag{2.3.19}$$

Since:

$$\begin{aligned} \|\overline{\mathbf{R}}_{n_k}(\beta)\overline{\mathbf{E}}^*_{n_k-1}(\beta) - \mathbf{I}\| &= \|\overline{\mathbf{R}}_{n_k}(\beta) \cdot [\overline{\mathbf{E}}^*_{n_k-1}(\beta) - \overline{\mathbf{R}}_{n_k}(\beta)^{-1}]\| \\ &\leq \|\overline{\mathbf{R}}_{n_k}(\beta)\| \cdot \|\overline{\mathbf{E}}^*_{n_k-1}(\beta) -^{-1} \|, \end{aligned}$$

using (2.3.14) and (2.3.19) we conclude that:

$$\|\overline{\mathbf{R}}_{n_k}(\beta)\overline{\mathbf{E}}^*_{n_k-1}(\beta) - \mathbf{I}\| \to 0, \text{ as } k \to \infty.$$

Arguing by contradiction, as in the proof of (2.3.8), we conclude that (2.3.9) holds.

We now turn to the proof of (2.3.6) and (2.3.7). Using relations (2.3.8), (2.3.9) and Theorem 3.14 in [33], for $\varepsilon > 0$, there exists an integer n_0 (depending on ε and β) such that for any $i \geq n_0$:

$$1 - \varepsilon \leq \lambda_{\min}[\overline{\mathbf{R}}_n(\beta)\mathbf{E}^*_{n-1}(\beta)] \leq \lambda_{\max}[\overline{\mathbf{R}}_n(\beta)\mathbf{E}^*_{n-1}(\beta)] \leq 1 + \varepsilon, \tag{2.3.20}$$

$$1 - \varepsilon \leq \lambda_{\min}[\overline{\mathbf{R}}_n(\beta)\overline{\mathbf{E}}^*_{n-1}(\beta)] \leq \lambda_{\max}[\overline{\mathbf{R}}_n(\beta)\overline{\mathbf{E}}^*_{n-1}(\beta)] \leq 1 + \varepsilon. \tag{2.3.21}$$

Denote:

$$\begin{aligned} \overline{\mathbf{M}}_{n_0,n}(\beta) &= \sum_{i=n_0}^{n} \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2}\overline{\mathbf{R}}_i(\beta)^{-1} = \overline{\mathbf{R}}_i(\beta)^{-1}\mathbf{A}_i(\beta)^{1/2}\mathbf{X}_i, \\ \mathbf{M}^*_{n_0,n}(\beta) &= \sum_{i=n_0}^{n} \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2}\overline{\mathbf{E}}^*_{i-1}(\beta)\mathbf{A}_i(\beta)^{1/2}\mathbf{X}_i, \\ \mathbf{H}^*_{n_0,n}(\beta) &= \sum_{i=n_0}^{n} \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2}\mathbf{E}^*_{i-1}(\beta)\mathbf{A}_i(\beta)^{1/2}\mathbf{X}_i. \end{aligned}$$

By writing $\mathbf{E}^*_{i-1}(\beta) = \overline{\mathbf{R}}_i(\beta)^{-1/2}[\overline{\mathbf{R}}_i(\beta)^{1/2}\mathbf{E}^*_{i-1}(\beta)\overline{\mathbf{R}}_i(\beta)^{1/2}]\overline{\mathbf{R}}_i(\beta)^{-1/2}$ and using the fact that the eigenvalues of $\overline{\mathbf{R}}_i(\beta)^{1/2}\mathbf{E}^*_{i-1}(\beta)\overline{\mathbf{R}}_i(\beta)^{1/2}$ are the same as the eigenvalues of $\overline{\mathbf{R}}_i(\beta)\mathbf{E}^*_{i-1}(\beta)$ we obtain:

$$\lambda_{\min}[\overline{\mathbf{R}}_i(\beta)\mathbf{E}^*_{i-1}(\beta)]\overline{\mathbf{R}}_i(\beta)^{-1} \leq \mathbf{E}^*_{i-1}(\beta) \leq \lambda_{\max}[\overline{\mathbf{R}}_i(\beta)\mathbf{E}^*_{i-1}(\beta)]\overline{\mathbf{R}}_i(\beta)^{-1}, \ \forall i \geq 1.$$

We multiply these inequalities by $\mathbf{X}_i^T\mathbf{A}_i(\beta)^{1/2}$ to the left and by $\mathbf{A}_i(\beta)^{1/2}\mathbf{X}_i$ to the right. Taking the sum over $i = n_0, \ldots n$ we obtain:

$$\begin{aligned}
\min_{n_0<i\leq n} \lambda_{\min}[\overline{\mathbf{R}}_i(\beta)^{-1}\mathbf{E}^*_{i-1}(\beta)]\overline{\mathbf{M}}_{n_0,n}(\beta) &\leq \mathbf{H}^*_{n_0,n}(\beta) \\
&\leq \max_{n_0<i\leq n} \lambda_{\max}[\overline{\mathbf{R}}_i(\beta)^{-1}\mathbf{E}^*_{i-1}(\beta)]\overline{\mathbf{M}}_{n_0,n}(\beta).
\end{aligned}$$

Using (2.3.20), this gives:

$$(1-\varepsilon)\overline{\mathbf{M}}_{n_0,n}(\beta) \leq \mathbf{H}^*_{n_0,n}(\beta) \leq (1+\varepsilon)\overline{\mathbf{M}}_{n_0,n}(\beta). \tag{2.3.22}$$

Taking the determinant and using the fact that the determinant is an increasing function with respect to the order of non-negative definite matrices we obtain:

$$(1-\varepsilon)^p \leq \frac{\det[\mathbf{H}^*_{n_0,n}(\beta)]}{\det[\overline{\mathbf{M}}_{n_0,n}(\beta)]} \leq (1+\varepsilon)^p. \tag{2.3.23}$$

By writing $\overline{\mathbf{E}}^*_{i-1}(\beta) = \overline{\mathbf{R}}_i(\beta)^{-1/2}[\overline{\mathbf{R}}_i(\beta)^{1/2}\overline{\mathbf{E}}^*_{i-1}(\beta)\overline{\mathbf{R}}_i(\beta)^{1/2}]\overline{\mathbf{R}}_i(\beta)^{-1/2}$ and using the fact that the eigenvalues of $\overline{\mathbf{R}}_i(\beta)^{1/2}\overline{\mathbf{E}}^*_{i-1}(\beta)\overline{\mathbf{R}}_i(\beta)^{1/2}$ are the same as the eigenvalues of $\overline{\mathbf{R}}_i(\beta)\overline{\mathbf{E}}^*_{i-1}(\beta)$ we obtain:

$$\lambda_{\min}[\overline{\mathbf{R}}_i(\beta)\overline{\mathbf{E}}^*_{i-1}(\beta)]\overline{\mathbf{R}}_i(\beta)^{-1} \leq \overline{\mathbf{E}}^*_{i-1}(\beta) \leq \lambda_{\max}[\overline{\mathbf{R}}_i(\beta)\overline{\mathbf{E}}^*_{i-1}(\beta)]\overline{\mathbf{R}}_i(\beta)^{-1}, \ \forall i \geq 1.$$

We multiply these inequalities by $\mathbf{X}_i^T\mathbf{A}_i(\beta)^{1/2}$ to the left and by $\mathbf{A}_i(\beta)^{1/2}\mathbf{X}_i$ to the right. Taking the sum over $i = n_0, \ldots n$ we obtain:

$$\begin{aligned}
\min_{n_0<i\leq n} \lambda_{\min}[\overline{\mathbf{R}}_i(\beta)\overline{\mathbf{E}}^*_{i-1}(\beta)]\overline{\mathbf{M}}_{n_0,n}(\beta) &\leq \mathbf{M}^*_{n_0,n}(\beta) \\
&\leq \max_{n_0<i\leq n} \lambda_{\max}[\overline{\mathbf{R}}_i(\beta)\overline{\mathbf{E}}^*_{i-1}(\beta)]\overline{\mathbf{M}}_{n_0,n}(\beta).
\end{aligned}$$

Using (2.3.21) this gives:

$$(1-\varepsilon)\overline{\mathbf{M}}_{n_0,n}(\beta) \le \mathbf{M}^*_{n_0,n}(\beta) \le (1+\varepsilon)\overline{\mathbf{M}}_{n_0,n}(\beta), \tag{2.3.24}$$

which implies, similarly to (2.3.23):

$$(1-\varepsilon)^p \le \frac{\det[\mathbf{M}^*_{n_0,n}(\beta)]}{\det[\overline{\mathbf{M}}_{n_0,n}(\beta)]} \le (1+\varepsilon)^p. \tag{2.3.25}$$

Since $\overline{\mathbf{R}}_i(\beta)^{-1} \ge m^{-1}\mathbf{I}$, using (2.3.1) and Lemma A.1.14(i), we have the following estimation

$$\overline{\mathbf{M}}_n(\beta) \;\ge\; \min_{i\le n}\lambda_{\min}[\overline{\mathbf{R}}_i(\beta)^{-1}]\mathbf{H}^{\mathrm{indep}}_n(\beta) \ge m^{-1}\mathbf{H}^{\mathrm{indep}}_n(\beta),$$

By condition (D), it follows that $\lambda_{\min}[\overline{\mathbf{M}}_n(\beta)] \to \infty$. Since $\lambda_{\min}(\mathbf{A}+\mathbf{B}) \ge \lambda_{\min}(\mathbf{A}) + \lambda_{\min}(\mathbf{B})$, by writing $\overline{\mathbf{M}}_{n_0,n}(\beta) = \overline{\mathbf{M}}_n(\beta) - \overline{\mathbf{M}}_{n_0-1}(\beta)$, we have the following relation:

$$\lambda_{\min}[\overline{\mathbf{M}}_n(\beta)] + \lambda_{\min}[-\overline{\mathbf{M}}_{n_0-1}(\beta)] \le \lambda_{\min}[\overline{\mathbf{M}}_{n_0,n}(\beta)].$$

Hence $\lambda_{\min}[\overline{\mathbf{M}}_{n_0,n}(\beta)] \to \infty$, and there exists $n_1 > n_0$ such that $\lambda_{\min}[\overline{\mathbf{M}}_{n_0,n}(\beta)] \ge \varepsilon^{-1}\lambda_{\max}[\overline{\mathbf{M}}_{n_0-1}(\beta)]$ for any $n \ge n_1$, since n_0 is fixed. It follows that for any $p \times 1$ vector $\mathbf{x}$ and any $n \ge n_1$:

$$\mathbf{x}^T\overline{\mathbf{M}}_{n_0-1}(\beta)\mathbf{x} \le \lambda_{\max}[\overline{\mathbf{M}}_{n_0-1}(\beta)] \le \varepsilon\lambda_{\min}[\overline{\mathbf{M}}_{n_0,n}(\beta)] \le \varepsilon\mathbf{x}^T\overline{\mathbf{M}}_{n_0,n}(\beta)\mathbf{x}$$

Hence $\overline{\mathbf{M}}_{n_0-1}(\beta) \le \varepsilon\overline{\mathbf{M}}_{n_0,n}(\beta)$ for any $n \ge n_1$ and therefore:

$$\overline{\mathbf{M}}_{n_0,n}(\beta) \le \overline{\mathbf{M}}_n(\beta) \le (1+\varepsilon)\overline{\mathbf{M}}_{n_0,n}(\beta),\ \forall n \ge n_1. \tag{2.3.26}$$

Using relations (2.3.22) and (2.3.24) and the fact that $\lambda_{\min}[\overline{\mathbf{M}}_{n_0,n}(\beta)] \to \infty$, we obtain that $\lambda_{\min}[\mathbf{H}^*_{n_0,n}(\beta)] \to \infty$, as well as $\lambda_{\min}[\mathbf{M}^*_{n_0,n}(\beta)] \to \infty$. An argument similar to the one leading to (2.3.26), applied to $\mathbf{H}^*_n(\beta)$ and $\mathbf{M}^*_n(\beta)$, allows us to conclude there exist $n_2 > n_1$ and $n_3 > n_1$ such that:

$$\mathbf{H}^*_{n_0,n}(\beta) \;\le\; \mathbf{H}^*_n(\beta) \le (1+\varepsilon)\mathbf{H}^*_{n_0,n}(\beta),\ \forall n \ge n_2. \tag{2.3.27}$$

$$\mathbf{M}^*_{n_0,n}(\beta) \leq \mathbf{M}^*_n(\beta) \leq (1+\varepsilon)\mathbf{M}^*_{n_0,n}(\beta),\ \forall n \geq n_3. \tag{2.3.28}$$

By taking the determinant, and using (2.3.26), (2.3.27) and (2.3.28), we obtain:

$$\begin{aligned}
(1+\varepsilon)^{-p}\frac{\det[\mathbf{H}^*_{n_0,n}(\beta)]}{\det[\overline{\mathbf{M}}_{n_0,n}(\beta)]} &\leq \frac{\det[\mathbf{H}^*_n(\beta)]}{\det[\overline{\mathbf{M}}_n(\beta)]} \leq (1+\varepsilon)^p\frac{\det[\mathbf{H}^*_{n_0,n}(\beta)]}{\det[\overline{\mathbf{M}}_{n_0,n}(\beta)]},\ \forall n \geq n_2,\\
(1+\varepsilon)^{-p}\frac{\det[\mathbf{M}^*_{n_0,n}(\beta)]}{\det[\overline{\mathbf{M}}_{n_0,n}(\beta)]} &\leq \frac{\det[\mathbf{M}^*_n(\beta)]}{\det[\overline{\mathbf{M}}_n(\beta)]} \leq (1+\varepsilon)^p\frac{\det[\mathbf{M}^*_{n_0,n}(\beta)]}{\det[\overline{\mathbf{M}}_{n_0,n}(\beta)]},\ \forall n \geq n_3.
\end{aligned}$$

These relations together with (2.3.23) and (2.3.25) imply:

$$\begin{aligned}
(1-\varepsilon)^p(1+\varepsilon)^{-p} &\leq \frac{\det[\mathbf{H}^*_n(\beta)]}{\det[\overline{\mathbf{M}}_n(\beta)]} \leq (1+\varepsilon)^{2p},\ \forall n \geq n_2,\\
(1-\varepsilon)^p(1+\varepsilon)^{-p} &\leq \frac{\det[\mathbf{M}^*_n(\beta)]}{\det[\overline{\mathbf{M}}_n(\beta)]} \leq (1+\varepsilon)^{2p},\ \forall n \geq n_3.
\end{aligned}$$

This proves (2.3.6) and (2.3.7).□

2.4 An Example

Throughout this section we assume that the correlation matrices $\overline{\mathbf{R}}_i(\beta)$ are the same for all individuals.

We consider the following sequence:

$$\mathcal{R}^*_n(\beta) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{A}_i(\beta)^{-1/2}(\mathbf{y}_i - \mu_i(\beta))(\mathbf{y}_i - \mu_i(\beta))^T\mathbf{A}_i(\beta)^{-1/2},\ \beta \in \mathcal{T}, n \geq 1.$$

We show that this sequence satisfies condition (C) of Theorem 2.3.4 and is equicontinuous at β_0, where $\beta_0 \in \mathcal{T}$ is a fixed value of the parameter.

Proposition 2.4.1 *Assume that for any $\beta \in \mathcal{T}$ and $i \geq 1$:*

$$\overline{\boldsymbol{R}}_i(\beta) = \overline{\boldsymbol{R}}(\beta), \tag{2.4.1}$$

and that the following condition holds:

(N_δ) *for any $\beta \in \mathcal{T}$ there exists some constants $\delta_\beta > 0$ and $C_\beta > 0$, such that:*

$$\mathrm{E}_\beta\|\boldsymbol{A}_i(\beta)^{-1/2}(\boldsymbol{y}_i - \mu_i(\beta))\|^{2+\delta_\beta} \leq C_\beta.\ \textit{for all } i \geq 1.$$

Then the sequence $\{\mathcal{R}_n^*(\beta)\}_{n\geq 1}$ satisfies condition (C) of Theorem 2.3.4. More precisely for all $\beta \in \mathcal{T}$ we have:

$$\mathcal{R}_n^*(\beta) - \overline{\boldsymbol{R}}(\beta) \overset{a.s \text{ and } L^1}{\longrightarrow} 0 \quad \textit{element-wise, with respect to } P_\beta.$$

Proof. Let $\beta \in \mathcal{T}$ be fixed. For any $n \geq 1$ and $j, k \in \{1, \ldots, m\}$, we denote by $r_{jk}^{*(n)}(\beta)$, $\overline{r}_{jk}(\beta)$ the elements of the matrices $\mathcal{R}_n^*(\beta)$ and $\overline{\mathbf{R}}(\beta)$, respectively.

Note that:

$$\begin{aligned} r_{jk}^{*(n)}(\beta) - \overline{r}_{jk}(\beta) &= \frac{1}{n}\sum_{i=1}^{n} \sigma_{ij}(\beta)^{-1}\{(y_{ij} - \mu_{ij}(\beta))(y_{ik} - \mu_{ik}(\beta)) - \\ &\quad - \mathrm{E}_\beta[(y_{ij} - \mu_{ij}(\beta))(y_{ik} - \mu_{ik}(\beta))]\}\sigma_{ik}(\beta)^{-1}. \end{aligned}$$

By the Strong Law of Large Numbers, we conclude that $\{r_{jk}^{*(n)}(\beta) - \overline{r}_{jk}(\beta)\}_{n\geq 1}$ converges to 0 a.s and in $L^1(P_\beta)$. $\square$

For the next theorem we fix a value $\beta_0 \in \mathcal{T}$ and we omit the argument β_0 in the corresponding functions. Note also that the a.s. relations are considered with respect to the probability measure P_{β_0}.

We denote by $B_r = B_r(\beta_0) = \{\beta \in \mathcal{T}; \|\beta - \beta_0\| \leq r\}$ and

$$\eta_n(r) = \sup_{\beta, \beta' \in B_r} \max_{i \leq n, j \leq m} \left| \left[\frac{\mu'(\mathbf{x}_{ij}^T \beta')}{\mu'(\mathbf{x}_{ij}^T \beta)} \right]^{1/2} - 1 \right|.$$

Note that in the case of linear regression (i.e. $\mu(x) = x$), $\eta_n(r) = 0$ for all $r > 0$ and $n \geq 1$.

The next result gives sufficient conditions for the sequence $\mathcal{R}_n^*(\beta)$ to be equicontinuous at β_0.

Proposition 2.4.2 *Assume that* (2.4.1) *holds as well as the following three conditions:*

(K') $\lim_{r\to 0} \limsup_{n\to\infty} \eta_n(r) = 0$,

(G) *there exists a constant* $C > 0$ *such that* $\lambda_{\max}(\boldsymbol{H}_n^{\text{indep}}) \leq Cn, \ \forall n \geq 1$.

(N'_δ) *there exist some constants* $\delta > 0$ *and* $C > 0$ *such that*

$E\|\boldsymbol{A}_i^{-1/2}\varepsilon_i\|^{2+\delta} < C,$ *for all* $i \geq 1$.

Then,

$$\lim_{r\to 0}\limsup_{n\to\infty}\sup_{\beta\in B_r}\|\mathcal{R}_n^*(\beta) - \mathcal{R}_n^*\| = 0 \quad a.s. \tag{2.4.2}$$

Remark 2.4.3 Note that condition (N'_δ) is weaker that condition (N_δ) encountered in Proposition 2.4.1.

Remark 2.4.4 Assumption (G) is satisfied if $\sup_{i\geq 1}\|\mathbf{x}_i\| < \infty$ and μ' is a bounded function: in particular this is the case of the linear regression and logistic probit regression for binary data. (To see this note that $\lambda_{\max}(\mathbf{H}_n^{\mathrm{indep}}) \leq \max_{i\leq n}\lambda_{\max}(\mathbf{A}_i)\sum_{i=1}^n \mathbf{X}_i^T\mathbf{X}_i \leq Cn$.)

Proof of Proposition 2.4.2. The proof is similar to that of Proposition 2 of [2], by replacing $\widetilde{\beta}_n$ with β. We include it for the sake of completeness.

For each $i \leq n$ and $j \leq m$, let $\varepsilon_{ij}(\beta) = y_{ij} - \mu_{ij}(\beta)$ and $y_{ij}^*(\beta) = \varepsilon_{ij}(\beta)/\sigma_{ij}(\beta)$. The elements of the matrix $\mathcal{R}_n^*(\beta)$ are:

$$r_{jk}^{*(n)}(\beta) = \frac{1}{n}\sum_{i=1}^n \frac{\varepsilon_{ij}(\beta)\varepsilon_{ik}(\beta)}{\sigma_{ij}(\beta)\sigma_{ik}(\beta)} = \frac{1}{n}\sum_{i=1}^n y_{ij}^*(\beta)y_{ik}^*(\beta), \text{ where } i \leq n \text{ and } j,k \leq m.$$

For each $i \leq n$ and $j,k \leq m$ we denote:

$$\begin{aligned}
\delta_{i,jk}(\beta) &:= \frac{\sigma_{ij}\sigma_{ik}}{\sigma_{ij}(\beta)\sigma_{ik}(\beta)} - 1\\
\Delta\mu_{ij}(\beta) &:= \mu_{ij}(\beta) - \mu_{ij}\\
\Delta(\varepsilon_{ij}\varepsilon_{ik})(\beta) &:= \varepsilon_{ij}(\beta)\varepsilon_{ik}(\beta) - \varepsilon_{ij}\varepsilon_{ik}\\
&= \Delta\mu_{ij}(\beta)\Delta\mu_{ik}(\beta) - \Delta\mu_{ij}(\beta)\varepsilon_{ik} - \Delta\mu_{ik}(\beta)\varepsilon_{ij}.
\end{aligned}$$

Let $j,k \in \{1,\ldots,m\}$ be fixed. With the above notation we have:

$$r_{jk}^{*(n)}(\beta) - r_{jk}^{*(n)} = \frac{1}{n}\sum_{i=1}^n \frac{\varepsilon_{ij}(\beta)\varepsilon_{ik}(\beta)}{\sigma_{ij}(\beta)\sigma_{ik}(\beta)} - \frac{1}{n}\sum_{i=1}^n \frac{\varepsilon_{ij}\varepsilon_{ik}}{\sigma_{ij}\sigma_{ik}}$$

$$= \frac{1}{n}\sum_{i=1}^{n}\frac{\Delta(\varepsilon_{ij}\varepsilon_{ik})(\beta)}{\sigma_{ij}\sigma_{ik}} + \frac{1}{n}\sum_{i=1}^{n}\frac{\Delta(\varepsilon_{ij}\varepsilon_{ik})(\beta)}{\sigma_{ij}\sigma_{ik}}\delta_{i,jk}(\beta) + \frac{1}{n}\sum_{i=1}^{n}\frac{\varepsilon_{ij}\varepsilon_{ik}}{\sigma_{ij}\sigma_{ik}}\delta_{i,jk}(\beta).$$

After some algebraic manipulations, we obtain that:

$$|r_{jk}^{*(n)}(\beta) - r_{jk}^{*(n)}| \le U_{n,jk}(\beta) + \max_{i\le n}|\delta_{i,jk}(\beta)|\cdot[U_{n,jk}(\beta) + V_{n,jk}], \tag{2.4.3}$$

where:

$$\begin{aligned} V_{n,jk} &= \frac{1}{n}\sum_{i=1}^{n}|y_{ij}^{*}y_{ik}^{*}|, \\ U_{n,jk}(\beta) &= \frac{1}{n}\sum_{i=1}^{n}\frac{|\Delta\mu_{ij}(\beta)||\Delta\mu_{ik}(\beta)|}{\sigma_{ij}\sigma_{ik}} + \frac{1}{n}\sum_{i=1}^{n}\frac{|\Delta\mu_{ij}(\beta)|}{\sigma_{ij}}|y_{ik}^{*}| + \frac{1}{n}\sum_{i=1}^{n}\frac{|\Delta\mu_{ik}(\beta)|}{\sigma_{ik}}|y_{ij}^{*}| \\ &:= U_{n,jk}^{(1)}(\beta) + U_{n,jk}^{(2)}(\beta) + U_{n,jk}^{(3)}(\beta). \end{aligned}$$

Let $\varepsilon > 0$ be arbitrary. We want to prove that with probability 1, there exist some random numbers $r_0(\varepsilon) > 0$ and $n_0(\varepsilon) \ge 1$ such that:

$$|r_{jk}^{*(n)}(\beta) - r_{jk}^{*(n)}| \le \varepsilon,\ \forall r \in (0, r_0(\varepsilon)),\ \forall n \ge n_0(\varepsilon).$$

We treat separately the terms $\max_{i\le n}|\delta_{i,jk}(\beta)|$, $V_{n,jk}$ and $U_{n,jk}(\beta)$.

We begin with the term $\max_{i\le n}|\delta_{i,jk}(\beta)|$. By the definition of $\eta_n(r)$, for any $\beta \in B_r$, $i \le n$, $j \le m$, we have the following relation:

$$\begin{aligned} |\delta_{i,jk}(\beta)| &= \left|\frac{\sigma_{ij}\sigma_{ik}}{\sigma_{ij}(\beta)\sigma_{ik}(\beta)} - 1\right| \le \left|\frac{\sigma_{ij}}{\sigma_{ij}(\beta)} - 1\right|\cdot\left|\frac{\sigma_{ik}}{\sigma_{ik}(\beta)} - 1\right| + \left|\frac{\sigma_{ij}}{\sigma_{ij}(\beta)} - 1\right| \\ &+ \left|\frac{\sigma_{ik}}{\sigma_{ik}(\beta)} - 1\right| \le \eta_n^2(r) + 2\eta_n(r). \end{aligned}$$

By (K'). there exist $r_0(\varepsilon) \in (0, \varepsilon)$ and $n_1(\varepsilon)$ such that:

$$\eta_n(r) \le \varepsilon,\ \forall r \in (0, r_0(\varepsilon)),\ \forall n \ge n_1(\varepsilon). \tag{2.4.4}$$

Hence:

$$\max_{i\le n}|\delta_{i,jk}(\beta)| \le \varepsilon^2 + 2\varepsilon,\ \forall \beta \in B_{r_0(\varepsilon)},\ \forall n \ge n_1(\varepsilon). \tag{2.4.5}$$

Next, we treat the term $V_{n,jk}$. By Cauchy-Schwarz inequality:

$$V_{n,jk} \leq \left[\frac{1}{n}\sum_{i=1}^{n}|y_{ij}^{*}|^{2}\right]^{1/2} \cdot \left[\frac{1}{n}\sum_{i=1}^{n}|y_{ik}^{*}|^{2}\right]^{1/2} = \left(r_{jj}^{*(n)}\right)^{1/2}\left(r_{kk}^{*(n)}\right)^{1/2}.$$

Using an argument similar to the one used in the proof of Proposition 2.4.1 (and using assumption $(N_{\delta}^{\prime})$), it follows that:

$$|r_{jj}^{*(n)} - \overline{r}_{jj}| \xrightarrow{a.s.} 0.$$

Since, $\overline{r}_{jj} \leq 1$, it follows that with probability 1, there exists a random number $n_{2,j}(\varepsilon)$ such that:

$$r_{jj}^{*(n)} \leq 1+\varepsilon \text{ a.s. } \forall n \geq n_{2,j}(\varepsilon). \tag{2.4.6}$$

By letting $n_{2}(\varepsilon) = \max_{j\leq m} n_{2,j}(\varepsilon)$ we obtain:

$$V_{n,jk} \leq 1+\varepsilon \text{ a.s } \forall n \geq n_{2}(\varepsilon) \tag{2.4.7}$$

Let $r>0$ and $\beta \in B_{r}$ by arbitrary. We now treat the term $U_{n,jk}(\beta)$. Applying Taylor's formula to the function $\mu_{ij}(\beta)$, we obtain that there exists $\overline{\beta}_{ij}$ between β and β_{0} such that $\mu_{ij}(\beta)-\mu_{ij} = \sigma_{ij}^{2}(\overline{\beta}_{ij})\mathbf{x}_{ij}^{T}(\beta-\beta_{0})$. Hence, for any $n \geq 1$ we have:

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\frac{|\Delta\mu_{ij}(\beta)|^{2}}{\sigma_{ij}^{2}} &= n^{-1}(\beta-\beta_{0})^{T}\left\{\sum_{i=1}^{n}\left[\frac{\sigma_{ij}^{2}(\overline{\beta}_{ij})}{\sigma_{ij}^{2}}\right]^{2}\sigma_{ij}^{2}\mathbf{x}_{ij}\mathbf{x}_{ij}^{T}\right\}(\beta-\beta_{0}) \\
&\leq n^{-1}(\beta-\beta_{0})^{T}\left\{\sum_{i=1}^{n}\mathbf{X}_{i}^{T}\mathbf{A}_{i}^{1/2}[\mathbf{A}_{i}(\overline{\beta}_{ij})\mathbf{A}_{i}^{-1}]\mathbf{A}_{i}^{1/2}\mathbf{X}_{i}\right\}(\beta-\beta_{0}) \\
&\leq n^{-1}\max_{i\leq n}\lambda_{\max}^{2}[\mathbf{A}_{i}(\overline{\beta}_{ij})\mathbf{A}_{i}^{-1}](\beta-\beta_{0})^{T}\mathbf{H}_{n}^{\text{indep}}(\beta-\beta_{0}) \\
&\leq n^{-1}\lambda_{\max}(\mathbf{H}_{n}^{\text{indep}})\max_{i\leq n}\lambda_{\max}^{2}[\mathbf{A}_{i}(\overline{\beta}_{ij})\mathbf{A}_{i}^{-1}]\|(\beta-\beta_{0})\|^{2} \\
&\leq C\max_{i\leq n}\lambda_{\max}^{2}[\mathbf{A}_{i}(\overline{\beta}_{ij})\mathbf{A}_{i}^{-1}]r^{2}.
\end{aligned}$$

where we used assumption (G) for the last inequality.

Note that for any $i \leq n$, $j \leq m$ and $\beta \in B_{r}$, $n \geq 1$:

$$\frac{\mu^{\prime}(\mathbf{x}_{ij}^{T}\overline{\beta}_{ij})}{\mu^{\prime}(\mathbf{x}_{ij}^{T}\beta_{0})} = \left\{1+\left[\frac{\mu^{\prime}(\mathbf{x}_{ij}^{T}\overline{\beta}_{ij})}{\mu^{\prime}(\mathbf{x}_{ij}^{T}\beta_{0})}\right]^{1/2}-1\right\}^{2} \leq [1+\eta_{n}(r)]^{2}.$$

Hence:

$$\max_{i\leq n}\lambda^2_{\max}[\mathbf{A}_i(\beta)\mathbf{A}_i^{-1}]\leq[1+\eta_n(r)]^4,\ \forall n\geq 1,\ \forall\beta\in B_r.$$

We conclude that for any $\beta\in B_r$, $n\geq 1$:

$$\frac{1}{n}\sum_{i=1}^n\frac{|\Delta\mu_{ij}(\beta)|^2}{\sigma^2_{ij}}\leq Cr^2[1+\eta_n(r)]^4.$$

By (2.4.4) it follows that:

$$\frac{1}{n}\sum_{i=1}^n\frac{|\Delta\mu_{ij}(\beta)|^2}{\sigma^2_{ij}}\leq Cr_0(\varepsilon)^2(1+\varepsilon)^4\leq C\varepsilon^2(1+\varepsilon)^4,\ \forall n\geq n_1(\varepsilon).\ \forall\beta\in B_{r_0(\varepsilon)}.$$

Using Cauchy-Schwarz inequality, it follows that with probability equal to 1:

$$\begin{aligned}
U^{(1)}_{n,jk}(\beta) &\leq \left(\frac{1}{n}\sum_{i=1}^n\frac{|\Delta\mu_{ij}(\beta)|^2}{\sigma^2_{ij}}\right)^{1/2}\left(\frac{1}{n}\sum_{i=1}^n\frac{|\Delta\mu_{ik}(\beta)|^2}{\sigma^2_{ik}}\right)^{1/2}\leq C\varepsilon^2(1+\varepsilon)^4,\\
U^{(2)}_{n,jk}(\beta) &\leq \left(\frac{1}{n}\sum_{i=1}^n\frac{|\Delta\mu_{ij}(\beta)|^2}{\sigma^2_{ij}}\right)^{1/2}\left(r^{*(n)}_{kk}\right)^{1/2}\leq C\varepsilon(1+\varepsilon)^{5/2},\\
U^{(3)}_{n,jk}(\beta) &\leq \left(\frac{1}{n}\sum_{i=1}^n\frac{|\Delta\mu_{ik}(\beta)|^2}{\sigma^2_{ik}}\right)^{1/2}\left(r^{*(n)}_{jj}\right)^{1/2}\leq C\varepsilon(1+\varepsilon)^{5/2},
\end{aligned}$$

for any $n\geq n_3(\varepsilon)$ and $\beta\in B_{r_0(\varepsilon)}$, where $n_3(\varepsilon):=\max\{n_1(\varepsilon),n_2(\varepsilon)\}$ is a random number.

Hence, with probability equal to 1:

$$U_{n,jk}(\beta)\leq 3C\varepsilon(1+\varepsilon)^4,\ \forall n\geq n_3(\varepsilon),\ \forall\beta\in B_{r_0(\varepsilon)}. \tag{2.4.8}$$

Returning to (2.4.3) and using (2.4.5), (2.4.7) and (2.4.8), we conclude that, with probability 1, for any $n\geq n_0(\varepsilon)$ and for any $\beta\in B_{r_0(\varepsilon)}$,

$$|r^{*(n)}_{jk}(\beta)-r^{*(n)}_{jk}|\leq 3C\varepsilon(1+\varepsilon)^4+(\varepsilon^2+2\varepsilon)[3C\varepsilon(1+\varepsilon)^4+1+\varepsilon]\leq\varepsilon.$$

Here $n_0(\varepsilon)>n_3(\varepsilon)$ is chosen sufficiently large so that the second equality above holds. This concludes the proof of (2.4.4). □

Chapter 3

Strong Consistency

In this chapter we fix β_0, assumed to be the "true", unknown value of the parameter β, and prove the asymptotic existence of a sequence $\{\widehat{\beta}_n\}_{n\geq 1}$ of strongly consistent estimators of β_0. To do this, we consider a sequence of estimating functions which form a zero-mean square integrable martingale with respect to P_{β_0}.

In Section 3.1 we prove a general theorem which guarantees the existence and a.s. convergence of a sequence of estimators of β_0, defined as roots of generalized estimating equations which are martingales. Some preliminary results needed for computational purposes are given in Section 3.2. The asymptotic behavior of the terms of the derivative of the estimating function is treated in Section 3.3. In Section 3.4 we find simplified conditions under which a strongly consistent sequence of estimators for β_0 exits.

Throughout this chapter, the almost sure relations are considered with respect to the probability measure P_{β_0}. For simplicity, we omit writing the argument β_0, i.e. we write P, $\mathrm{E}(\cdot)$ instead of P_{β_0}, $\mathrm{E}_{\beta_0}(\cdot)$, etc. We denote by C a general constant which can differ from one case to another.

3.1 A General Theorem

In what follows we consider $\beta_0 \in \mathcal{T}$ to be arbitrary (but fixed). All the random variables are defined on the probability space $(\Omega, \mathcal{F}, P_{\beta_0})$.

For each $n \geq 1$ and $\beta \in \mathcal{T}$ let $\mathbf{g}_n(\beta)$ be a $p \times 1$ random vector such that the function $\mathbf{g}_n$ is continuously differentiable with respect to β. We use the notations $B_r = B_r(\beta_0) = \{\beta \in \mathcal{T}; \|\beta - \beta_0\| \leq r\}$ and $\partial B_r = \{\beta \in \mathcal{T}; \|\beta - \beta_0\| = r\}$ with $r > 0$.

Recall the notation for $\beta \in \mathcal{T}$:

$$\begin{aligned} \mathcal{D}_n(\beta) &= -\frac{\partial \mathbf{g}_n(\beta)}{\partial \beta^T} \\ \mathbf{M}_n(\beta) &= \mathrm{Cov}_\beta[\mathbf{g}_n(\beta)]. \end{aligned}$$

The following theorem gives sufficient conditions for the existence of a strongly consistent sequence of estimators of β_0. Note that some of the conditions are similar to those of Theorem 7 of [37]. Note that here $\{\mathbf{g}_n, \mathcal{F}_n\}_{n\geq 1}$ is a martingale and we do not assume that $\mathcal{D}_n$ is positive definite.

Theorem 3.1.1 *Let $\{\boldsymbol{g}_n, \mathcal{F}_n\}_{n\geq 1}$ be a zero-mean square integrable martingale defined on a probability space $(\Omega, \mathcal{F}, P_{\beta_0})$.*

Let $(\alpha_n)_{n\geq 1}$ be a sequence of constants such that there exists $C > 0$ and $N \geq 1$ with:

$$C\lambda_{\max}[\boldsymbol{M}_n] \leq \alpha_n, \ \forall n \geq N. \tag{3.1.1}$$

Suppose the following conditions hold:

(I) $\lambda_{\min}[\boldsymbol{M}_n] \to \infty$,

(S) there exist some constants $\delta > 0$, $c_0 > 0$ such that with probability 1, there exist some random numbers $r_0 > 0$ and $n_0 \geq 1$ with:

(i) $|\boldsymbol{x}^T \mathcal{D}_n(\beta)\boldsymbol{x}| > 0$ for any $p \times 1$ vector, $\boldsymbol{x}$, $\|\boldsymbol{x}\| = 1$, $\beta \in B_{r_0}$ and $n \geq n_0$,

(ii) $\lim_{r\to 0} \limsup_{n\to\infty} \alpha_n^{-1/2-\delta} \sup_{\beta \in B_r} \|\mathcal{D}_n(\beta) - \mathcal{D}_n\| = 0$,

(iii) $|\boldsymbol{x}^T \mathcal{D}_n \boldsymbol{x}| \geq c_0 \alpha_n^{1/2+\delta}$ for any $p \times 1$ vector, $\boldsymbol{x}$, $\|\boldsymbol{x}\| = 1$ and $n \geq n_0$.

Then, there exists a sequence of random vectors $(\widehat{\beta}_n)_{n\geq 1}$ *and a random number* n_1 *such that:*

$$(a)\ P_\beta(\boldsymbol{g}_n(\widehat{\beta}_n) = 0,\ \textit{for any } n \geq n_1) = 1,$$
$$(b)\ \widehat{\beta}_n \xrightarrow{a.s.} \beta_0.$$

Proof. Since

$$\frac{\partial[\alpha_n^{-1/2-\delta}\mathbf{g}_n(\beta)]}{\partial\beta^T} = -\alpha_n^{-1/2-\delta}\mathcal{D}_n(\beta)$$

and by condition (S)(i), $|\mathbf{x}^T\mathcal{D}_n(\beta)\mathbf{x}| > 0$ for all vectors $\mathbf{x}$ of norm 1, it follows that with probability 1, for all $\beta \in B_{r_0}$ and $n \geq n_0$, $\mathcal{D}_n(\beta)$ is non-singular and hence the derivative of $\alpha_n^{-1/2-\delta}\mathbf{g}_n(\beta)$ is non-singular (i.e. $\alpha_n^{-1/2-\delta}\mathbf{g}_n(\beta)$ is one-to-one).

We show that with probability 1, for any $\varepsilon > 0$, there exist some random numbers $r_\varepsilon \in (0,\varepsilon)$, $r_\varepsilon < r_0$ and $N_\varepsilon > n_0$ such that:

$$\alpha_n^{-1/2-\delta}\|\mathbf{g}_n\| \leq \alpha_n^{-1/2-\delta}\inf_{\beta\in\partial B_{r_\varepsilon}}\|\mathbf{g}_n(\beta) - \mathbf{g}_n\|, \forall n \geq N_\varepsilon. \tag{3.1.2}$$

Applying Theorem C.0.10 to the function $\mathbf{T}_n(\beta) = \alpha_n^{-1/2-\delta}\mathbf{g}_n(\beta)$, this implies that with probability 1, for any $\varepsilon > 0$ there exist some random numbers $r_\varepsilon \in (0,\varepsilon)$, $r_\varepsilon < r_0$ and $N_\varepsilon > n_0$ such that for any $n \geq N_\varepsilon$, there exists a random vector $\widehat{\beta}_n \in B_{r_\varepsilon}$ with $\mathbf{g}_n(\widehat{\beta}_n) = 0$. Part (b) follows from the definition of β_n since, with probability 1, $\|\widehat{\beta}_n - \beta_0\| \leq r_\varepsilon < \varepsilon$, for any $n \geq N_\varepsilon$.

To prove (3.1.2), it follows from (S)(ii) that, for any $\varepsilon > 0$, there exist some random numbers $r_\varepsilon \in (0,\varepsilon)$, $r_\varepsilon < r_0$ and $n_\varepsilon > n_0$ such that:

$$\alpha_n^{-1/2-\delta}\|\mathcal{D}_n(\beta) - \mathcal{D}_n\| < \varepsilon,$$

for any $\beta \in \partial B_{r_\varepsilon}$ and $n \geq n_\varepsilon$. Hence, $\alpha_n^{-1/2-\delta}|\mathbf{x}^T[\mathcal{D}_n(\beta) - \mathcal{D}_n]\mathbf{x}| < \varepsilon$, for any vector $\mathbf{x}$ of norm 1 . Therefore, for any $n \geq n_\varepsilon$ and $\beta \in \partial B_{r_\varepsilon}$ we have:

$$\alpha_n^{-1/2-\delta}|\mathbf{x}^T\mathcal{D}_n(\beta)\mathbf{x}| > \alpha_n^{-1/2-\delta}|\mathbf{x}^T\mathcal{D}_n\mathbf{x}| - \varepsilon. \tag{3.1.3}$$

Let $n \geq n_\varepsilon$ and $\beta \in \partial B_{r_\varepsilon}$. We apply the Mean Value Theorem to $\mathbf{T}_n(\beta)$ and hence there exists $\overline{\beta}_n \in B_{r_\varepsilon}$ such that:

$$\alpha_n^{-1/2-\delta}\|\mathbf{g}_n(\beta) - \mathbf{g}_n\| = \alpha_n^{-1/2-\delta}\|\mathcal{D}_n(\overline{\beta}_n)(\beta - \beta_0)\|.$$

Therefore, using Lemma A.1.13 with $\mathbf{A} = \mathcal{D}_n(\overline{\beta}_n)$, $\mathbf{v} = \beta - \beta_0$ and $\mathbf{x}_n$ - an eigenvector of norm 1 corresponding to $\lambda_{\min}[\mathcal{D}_n(\overline{\beta}_n)^T\mathcal{D}_n(\overline{\beta}_n)]$, we obtain that for any $\beta \in \partial B_{r_\varepsilon}$:

$$\begin{aligned} \alpha_n^{-1-2\delta}\|\mathbf{g}_n(\beta) - \mathbf{g}_n\|^2 &= \alpha_n^{-1-2\delta}\|\mathcal{D}_n(\overline{\beta}_n)(\beta - \beta_0)\|^2 \\ &\geq \alpha_n^{-1-2\delta}[\mathbf{x}_n^T\mathcal{D}_n(\overline{\beta}_n)\mathbf{x}_n]^2\|(\beta - \beta_0)\|^2 \end{aligned}$$

Hence, using (3.1.3),

$$\begin{aligned} \alpha_n^{-1/2-\delta}\|\mathbf{g}_n(\beta) - \mathbf{g}_n\| &\geq \alpha_n^{-1/2-\delta}|\mathbf{x}_n^T\mathcal{D}_n(\overline{\beta}_n)\mathbf{x}_n|\|(\beta - \beta_0)\| \\ &\geq [\alpha_n^{-1/2-\delta}|\mathbf{x}_n^T\mathcal{D}_n\mathbf{x}_n| - \varepsilon]r_\varepsilon \\ &\geq (c_0 - \varepsilon)r_\varepsilon, \quad \text{a.s. }, \end{aligned}$$

where for the last inequality we have used condition (S) (iii). Hence with probability 1, for any $\beta \in \partial B_{r_\varepsilon}$ and $n \geq n_\varepsilon$

$$\alpha_n^{-1/2-\delta}\inf_{\beta \in \partial B_{r_\varepsilon}}\|\mathbf{g}_n(\beta) - \mathbf{g}_n\| \geq (c_0 - \varepsilon)r_\varepsilon. \tag{3.1.4}$$

We next show that for any p-dimensional vector λ of norm 1:

$$\alpha_n^{-1/2-\delta}\lambda^T\mathbf{g}_n \to 0, \quad \text{a.s.} \tag{3.1.5}$$

Since $\{\mathbf{g}_n, \mathcal{F}_n\}_{n\geq1}$ is a zero-mean square integrable martingale, it follows that $\{\lambda^T\mathbf{g}_n, \mathcal{F}_n\}_{n\geq1}$ is also a zero-mean square integrable martingale. To see this, denote $\lambda = (\lambda_1, \ldots, \lambda_p)^T$ and $\mathbf{g}_n = [\mathbf{g}_n^1, \ldots, \mathbf{g}_n^p]^T$. Then $\mathrm{E}[\lambda^T\mathbf{g}_n] = \mathrm{E}[\sum_{i=1}^p \lambda_i\mathbf{g}_n^i] = \sum_{i=1}^p \lambda_i\mathrm{E}[\mathbf{g}_n^i] = 0$, and:

$$\mathrm{E}[\lambda^T\mathbf{g}_n|\mathcal{F}_{n-1}] = \mathrm{E}[\sum_{i=1}^p \lambda_i\mathbf{g}_n^i|\mathcal{F}_{n-1}] = \sum_{i=1}^p \lambda_i\mathrm{E}[\mathbf{g}_n^i|\mathcal{F}_{n-1}] = \sum_{i=1}^p \lambda_i\mathbf{g}_{n-1}^i = \lambda^T\mathbf{g}_{n-1},$$

where we have used the fact that $\{\mathbf{g}_n^i, \mathcal{F}_n\}_{n\geq 1}$ is a martingale. Note that:

$$\mathrm{Var}[\lambda^T \mathbf{g}_n] = \lambda^T \mathrm{E}[\mathbf{g}_n \mathbf{g}_n^T]\lambda = \lambda^T \mathbf{M}_n \lambda \leq \lambda_{\max}[\mathbf{M}_n]\lambda^T\lambda = \lambda_{\max}[\mathbf{M}_n] \leq C^{-1}\alpha_n,$$

by (3.1.1). As condition (I) holds, by applying Theorem B.0.9 we conclude that (3.1.5) holds.

For any $\varepsilon > 0$, let r_ε, n_ε as defined above. From (3.1.5) it follows that with probability 1, there exists a random number $N_\varepsilon > n_\varepsilon$, such that:

$$\alpha_n^{-1/2-\delta}\|\mathbf{g}_n\| \leq (c_0 - \varepsilon)r_\varepsilon, \quad \forall n \geq N_\varepsilon. \tag{3.1.6}$$

From (3.1.4) and (3.1.6) we obtain (3.1.2).□

3.2 Preliminary Results

We use the following notations:

$$k_n^{[2]}(\beta) := \max_{i\leq n, j\leq m}\left|\frac{\mu''(\mathbf{x}_{ij}^T\beta)}{\mu'(\mathbf{x}_{ij}^T\beta)}\right|, \; k_n^{[3]}(\beta) := \max_{i\leq n, j\leq m}\left|\frac{\mu'''(\mathbf{x}_{ij}^T\beta)}{\mu'(\mathbf{x}_{ij}^T\beta)}\right|,$$

$$\gamma_n^{(0),\mathrm{indep}} = \max_{i\leq n, j\leq m} \mathbf{x}_{ij}^T(\mathbf{H}_n^{\mathrm{indep}})^{-1}\mathbf{x}_{ij}. \; a_n := \gamma_n^{(0),\mathrm{indep}} \cdot \lambda_{\max}(\mathbf{H}_n^{\mathrm{indep}}).$$

We introduce the following assumptions:

(AH) there exists $C > 0$ such that for $l = 2, 3$, we have:

$$k_{n,r}^{[l]} := \sup_{\beta\in B_r} k_{n,r}^{[l]}(\beta) \leq C, \; \forall r > 0, \quad \forall n \geq 1.$$

(K) $\lim_{r\to 0}\limsup_{n\to\infty} r a_n^{1/2} = 0$.

Note that assumption (K) is equivalent to $\limsup_{n\to\infty} a_n < \infty$.

The next result shows that assumption (K) is stronger than assumption (K') encountered in Proposition 2.4.2. The result is similar to Lemma B.1 of [37].

Lemma 3.2.1 *Suppose assumptions (AH) and (K) hold. Then, there exist some constants $r_0 > 0$ and $n_0 \geq 1$ such that for every $r \in (0, r_0)$ and $n \geq n_0$, we have:*

$$\eta_n(r) \leq C r a_n^{1/2}.$$

Proof. Using the fact that for any $x > 0$. $|x-1| \le |x-1|(x+1) = |x^2-1|$, we obtain:

$$\eta_n(r) \;\le\; \sup_{\beta,\beta' \in B_r} \max_{i \le n, j \le m} \left| \frac{\mu'(\mathbf{x}_{ij}^T \beta')}{\mu'(\mathbf{x}_{ij}^T \beta)} - 1 \right| := \psi_n(r). \tag{3.2.1}$$

By Taylor's formula, applied to the function $\phi_{ij}(\beta) = \mu'(\mathbf{x}_{ij}^T \beta)$, there exists a $\overline{\beta}_{ij}$ between β and β' such that:

$$\mu'(\mathbf{x}_{ij}^T \beta') - \mu'(\mathbf{x}_{ij}^T \beta) = \mu''(\mathbf{x}_{ij}^T \overline{\beta}_{ij}) x_{ij}^T (\beta' - \beta).$$

Hence for all $\beta, \beta' \in B_r$ and $i \le n, j \le m$:

$$\begin{aligned} \left| \frac{\mu'(\mathbf{x}_{ij}^T \beta')}{\mu'(\mathbf{x}_{ij}^T \beta)} - 1 \right| &= \left| \frac{\mu''(\mathbf{x}_{ij}^T \overline{\beta}_{ij})}{\mu'(\mathbf{x}_{ij}^T \beta)} \right| \cdot |\mathbf{x}_{ij}^T (\beta' - \beta)| \\ &\le \left| \frac{\mu''(\mathbf{x}_{ij}^T \overline{\beta}_{ij})}{\mu'(\mathbf{x}_{ij}^T \overline{\beta}_{ij})} \right| \left| \frac{\mu'(\mathbf{x}_{ij}^T \overline{\beta}_{ij})}{\mu'(\mathbf{x}_{ij}^T \beta)} \right| \cdot |\mathbf{x}_{ij}^T (\beta' - \beta)|. \end{aligned}$$

Using assumption (AH), it follows that:

$$\begin{aligned} \psi_n(r) &\le \sup_{\beta \in B_r} \max_{i \le n, j \le m} \left| \frac{\mu''(\mathbf{x}_{ij}^T \beta)}{\mu'(\mathbf{x}_{ij}^T \beta)} \right| \sup_{\beta,\beta' \in B_r} \max_{i \le n, j \le m} \left| \frac{\mu'(\mathbf{x}_{ij}^T \beta')}{\mu'(\mathbf{x}_{ij}^T \beta)} \right| \sup_{\beta,\beta' \in B_r} \max_{i \le n, j \le m} |\mathbf{x}_{ij}^T (\beta' - \beta)| \\ &\le C \left[\sup_{\beta,\beta' \in B_r} \max_{i \le n, j \le m} \left| \frac{\mu'(\mathbf{x}_{ij}^T \beta')}{\mu'(\mathbf{x}_{ij}^T \beta)} - 1 \right| + 1 \right] \sup_{\beta,\beta' \in B_r} \max_{i \le n, j \le m} |\mathbf{x}_{ij}^T (\beta' - \beta)| \\ &= C\psi_n(r) \sup_{\beta,\beta' \in B_r} \max_{i \le n, j \le m} |\mathbf{x}_{ij}^T (\beta' - \beta)| + C \sup_{\beta,\beta' \in B_r} \max_{i \le n, j \le m} |\mathbf{x}_{ij}^T (\beta' - \beta)|. \end{aligned} \tag{3.2.2}$$

By the definition of $\gamma_n^{(0)\text{ indep}}$, for any $\beta, \beta' \in B_r$ we have:

$$\begin{aligned} |\mathbf{x}_{ij}^T (\beta' - \beta)|^2 &= |\mathbf{x}_{ij}^T (\mathbf{H}_n^{\text{indep}})^{-1/2} (\mathbf{H}_n^{\text{indep}})^{1/2} (\beta' - \beta)|^2 \\ &\le \|\mathbf{x}_{ij}^T (\mathbf{H}_n^{\text{indep}})^{-1/2}\|^2 \cdot \|(\mathbf{H}_n^{\text{indep}})^{1/2} (\beta' - \beta)\|^2 \\ &= \mathbf{x}_{ij}^T (\mathbf{H}_n^{\text{indep}})^{-1} \mathbf{x}_{ij} \cdot (\beta' - \beta)^T \mathbf{H}_n^{\text{indep}} (\beta' - \beta) \\ &\le \gamma_n^{(0),\text{indep}} \lambda_{\max}(\mathbf{H}_n^{\text{indep}}) r^2 = a_n r^2. \end{aligned} \tag{3.2.3}$$

Therefore,

$$\sup_{\beta,\beta' \in B_r} \max_{i \le n, j \le m} |\mathbf{x}_{ij}^T (\beta' - \beta)| \le a_n^{1/2} r, \; \forall n \ge 1, \; \forall r > 0.$$

Relation (3.2.2) becomes:

$$\psi_n(r) \leq C\psi_n(r)a_n^{1/2}r + Ca_n^{1/2}r. \ \forall n \geq 1, \ \forall r > 0,$$

i.e.

$$\psi_n(r)(1 - Ca_n^{1/2}r) \leq Ca_n^{1/2}r, \ \forall n \geq 1, \ \forall r > 0.$$

By assumption (K). there exists some constants $r_0 > 0$ and $n_0 \geq 1$ such that $a_n^{1/2}r \leq 1/(2C)$ for all $r \in (0, r_0)$ and $n \geq n_0$. Hence $\psi_n(r) \leq 2Ca_n^{1/2}r$, for all $r \in (0, r_0)$ and $n \geq n_0$.

The conclusion follows from (3.2.1). □

The next lemma is needed for computational purposes in Section 3.3. The result is similar to Lemma B.2 of [37].

Lemma 3.2.2 *Suppose assumptions (AH) and (K) hold. Then, there exist some constants $r_0 > 0$ and $n_0 \geq 1$ such that for every $r \in (0, r_0)$ and $n \geq n_0$, we have:*

$$(i)\ \nu_n(r) = \sup_{\beta\in B_r} \max_{i\leq n, j\leq m} \left\{ \left| \frac{\mu''(\boldsymbol{x}_{ij}^T\beta)}{\mu'(\boldsymbol{x}_{ij}^T\beta)^{1/2}} - \frac{\mu''(\boldsymbol{x}_{ij}^T\beta_0)}{\mu'(\boldsymbol{x}_{ij}^T\beta_0)^{1/2}} \right| \cdot |\mu'(\boldsymbol{x}_{ij}^T\beta_0)|^{-1/2} \right\} \leq Ca_n^{1/2}r,$$

$$(ii)\ \xi_n(r) = \sup_{\beta\in B_r} \max_{i\leq n, j\leq m} \left\{ \left| \frac{\mu''(\boldsymbol{x}_{ij}^T\beta)}{\mu'(\boldsymbol{x}_{ij}^T\beta)^{3/2}} - \frac{\mu''(\boldsymbol{x}_{ij}^T\beta_0)}{\mu'(\boldsymbol{x}_{ij}^T\beta_0)^{3/2}} \right| \cdot |\mu'(\boldsymbol{x}_{ij}^T\beta_0)|^{1/2} \right\} \leq Ca_n^{1/2}r.$$

Proof. (i) Let $h_{ij}(\beta) = \dfrac{\mu''(\mathbf{x}_{ij}^T\beta)}{\mu'(\mathbf{x}_{ij}^T\beta)^{1/2}}$. The function $h_{ij}(\beta)$ is differentiable on B_r, with:

$$h'_{ij}(\beta) = \left[\mu'''(\mathbf{x}_{ij}^T\beta)\mu'(\mathbf{x}_{ij}^T\beta)^{-1/2} - \frac{1}{2}\mu''(\mathbf{x}_{ij}^T\beta)^2\mu'(\mathbf{x}_{ij}^T\beta)^{-3/2}\right]\mathbf{x}_{ij}^T.$$

by Taylor's formula, there exist a $\overline{\beta}_{ij}$ between β and β_0 such that:

$$h_{ij}(\beta) - h_{ij}(\beta_0) = h'_{ij}(\overline{\beta}_{ij})(\beta - \beta_0).$$

Hence,

$$|h_{ij}(\beta) - h_{ij}(\beta_0)| \cdot |\mu'(\mathbf{x}_{ij}^T\beta_0)|^{-1/2} = |h'_{ij}(\overline{\beta}_{ij})\mu'(\mathbf{x}_{ij}^T\beta_0)^{-1/2}(\beta - \beta_0)|$$

$$\leq \left[\left| \frac{\mu'''(\mathbf{x}_{ij}^T\overline{\beta}_{ij})}{\mu'(\mathbf{x}_{ij}^T\overline{\beta}_{ij})} \right| \cdot \left| \frac{\mu'(\mathbf{x}_{ij}^T\overline{\beta}_{ij})}{\mu'(\mathbf{x}_{ij}^T\beta_0)} \right|^{1/2} + \frac{1}{2} \left| \frac{\mu''(\mathbf{x}_{ij}^T\overline{\beta}_{ij})}{\mu'(\mathbf{x}_{ij}^T\overline{\beta}_{ij})} \right|^2 \cdot \left| \frac{\mu'(\mathbf{x}_{ij}^T\overline{\beta}_{ij})}{\mu'(\mathbf{x}_{ij}^T\beta_0)} \right|^{1/2} \right] |\mathbf{x}_{ij}^T(\beta - \beta_0)|$$
$$:= \phi_{ij}(\overline{\beta}_{ij})|\mathbf{x}_{ij}^T(\beta - \beta_0)|.$$

By taking $\sup_{\beta \in B_r} \max_{i \leq n, j \leq m}$ in the above inequality, we obtain:

$$\nu_n(r) \leq \sup_{\beta \in B_r} \max_{i \leq n, j \leq m} \phi_{ij}(\overline{\beta}_{ij})|\mathbf{x}_{ij}^T(\beta - \beta_0)|.$$

We have:

$$\left[\frac{\mu'(\mathbf{x}_{ij}^T\overline{\beta}_{ij})}{\mu'(\mathbf{x}_{ij}^T\beta_0)} \right]^{1/2} \leq \eta_n(r) + 1 \leq Cra_n^{1/2} + 1 \leq C, \ \forall r \in (0, r_0), \ \forall n \geq n_0,$$

by Lemma 3.2.1 and assumption (K). Using assumption (AH), this implies that:

$$\sup_{\beta \in B_r} \max_{i \leq n, j \leq m} \phi_{ij}(\overline{\beta}_{ij}) \leq C, \ \forall r \in (0, r_0), \ \forall n \geq n_0.$$

Therefore,

$$\nu_n(r) \leq C \sup_{\beta \in B_r} \max_{i \leq n, j \leq m} |\mathbf{x}_{ij}^T(\beta - \beta_0)|, \ \forall r \in (0, r_0), \ \forall n \geq n_0,$$

and the conclusion follows from (3.2.3), applied to $\beta' = \beta_0$.

(ii) In this case, $h_{ij}(\beta) = \dfrac{\mu''(\mathbf{x}_{ij}^T\beta)}{\mu'(\mathbf{x}_{ij}^T\beta)^{3/2}}$. The proof is very similar to part (i) and is omitted. □

3.3 Asymptotic Results for $\mathcal{D}_n(\beta)$

This section gives sufficient conditions for $(S)(ii)$ in Theorem 3.1.1 to be satisfied, in the case of estimating functions of form (2.2.1). Throughout this section we assume that the sequence of random matrices $\mathcal{R}_n(\beta)$ satisfies assumptions (A) and (B).

We denote:

$$\mathcal{D}_n(\beta) = -\frac{\partial \mathbf{g}_n(\beta)}{\partial \beta^T}. \tag{3.3.1}$$

To proceed with the calculation of $\mathcal{D}_n(\beta)$, we write:

$$\begin{aligned}
\mathbf{g}_n(\beta) &= \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2} \mathcal{R}_{i-1}(\beta)^{-1} \mathbf{A}_i(\beta)^{-1/2} [\mathbf{y}_i - \mu_i(\beta)] \\
&= \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2} \mathcal{R}_{i-1}(\beta)^{-1} \mathbf{A}_i(\beta)^{-1/2} [\mu_i - \mu_i(\beta)] \\
&+ \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2} \mathcal{R}_{i-1}(\beta)^{-1} \mathbf{A}_i(\beta)^{-1/2} (\mathbf{y}_i - \mu_i) \\
&:= g_n^{(1)}(\beta) + g_n^{(2)}(\beta).
\end{aligned} \tag{3.3.2}$$

The derivative of the first term gives rise to four terms. More precisely,

$$\frac{\partial \mathbf{g}_n^{(1)}(\beta)}{\partial \beta^T} = \mathbf{B}_n^{(1)}(\beta) + \mathbf{B}_n^{(2)}(\beta) + \mathbf{B}_n^{(3)}(\beta) - \mathbf{H}_n(\beta) := \mathbf{B}_n(\beta) - \mathbf{H}_n(\beta), \tag{3.3.3}$$

where:

$$\begin{aligned}
\mathbf{H}_n(\beta) &= \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2} \mathcal{R}_{i-1}(\beta)^{-1} \mathbf{A}_i(\beta)^{1/2} \mathbf{X}_i \\
\mathbf{B}_n^{(1)}(\beta) &= \sum_{i=1}^{n} \mathbf{X}_i^T \mathrm{diag}\{\mathcal{R}_{i-1}(\beta)^{-1} \mathbf{A}_i(\beta)^{-1/2} [\mu_i - \mu_i(\beta)]\} \mathbf{G}_i^{(1)}(\beta) \mathbf{X}_i \\
\mathbf{B}_n^{(2)}(\beta) &= \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2} \mathcal{R}_{i-1}(\beta)^{-1} \mathrm{diag}\{\mu_i - \mu_i(\beta)\} \mathbf{G}_i^{(2)}(\beta) \mathbf{X}_i \\
\mathbf{b}_n^{(l)}(\beta) &= \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2} \mathbf{T}_{i-1}^{(l)}(\beta) \mathbf{A}_i(\beta)^{-1/2} [\mu_i - \mu_i(\beta)].
\end{aligned} \tag{3.3.4}$$

Here, $\mathbf{b}_n^{(l)}(\beta)$ is the l-th column of $\mathbf{B}_n^{(3)}(\beta)$ and

$$\mathbf{T}_{i-1}^{(l)}(\beta) = \frac{\partial \mathcal{R}_{i-1}(\beta)^{-1}}{\partial \beta_l^T} = -\mathcal{R}_{i-1}(\beta)^{-1} \mathbf{Q}_{i-1}^{(l)}(\beta) \mathcal{R}_{i-1}(\beta)^{-1},$$

with $\mathbf{Q}_{i-1}^{(l)}(\beta) = \dfrac{\partial \mathcal{R}_{i-1}(\beta)}{\partial \beta_l^T}$ and $l = 1, \ldots p$. The derivative of the second term is:

$$\frac{\partial \mathbf{g}_n^{(2)}(\beta)}{\partial \beta^T} = \mathcal{E}_n^{(1)}(\beta) + \mathcal{E}_n^{(2)}(\beta) + \mathcal{E}_n^{(3)}(\beta) := \mathcal{E}_n(\beta), \tag{3.3.5}$$

where:

$$\mathcal{E}_n^{(1)}(\beta) = \sum_{i=1}^{n} \mathbf{X}_i^T \mathrm{diag}\{\mathcal{R}_{i-1}(\beta)^{-1} \mathbf{A}_i(\beta)^{-1/2} (\mathbf{y}_i - \mu_i)\} \mathbf{G}_i^{(1)}(\beta) \mathbf{X}_i$$

$$\mathcal{E}_n^{(2)}(\beta) = \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2} \mathcal{R}_{i-1}(\beta)^{-1} \mathrm{diag}\{\mathbf{y}_i - \mu_i\} \mathbf{G}_i^{(2)}(\beta) \mathbf{X}_i \quad (3.3.6)$$

$$e_n^{(l)}(\beta) = \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2} \mathbf{T}_{i-1}^{(l)}(\beta) \mathbf{A}_i(\beta)^{-1/2} (\mathbf{y}_i - \mu_i).$$

Here $e_n^{(l)}(\beta)$ is the l-th column of $\mathcal{E}_n^{(3)}(\beta)$. The matrices $\mathbf{G}_i^{(1)}(\beta)$ and $\mathbf{G}_i^{(2)}(\beta)$ are defined by: (see also Appendix A of [37])

$$\mathbf{G}_i^{(1)}(\beta) = \mathrm{diag}\left\{\frac{1}{2}\mu''(\mathbf{x}_{i1}^T\beta)\mu'(\mathbf{x}_{i1}^T\beta)^{-1/2}, \ldots, \frac{1}{2}\mu''(\mathbf{x}_{im}^T\beta)\mu'(\mathbf{x}_{i1}^T\beta)^{-1/2}\right\}$$

$$\mathbf{G}_i^{(2)}(\beta) = \mathrm{diag}\left\{-\frac{1}{2}\mu''(\mathbf{x}_{i1}^T\beta)\mu'(\mathbf{x}_{i1}^T\beta)^{-3/2}, \ldots, -\frac{1}{2}\mu''(\mathbf{x}_{im}^T\beta)\mu'(\mathbf{x}_{im}^T\beta)^{-3/2}\right\}$$

From (3.3.1), (3.3.2), (3.3.3), (3.3.5), we conclude that:

$$\mathcal{D}_n(\beta) = \mathbf{H}_n(\beta) - \mathbf{B}_n(\beta) - \mathcal{E}_n(\beta).$$

For any $r > 0$ and any $n \geq 1$, we define:

$$\pi_n(r) = \sup_{\beta \in B_r} \max_{i \leq n} \lambda_{\max}[\mathcal{R}_{i-1}^{1/2} \mathcal{R}_{i-1}(\beta)^{-1} \mathcal{R}_{i-1}^{1/2}]$$

$$\rho_n(r) = \sup_{\beta \in B_r} \max_{i \leq n} \lambda_{\max}[|\mathcal{R}_{i-1}^{1/2} \mathcal{R}_{i-1}(\beta)^{-1} \mathcal{R}_{i-1}^{1/2} - \mathbf{I}|].$$

Note that $\pi_n(r)$ and $\rho_n(r)$ are random variables since they depend on the random matrix $\mathcal{R}_i(\beta)$. Also, $\rho_n(r) \geq 0$ a.s.. for any $r > 0$ and $n \geq 1$. Furthermore,

$$\pi_n(r) \geq 1 \text{ a.s. for any } r > 0 \text{ and } n \geq 1.$$

(To see this, note that $\pi_n(r) \geq \max_{i \leq n} \lambda_{\max}[\mathcal{R}_{i-1}^{1/2} \mathcal{R}_{i-1}(\beta)^{-1} \mathcal{R}_{i-1}^{1/2}]$, for any $\beta \in B_r$. In particular, for $\beta = \beta_0$, we have $\pi_n(r) \geq \max_{i \leq n} \lambda_{\max}(\mathbf{I}) = 1$.)

Remark 3.3.1 If the sequence $\{\mathcal{R}_n(\beta)\}_{n \geq 1}$ is a.s. equicontinuous at β_0 a.s., i.e.

$$\lim_{r \to 0} \limsup_{n \to \infty} \sup_{\beta \in B_r} \|\mathcal{R}_n(\beta) - \mathcal{R}_n\| = 0 \text{ a.s.,} \quad (3.3.7)$$

then:

$$\lim_{r \to 0} \limsup_{n \to \infty} \rho_n(r) = 0 \text{ a.s. and } \lim_{r \to 0} \limsup_{n \to \infty} \pi_n(r) = 1 \text{ a.s.}$$

Proposition 2.4.2 shows that under some conditions, the sequence $\{\mathcal{R}_n^*(\beta)\}_{n\geq 1}$ introduced in Example 2.2.4 is equicontinuous at β_0.

We introduce the following assumption:

(R') there exists a constant $C > 0$ such that $\lambda_{\min}(\mathcal{R}_n) \geq C$, for all $n \geq 1$,a.s.

Note that assumption (R') is weaker than assumption (R) encountered in Theorem 2.3.4.

Remark 3.3.2 By the definition of $\pi_n(r)$ and assumption (R'), it follows that for any $n \geq 1$ and $r > 0$:

$$\lambda_{\max}[\mathcal{R}_{i-1}(\beta)^{-1}] \leq C\pi_n(r), \ \forall \beta \in B_r, \forall i \leq n, \forall n \geq 1, \ \text{a.s.} \tag{3.3.8}$$

(To see this, note that for any $p \times 1$ vector $\mathbf{x}$ of norm 1, for any $i \leq n$ and $\beta \in B_r$, we have:

$$\begin{aligned}
\mathbf{x}^T\mathcal{R}_{i-1}(\beta)^{-1}\mathbf{x} &= \mathbf{x}^T\mathcal{R}_{i-1}^{-1/2}\mathcal{R}_{i-1}^{1/2}\mathcal{R}_{i-1}(\beta)^{-1}\mathcal{R}_{i-1}^{1/2}\mathcal{R}_{i-1}^{-1/2}\mathbf{x} \\
&\leq \max_{i\leq n}\lambda_{\max}[\mathcal{R}_{i-1}^{1/2}\mathcal{R}_{i-1}(\beta)^{-1}\mathcal{R}_{i-1}^{1/2}]\mathbf{x}^T\mathcal{R}_{i-1}^{-1}\mathbf{x} \\
&\leq \pi_n(r)\lambda_{\max}(\mathcal{R}_{i-1}^{-1}) \leq C\pi_n(r).)
\end{aligned}$$

Recall that $\mathbf{H}_n^{\text{indep}} = \sum_{i=1}^n \mathbf{X}_i^T\mathbf{A}_i\mathbf{X}_i$ and denote

$$\delta_n := \alpha_n^{-1/2-\delta}\lambda_{\max}(\mathbf{H}_n^{\text{indep}}),$$

where $(\alpha_n)_{n\geq 1}$ is a non-decreasing sequence of constants (to be specified later) which satsifies $\lim_{n\to\infty}\alpha_n = \infty$.

The following inequalities will be used repeatedly in the proofs of Lemmas 3.3.3, 3.3.3 and 3.3.6.

(a) Cauchy-Schwarz inequality:

$$\left|\sum_{i=1}^n \mathbf{a}_i^T\mathbf{b}_i\right| \leq \left[\sum_{i=1}^n \mathbf{a}_i^T\mathbf{a}_i\right]^{1/2} \cdot \left[\sum_{i=1}^n \mathbf{b}_i^T\mathbf{b}_i\right]^{1/2} \tag{3.3.9}$$

(b)

$$\min_{i\leq n} \lambda_{\min}(\mathbf{A}_i) \sum_{i=1}^{n} \mathbf{x}_i^T \mathbf{x}_i \leq \sum_{i=1}^{n} \mathbf{x}_i^T \mathbf{A}_i \mathbf{x}_i \leq \max_{i\leq n} \lambda_{\max}(\mathbf{A}_i) \sum_{i=1}^{n} \mathbf{x}_i^T \mathbf{x}_i \tag{3.3.10}$$

Lemma 3.3.3 *Assume that* (AH), (K') *and* (R') *hold. In addition, if:*

$$(C_1') : \lim_{r\to 0} \limsup_{n\to\infty} \delta_n \pi_n(r) \eta_n(r) = 0 \ a.s.$$
$$(C_2) : \lim_{r\to 0} \limsup_{n\to\infty} \delta_n \rho_n(r) = 0 \ a.s.$$

then

$$\lim_{r\to 0} \limsup_{n\to\infty} \alpha_n^{-1/2-\delta} \sup_{\beta\subset B_r} \|\boldsymbol{H}_n(\beta) - \boldsymbol{H}_n\| = 0 \ a.s.$$

Proof. Note that

$$\mathbf{H}_n(\beta) - \mathbf{H}_n = \mathbf{H}_n^{(1)}(\beta) + \mathbf{H}_n^{(2)}(\beta) + \mathbf{H}_n^{(3)}(\beta), \tag{3.3.11}$$

where

$$\begin{aligned}
\mathbf{H}_n^{(1)}(\beta) &= \sum_{i=1}^{n} \mathbf{X}_i^T [\mathbf{A}_i(\beta)^{1/2} - \mathbf{A}_i^{1/2}] \mathcal{R}_{i-1}(\beta)^{-1} \mathbf{A}_i(\beta)^{1/2} \mathbf{X}_i \\
\mathbf{H}_n^{(2)}(\beta) &= \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i^{1/2} [\mathcal{R}_{i-1}(\beta)^{-1} - \mathcal{R}_{i-1}^{-1}] \mathbf{A}_i(\beta)^{1/2} \mathbf{X}_i \\
\mathbf{H}_n^{(3)}(\beta) &= \sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{A}_i^{1/2} \mathcal{R}_{i-1}^{-1} [\mathbf{A}_i(\beta)^{1/2} - \mathbf{A}_i^{1/2}] \mathbf{X}_i.
\end{aligned}$$

Let $\mathbf{x}$ be a $p \times 1$ vector of norm 1 and $r > 0$ be arbitrary. All the inequalities obtained bellow are valid with probability 1.

We begin treating the term $\mathbf{H}_n^{(1)}(\beta)$. Using inequality (3.3.9) with:

$$\begin{aligned}
\mathbf{a}_i^T &= \mathbf{x}^T \mathbf{X}_i^T \mathbf{A}_i^{1/2} [\mathbf{A}_i^{-1/2} \mathbf{A}_i(\beta)^{1/2} - \mathbf{I}] \mathcal{R}_{i-1}(\beta)^{-1/2} \\
\mathbf{b}_i &= \mathcal{R}_{i-1}(\beta)^{-1/2} \mathbf{A}_i(\beta)^{1/2} \mathbf{X}_i \mathbf{x},
\end{aligned}$$

we have:

$$|\mathbf{x}^T \mathbf{H}_n^{(1)}(\beta) \mathbf{x}| = \left| \sum_{i=1}^{n} \mathbf{x}^T \mathbf{X}_i^T \mathbf{A}_i^{1/2} [\mathbf{A}_i^{-1/2} \mathbf{A}_i(\beta)^{1/2} - \mathbf{I}] \mathcal{R}_{i-1}(\beta)^{-1} \mathbf{A}_i(\beta)^{1/2} \mathbf{X}_i \mathbf{x} \right|$$

$$\leq \left\{ \sum_{i=1}^{n} \mathbf{x}^T \mathbf{X}_i^T \mathbf{A}_i^{1/2} [\mathbf{A}_i^{-1/2} \mathbf{A}_i(\beta)^{1/2} - \mathbf{I}] \mathcal{R}_{i-1}(\beta)^{-1} [\mathbf{A}_i^{-1/2} \mathbf{A}_i(\beta)^{1/2} - \mathbf{I}] \mathbf{A}_i^{1/2} \mathbf{X}_i \mathbf{x} \right\}^{1/2}$$

$$\cdot \left\{ \sum_{i=1}^{n} \mathbf{x}^T \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2} \mathcal{R}_{i-1}(\beta)^{-1} \mathbf{A}_i(\beta)^{1/2} \mathbf{X}_i \mathbf{x} \right\}^{1/2}$$

$$:= I_1(\beta, \mathbf{x})^{1/2} \cdot I_2(\beta, \mathbf{x})^{1/2}.$$

We obtain the following estimates:

$$\begin{aligned} I_1(\beta, \mathbf{x}) &\leq \max_{i \leq n} \lambda_{\max}[\mathcal{R}_{i-1}(\beta)^{-1}] \max_{i \leq n} \lambda_{\max}[\mathbf{A}_i^{-1/2} \mathbf{A}_i(\beta)^{1/2} - \mathbf{I}]^2 \mathbf{x}^T \mathbf{H}_n^{\mathrm{indep}} \mathbf{x} \\ &\leq C \pi_n(r) \eta_n(r)^2 \lambda_{\max}(\mathbf{H}_n^{\mathrm{indep}}), \qquad (3.3.12) \\ I_2(\beta, \mathbf{x}) &\leq \max_{i \leq n} \lambda_{\max}[\mathcal{R}_{i-1}(\beta)^{-1}] \cdot \max_{i \leq n} \lambda_{\max}[\mathbf{A}_i^{-1} \mathbf{A}_i(\beta)] \mathbf{x}^T \mathbf{H}_n^{\mathrm{indep}} \mathbf{x} \\ &\leq C \pi_n(r) \cdot \max_{i \leq n} \lambda_{\max}[\mathbf{A}_i^{-1} \mathbf{A}_i(\beta)] \lambda_{\max}(\mathbf{H}_n^{\mathrm{indep}}), \qquad (3.3.13) \end{aligned}$$

where for the previous inequalities we used (3.3.8).

Note that, by Lemma 3.2.1, there exist some constants $r_0 > 0$ and $n \geq n_0$ such that for any $r \in (0, r_0)$ and for any $n \geq n_0$ we have:

$$\max_{i \leq n} \lambda_{\max}[\mathbf{A}_i^{-1} \mathbf{A}_i(\beta)] = \max_{i \leq n, j \leq m} \frac{\mu'(\mathbf{x}_{ij}^T \beta)}{\mu'(\mathbf{x}_{ij}^T \beta_0)} \leq [\eta_n(r) + 1]^2 \leq C. \qquad (3.3.14)$$

Hence (3.3.13) becomes $I_2(\beta, \mathbf{x}) \leq C \pi_n(r) \lambda_{\max}(\mathbf{H}_n^{\mathrm{indep}})$, for any $r \in (0, r_0)$ and any $n \geq n_0$. It follows that $|\mathbf{x}^T \mathbf{H}_n^{(1)}(\beta) \mathbf{x}| \leq C \pi_n(r) \eta_n(r) \cdot \lambda_{\max}(\mathbf{H}_n^{\mathrm{indep}})$, for any $r \in (0, r_0)$ and any $n \geq n_0$ and hence:

$$\alpha_n^{-1/2-\delta} \sup_{\beta \in B_r} \sup_{\|\mathbf{x}\|=1} |\mathbf{x}^T \mathbf{H}_n^{(1)}(\beta) \mathbf{x}| \leq C \delta_n \pi_n(r) \eta_n(r), \ \forall r \in (0, r_0), \ \forall n \geq n_0.$$

From (C_1'), we obtain that $\lim_{r \to 0} \limsup_{n \to \infty} \alpha_n^{-1/2-\delta} \sup_{\beta \in B_r} \sup_{\|\mathbf{x}\|=1} |\mathbf{x}^T \mathbf{H}_n^{(1)}(\beta) \mathbf{x}| = 0$ a.s.

The term $\mathbf{H}_n^{(2)}(\beta)$ is treated similarly. Using inequality (3.3.9) with:

$$\begin{aligned} \mathbf{a}_i^T &= \mathbf{x}^T \mathbf{X}_i^T \mathbf{A}_i^{1/2} \mathcal{R}_{i-1}^{-1/2}, \\ \mathbf{b}_i &= [\mathcal{R}_{i-1}^{1/2} \mathcal{R}_{i-1}(\beta)^{-1} \mathcal{R}_{i-1}^{1/2} - \mathbf{I}] \mathcal{R}_{i-1}^{-1/2} \mathbf{A}_i(\beta)^{1/2} \mathbf{X}_i \mathbf{x}, \end{aligned}$$

we have:

$$
\begin{aligned}
&|\mathbf{x}^T\mathbf{H}_n^{(2)}(\beta)\mathbf{x}| = \\
&\left|\sum_{i=1}^{n}\mathbf{x}^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}\mathcal{R}_{i-1}^{-1/2}[\mathcal{R}_{i-1}^{1/2}\mathcal{R}_{i-1}(\beta)^{-1}\mathcal{R}_{i-1}^{1/2}-\mathbf{I}]\mathcal{R}_{i-1}^{-1/2}\mathbf{A}_i(\beta)^{1/2}\mathbf{X}_i\mathbf{x}\right| \\
&\leq \left\{\sum_{i=1}^{n}\mathbf{x}^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}\mathcal{R}_{i-1}^{-1}\mathbf{A}_i^{1/2}\mathbf{X}_i\mathbf{x}\right\}^{1/2} \\
&\cdot\{\sum_{i=1}^{n}\mathbf{x}^T\mathbf{X}_i^T\mathbf{A}_i(\beta)^{1/2}\mathcal{R}_{i-1}^{-1/2}[\mathcal{R}_{i-1}^{1/2}\mathcal{R}_{i-1}(\beta)^{-1}\mathcal{R}_{i-1}^{1/2}-\mathbf{I}]^2\mathcal{R}_{i-1}^{-1/2}\cdot \\
&\cdot\mathbf{A}_i(\beta)^{1/2}\mathbf{X}_i\mathbf{x}\}^{1/2} \\
&:= I_1'(\mathbf{x})^{1/2}\cdot I_2'(\beta,\mathbf{x})^{1/2}.
\end{aligned}
$$

By assumption (R') we obtain:

$$
\begin{aligned}
I_1'(\mathbf{x}) &\leq \max_{i\leq n}\lambda_{\max}(\mathcal{R}_{i-1}^{-1})\mathbf{x}^T\mathbf{H}_n^{\mathrm{indep}}\mathbf{x}\leq C\lambda_{\max}(\mathbf{H}_n^{\mathrm{indep}}) && (3.3.15)\\
I_2'(\beta,\mathbf{x}) &\leq \max_{i\leq n}\lambda_{\max}[\mathcal{R}_{i-1}^{1/2}\mathcal{R}_{i-1}(\beta)^{-1}\mathcal{R}_{i-1}^{1/2}-\mathbf{I}]^2\cdot\max_{i\leq n}\lambda_{\max}(\mathcal{R}_{i-1}^{-1})\cdot \\
&\cdot\ \max_{i\leq n}\lambda_{\max}[\mathbf{A}_i^{-1}\mathbf{A}_i(\beta)]\mathbf{x}^T\mathbf{H}_n^{\mathrm{indep}}\mathbf{x} \\
&\leq C\rho_n(r)^2\max_{i\leq n}\lambda_{\max}[\mathbf{A}_i^{-1}\mathbf{A}_i(\beta)]\lambda_{\max}(\mathbf{H}_n^{\mathrm{indep}}). && (3.3.16)
\end{aligned}
$$

As in the proof of (3.3.14), there exist some constants $r_0>0$ and $n\geq n_0$ such that (3.3.16) can be written as $I_2'(\beta,\mathbf{x})\leq C\rho_n(r)^2\lambda_{\max}(\mathbf{H}_n^{\mathrm{indep}})$, for any $r\in(0,r_0)$ and for any $n\geq n_0$. Using (3.3.15) we obtain $|\mathbf{x}^T\mathbf{H}_n^{(2)}(\beta)\mathbf{x}|\leq C\rho_n(r)\cdot\lambda_{\max}(\mathbf{H}_n^{\mathrm{indep}})$, for any $r\in(0,r_0)$ and for any $n\geq n_0$, and:

$$\alpha_n^{-1/2-\delta}\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1}|\mathbf{x}^T\mathbf{H}_n^{(2)}(\beta)\mathbf{x}|\leq C\delta_n\rho_n(r),\ \forall r\in(0,r_0),\ \forall n\geq n_0.$$

Using (C_2'), it follows that $\lim_{r\to 0}\limsup_{n\to\infty}\alpha_n^{-1/2-\delta}\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1}|\mathbf{x}^T\mathbf{H}_n^{(2)}(\beta)\mathbf{x}|=0$ a.s.

To treat the term $\mathbf{H}_n^{(3)}(\beta)$ we are using inequality (3.3.9) with:

$$
\begin{aligned}
\mathbf{a}_i^T &= \mathbf{x}^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}\mathcal{R}_{i-1}^{-1}[\mathbf{A}_i(\beta)^{1/2}\mathbf{A}_i^{-1/2}-\mathbf{I}],\\
\mathbf{b}_i &= \mathbf{A}_i^{1/2}\mathbf{X}_i\mathbf{x}.
\end{aligned}
$$

We have:

$$
\begin{aligned}
|\mathbf{x}^T\mathbf{H}_n^{(3)}(\beta)\mathbf{x}| &= \left|\sum_{i=1}^{n} \mathbf{x}^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}\mathcal{R}_{i-1}^{-1}[\mathbf{A}_i(\beta)^{1/2}\mathbf{A}_i^{-1/2}-\mathbf{I}]\mathbf{A}_i^{1/2}\mathbf{X}_i\mathbf{x}\right| \\
&\le \{\sum_{i=1}^{n} \mathbf{x}^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}\mathcal{R}_{i-1}^{-1}[\mathbf{A}_i(\beta)^{1/2}\mathbf{A}_i^{-1/2}-\mathbf{I}][\mathbf{A}_i(\beta)^{1/2}\mathbf{A}_i^{-1/2}-\mathbf{I}]\cdot \\
&\cdot\mathcal{R}_{i-1}^{-1}\mathbf{A}_i^{1/2}\mathbf{X}_i\mathbf{x}\}^{1/2}\cdot\left\{\sum_{i=1}^{n}\mathbf{x}^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}\mathbf{A}_i^{1/2}\mathbf{X}_i\mathbf{x}\right\}^{1/2} \\
&:= I_3(\beta,\mathbf{x})^{1/2}\cdot(\mathbf{x}^T\mathbf{H}_n^{\text{indep}}\mathbf{x})^{1/2}.
\end{aligned}
$$

Using (R'), it follows that:

$$
\begin{aligned}
I_3(\beta,\mathbf{x}) &\le \max_{i\le n}\lambda_{\max}[\mathbf{A}_i(\beta)^{1/2}\mathbf{A}_i^{-1/2}-\mathbf{I}]^2\max_{i\le n}\lambda_{\max}[\mathcal{R}_{i-1}^{-2}]\mathbf{x}^T\mathbf{H}_n^{\text{indep}}\mathbf{x} \\
&\le C\eta_n(r)^2\lambda_{\max}(\mathbf{H}_n^{\text{indep}}),
\end{aligned}
$$

and hence $|\mathbf{x}^T\mathbf{H}_n^{(3)}(\beta)\mathbf{x}|\le C\eta_n(r)\lambda_{\max}(\mathbf{H}_n^{\text{indep}})$ for any $n\ge 1$. It follows that:

$$
\alpha_n^{-1/2-\delta}\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1}|\mathbf{x}^T\mathbf{H}_n^{(3)}(\beta)\mathbf{x}|\le C\delta_n\eta_n(r)\le C\delta_n\pi_n(r)\eta_n(r),\ \forall r>0,\ \forall n\ge 1.
$$

since $\pi_n(r)\ge 1$.

Using (C_1'), it follows that $\lim_{r\to 0}\limsup_{n\to\infty}\alpha_n^{-1/2-\delta}\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1}|\mathbf{x}^T\mathbf{H}_n^{(3)}(\beta)\mathbf{x}|=0$ a.s. This concludes the proof. □

Let:

$$
\begin{aligned}
q_n^{(l)}(r) &:= \sup_{\beta\in B_r}\max_{i\le n}\lambda_{\max}[\mathbf{Q}_{i-1}^{(l)}(\beta)],\ l=1,\dots p \\
q_n(r) &:= \max_{l\le p} q_n^{(l)}(r).
\end{aligned}
$$

Lemma 3.3.4 *Assume that (AH), (K) and (R') hold. In addition if:*

$$
\begin{aligned}
(C_1) &: \lim_{r\to 0}\limsup_{n\to\infty} r\delta_n\pi_n(r)a_n^{1/2}=0\ a.s. \\
(C_3) &: \lim_{r\to 0}\limsup_{n\to\infty} r\delta_n\pi_n^2(r)q_n(r)=0\ a.s.,
\end{aligned}
$$

then

$$
\lim_{r\to 0}\limsup_{n\to\infty}\alpha_n^{-1/2-\delta}\sup_{\beta\in B_r}\|\boldsymbol{B}_n(\beta)\|=0\ a.s.
$$

Remark 3.3.5 Since $\eta_n(r) \leq Cra_n^{1/2}$ by Lemma 3.2.1, condition (C_1) is stronger than condition (C_1') encountered in Lemma 3.3.3.

Proof of Lemma 3.3.4. Recall that $\mathbf{B}_n(\beta) = \mathbf{B}_n^{(1)}(\beta) + \mathbf{B}_n^{(2)}(\beta) + \mathbf{B}_n^{(3)}(\beta)$, where $\mathbf{B}_n^{(1)}(\beta)$, $\mathbf{B}_n^{(2)}(\beta)$ and $\mathbf{B}_n^{(3)}(\beta)$ are given by (3.3.4).

We use the following property:

$$\mathrm{diag}\{\mathbf{v}\}\mathbf{D}\mathbf{u} = \mathbf{D}\mathrm{diag}\{\mathbf{u}\}\mathbf{v}, \tag{3.3.17}$$

where D is any diagonal matrix and $\mathbf{u}$, $\mathbf{v}$ are any vectors.

Let $\mathbf{x}$ be a p-dimensional vector with $\mathbf{x}$ with $\|\mathbf{x}\| = 1$. We begin treating the term which involves $\mathbf{B}_n^{(1)}(\beta)$. Using (3.3.17), and relation (3.3.9) with:

$$\begin{aligned}
\mathbf{a}_i^T &= \mathbf{x}^T\mathbf{X}_i^T\mathbf{G}_i^{(1)}(\beta)\mathrm{diag}(\mathbf{X}_i\mathbf{x})\mathcal{R}_{i-1}(\beta)^{\ 1/2} \\
\mathbf{b}_i &= \mathcal{R}_{i-1}(\beta)^{-1/2}\mathbf{A}_i(\beta)^{-1/2}[\mu_i - \mu_i(\beta)].
\end{aligned}$$

we obtain:

$$\begin{aligned}
&|\mathbf{x}^T\mathbf{B}_n^{(1)}(\beta)\mathbf{x}| = |\sum_{i=1}^{n}\mathbf{x}^T\mathbf{X}_i^T\mathbf{G}_i^{(1)}(\beta)\mathrm{diag}(\mathbf{X}_i\mathbf{x})\mathcal{R}_{i-1}(\beta)^{-1}\mathbf{A}_i(\beta)^{-1/2}[\mu_i - \mu_i(\beta)]| \\
&\leq \left[\sum_{i=1}^{n}\mathbf{x}^T\mathbf{X}_i^T\mathbf{G}_i^{(1)}(\beta)\mathrm{diag}(\mathbf{X}_i\mathbf{x})\mathcal{R}_{i-1}(\beta)^{-1}\mathrm{diag}(\mathbf{X}_i\mathbf{x})\mathbf{G}_i^{(1)}(\beta)\mathbf{X}_i^T\mathbf{x}^T\right]^{1/2} \cdot \\
&\cdot\left[\sum_{i=1}^{n}[\mu_i - \mu_i(\beta)]^T\mathbf{A}_i(\beta)^{-1/2}\mathcal{R}_{i-1}(\beta)^{-1}\mathbf{A}_i(\beta)^{-1/2}[\mu_i - \mu_i(\beta)]\right]^{1/2} \\
&:= I_1(\beta,\mathbf{x})^{1/2}\cdot I_2(\beta)^{1/2}
\end{aligned} \tag{3.3.18}$$

The first term of the product is estimated as follows:

$$\begin{aligned}
I_1(\beta,\mathbf{x}) &\leq \max_{i\leq n}\lambda_{\max}[\mathcal{R}_{i-1}(\beta)^{-1}]\cdot\max_{i\leq n}\lambda_{\max}[\mathrm{diag}^2(\mathbf{X}_i\mathbf{x})] \\
&\cdot \sum_{i=1}^{n}\mathbf{x}_i^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}\mathbf{A}_i^{-1/2}\mathbf{G}_i^{(1)}(\beta)^2\mathbf{A}_i^{-1/2}\mathbf{A}_i^{1/2}\mathbf{X}_i\mathbf{x}_i \\
&\leq \max_{i\leq n}\lambda_{\max}[\mathcal{R}_{i-1}(\beta)^{-1}]\cdot\max_{i\leq n}\lambda_{\max}[\mathrm{diag}^2(\mathbf{X}_i\mathbf{x})]
\end{aligned}$$

$$
\begin{aligned}
& \max_{i\le n} \lambda^2_{\max}[\mathbf{A}_i^{-1/2}\mathbf{G}_i^{(1)}(\beta)]\mathbf{x}^T\mathbf{H}_n^{\text{indep}}\mathbf{x} \\
\le\ & C\pi_n(r)\max_{i\le n}\lambda_{\max}[\text{diag}^2(\mathbf{X}_i\mathbf{x})]\cdot\max_{i\le n}\lambda^2_{\max}[\mathbf{A}_i^{-1/2}\mathbf{G}_i^{(1)}(\beta)] \\
\cdot\ & \lambda_{\max}(\mathbf{H}_n^{\text{indep}}).
\end{aligned}
$$

where we used (3.3.8) for the last inequality.

We also have the following evaluations, for each $j\le m$ and $i\le n$:

$$
\begin{aligned}
|(\mathbf{X}_i\mathbf{x})_j| &= |\mathbf{x}_{ij}^T\mathbf{x}| = |\mathbf{x}_{ij}^T(\mathbf{H}_n^{\text{indep}})^{-1/2}(\mathbf{H}_n^{\text{indep}})^{1/2}\mathbf{x}| \\
&\le \|\mathbf{x}_{ij}^T(\mathbf{H}_n^{\text{indep}})^{-1/2}\|\,\|(\mathbf{H}_n^{\text{indep}})^{1/2}\mathbf{x}\| = |\mathbf{x}_{ij}^T(\mathbf{H}_n^{\text{indep}})^{-1}\mathbf{x}_{ij}|^{1/2}|\mathbf{x}^T\mathbf{H}_n^{\text{indep}}\mathbf{x}|^{1/2} \\
&\le (\gamma_n^{(0),\text{indep}})^{1/2}[\lambda_{\max}(\mathbf{H}_n^{\text{indep}})]^{1/2} = a_n^{1/2}.
\end{aligned}
$$

Hence, $|(\mathbf{X}_i\mathbf{x})_j|^2\le a_n$, $\forall j\le m$ and $\lambda_{\max}[\text{diag}^2(\mathbf{X}_i\mathbf{x})]=\max_{j\le m}|(\mathbf{X}_i\mathbf{x})_j|^2\le a_n$, $\forall i\le n$. We conclude that:

$$
\max_{i\le n}\lambda_{\max}[\text{diag}^2(\mathbf{X}_i\mathbf{x})]\le a_n. \tag{3.3.19}
$$

Note that using assumption (AH) and Lemma 3.2.1, there exist some numbers $r_1>0$ and $n_1>0$ such that for any $r\in(0,r_1)$, and $n\ge n_1$, we have:

$$
\begin{aligned}
[\mathbf{A}_i^{-1/2}\mathbf{G}_i^{(1)}(\beta)]_{jj} &= \frac{1}{2}\mu'(\mathbf{x}_{ij}^T\beta_0)^{-1/2}\frac{\mu''(\mathbf{x}_{ij}^T\beta)}{\mu'(\mathbf{x}_{ij}^T\beta)^{1/2}} = \frac{1}{2}\frac{\mu''(\mathbf{x}_{ij}^T\beta)}{\mu'(\mathbf{x}_{ij}^T\beta)}\cdot\left[\frac{\mu'(\mathbf{x}_{ij}^T\beta)}{\mu'(\mathbf{x}_{ij}^T\beta_0)}\right]^{1/2} \\
&\le C[\eta_n(r)+1]\le C,
\end{aligned}
$$

for any $i\le n$, $j\le m$ and $\beta\in B_r$.

Hence, the first term of the product in (3.3.18) is estimated as follows:

$$
I_1(\beta,\mathbf{x})\le C\pi_n(r)a_n\lambda_{\max}(\mathbf{H}_n^{\text{indep}}),\ \forall\beta\in B_r,\ \forall r\in(0,r_1),\ \forall n\ge n_1. \tag{3.3.20}
$$

We begin now to treat the second term of the product in (3.3.18). By Taylor's expansion, there exists $\overline{\beta}_i$ between β and β_0 such that $\mu_i(\beta)-\mu_i=\mathbf{A}_i(\overline{\beta}_i)\mathbf{X}_i(\beta-\beta_0)$. Hence, the term $I_2(\beta)$ can be evaluated as follows:

$$
I_2(\beta)\le\max_{i\le n}\lambda_{\max}[\mathcal{R}_{i-1}(\beta)^{-1}]\cdot[\sum_{i=1}^n(\beta-\beta_0)^T\mathbf{X}_i^T\mathbf{A}_i(\beta)^{1/2}\mathbf{A}_i(\beta)^{-1/2}\mathbf{A}_i(\overline{\beta}_i)\cdot
$$

$$
\begin{aligned}
&\cdot\ \mathbf{A}_i(\beta)^{-1/2}\mathbf{A}_i(\beta)^{-1/2}\mathbf{A}_i(\overline{\beta}_i)\mathbf{A}_i(\beta)^{-1/2}\mathbf{A}_i(\beta)^{1/2}\mathbf{X}_i(\beta-\beta_0)] \\
&\leq \max_{i\leq n}\lambda_{\max}[\mathcal{R}_{i-1}(\beta)^{-1}]\max_{i\leq n}\lambda_{\max}[\mathbf{A}_i(\beta)^{-1/2}\mathbf{A}_i(\overline{\beta}_i)\mathbf{A}_i(\beta)^{-1/2}]\cdot \\
&\cdot\ \sum_{i=1}^{n}(\beta-\beta_0)^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}\mathbf{A}_i^{-1/2}\mathbf{A}_i(\beta)^{1/2}\mathbf{A}_i(\beta)^{1/2}\mathbf{A}_i^{-1/2}\mathbf{A}_i^{1/2}\mathbf{X}_i(\beta-\beta_0) \\
&\leq \max_{i\leq n}\lambda_{\max}[\mathcal{R}_{i-1}(\beta)^{-1}]\cdot\max_{i\leq n}\lambda_{\max}[\mathbf{A}_i(\beta)^{-1/2}\mathbf{A}_i(\overline{\beta}_i)\mathbf{A}_i(\beta)^{-1/2}]\cdot \\
&\cdot\ \max_{i\leq n}\lambda_{\max}[\mathbf{A}_i^{-1/2}\mathbf{A}_i(\beta)\mathbf{A}_i^{-1/2}]\lambda_{\max}(\mathbf{H}_n^{\text{indep}})\cdot(\beta-\beta_0)^T(\beta-\beta_0) \\
&\leq C\pi_n(r)\cdot\max_{i\leq n}\lambda_{\max}[\mathbf{A}_i(\beta)^{-1/2}\mathbf{A}_i(\overline{\beta}_i)\mathbf{A}_i(\beta)^{-1/2}]\max_{i\leq n}\lambda_{\max}[\mathbf{A}_i^{-1/2}\mathbf{A}_i(\beta)\mathbf{A}_i^{-1/2}]\cdot \\
&\cdot\ \lambda_{\max}(\mathbf{H}_n^{\text{indep}})r^2,
\end{aligned}
$$

were we used (3.3.8) for the last inequality.

Note that, by Lemma 3.2.1, for any $0 < r \leq r_1$ and $n \geq n_1$, and for any $\overline{\beta}_i \in B_r$, we have:

$$
[\mathbf{A}_i(\beta)^{-1/2}\mathbf{A}_i(\overline{\beta}_i)\mathbf{A}_i(\beta)^{-1/2}]_{jj} = \frac{\mu'(\mathbf{x}_{ij}^T\overline{\beta}_i)}{\mu'(\mathbf{x}_{ij}^T\beta)} \leq [\eta_n(r)+1]^2 \leq C,
$$

for any $j \leq m$ and for any $i \leq n$. Hence, $\lambda_{\max}[\mathbf{A}_i(\beta)^{-1/2}\mathbf{A}_i(\overline{\beta}_i)\mathbf{A}_i(\beta)^{-1/2}] = \max_{j\leq m}[\mathbf{A}_i(\beta)^{-1/2}\mathbf{A}_i(\overline{\beta}_i)\mathbf{A}_i(\beta)^{-1/2}]_{jj} \leq C$, for all $i \leq n$. We obtain that:

$$
\max_{i\leq n}\lambda_{\max}[\mathbf{A}_i(\beta)^{-1/2}\mathbf{A}_i(\overline{\beta}_i)\mathbf{A}_i(\beta)^{-1/2}] \leq C,\ \forall r\in(0,r_1),\ \forall n\geq n_1,\ \forall\beta\in B_r.
$$

Similarly, $\max_{i\leq n, j\leq m}\frac{\mu_{ij}(\mathbf{x}_{ij}^T\beta)}{\mu_{ij}(\mathbf{x}_{ij}^T\beta_0)} \leq [\eta_n(r)+1]^2$ and by Lemma 3.2.1 it follows that, for any $r\in(0,r_1)$, for any $n\geq n_1$ and for any $\beta\in B_r$, we have

$$
\max_{i\leq n}\lambda_{\max}[\mathbf{A}_i^{-1/2}\mathbf{A}_i(\beta)\mathbf{A}_i^{-1/2}] = \max_{i\leq n,j\leq m}\frac{\mu_{ij}(\mathbf{x}_{ij}^T\beta)}{\mu_{ij}(\mathbf{x}_{ij}^T\beta_0)} \leq C. \tag{3.3.21}
$$

Therefore, the second term of the product in (3.3.18) is estimated as follows:

$$
I_2(\beta) \leq Cr^2\pi_n(r)\lambda_{\max}(\mathbf{H}_n^{\text{indep}}),\ \forall\beta\in B_r,\ \forall r\in(0,r_1),\ \forall n\geq n_1 \tag{3.3.22}
$$

From (3.3.18), (3.3.20) and (3.3.22), we conclude that:

$$
|\mathbf{x}^T\mathbf{B}_n^{(1)}(\beta)\mathbf{x}| \leq Cr\pi_n(r)a_n^{1/2}\lambda_{\max}(\mathbf{H}_n^{\text{indep}}),
$$

for any $\beta \in B_r$, $r \in (0, r_1), \forall n \geq n_1$. Multiplying by $\alpha_n^{-1/2-\delta}$ we get:

$$\alpha_n^{-1/2-\delta} \sup_{\beta \in B_r} \sup_{\|\mathbf{x}\|=1} |\mathbf{x}^T \mathbf{B}_n^{(1)}(\beta)\mathbf{x}| \leq Cr\delta_n \pi_n(r) a_n^{1/2},$$

for any $r \in (0, r_1)$ nd for any $n \geq n_1$. Using condition (C_1), it follows that:

$$\lim_{r \to 0} \limsup_{n \to \infty} \alpha_n^{-1/2-\delta} \sup_{\beta \in B_r} \sup_{\|\mathbf{x}\|=1} |\mathbf{x}^T \mathbf{B}_n^{(1)}(\beta)\mathbf{x}| = 0 \text{ a.s.} \tag{3.3.23}$$

We begin now to treat the term which involves $\mathbf{B}_n^{(2)}(\beta)$. Using (3.3.17) and relation (3.3.9) with:

$$\begin{aligned} \mathbf{a}_i^T &= \mathbf{x}^T \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2} \mathcal{R}_{i-1}(\beta)^{-1/2} \\ \mathbf{b}_i &= \mathcal{R}_{i-1}(\beta)^{-1/2} \text{diag}(\mathbf{X}_i \mathbf{x}) \mathbf{G}_i^{(2)}(\beta)[\mu_i - \mu_i(\beta)], \end{aligned}$$

we obtain:

$$\begin{aligned} &|\mathbf{x}^T \mathbf{B}_n^{(2)}(\beta)\mathbf{x}| = |\sum_{i=1}^n \mathbf{x}^T \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2} \mathcal{R}_{i-1}(\beta)^{-1} \text{diag}\{\mu_i - \mu_i(\beta)\} \mathbf{G}_i^{(2)}(\beta) \mathbf{X}_i \mathbf{x}| \\ &= |\sum_{i=1}^n \mathbf{x}^T \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2} \mathcal{R}_{i-1}(\beta)^{-1} \text{diag}(\mathbf{X}_i \mathbf{x}) \mathbf{G}_i^{(2)}(\beta)[\mu_i - \mu_i(\beta)]| \\ &\leq \left[\sum_{i=1}^n \mathbf{x}^T \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2} \mathcal{R}_{i-1}(\beta)^{-1} \mathbf{A}_i(\beta)^{1/2} \mathbf{X}_i \mathbf{x} \right]^{1/2} \cdot \\ &\cdot \left[\sum_{i=1}^n [\mu_i - \mu_i(\beta)]^T \mathbf{G}_i^{(2)}(\beta) \text{diag}(\mathbf{X}_i \mathbf{x}) \mathcal{R}_{i-1}(\beta)^{-1} \text{diag}(\mathbf{X}_i \mathbf{x}) \mathbf{G}_i^{(2)}(\beta)[\mu_i - \mu_i(\beta)] \right]^{1/2} \\ &:= I_1'(\beta, \mathbf{x})^{1/2} \cdot I_2'(\beta, \mathbf{x})^{1/2}. \end{aligned} \tag{3.3.24}$$

The first term of the product in (3.3.24) is estimated using (3.3.8) as follows:

$$\begin{aligned} I_1'(\beta, \mathbf{x}) &\leq \max_{i \leq n} \lambda_{\max}[\mathcal{R}_{i-1}(\beta)^{-1}] \max_{i \leq n} \lambda_{\max}[\mathbf{A}_i^{-1/2} \mathbf{A}_i(\beta) \mathbf{A}_i^{-1/2}] \cdot \mathbf{x}^T \mathbf{H}_n^{\text{indep}} \mathbf{x} \\ &\leq C\pi_n(r) \lambda_{\max}[\mathbf{A}_i^{-1/2} \mathbf{A}_i(\beta) \mathbf{A}_i^{-1/2}] \lambda_{\max}(\mathbf{H}_n^{\text{indep}}), \end{aligned}$$

where we used (3.3.8) for the last inequality. By (3.3.21), $\lambda_{\max}[\mathbf{A}_i^{-1/2} \mathbf{A}_i(\beta) \mathbf{A}_i^{-1/2}] \leq C$ for any $\beta \in B_r$, any $r \in (0, r_1)$ and any $n \geq n_1$. Hence:

$$I_1'(\beta, \mathbf{x}) \leq C\pi_n(r) \lambda_{\max}(\mathbf{H}_n^{\text{indep}}), \ \forall \beta \in B_r, \ \forall r \in (0, r_1), \ \forall n \geq n_1. \tag{3.3.25}$$

The second term in the product (3.3.24) is estimated as:

$$\begin{aligned}
I_2'(\beta,\mathbf{x}) &\le \max_{i\le n}\lambda_{\max}[\mathcal{R}_{i-1}(\beta)^{-1}]\max_{i\le n}\lambda_{\max}[\mathrm{diag}^2(\mathbf{X}_i\mathbf{x})] \\
&\cdot \sum_{i=1}^n[\mu_i-\mu_i(\beta)]^T\mathbf{G}_i^{(2)}(\beta)^2[\mu_i-\mu_i(\beta)] \\
&\le C\pi_n(r)a_n\sum_{i=1}^n[\mu_i-\mu_i(\beta)]^T\mathbf{G}_i^{(2)}(\beta)^2[\mu_i-\mu_i(\beta)],
\end{aligned}$$

where we used (3.3.8) and (3.3.19) for the first inequality. By using Taylor's expansion again, there exists $\overline{\beta}_i$ between β and β_0 such that $\mu_i(\beta)-\mu_i=\mathbf{A}_i(\overline{\beta}_i)\mathbf{X}_i(\beta-\beta_0)$. We obtain:

$$\begin{aligned}
I_2'(\beta,\mathbf{x}) &\le C\pi_n(r)a_n\sum_{i=1}^n(\beta-\beta_0)^T\mathbf{X}_i^T\mathbf{A}_i(\overline{\beta}_i)\mathbf{G}_i^{(2)}(\beta)^2\mathbf{A}_i(\overline{\beta}_i)\mathbf{X}_i(\beta-\beta_0) \\
&\le C\pi_n(r)a_n\max_{i\le n}\lambda_{\max}[\mathbf{A}_i(\overline{\beta}_i)^{1/2}\mathbf{G}_i^{(2)}(\beta)^2\mathbf{A}_i(\overline{\beta}_i)^{1/2}] \\
&\cdot \sum_{i=1}^n(\beta-\beta_0)^T\mathbf{X}_i^T\mathbf{A}_i(\overline{\beta}_i)\mathbf{X}_i(\beta-\beta_0).
\end{aligned}$$

Note that using condition (AH) and Lemma 3.2.1, for any $r\in(0,r_1)$, any $\beta\in B_r$, any $n\ge n_1$, $i\le n$, $j\le m$, we have:

$$\begin{aligned}
[\mathbf{A}_i(\overline{\beta}_i)^{1/2}\mathbf{G}_i^{(2)}(\beta)^2\mathbf{A}_i(\overline{\beta}_i)^{1/2}]_{jj} &= \left[-\frac{1}{2}\frac{\mu''(\mathbf{x}_{ij}^T\beta)}{\mu'(\mathbf{x}_{ij}^T\beta)^{3/2}}\right]^2\cdot\mu'(\mathbf{x}_{ij}^T\overline{\beta}_{ij}) \\
&= \left[\frac{1}{2}\frac{\mu''(\mathbf{x}_{ij}^T\beta)}{\mu'(\mathbf{x}_{ij}^T\beta)}\right]^2\frac{\mu'(\mathbf{x}_{ij}^T\overline{\beta}_{ij})}{\mu'(\mathbf{x}_{ij}^T\beta)}\le C[\eta_n(r)+1]^2\le C.
\end{aligned}$$

Hence for any $r\in(0,r_1)$, any $\beta\in B_r$, any $n\ge n_1$:

$$\begin{aligned}
I_2'(\beta,\mathbf{x}) &\le C\pi_n(r)a_n\max_{i\le n,}[\mathbf{A}_i^{-1/2}\mathbf{A}_i(\overline{\beta}_i)\mathbf{A}_i^{-1/2}]\sum_{i=1}^n(\beta-\beta_0)^T\mathbf{X}_i^T\mathbf{A}_i\mathbf{X}_i(\beta-\beta_0) \\
&\le C\pi_n(r)a_n\max_{i\le n}\lambda_{\max}[\mathbf{A}_i^{-1/2}\mathbf{A}_i(\overline{\beta}_i)\mathbf{A}_i^{-1/2}]\lambda_{\max}(\mathbf{H}_n^{\mathrm{indep}})(\beta-\beta_0)^T(\beta-\beta_0).
\end{aligned}$$

Using Lemma 3.2.1, we obtain that for any $r\in(0,r_1)$ and for any $n\ge n_1$ we have:

$$\max_{i\le n}\lambda_{\max}[\mathbf{A}_i^{-1/2}\mathbf{A}_i(\overline{\beta}_i)\mathbf{A}_i^{-1/2}]=\max_{i\le n,j\le m}\frac{\mu'(\mathbf{x}_{ij}^T\overline{\beta}_i)}{\mu'(\mathbf{x}_{ij}^T\beta_0)}\le[\eta_n(r)+1]^2\le C.$$

It follows that:

$$I_2'(\beta, \mathbf{x}) \leq C\pi_n(r)a_n\lambda_{\max}(\mathbf{H}_n^{\text{indep}})r^2, \forall\beta \in B_r, \ \forall r \in (0, r_1), \ \forall n \geq n_1.$$

From (3.3.24), (3.3.25) and (3.3.26) we conclude that:

$$|\mathbf{x}^T\mathbf{B}_n^{(2)}(\beta)\mathbf{x}| \leq Cr\pi_n(r)a_n^{1/2} \cdot \lambda_{\max}(\mathbf{H}_n^{\text{indep}}),$$

for any $\beta \in B_r$, $r \in (0, r_1)$, $n \geq n_1$ Multiplying this last inequality by $\alpha_n^{-1/2-\delta}$, it follows that for any $n \geq n_1$, $r \in (0, r_1)$:

$$\alpha_n^{-1/2-\delta} \sup_{\beta\in B_r} \sup_{\|\mathbf{x}\|=1} |\mathbf{x}^T\mathbf{B}_n^{(2)}(\beta)\mathbf{x}| \ \leq \ Cr\delta_n\pi_n(r)a_n^{1/2}.$$

Using condition (C_1), we conclude that:

$$\lim_{r\to 0}\limsup_{n\to\infty} \alpha_n^{-1/2-\delta} \sup_{\beta\in B_r} \sup_{\|\mathbf{x}\|=1} |\mathbf{x}^T\mathbf{B}_n^{(2)}(\beta)\mathbf{x}| = 0 \text{ a.s.} \tag{3.3.26}$$

We are now treating the term which involves $\mathbf{B}_n^{(3)}(\beta)$. Recall that $\mathbf{b}_n^{(l)}(\beta)$ the l-th column of $\mathbf{B}_n^{(3)}(\beta)$, for $l = 1, \ldots p$. Using Lemma A.1.16, we have: for any $\mathbf{x}$ vector, $\|\mathbf{x}\| = 1$,

$$|\mathbf{x}^T\mathbf{B}_n^{(3)}(\beta)\mathbf{x}| \leq \left\{\sum_{l=1}^{p} |\mathbf{x}^T\mathbf{b}_n^{(l)}(\beta)|^2\right\}^{1/2}. \tag{3.3.27}$$

Hence, if we find a common upper bound for all $|\mathbf{x}^T\mathbf{b}_n^{(l)}(\beta)|$, with $l = 1 \ldots p$, this is also an upper bound for $\mathbf{x}^T\mathbf{B}_n^{(3)}(\beta)\mathbf{x}$.

Applying Theorem A.2.4, we obtain:

$$\mathbf{T}_{i-1}^{(l)}(\beta) \ = \ \frac{\partial\mathcal{R}_{i-1}(\beta)^{-1}}{\partial\beta_l^T} = -\mathcal{R}_{i-1}(\beta)^{-1}\mathbf{Q}_{i-1}^{(l)}(\beta)\mathcal{R}_{i-1}(\beta)^{-1}$$

where $\mathbf{Q}_{i-1}^{(l)}(\beta) = \dfrac{\partial\mathcal{R}_{i-1}(\beta)}{\partial\beta_l^T}$. Notice that $\mathbf{Q}_{i-1}^{(l)}(\beta)$ is a symmetric matrix.

Let $l \in \{1, \ldots p\}$ be fixed. To estimate the term $|\mathbf{x}^T\mathbf{b}_n^{(l)}(\beta)|$, we proceed as for $\mathbf{x}^T\mathbf{B}_n^{(1)}(\beta)\mathbf{x}$ and $\mathbf{x}^T\mathbf{B}_n^{(2)}(\beta)\mathbf{x}$. Using (3.3.9) with:

$$\mathbf{a}_i^T \ = \ \sum_{i=1}^{n} \mathbf{X}_i^T\mathbf{A}_i(\beta)^{1/2}\mathcal{R}_{i-1}(\beta)^{-1}\mathbf{Q}_{i-1}^{(l)}(\beta)\mathcal{R}_{i-1}(\beta)^{-1/2}$$

$$\mathbf{b}_i = \sum_{i-1}^{n} \mathcal{R}_{i-1}(\beta)^{-1/2}\mathbf{A}_i(\beta)^{-1/2}[\mu_i - \mu_i(\beta)].$$

we obtain:

$$\begin{aligned}
|\mathbf{x}^T\mathbf{b}_n^{(l)}(\beta)| &\leq [\sum_{i-1}^{n} \mathbf{x}^T\mathbf{X}_i^T\mathbf{A}_i(\beta)^{1/2}\mathcal{R}_{i-1}(\beta) \ {}^1\mathbf{Q}_{i-1}^{(l)}(\beta)\mathcal{R}_{i-1}(\beta) \ {}^1\mathbf{Q}_{i-1}^{(l)}(\beta)\mathcal{R}_{i-1}(\beta)^{-1} \cdot \\
&\cdot \ \mathbf{A}_i(\beta)^{1/2}\mathbf{X}_i\mathbf{x}]^{1/2} \cdot \\
&\cdot \ \left[\sum_{i-1}^{n}(\mu_i - \mu_i(\beta))^T\mathbf{A}_i(\beta)^{-1/2}\mathcal{R}_{i-1}(\beta)^{-1}\mathbf{A}_i(\beta)^{-1/2}(\mu_i - \mu_i(\beta))\right]^{1/2} \\
&:= I_1''(l,\beta,\mathbf{x})^{1/2} \cdot I_2(\beta)^{1/2}. \qquad (3.3.28)
\end{aligned}$$

Using (3.3.10), the first term in (3.3.28) can be estimated as:

$$\begin{aligned}
I_1''(l.\beta,\mathbf{x}) &\leq \max_{i\leq n}\lambda_{\max}[\mathcal{R}_{i-1}(\beta)^{-1}]\sum_{i-1}^{n}\mathbf{x}^T\mathbf{X}_i^T\mathbf{A}_i(\beta)^{1/2}\mathcal{R}_{i-1}(\beta)^{-1}\mathbf{Q}_{i-1}^{(l)}(\beta)^2\mathcal{R}_{i-1}(\beta)^{-1} \\
&\cdot \ \mathbf{A}_i(\beta)^{1/2}\mathbf{X}_i\mathbf{x} \\
&\leq \max_{i\leq n}\lambda_{\max}[\mathcal{R}_{i-1}(\beta)^{-1}] \cdot \max_{i\leq n}\lambda_{\max}[\mathbf{Q}_{i-1}^{(l)}(\beta)^2] \cdot \\
&\cdot \ \sum_{i-1}^{n}\mathbf{x}^T\mathbf{X}_i^T\mathbf{A}_i(\beta)^{1/2}[\mathcal{R}_{i-1}(\beta)^{-1}]^2\mathbf{A}_i(\beta)^{1/2}\mathbf{X}_i\mathbf{x} \\
&\leq \max_{i\leq n}\lambda_{\max}[\mathcal{R}_{i-1}(\beta)^{-1}] \cdot \max_{i\leq n}\lambda_{\max}[\mathbf{Q}_{i-1}^{(l)}(\beta)^2] \cdot \max_{i\leq n}\lambda_{\max}^2[\mathcal{R}_{i-1}(\beta)^{-1}] \cdot \\
&\cdot \ \max_{i\leq n}[\mathbf{A}_i^{-1/2}\mathbf{A}_i(\beta)\mathbf{A}_i^{-1/2}]\sum_{i\ 1}^{n}\mathbf{x}^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}\mathbf{A}_i^{1/2}\mathbf{X}_i\mathbf{x} \\
&\leq \max_{i\leq n}\lambda_{\max}^3[\mathcal{R}_{i-1}(\beta)^{-1}] \cdot \max_{i\leq n}\lambda_{\max}[\mathbf{Q}_{i-1}^{(l)}(\beta)^2] \cdot \max_{i\leq n}[\mathbf{A}_i^{-1/2}\mathbf{A}_i(\beta)\mathbf{A}_i^{-1/2}] \cdot \\
&\cdot \ \cdot\mathbf{x}^T\mathbf{H}_n^{\text{indep}}\mathbf{x} \\
&= C\pi_n^3(r)q_n^2(r)\lambda_{\max}(\mathbf{H}_n^{\text{indep}}), \qquad (3.3.29)
\end{aligned}$$

for any $r \in (0.r_1)$, $\beta \in B_r$ and $n \geq n_1$. For the last inequality we have used (3.3.8) and (3.3.21).

Using the estimation (3.3.28), (3.3.29), and (3.3.22) and it follows that:

$$|\mathbf{x}^T\mathbf{b}_n^{(l)}(\beta)| \leq Cr\pi_n^2(r)q_n(r)\lambda_{\max}(\mathbf{H}_n^{\text{indep}}),$$

for any $\beta \in B_r$, $r \in (0, r_1)$ and $n \geq n_1$. Multiplying by $\alpha_n^{-1/2-\delta}$, we obtain for any $r \in (0, r_1)$ and any $n \geq n_1$:

$$\alpha_n^{-1/2-\delta} \sup_{\beta \in B_r} \sup_{\|\mathbf{x}\|=1} |\mathbf{x}^T \mathbf{b}_n^{(l)}(\beta)| \leq Cr\delta_n \pi_n^2(r) q_n(r).$$

Using condition (C_3), it follows that for any $l \in \{1, \ldots p\}$:

$$\lim_{r\to 0} \limsup_{n\to\infty} \alpha_n^{-1/2-\delta} \sup_{\beta \in B_r} \sup_{\|\mathbf{x}\|=1} |\mathbf{x}^T \mathbf{b}_n^{(l)}(\beta)| = 0 \text{ a.s.}$$

From (3.3.27) we conclude that:

$$\lim_{r\to 0} \limsup_{n\to\infty} \alpha_n^{-1/2-\delta} \sup_{\beta \in B_r} \sup_{\|\mathbf{x}\|=1} |\mathbf{x}^T \mathbf{B}_n^{(3)}(\beta)\mathbf{x}| = 0 \text{ a.s.}$$

□

Lemma 3.3.6 *Assume that (AH), (K), and (R') hold. If there exists a constant $C > 0$ and there exists $r_0 > 0$ (non-random) such that for any $r \in (0, r_0)$:*

$$(C_4) : \limsup_{n\to\infty} nE[\pi_n^2(r)]\tilde{a}_n \lambda_{\max}(\boldsymbol{H}_n^{\text{indep}}) < \infty, \text{ where } \tilde{a}_n = \max\{a_n, a_n^2\}$$
$$(C_5) : \limsup_{n\to\infty} nE[\pi_n^4(r) q_n^2(r)]\lambda_{\max}(\boldsymbol{H}_n^{\text{indep}}) < \infty,$$

then, for any $r \in (0, r_0)$:

$$\lim_{n\to\infty} \alpha_n^{-1/2-\delta} \sup_{\beta \in B_r} \|\mathcal{E}_n(\beta)\| = 0, \ a.s.$$

In particular, $\lim_{n\to\infty} \limsup_{n\to\infty} \alpha_n^{-1/2-\delta} \sup_{\beta \in B_r} \|\mathcal{E}_n(\beta)\| = 0$, *a.s.*

Proof. Recall that $\mathcal{E}_n(\beta) = \mathcal{E}_n^{(1)}(\beta) + \mathcal{E}_n^{(2)}(\beta) + \mathcal{E}_n^{(3)}(\beta)$, where $\mathcal{E}_n^{(1)}(\beta)$, $\mathcal{E}_n^{(2)}(\beta)$ and $e_n^{(l)}(\beta)$ are given by (3.3.6). Recall that $e_n^{(l)}(\beta)$ denotes the l-th column of $\mathcal{E}_n^{(3)}(\beta)$ and $|\mathbf{x}^T \mathcal{E}_n^{(3)}(\beta)\mathbf{x}| \leq \{\sum_{l=1}^p |\mathbf{x}^T e_n^{(l)}(\beta)|^2\}^{1/2}$.

Let $\mathbf{x}$ be a p-dimensional vector with $\|\mathbf{x}\| = 1$. Using (3.3.17), we can write:

$$\begin{aligned}
\mathbf{x}^T \mathcal{E}_n^{(1)}(\beta)\mathbf{x} &= U_n^{(1)}(\beta, \mathbf{x}) + U_n^{(2)}(\beta, \mathbf{x}) + U_n^{(3)}(\beta, \mathbf{x}), \\
\mathbf{x}^T \mathcal{E}_n^{(2)}(\beta)\mathbf{x} &= U_n^{(4)}(\beta, \mathbf{x}) + U_n^{(5)}(\beta, \mathbf{x}) + U_n^{(6)}(\beta, \mathbf{x}),
\end{aligned}$$

$$\mathbf{x}^T e_n^{(l)}(\beta) \;=\; U_n^{(7)}(\beta, \mathbf{x}),\ l \in \{1, \ldots, p\}$$

where

$$\begin{aligned}
U_n^{(1)}(\beta.\mathbf{x}) &= \sum_{i=1}^{n} \mathbf{x}^T \mathbf{X}_i^T \mathbf{G}_i^{(1)} \mathrm{diag}(\mathbf{X}_i \mathbf{x}) \mathcal{R}_{i-1}(\beta)^{-1} \mathbf{A}_i^{-1/2} (\mathbf{y}_i - \mu_i) \\
U_n^{(2)}(\beta, \mathbf{x}) &= \sum_{i=1}^{n} \mathbf{x}^T \mathbf{X}_i^T \mathbf{G}_i^{(1)}(\beta) \mathrm{diag}(\mathbf{X}_i \mathbf{x}) \mathcal{R}_{i-1}(\beta)^{-1} [\mathbf{A}_i(\beta)^{-1/2} - \mathbf{A}_i^{-1/2}] (\mathbf{y}_i - \mu_i) \\
U_n^{(3)}(\beta, \mathbf{x}) &= \sum_{i=1}^{n} \mathbf{x}^T \mathbf{X}_i^T [\mathbf{G}_i^{(1)}(\beta) - \mathbf{G}_i^{(1)}] \mathrm{diag}(\mathbf{X}_i \mathbf{x}) \mathcal{R}_{i-1}(\beta)^{-1} \mathbf{A}_i^{-1/2} (\mathbf{y}_i - \mu_i),
\end{aligned}$$

$$\begin{aligned}
U_n^{(4)}(\beta.\mathbf{x}) &= \sum_{i=1}^{n} \mathbf{x}^T \mathbf{X}_i^T \mathbf{A}_i^{1/2} \mathcal{R}_{i-1}(\beta)^{-1} \mathrm{diag}(\mathbf{X}_i \mathbf{x}) \mathbf{G}_i^{(2)} (\mathbf{y}_i - \mu_i) \\
U_n^{(5)}(\beta, \mathbf{x}) &= \sum_{i=1}^{n} \mathbf{x}^T \mathbf{X}_i^T [\mathbf{A}_i(\beta)^{1/2} - \mathbf{A}_i^{1/2}] \mathcal{R}_{i-1}(\beta)^{-1} \mathrm{diag}(\mathbf{X}_i \mathbf{x}) \mathbf{G}_i^{(2)}(\beta) (\mathbf{y}_i - \mu_i) \\
U_n^{(6)}(\beta, \mathbf{x}) &= \sum_{i=1}^{n} \mathbf{x}^T \mathbf{X}_i^T \mathbf{A}_i^{1/2} \mathcal{R}_{i-1}(\beta)^{-1} \mathrm{diag}(\mathbf{X}_i \mathbf{x}) [\mathbf{G}_i^{(2)}(\beta) - \mathbf{G}_i^{(2)}] (\mathbf{y}_i - \mu_i)
\end{aligned}$$

and

$$U_n^{(7)}(l, \beta, \mathbf{x}) \;=\; \sum_{i=1}^{n} \mathbf{x}^T \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2} \mathcal{R}_{i-1}(\beta)^{-1} \mathbf{Q}_{i-1}^{(l)}(\beta) \mathcal{R}_{i-1}(\beta)^{-1} \mathbf{A}_i(\beta)^{-1/2} (\mathbf{y}_i - \mu_i).$$

For any $s \in \{1, \ldots 6\}$, and for any $\beta \in B_r$, $r > 0$. $\{U_n^{(s)}(\beta), \mathcal{F}_n\}_{n \geq 1}$ is a martingale with respect to P_{β_0}:

$$\begin{aligned}
&\mathrm{E}[U_n^{(s)}(\beta, \mathbf{x}) | \mathcal{F}_{n-1}] = \mathrm{E}\left[\sum_{i=1}^{n} (\mathbf{y}_i^{(s)})^T \mathcal{R}_{i-1}(\beta)^{-1} \mathbf{Y}_i^{(s)} (\mathbf{y}_i - \mu_i) | \mathcal{F}_{n-1}\right] \\
&= \mathrm{E}\left[\sum_{i=1}^{n-1} (\mathbf{y}_i^{(s)})^T \mathcal{R}_{i-1}(\beta)^{-1} \mathbf{Y}_i^{(s)} (\mathbf{y}_i - \mu_i) | \mathcal{F}_{n-1}\right] + \mathrm{E}\left[(\mathbf{y}_n^{(s)})^T \mathcal{R}_{n-1}^{-1}(\beta) \mathbf{Y}_n^{(s)} \varepsilon_n | \mathcal{F}_{n-1}\right] \\
&= \sum_{i=1}^{n-1} (\mathbf{y}_i^{(s)})^T \mathcal{R}_{i-1}(\beta)^{-1} \mathbf{Y}_i^{(s)} (\mathbf{y}_i - \mu_i) + (\mathbf{y}_n^{(s)})^T \mathcal{R}_{n-1}^{-1}(\beta) \mathbf{Y}_n^{(s)} \mathrm{E}[\varepsilon_n | \mathcal{F}_{n-1}] \\
&= \mathbf{U}_{n-1}^{(s)}(\beta, \mathbf{x}).
\end{aligned}$$

where $\mathbf{y}_i^{(s)}$ is an $m \times 1$ vector of constants and $\mathbf{Y}_i^{(s)}$ is an $m \times m$ matrix of constants. For the equalities above we have used the fact that $\mathcal{R}_i(\beta)^{-1}$ is $\mathcal{F}_i$ measurable and $\mathrm{E}[\varepsilon_n|\mathcal{F}_{n-1}] = 0$.

It follows that $\{|U_n^{(s)}(\beta, \mathbf{x})|, \mathcal{F}_n\}_{n\geq 1}$ is a submartingale, for any $\beta \in B_r$. Hence, for any $\beta \in B_r$ and any $p \times 1$ vector $\mathbf{x}$ of norm 1, we have:

$$\mathrm{E}[\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1}|U_n^{(s)}(\beta, \mathbf{x})||\mathcal{F}_{n-1}] \geq \mathrm{E}[|U_n^{(s)}(\beta, \mathbf{x})||\mathcal{F}_{n-1}] \geq |U_{n-1}^{(s)}(\beta, \mathbf{x})|$$

From where it follows that:

$$\mathrm{E}_{\beta_0}[\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1}|U_n^{(s)}(\beta, \mathbf{x})||\mathcal{F}_{n-1}] \geq \sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1}|U_{n-1}^{(s)}(\beta, \mathbf{x})| \text{ a.s. } .\forall n \geq 1.$$

i.e $\{\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1}|U_n^{(s)}(\beta, \mathbf{x})|, \mathcal{F}_n\}_{n\geq 1}$ is a submartingale with respect to P_{β_0}, for any $r > 0$.

Similarly, for any $l \in \{1, \ldots p\}$ and for any $\beta \in B_r$, $r > 0$, $\{U_n^{(7)}(\beta), \mathcal{F}_n\}_{n\geq 1}$ is a martingale with respect to P_{β_0}, since the components of the vector

$$(\mathbf{y}_i^{(7)})^T := \mathbf{x}^T\mathbf{X}_i^T\mathbf{A}_i(\beta)^{1/2}\mathcal{R}_{i-1}(\beta)^{-1}\mathbf{Q}_{i-1}^{(l)}(\beta)$$

are $\mathcal{F}_{i-1}$ measurable. Hence, $\{\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1}|U_n^{(7)}(l, \beta, \mathbf{x})|, \mathcal{F}_n\}_{n\geq 1}$ is a submartingale with respect to P_{β_0}.

Note that by Lemma 3.2.1, there exist some $r_0 > 0$, $n_0 \geq 1$ such that:

$$\eta_n(r) \leq C, \ \forall r \in (0, r_0), \ \forall n \geq n_0. \tag{3.3.30}$$

We prove that for any $r \in (0, r_0)$, there exists a constant $C > 0$ (depending on r) such that:

$$\mathrm{E}[\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1}|U_n^{(s)}(\beta, \mathbf{x})|] \leq C, \ \forall n \geq n_0, \ \forall s \in \{1, \ldots, 6\}, \tag{3.3.31}$$

$$\mathrm{E}[\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1}|U_n^{(7)}(\beta, \mathbf{x})|] \leq C, \ \forall n \geq n_0. \tag{3.3.32}$$

By the Martingale Convergence Theorem (Theorem 35.5 in [3]), it follows that the sequence $\{\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1}|U_n^{(s)}(\beta,\mathbf{x})|\}_{n\geq 1}$ converges a.s. For any $s\in\{1,\dots 7\}$ and for any $r>0$ since $\alpha_n\to\infty$, we obtain:

$$\lim_{n\to\infty}\alpha_n^{-1/2-\delta}\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1}|U_n^{(s)}(\beta,\mathbf{x})|=0 \text{ a.s.}$$

From here, it will follow that for any $r>0$:

$$\lim_{n\to\infty}\alpha_n^{-1/2-\delta}\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1}|\mathbf{x}^T\mathcal{E}_n^{(k)}\mathbf{x}|=0 \text{ a.s.}, \ \forall k\in\{1,2,3\}.$$

The remaining part of the proof is dedicated to relations (3.3.31) and (3.3.32). We begin with the term $U_n^{(1)}(\beta,\mathbf{x})$. Using (3.3.9) with:

$$\begin{aligned}
\mathbf{a}_i^T &= \sum_{i-1}^{n}\mathbf{x}^T\mathbf{X}_i^T\mathbf{G}_i^{(1)}\mathrm{diag}(\mathbf{X}_i\mathbf{x})\mathcal{R}_{i-1}(\beta)^{-1}\\
\mathbf{b}_i &= \sum_{i-1}^{n}\mathbf{A}_i^{-1/2}(\mathbf{y}_i-\mu_i),
\end{aligned}$$

we obtain:

$$\begin{aligned}
|U_n^{(1)}(\beta,\mathbf{x})| &\leq \left[\sum_{i=1}^{n}\mathbf{x}^T\mathbf{X}_i^T\mathbf{G}_i^{(1)}\mathrm{diag}(\mathbf{X}_i\mathbf{x})[\mathcal{R}_{i-1}(\beta)^{-1}]^2\mathrm{diag}(\mathbf{X}_i\mathbf{x})\mathbf{G}_i^{(1)}\mathbf{X}_i\mathbf{x}\right]^{1/2}\\
&\cdot \left[\sum_{i=1}^{n}(\mathbf{y}_i-\mu_i)^T\mathbf{A}_i^{-1}(\mathbf{y}_i-\mu_i)\right]^{1/2}\\
&:= J_1(\beta,\mathbf{x})^{1/2}\cdot J_2^{1/2}. \qquad (3.3.33)
\end{aligned}$$

The first term in (3.3.33) can be estimated as:

$$\begin{aligned}
J_1(\beta,\mathbf{x}) &\leq \max_{i\leq n}\lambda_{\max}^2[\mathcal{R}_{i-1}(\beta)^{-1}]\cdot\max_{i\leq n}\lambda_{\max}[\mathrm{diag}^2(\mathbf{X}_i\mathbf{x})]\cdot\max_{i\leq n}\lambda_{\max}^2[\mathbf{A}_i^{-1/2}\mathbf{G}_i^{(1)}]\cdot\\
&\cdot\left[\sum_{i=1}^{n}\mathbf{x}^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}\mathbf{A}_i^{1/2}\mathbf{X}_i\mathbf{x}\right]\\
&= \max_{i\leq n}\lambda_{\max}^2[\mathcal{R}_{i-1}(\beta)^{-1}]\cdot\max_{i\leq n}\lambda_{\max}[\mathrm{diag}^2(\mathbf{X}_i\mathbf{x})]\cdot\max_{i\leq n}\lambda_{\max}^2[\mathbf{A}_i^{-1/2}\mathbf{G}_i^{(1)}]\cdot\\
&\cdot (\mathbf{x}^T\mathbf{H}_n^{\mathrm{indep}}\mathbf{x})
\end{aligned}$$

$$\leq \; C\pi_n^2(r)a_n \cdot \max_{i\leq n} \lambda_{\max}^2[\mathbf{A}_i^{-1/2}\mathbf{G}_i^{(1)}] \cdot \lambda_{\max}(\mathbf{H}_n^{\text{indep}}),$$

where we used (3.3.8) and (3.3.19) for the last inequality. Using assumption (AH) we have the following estimation:

$$[\mathbf{A}_i^{-1/2}\mathbf{G}_i^{(1)}]_{jj} \;=\; \frac{1}{2}\mu'(\mathbf{x}_{ij}^T\beta_0)^{-1/2}\frac{\mu''(\mathbf{x}_{ij}^T\beta_0)}{\mu'(\mathbf{x}_{ij}^T\beta_0)^{1/2}} = \frac{1}{2}\frac{\mu''(\mathbf{x}_{ij}^T\beta_0)}{\mu'(\mathbf{x}_{ij}^T\beta_0)} \leq \frac{1}{2}C, \; \forall j \leq m, \; \forall i \geq 1.$$

Therefore, $\max_{i\leq n} \lambda_{\max}^2[\mathbf{A}_i^{-1/2}\mathbf{G}_i^{(1)}] \leq C$, for any $n \geq 1$. Hence,

$$E[\sup_{\beta\in B_r} J_1(\beta,\mathbf{x})] \leq C E[\pi_n^2(r)]a_n\lambda_{\max}(\mathbf{H}_n^{\text{indep}}), \; \forall n \geq 1. \tag{3.3.34}$$

As for the second term in (3.3.33), we observe that:

$$E(J_2) \;=\; E\left[\sum_{i=1}^n(\mathbf{y}_i - \mu_i)^T\mathbf{A}_i^{-1}(\mathbf{y}_i - \mu_i)\right] = nm. \tag{3.3.35}$$

To see this, we use the fact that the elements on the diagonal of $\mathbf{A}_i$ are non-random. More precisely, we have:

$$E\left[\sum_{i=1}^n(\mathbf{y}_i - \mu_i)^T\mathbf{A}_i^{-1}(\mathbf{y}_i - \mu_i)\right] = \sum_{i=1}^n\sum_{j=1}^m E[\varepsilon_{ij}^2(\sigma_{ij}^2)^{-1}]$$

$$= \sum_{i=1}^n\sum_{j=1}^m E[(\sigma_{ij}^2)^{-1}E(\varepsilon_{ij}^2|\mathcal{F}_{i-1})] = \sum_{i=1}^n\sum_{j=1}^m E[(\sigma_{ij}^2)^{-1}\sigma_{ij}^2] = mn.$$

Using the Cauchy-Schwarz inequality, and relations (3.3.33), (3.3.34) and (3.3.35) we obtain:

$$\begin{aligned} E[\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1}|U_n^{(1)}(\beta,\mathbf{x})|] \;&\leq\; \{E[\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1} J_1(\beta,\mathbf{x})]\}^{1/2}\cdot[E(J_2)]^{1/2} \\ &\leq\; C\{nE[\pi_n^2(r)]a_n\lambda_{\max}(\mathbf{H}_n^{\text{indep}})\}^{1/2}, \; \forall n \geq 1, \; \forall r > 0. \end{aligned}$$

Relation (3.3.31) with $s = 1$ follows from condition (C_4).

We now treat the term involving $U_n^{(2)}(\beta,\mathbf{x})$. Using (3.3.9) with:

$$\mathbf{a}_i^T \;=\; \sum_{i-1}^n \mathbf{x}^T\mathbf{X}_i^T\mathbf{G}_i^{(1)}(\beta)\text{diag}(\mathbf{X}_i\mathbf{x})\mathcal{R}_{i-1}(\beta)^{-1}$$

$$\mathbf{b}_i = \sum_{i-1}^{n} [\mathbf{A}_i(\beta)^{-1/2} - \mathbf{A}_i^{-1/2}](\mathbf{y}_i - \mu_i),$$

we obtain:

$$\begin{aligned}
|U_n^{(2)}(\beta,\mathbf{x})| &\leq \left[\sum_{i=1}^{n} \mathbf{x}^T\mathbf{X}_i^T\mathbf{G}_i^{(1)}(\beta)\mathrm{diag}(\mathbf{X}_i\mathbf{x})[\mathcal{R}_{i-1}(\beta)^{-1}]^2\mathrm{diag}(\mathbf{X}_i\mathbf{x})\mathbf{G}_i^{(1)}(\beta)\mathbf{X}_i\mathbf{x}\right]^{1/2} \\
&\cdot \left[\sum_{i=1}^{n}(\mathbf{y}_i - \mu_i)^T[\mathbf{A}_i(\beta)^{-1/2} - \mathbf{A}_i^{-1/2}][\mathbf{A}_i(\beta)^{-1/2} - \mathbf{A}_i^{-1/2}](\mathbf{y}_i - \mu_i)\right]^{1/2} \\
&:= J_3(\beta,\mathbf{x})^{1/2} \cdot J_4(\beta)^{1/2} \qquad (3.3.36)
\end{aligned}$$

The first term in (3.3.36) can be estimated as

$$\begin{aligned}
J_3(\beta,\mathbf{x}) &\leq \max_{i\leq n}\lambda_{\max}^2[\mathcal{R}_{i-1}(\beta)^{-1}] \cdot \max_{i\leq n}\lambda_{\max}[\mathrm{diag}^2(\mathbf{X}_i\mathbf{x})] \cdot \max_{i\leq n}\lambda_{\max}^2[\mathbf{A}_i^{-1/2}\mathbf{G}_i^{(1)}(\beta)] \cdot \\
&\cdot \left[\sum_{i=1}^{n}\mathbf{x}^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}\mathbf{A}_i^{1/2}\mathbf{X}_i\mathbf{x}\right] \\
&= \max_{i\leq n}\lambda_{\max}^2[\mathcal{R}_{i-1}(\beta)^{-1}] \cdot \max_{i\leq n}\lambda_{\max}[\mathrm{diag}^2(\mathbf{X}_i\mathbf{x})] \cdot \max_{i\leq n}\lambda_{\max}^2[\mathbf{A}_i^{-1/2}\mathbf{G}_i^{(1)}(\beta)] \cdot \\
&\cdot \mathbf{x}^T\mathbf{H}_n^{\mathrm{indep}}\mathbf{x} \\
&\leq C\pi_n^2(r)a_n\max_{i\leq n}\lambda_{\max}^2[\mathbf{A}_i^{-1/2}\mathbf{G}_i^{(1)}(\beta)] \cdot \lambda_{\max}(\mathbf{H}_n^{\mathrm{indep}}),
\end{aligned}$$

where for the last inequality we used (3.3.8) and (3.3.19).

Using (AH) (3.3.30) we have:

$$\begin{aligned}
[\mathbf{A}_i^{-1/2}\mathbf{G}_i^{(1)}(\beta)]_{jj} &= \frac{1}{2}\mu'(\mathbf{x}_{ij}^T\beta_0)^{-1/2}\frac{\mu''(\mathbf{x}_{ij}^T\beta)}{\mu'(\mathbf{x}_{ij}^T\beta)^{1/2}} = \frac{1}{2}\left[\frac{\mu'(\mathbf{x}_{ij}^T\beta)}{\mu'(\mathbf{x}_{ij}^T\beta_0)}\right]^{1/2}\frac{\mu''(\mathbf{x}_{ij}^T\beta)}{\mu'(\mathbf{x}_{ij}^T\beta)} \\
&\leq \frac{1}{2}C[\eta_n(r)+1] \leq C,\ \forall i\leq n,\ \forall j\leq m
\end{aligned}$$

for any $\beta \in B_r$, any $r \in (0, r_0)$ and any $n \geq n_0$. Therefore:

$$\max_{i\leq n}\lambda_{\max}^2[\mathbf{A}_i^{-1/2}\mathbf{G}_i^{(1)}(\beta)] \leq C,\ \forall\beta\in B_r,\ \forall r\in(0,r_0),\ \forall n\geq n_0.$$

It follows that:

$$E[\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1} J_3(\beta,\mathbf{x})] \leq CE[\pi_n^2(r)]a_n\lambda_{\max}(\mathbf{H}_n^{\mathrm{indep}}),\ \forall r\in(0,r_0),\ \forall n\geq n_0. \quad (3.3.37)$$

The second term in (3.3.36) is:

$$\begin{aligned} J_4(\beta) &= \sum_{i=1}^{n}(\mathbf{y}_i-\mu_i)^T\mathbf{A}_i^{-1/2}[\mathbf{A}_i(\beta)^{-1/2}\mathbf{A}_i^{1/2}-\mathbf{I}]^2\mathbf{A}_i^{-1/2}(\mathbf{y}_i-\mu_i) \\ &\leq \max_{i\leq n}\lambda_{\max}[\mathbf{A}_i(\beta)^{-1/2}\mathbf{A}_i^{1/2}-\mathbf{I}]^2\sum_{i\leq n}(\mathbf{y}_i-\mu_i)^T\mathbf{A}_i^{-1}(\mathbf{y}_i-\mu_i). \end{aligned}$$

Note that for any $\beta\in B_r$, any $r>0$ and $n\geq 1$:

$$[\mathbf{A}_i(\beta)^{-1/2}\mathbf{A}_i^{1/2}-\mathbf{I}]_{jj} = \left[\frac{\mu'(\mathbf{x}_{ij}^T\beta_0)}{\mu'(\mathbf{x}_{ij}^T\beta)}\right]^{1/2}-1\leq\eta_n(r). \tag{3.3.38}$$

Hence, by Lemma 3.2.1 $\max_{i\leq n}\lambda_{\max}[\mathbf{A}_i(\beta)^{-1/2}\mathbf{A}_i^{1/2}-\mathbf{I}]^2\leq a_n r^2$, for any $\beta\in B_r$, any $r\in(0,r_0)$ and any $n\geq n_0$. Therefore, $\sup_{\beta\in B_r}J_4(\beta)\leq a_n r^2\sum_{i\leq n}(\mathbf{y}_i-\mu_i)^T\mathbf{A}_i^{-1}(\mathbf{y}_i-\mu_i)$, for any $r\in(0,r_0)$ and any $n\geq n_0$. Using (3.3.35) it follows that:

$$E[\sup_{\beta\in B_r}J_4(\beta)]\leq Cna_nr^2,\ \forall r\in(0,r_0),\ \forall n\geq n_0. \tag{3.3.39}$$

where $C=m$. Using Cauchy-Schwarz inequality and relations (3.3.36), (3.3.37) and (3.3.39), we obtain:

$$\begin{aligned} E[\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1}|U_n^{(2)}(\beta,\mathbf{x})|] &\leq \{E[\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1}J_3(\beta,\mathbf{x})]\}^{1/2}\cdot\{E[\sup_{\beta\in B_r}J_4(\beta)]\}^{1/2} \\ &\leq C\{nE[\pi_n^2(r)]a_n^2r^2\lambda_{\max}(\mathbf{H}_n^{\text{indep}})\}^{1/2}, \end{aligned}$$

for any $r\in(0,r_0)$ and any $n\geq n_0$. Relation (3.3.31) with $s=2$ follows from condition (C_4).

In order to evaluate $U_n^{(3)}(\beta,\mathbf{x})$, we use again (3.3.9) with:

$$\begin{aligned} \mathbf{a}_i^T &= \sum_{i-1}^{n}\mathbf{x}^T\mathbf{X}_i^T[\mathbf{G}_i^{(1)}(\beta)-\mathbf{G}_i^{(1)}]\text{diag}(\mathbf{X}_i\mathbf{x})\mathcal{R}_{i-1}(\beta)^{-1} \\ \mathbf{b}_i &= \sum_{i-1}^{n}\mathbf{A}_i^{-1/2}(\mathbf{y}_i-\mu_i), \end{aligned}$$

and we obtain:

$$|U_n^{(3)}(\beta,\mathbf{x})| \leq [\sum_{i=1}^{n}\mathbf{x}^T\mathbf{X}_i^T[\mathbf{G}_i^{(1)}(\beta)-\mathbf{G}_i^{(1)}]\text{diag}(\mathbf{X}_i\mathbf{x})\mathcal{R}_{i-1}(\beta)^{-2}\text{diag}(\mathbf{X}_i\mathbf{x})$$

$$
\begin{aligned}
&\cdot \quad \mathbf{G}_i^{(1)}(\beta)\mathbf{X}_i\mathbf{x}]^{1/2}\left[\sum_{i=1}^{n}(\mathbf{y}_i-\mu_i)^T\mathbf{A}_i^{-1}(\mathbf{y}_i-\mu_i)\right]^{1/2} \\
&:= \quad J_5(\beta,\mathbf{x})^{1/2}\cdot J_2^{1/2} \qquad (3.3.40)
\end{aligned}
$$

The first term in (3.3.40) can be estimated as:

$$
\begin{aligned}
J_5(\beta,\mathbf{x}) \quad &\le \quad \max_{i\le n}\lambda_{\max}^2[\mathcal{R}_{i-1}(\beta)^{-1}]\cdot\max_{i\le n}\lambda_{\max}[\mathrm{diag}^2(\mathbf{X}_i\mathbf{x})] \\
&\cdot \quad \max_{i\le n}\lambda_{\max}\{\mathbf{A}_i^{-1/2}[\mathbf{G}_i^{(1)}(\beta)-\mathbf{G}_i^{(1)}]\}^2\left[\sum_{i=1}^{n}\mathbf{x}^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}\mathbf{A}_i^{1/2}\mathbf{X}_i\mathbf{x}\right] \\
&= \quad \max_{i\le n}\lambda_{\max}^2[\mathcal{R}_{i-1}(\beta)^{-1}]\cdot\max_{i\le n}\lambda_{\max}[\mathrm{diag}^2(\mathbf{X}_i\mathbf{x})] \\
&\cdot \quad \max_{i\le n}\lambda_{\max}\{\mathbf{A}_i^{-1/2}[\mathbf{G}_i^{(1)}(\beta)-\mathbf{G}_i^{(1)}]\}^2\mathbf{x}^T\mathbf{H}_n^{\mathrm{indep}}\mathbf{x} \\
&\le \quad C\pi_n^2(r)a_n\max_{i\le n}\lambda_{\max}\{\mathbf{A}_i^{-1/2}[\mathbf{G}_i^{(1)}(\beta)-\mathbf{G}_i^{(1)}]\}^2\lambda_{\max}(\mathbf{H}_n^{\mathrm{indep}}),
\end{aligned}
$$

where we used (3.3.8) and (3.3.19) for the last inequality. We have the following estimation which follows from Lemma 3.2.2(i):

$$
\begin{aligned}
\{\mathbf{A}_i^{-1/2}[\mathbf{G}_i^{(1)}(\beta)-\mathbf{G}_i^{(1)}]\}_{jj} \quad &= \quad \frac{1}{2}\mu'(\mathbf{x}_{ij}^T\beta_0)^{-1/2}\left[\frac{\mu''(\mathbf{x}_{ij}^T\beta)}{\mu'(\mathbf{x}_{ij}^T\beta)^{1/2}}-\frac{\mu''(\mathbf{x}_{ij}^T\beta_0)}{\mu'(\mathbf{x}_{ij}^T\beta_0)^{1/2}}\right] \\
&\le \quad \nu_n\le a_n^{1/2}r,
\end{aligned}
$$

for any $\beta\in B_r$, any $r\in(0,r_0)$ and any $n\ge n_0$. It follows that:

$$
\max_{i\le n}\lambda_{\max}\{\mathbf{A}_i^{-1/2}[\mathbf{G}_i^{(1)}(\beta)-\mathbf{G}_i^{(1)}]\}^2\le a_nr^2,\ \forall\beta\in B_r,\ \forall r\in(0,r_0),\ \forall n\ge n_0.
$$

Hence:

$$
\mathrm{E}[\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1}J_5(\beta,\mathbf{x})] \quad \le \quad C\mathrm{E}[\pi_n^2(r)]a_n^2r^2\lambda_{\max}(\mathbf{H}_n^{\mathrm{indep}}), \qquad (3.3.41)
$$

for any $r\in(0,r_0)$ and $n\ge n_0$. Using Cauchy-Schwarz inequality and relations (3.3.40), (3.3.41) and (3.3.35), we obtain:

$$
\mathrm{E}[\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1}|U_n^{(3)}(\beta,\mathbf{x})|] \quad \le \quad \{\mathrm{E}[\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1}J_5(\beta,\mathbf{x})]\}^{1/2}\cdot[\mathrm{E}(J_2)]^{1/2}
$$

$$\leq \quad C\{n\mathrm{E}[\pi_n^2(r)]a_n^2 r^2 \lambda_{\max}(\mathbf{H}_n^{\mathrm{indep}})\}^{1/2},$$

for any $r \in (0, r_0)$ and any $n \geq n_0$. Relation (3.3.31) with $s = 3$ follows from condition (C_4).

To obtain an estimation for $|U_n^{(4)}(\beta, \mathbf{x})|$ we use again (3.3.9) with:

$$\begin{aligned}
\mathbf{a}_i^T &= \sum_{i-1}^{n} \mathbf{x}^T \mathbf{X}_i^T \mathbf{A}_i^{1/2} \mathcal{R}_{i-1}(\beta)^{-1} \\
\mathbf{b}_i &= \sum_{i-1}^{n} \mathrm{diag}(\mathbf{X}_i\mathbf{x}) \mathbf{G}_i^{(2)}(\mathbf{y}_i - \mu_i),
\end{aligned}$$

and we obtain:

$$\begin{aligned}
|U_n^{(4)}(\beta, \mathbf{x})| &\leq \left[\sum_{i=1}^{n} \mathbf{x}^T \mathbf{X}_i^T \mathbf{A}_i^{1/2} [\mathcal{R}_{i-1}(\beta)^{-1}]^2 \mathbf{A}_i^{1/2} \mathbf{X}_i \mathbf{x}\right]^{1/2} \\
&\cdot \left[\sum_{i=1}^{n} (\mathbf{y}_i - \mu_i)^T \mathbf{G}_i^{(2)} \mathrm{diag}(\mathbf{X}_i\mathbf{x}) \mathrm{diag}(\mathbf{X}_i\mathbf{x}) \mathbf{G}_i^{(2)} (\mathbf{y}_i - \mu_i)\right]^{1/2} \\
&:= J_6(\beta, \mathbf{x})^{1/2} \cdot J_7(\mathbf{x})^{1/2}. \qquad (3.3.42)
\end{aligned}$$

The first term in (3.3.42) can be estimated as:

$$\begin{aligned}
J_6(\beta, \mathbf{x}) &\leq \max_{i \leq n} \lambda_{\max}[\mathcal{R}_{i-1}(\beta)^{-1}] \cdot \sum_{i=1}^{n} \mathbf{x}^T \mathbf{X}_i^T \mathbf{A}_i^{1/2} \mathbf{A}_i^{1/2} \mathbf{X}_i \mathbf{x} \\
&= \max_{i \leq n} \lambda_{\max}^2 [\mathcal{R}_{i-1}(\beta)^{-1}] \cdot \mathbf{x}^T \mathbf{H}_n^{\mathrm{indep}} \mathbf{x} \\
&\leq C\pi_n^2(r) \lambda_{\max}(\mathbf{H}_n^{\mathrm{indep}}),
\end{aligned}$$

where for the last inequality we used (3.3.8). Hence:

$$\mathrm{E}[\sup_{\beta \in B_r} \sup_{\|\mathbf{x}\|=1} J_6(\beta. \mathbf{x})] \leq C\mathrm{E}[\pi_n^2(r)] \lambda_{\max}(\mathbf{H}_n^{\mathrm{indep}}). \qquad (3.3.43)$$

For the second term of the product in (3.3.42). we have:

$$\begin{aligned}
J_7(\mathbf{x}) &\leq \max_{i \leq n} \lambda_{\max}[\mathrm{diag}^2(\mathbf{X}_i\mathbf{x})] \cdot \max_{i \leq n} \lambda_{\max}^2 [\mathbf{A}_i^{1/2} \mathbf{G}_i^{(2)}] \sum_{i \leq n} (\mathbf{y}_i - \mu_i)^T \mathbf{A}_i^{-1} (\mathbf{y}_i - \mu_i) \\
&\leq a_n \cdot \max_{i \leq n} \lambda_{\max}^2 [\mathbf{A}_i^{1/2} \mathbf{G}_i^{(2)}] \sum_{i \leq n} (\mathbf{y}_i - \mu_i)^T \mathbf{A}_i^{-1} (\mathbf{y}_i - \mu_i),
\end{aligned}$$

using (3.3.19) for the last inequality. By (AH), for any $i \leq n$ and $j \leq m$:

$$[\mathbf{A}_i^{1/2}\mathbf{G}_i^{(2)}]_{jj} = -\mu'(\mathbf{x}_{ij}^T\beta_0)^{1/2}\frac{1}{2}\frac{\mu''(\mathbf{x}_{ij}^T\beta_0)}{\mu'(\mathbf{x}_{ij}^T\beta_0)^{3/2}} = -\frac{1}{2}\frac{\mu''(\mathbf{x}_{ij}^T\beta_0)}{\mu'(\mathbf{x}_{ij}^T\beta_0)} \leq C.$$

It follows that $\max_{i\leq n}\lambda_{\max}^2[\mathbf{A}_i^{1/2}\mathbf{G}_i^{(2)}] \leq C$, for any $n \geq 1$. Therefore, $\sup_{\|\mathbf{x}\|=1} J_7(\mathbf{x}) \leq Ca_n \sum_{i\leq n}(\mathbf{y}_i - \mu_i)^T\mathbf{A}_i^{-1}(\mathbf{y}_i - \mu_i)$. By (3.3.35), it follows that:

$$E[\sup_{\|\mathbf{x}\|=1} J_7(\mathbf{x})] \leq Cna_n, \ \forall n \geq 1. \tag{3.3.44}$$

Using Cauchy-Schwarz inequality and relations (3.3.42), (3.3.43) and (3.3.44), we obtain:

$$\begin{aligned} E[\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1}|U_n^{(4)}(\beta,\mathbf{x})|] &\leq \{E[\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1} J_6(\beta,\mathbf{x})]\}^{1/2}\cdot\{E[\sup_{\|\mathbf{x}\|=1} J_7(\mathbf{x})]\}^{1/2} \\ &\leq C\{nE[\pi_n^2(r)]a_n\lambda_{\max}(\mathbf{H}_n^{\text{indep}})\}^{1/2}, \ \forall r > 0, \ \forall n \geq 1. \end{aligned}$$

Relation (3.3.31) with $s = 4$ follows from condition (C_4).

For $|U_n^{(5)}(\beta,\mathbf{x})|$, we use (3.3.9) with:

$$\begin{aligned} \mathbf{a}_i^T &= \sum_{i-1}^{n}\mathbf{x}^T\mathbf{X}_i^T[\mathbf{A}_i(\beta)^{1/2} - \mathbf{A}_i^{1/2}]\mathcal{R}_{i-1}(\beta)^{-1} \\ \mathbf{b}_i &= \sum_{i-1}^{n}\text{diag}(\mathbf{X}_i\mathbf{x})\mathbf{G}_i^{(2)}(\beta)(\mathbf{y}_i - \mu_i), \end{aligned}$$

and we obtain:

$$\begin{aligned} |U_n^{(5)}(\beta,\mathbf{x})| &\leq \left[\sum_{i=1}^{n}\mathbf{x}^T\mathbf{X}_i^T[\mathbf{A}_i(\beta)^{1/2} - \mathbf{A}_i^{1/2}][\mathcal{R}_{i-1}(\beta)^{-1}]^2[\mathbf{A}_i(\beta)^{1/2} - \mathbf{A}_i^{1/2}[\mathbf{X}_i\mathbf{x}\right]^{1/2} \\ &\cdot \left[\sum_{i=1}^{n}(\mathbf{y}_i - \mu_i)^T\mathbf{G}_i^{(2)}(\beta)\text{diag}(\mathbf{X}_i\mathbf{x})\text{diag}(\mathbf{X}_i\mathbf{x})\mathbf{G}_i^{(2)}(\beta)(\mathbf{y}_i - \mu_i)\right]^{1/2} \\ &:= J_8(\beta,\mathbf{x})^{1/2}\cdot J_9(\beta,\mathbf{x})^{1/2}. \end{aligned} \tag{3.3.45}$$

The first term in (3.3.45) can be estimated as:

$$J_8(\beta,\mathbf{x}) \leq \max_{i\leq n}\lambda_{\max}^2[\mathcal{R}_{i-1}(\beta)^{-1}]\cdot\max_{i\leq n}\lambda_{\max}[\mathbf{A}_i^{-1/2}(\mathbf{A}_i(\beta)^{1/2} - \mathbf{A}_i^{1/2})]^2$$

$$
\begin{aligned}
&\cdot \sum_{i=1}^{n} \mathbf{x}^T \mathbf{X}_i^T \mathbf{A}_i^{1/2} \mathbf{A}_i^{1/2} \mathbf{X}_i \mathbf{x} \\
&= \max_{i \le n} \lambda_{\max}^2 [\mathcal{R}_{i-1}(\beta)^{-1}] \cdot \max_{i \le n} \lambda_{\max}[\mathbf{A}_i^{-1/2} \mathbf{A}_i(\beta)^{1/2} - \mathbf{I}]^2 \cdot \mathbf{x}^T \mathbf{H}_n^{\mathrm{indep}} \mathbf{x} \\
&\le C\pi_n^2(r) \cdot \max_{i \le n} \lambda_{\max}[\mathbf{A}_i^{-1/2} \mathbf{A}_i(\beta)^{1/2} - \mathbf{I}]^2 \cdot \lambda_{\max}(\mathbf{H}_n^{\mathrm{indep}}),
\end{aligned}
$$

where we used (3.3.19) for the last inequality. Note that:

$$
[\mathbf{A}_i^{-1/2} \mathbf{A}_i(\beta)^{1/2} - \mathbf{I}]_{jj} = \left[\frac{\mu'(\mathbf{x}_{ij}^T \beta)}{\mu'(\mathbf{x}_{ij}^T \beta_0)} \right]^{1/2} - 1 \le \eta_n,
$$

and hence, by Lemma 3.2.1, for any $\beta \in B_r$, any $r \in (0, r_0)$ and any $n \ge n_0$, we have:

$$
\max_{i \le n} \lambda_{\max}[\mathbf{A}_i^{-1/2} \mathbf{A}_i(\beta)^{1/2} - \mathbf{I}]^2 \le a_n r^2.
$$

Therefore, for any $r \in (0, r_0)$ and for any $n \ge n_0$, we obtain

$$
\mathrm{E}[\sup_{\beta \in B_r} \sup_{\|\mathbf{x}\|=1} J_8(\beta, \mathbf{x})] \le C\mathrm{E}[\pi_n^2(r)] a_n r^2 \lambda_{\max}(\mathbf{H}_n^{\mathrm{indep}}). \tag{3.3.46}
$$

For the second term of the product in (3.3.45), we have

$$
J_9(\beta, \mathbf{x}) \le \max_{i \le n} \lambda_{\max}[\mathrm{diag}^2(\mathbf{X}_i \mathbf{x})] \max_{i \le n} \lambda_{\max}^2 [\mathbf{A}_i^{1/2} \mathbf{G}_i^{(2)}(\beta)] \sum_{i=1}^{n} (\mathbf{y}_i - \mu_i)^T \mathbf{A}_i^{-1} (\mathbf{y}_i - \mu_i).
$$

By (AH) we obtain:

$$
\begin{aligned}
[\mathbf{A}_i^{1/2} \mathbf{G}_i^{(2)}(\beta)]_{jj} &= -\mu'(\mathbf{x}_{ij}^T \beta_0)^{1/2} \frac{1}{2} \frac{\mu''(\mathbf{x}_{ij}^T \beta)}{\mu'(\mathbf{x}_{ij}^T \beta)^{3/2}} = -\frac{1}{2} \frac{\mu''(\mathbf{x}_{ij}^T \beta)}{\mu'(\mathbf{x}_{ij}^T \beta)} \left[\frac{\mu'(\mathbf{x}_{ij}^T \beta_0)}{\mu'(\mathbf{x}_{ij}^T \beta)} \right]^{1/2} \\
&\le C[\eta_n(r) + 1] \le C, \ \forall \beta \in B_r, \ \forall r \in (0, r_0), \ \forall n \ge n_0,
\end{aligned}
$$

where for the last inequality we applied Lemma 3.2.1. It follows that:

$$
\max_{i \le n} \lambda_{\max}^2 [\mathbf{A}_i^{1/2} \mathbf{G}_i^{(2)}(\beta)] \le C, \ \forall \beta \in B_r, \ \forall r \in (0, r_0), \ \forall n \ge n_0.
$$

Therefore, using (3.3.19),

$$
\sup_{\beta \in B_r} \sup_{\|\mathbf{x}\|=1} J_9(\beta, \mathbf{x}) \le Ca_n \cdot \sum_{i \le n} (\mathbf{y}_i - \mu_i)^T \mathbf{A}_i^{-1} (\mathbf{y}_i - \mu_i).
$$

By (3.3.35), it follows that:

$$E[\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1} J_9(\beta,\mathbf{x})] \le Cna_n. \tag{3.3.47}$$

Using cauchy-Schwarz inequality and relations (3.3.45), (3.3.46) and (3.3.47), we obtain:

$$\begin{aligned} E[\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1} |U_n^{(5)}(\beta,\mathbf{x})|] &\le \{E[\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1} J_8(\beta,\mathbf{x})]\}^{1/2}\cdot\{E[\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1} J_9(\beta,\mathbf{x})]\}^{1/2} \\ &\le C\{nE[\pi_n^2(r)]a_n^2r^2\lambda_{\max}(\mathbf{H}_n^{\mathrm{indep}})\}^{1/2}, \forall r\in(0,r_0),\ \forall n\ge n_0. \end{aligned}$$

Relation (3.3.31) with $s=5$ follows from condition (C_4).

To estimate $U_n^{(6)}(\beta,\mathbf{x})$, we use (3.3.9) with:

$$\begin{aligned} \mathbf{a}_i^T &= \sum_{i-1}^{n}\mathbf{x}^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}\mathcal{R}_{i-1}(\beta)^{-1} \\ \mathbf{b}_i &= \sum_{i-1}^{n}\mathrm{diag}(\mathbf{X}_i\mathbf{x})[\mathbf{G}_i^{(2)}(\beta)-\mathbf{G}_i^{(2)}](\mathbf{y}_i-\mu_i), \end{aligned}$$

and we obtain:

$$\begin{aligned} &|U_n^{(6)}(\beta,\mathbf{x})| \le \left[\sum_{i=1}^{n}\mathbf{x}^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}[\mathcal{R}_{i-1}(\beta)^{-1}]^2\mathbf{A}_i^{1/2}\mathbf{X}_i\mathbf{x}\right]^{1/2} \\ &\cdot\left[\sum_{i=1}^{n}(\mathbf{y}_i-\mu_i)^T[\mathbf{G}_i^{(2)}(\beta)-\mathbf{G}_i^{(2)}]\mathrm{diag}^2(\mathbf{X}_i\mathbf{x})[\mathbf{G}_i^{(2)}(\beta)-\mathbf{G}_i^{(2)}](\mathbf{y}_i-\mu_i)\right]^{1/2} \\ &:= J_6(\beta,\mathbf{x})^{1/2}\cdot J_{10}(\beta,\mathbf{x})^{1/2} \end{aligned} \tag{3.3.48}$$

For the second term of the product in (3.3.48), we have:

$$\begin{aligned} J_{10}(\beta,\mathbf{x}) &\le \max_{i\le n}\lambda_{\max}[\mathrm{diag}^2(\mathbf{X}_i\mathbf{x})]\cdot\max_{i\le n}\lambda_{\max}^2\{\mathbf{A}_i^{1/2}[\mathbf{G}_i^{(2)}(\beta)-\mathbf{G}_i^{(2)}]\}\cdot \\ &\cdot\ \sum_{i\le n}(\mathbf{y}_i-\mu_i)^T\mathbf{A}_i^{-1}(\mathbf{y}_i-\mu_i). \end{aligned}$$

By Lemma 3.2.2(ii), for any $\beta\in B_r$, any $r\in(0,r_0)$ and any $n\ge n_0$:

$$\{\mathbf{A}_i^{1/2}[\mathbf{G}_i^{(2)}(\beta)-\mathbf{G}_i^{(2)}]\}_{jj} = -\mu'(\mathbf{x}_{ij}^T\beta_0)^{1/2}\frac{1}{2}\left[\frac{\mu''(\mathbf{x}_{ij}^T\beta)}{\mu'(\mathbf{x}_{ij}^T\beta)^{3/2}}-\frac{\mu''(\mathbf{x}_{ij}^T\beta_0)}{\mu'(\mathbf{x}_{ij}^T\beta_0)^{3/2}}\right]$$

$$\leq \quad \xi_n \leq a_n^{1/2} r.$$

It follows that $\max_{i\leq n} \lambda^2_{\max}\{\mathbf{A}_i^{1/2}[\mathbf{G}_i^{(2)}(\beta) - \mathbf{G}_i^{(2)}]\} \leq a_n r^2$, for any $\beta \in B_r$, any $r \in (0, r_0)$ and any $n \geq n_0$. Therefore, using (3.3.19) we have:

$$\sup_{\beta\in B_r} \sup_{\|\mathbf{x}\|=1} J_{10}(\beta, \mathbf{x}) \leq C a_n^2 r^2 \sum_{i\leq n} (\mathbf{y}_i - \mu_i)^T \mathbf{A}_i^{-1} (\mathbf{y}_i - \mu_i), \ \forall r \in (0, r_0), \ \forall n \geq n_0.$$

By (3.3.35), it follows that:

$$\mathrm{E}[\sup_{\beta\in B_r} \sup_{\|\mathbf{x}\|=1} J_{10}(\beta, \mathbf{x})] \leq C a_n r^2 n, \ \forall r \in (0, r_0), \ \forall n \geq n_0. \tag{3.3.49}$$

Using Cauchy-Schwarz inequality and relations (3.3.48), (3.3.43) and (3.3.49), we obtain:

$$\begin{aligned} \mathrm{E}[\sup_{\beta\in B_r} \sup_{\|\mathbf{x}\|=1} |U_n^{(6)}(\beta, \mathbf{x})|] &\leq \{\mathrm{E}[\sup_{\beta\in B_r} \sup_{\|\mathbf{x}\|=1} J_6(\beta, \mathbf{x})]\}^{1/2} \cdot \{\mathrm{E}[\sup_{\beta\in B_r} \sup_{\|\mathbf{x}\|=1} J_{10}(\beta, \mathbf{x})]\}^{1/2} \\ &\leq C\{n\mathrm{E}[\pi_n^2(r)] a_n^2 r^2 \lambda_{\max}(\mathbf{H}_n^{\mathrm{indep}})\}^{1/2}, \end{aligned}$$

for any $r \in (0, r_0)$ any any $n \geq n_0$. Relation (3.3.31) with $s = 6$ follows from condition (C_4).

Finally, we prove (3.3.32). Let $l \in \{1, \ldots, p\}$ be fixed. Using (3.3.9) with:

$$\begin{aligned} \mathbf{a}_i^T &= \sum_{i-1}^{n} \alpha_i^{-1/2-\delta} \mathbf{x}^T \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2} \mathcal{R}_{i-1}(\beta)^{-1} \mathbf{Q}_{i-1}^{(l)}(\beta) \mathcal{R}_{i-1}(\beta)^{-1} \\ \mathbf{b}_i &= \sum_{i-1}^{n} \mathbf{A}_i(\beta)^{-1/2} (\mathbf{y}_i - \mu_i), \end{aligned}$$

we obtain:

$$\begin{aligned} |U_n^{(7)}(l, \beta, \mathbf{x})| &\leq [\sum_{i-1}^{n} \mathbf{x}^T \mathbf{X}_i^T \mathbf{A}_i(\beta)^{1/2} \mathcal{R}_{i-1}(\beta)^{-1} \mathbf{Q}_{i-1}^{(l)}(\beta) [\mathcal{R}_{i-1}(\beta)^{-1}]^2 \mathbf{Q}_{i-1}^{(l)}(\beta) \\ &\quad \cdot \ \mathcal{R}_{i-1}(\beta)^{-1} \mathbf{A}_i(\beta)^{1/2} \mathbf{X}_i \mathbf{x}]^{1/2} \left[\sum_{i-1}^{n} (\mathbf{y}_i - \mu_i)^T \mathbf{A}_i(\beta)^{-1} (\mathbf{y}_i - \mu_i)\right]^{1/2} \\ &:= J_{11}(\beta, \mathbf{x})^{1/2} \cdot J_{12}(\beta)^{1/2}. \end{aligned} \tag{3.3.50}$$

For the first term of the product in (3.3.50), we have:

$$\begin{aligned}
J_{11}(\beta, \mathbf{x}) &\le \max_{i\le n} \lambda^4_{\max}[\mathcal{R}_{i-1}(\beta)^{-1}] \cdot \max_{i\le n} \lambda^2_{\max}[\mathbf{Q}^{(l)}_{i-1}(\beta)] \cdot \max_{i\le n} \lambda^2_{\max}[\mathbf{A}_i^{-1/2}\mathbf{A}_i(\beta)^{1/2}] \\
&\quad \cdot \sum_{i=1}^{n} \mathbf{x}^T \mathbf{X}_i^T \mathbf{A}_i \mathbf{X}_i \mathbf{x} \\
&\le C\pi_n^4(r) q_n^2(r) \max_{i\le n} \lambda^2_{\max}[\mathbf{A}_i^{-1/2}\mathbf{A}_i(\beta)^{1/2}] \lambda_{\max}(\mathbf{H}_n^{\text{indep}}),
\end{aligned}$$

where we used (3.3.19) and the definition of $q_n(r)$ for the last inequality. By Lemma 3.2.1, for any $\beta \in B_r$, any $r \in (0, r_0)$ and any $n \ge n_0$ we have:

$$[\mathbf{A}_i^{-1/2}\mathbf{A}_i(\beta)^{1/2}]_{jj} = \left[\frac{\mu'(\mathbf{x}_{ij}^T\beta)}{\mu'(\mathbf{x}_{ij}^T\beta_0)}\right]^{1/2} \le \eta_n(r) + 1 \le C,$$

and so $\max_{i\le n} \lambda^2_{\max}[\mathbf{A}_i^{-1/2}\mathbf{A}_i(\beta)^{1/2}] \le C$, for any $\beta \in B_r$, any $r \in (0, r_0)$ and any $n \ge n_0$. Therefore, for any $r \in (0, r_0)$ and any $n \ge n_0$, we obtain:

$$\mathrm{E}[\sup_{\beta\in B_r} \sup_{\|\mathbf{x}\|=1} J_{11}(\beta, \mathbf{x})] \le C\mathrm{E}[\pi_n^4(r) q_n^2(r)] \lambda_{\max}(\mathbf{H}_n^{\text{indep}}). \tag{3.3.51}$$

For the second term in product (3.3.50), we have:

$$\begin{aligned}
J_{12}(\beta, \mathbf{x}) &\le \sum_{i-1}^{n} (\mathbf{y}_i - \mu_i)^T \mathbf{A}_i^{-1/2} \mathbf{A}_i^{1/2} \mathbf{A}_i(\beta)^{-1/2} \mathbf{A}_i^{1/2} \mathbf{A}_i^{-1/2} (\mathbf{y}_i - \mu_i) \\
&\le \max_{i\le n} \lambda^2_{\max}[\mathbf{A}_i(\beta)^{-1/2}\mathbf{A}_i^{1/2}] \cdot \sum_{i-1}^{n} (\mathbf{y}_i - \mu_i)^T \mathbf{A}_i^{-1} (\mathbf{y}_i - \mu_i).
\end{aligned}$$

By Lemma 3.2.1, for any $\beta \in B_r$, any $r \in (0, r_0)$ and any $n \ge n_0$ we have:

$$[\mathbf{A}_i(\beta)^{-1/2}\mathbf{A}_i^{1/2}]_{jj} = \left[\frac{\mu'(\mathbf{x}_{ij}^T\beta_0)}{\mu'(\mathbf{x}_{ij}^T\beta)}\right]^{1/2} \le \eta_n(r) + 1 \le C,$$

and so $\max_{i\le n} \lambda^2_{\max}[\mathbf{A}_i(\beta)^{-1/2}\mathbf{A}_i^{1/2}] \le C$, for any $\beta \in B_r$, any $r \in (0, r_0)$ and any $n \ge n_0$. Therefore, $\sup_{\beta\in B_r} \sup_{\|\mathbf{x}\|=1} J_{12}(\beta, \mathbf{x}) \le C \sum_{i-1}^{n} (\mathbf{y}_i - \mu_i)^T \mathbf{A}_i^{-1} (\mathbf{y}_i - \mu_i)$. Using (3.3.35), we obtain:

$$\mathrm{E}[\sup_{\beta\in B_r} \sup_{\|\mathbf{x}\|=1} J_{12}(\beta, \mathbf{x})] \le Cn, \ \forall r \in (0, r_0), \ \forall n \ge n_0. \tag{3.3.52}$$

Using Cauchy-Schwarz inquality and relations (3.3.50), (3.3.51) and (3.3.52), we obtain:

$$\begin{aligned} \mathrm{E}[\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1}|U_n^{(7)}(\beta,\mathbf{x})|] &\leq \{\mathrm{E}[\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1}J_{11}(\beta.\mathbf{x})]\}^{1/2}\cdot\{\mathrm{E}[\sup_{\beta\in B_r}\sup_{\|\mathbf{x}\|=1}J_{12}(\beta,\mathbf{x})]\}^{1/2} \\ &\leq C\{n\mathrm{E}[\pi_n^4(r)q_n^2(r)]\lambda_{\max}(\mathbf{H}_n^{\mathrm{indep}})\}^{1/2}, \end{aligned}$$

for any $r \in (0, r_0)$ and any $n \geq n_0$.

Relation (3.3.32) follows from (C_5). □

3.4 The Strong Consistency Theorem

In this section prove the existence of a strongly consistent sequence of estimators of β_0, defined as roots of the equation (2.2.1).

We introduce the following assumptions:

(H') there exists a constant $C > 0$ such that $\lambda_{\min}(\overline{\mathbf{R}}_n) \geq C, \forall n \geq 1$,

(D') $\lambda_{\min}(\mathbf{H}_n^{\mathrm{indep}}) \to \infty$,

(S'_δ) there exist an integer $N \geq 1$ and some constants $\delta > 0, c_0 > 0$ such that:

$$\lambda_{\min}(\mathbf{H}_n^{\mathrm{indep}}) \geq c_0[\lambda_{\max}(\mathbf{H}_n^{\mathrm{indep}})]^{1/2+\delta}, \quad \forall n \geq N,$$

Theorem 3.4.1 *Suppose that (AH), (K) and (H'), hold. Let $\{\mathcal{R}_n(\cdot)\}_n$ be a sequence of random matrices which satisfy (A), (B), (R'), as well as:*

(E) there exists a constant $C > 0$ such that $\lambda_{\max}(\mathcal{R}_n) \leq C$, $\forall n \geq 1$, a.s.

Suppose that conditions (C_1)-(C_5) are satisfied with $\alpha_n = \lambda_{\max}(\mathbf{H}_n^{\mathrm{indep}})$. If (D') and (S'_δ) hold then there exists a sequence $\{\widehat{\beta}_n\}_n \subset \mathcal{T}$ and a random number n_0 such that:

(a) $P(\mathbf{g}_n(\widehat{\beta}_n) = 0,$ for all $n \geq n_0) = 1$;

(b) $\widehat{\beta}_n \to \beta_0$ a.s.

Proof. We apply Theorem 3.1.1 to the sequence $\{\mathbf{g}_n, \mathcal{F}_n\}_{n\geq 1}$. by taking $\alpha_n = \lambda_{\max}(\mathbf{H}_n^{\text{indep}})$.

We first prove $\{\mathbf{g}_n, \mathcal{F}_n\}_{n\geq 1}$ is a zero-mean martingale. By definition $\mathbf{g}_n$ is $\mathcal{F}_n$ - measurable. We have:

$$\begin{aligned}
\mathrm{E}(\mathbf{g}_n) &= \sum_{i=1}^{n} \mathrm{E}[\mathbf{X}_i^T \mathbf{A}_i^{1/2} \mathcal{R}_{i-1}^{-1} \mathbf{A}_i^{-1/2} (\mathbf{y}_i - \mu_i(\beta))] \\
&= \sum_{i=1}^{n} \mathrm{E}\{\mathrm{E}[\mathbf{X}_i^T \mathbf{A}_i^{1/2} \mathcal{R}_{i-1}^{-1} \mathbf{A}_i^{-1/2} (\mathbf{y}_i - \mu_i(\beta))] | \mathcal{F}_{i-1}\} \\
&= \sum_{i=1}^{n} \mathrm{E}[\mathbf{X}_i^T \mathbf{A}_i^{1/2} \mathcal{R}_{i-1}^{-1} \mathbf{A}_i^{-1/2} \mathrm{E}(\mathbf{y}_i - \mu_i(\beta) | \mathcal{F}_{i-1})] = 0,
\end{aligned}$$

where we used the fact that the entries of $\mathcal{R}_{i-1}$ are $\mathcal{F}_{i-1}$ - measurable and that $\mathrm{E}(\mathbf{y}_i - \mu_i(\beta) | \mathcal{F}_{i-1}) = \mathrm{E}(\mathbf{y}_i - \mu_i(\beta)) = 0$ for all $i \geq 1$.

Also, by the same arguments:

$$\begin{aligned}
\mathrm{E}(\mathbf{g}_n | \mathcal{F}_{n-1}) &= \mathrm{E}[\mathbf{g}_{n-1} + \mathbf{X}_n^T \mathbf{A}_n^{1/2} \mathcal{R}_{n-1}^{-1} \mathbf{A}_n^{-1/2} (\mathbf{y}_i - \mu_i(\beta)) | \mathcal{F}_{n-1}] \\
&= \mathbf{g}_{n-1} + \mathbf{X}_n^T \mathbf{A}_n^{1/2} \mathcal{R}_{n-1}^{-1} \mathbf{A}_n^{-1/2} \mathrm{E}(\mathbf{y}_i - \mu_i(\beta) | \mathcal{F}_{n-1}) \\
&= \mathbf{g}_{n-1}
\end{aligned}$$

Recall that $\mathbf{M}_n = \mathrm{Cov}(\mathbf{g}_n)$. Similarly to (2.3.2), (with $\mathcal{R}_{i-1}$ in lieu of $\mathcal{R}_{i-1}^*$), we have $\mathbf{M}_n = \sum_{i=1}^n \mathbf{X}_i \mathbf{A}_i^{1/2} \mathbf{E}_{i-1} \mathbf{A}_i^{1/2} \mathbf{X}_i$, where $\mathbf{E}_{i-1} = E(\mathcal{R}_{i-1}^{-1} \overline{\mathbf{R}}_i \mathcal{R}_{i-1}^{-1})$.

We prove that (3.1.1) holds Using (R') and the fact that $\overline{\mathbf{R}}_i \leq m\mathbf{I}$ for all $i \geq 1$, it follows that $\mathcal{R}_{i-1}^{-1} \overline{\mathbf{R}}_i \mathcal{R}_{i-1}^{-1} \leq C\mathbf{I}$ for all $i \geq 1$, a.s. Hence $\mathbf{E}_i \leq C\mathbf{I}$ for all $i \geq 1$ and

$$\mathbf{M}_n \leq C\mathbf{H}_n^{\text{indep}}, \quad \forall n \geq 1.$$

Hence $\lambda_{\max}(\mathbf{M}_n) \leq C\lambda_{\max}(\mathbf{H}_n^{\text{indep}})$, i.e. $\{\alpha_n\}_{n\geq 1}$ satisfies relation (3.1.1).

We now prove that (I) holds. For any $p \times 1$ vector $\mathbf{x}$ with $\|\mathbf{x}\| = 1$, we have:

$$\begin{aligned}
\sum_{i=1}^{n} \mathbf{x}^T \mathbf{X}_i^T \mathbf{A}_i^{1/2} \mathcal{R}_{i-1}^{-1} \overline{\mathbf{R}}_i \mathcal{R}_{i-1}^{-1} \mathbf{A}_i^{1/2} \mathbf{X}_i \mathbf{x} &\geq \min_{i\leq n} \lambda_{\min}(\overline{\mathbf{R}}_i) \min_{i\leq n} \lambda_{\min}(\mathcal{R}_{i-1}^{-2}) \cdot \\
&\cdot \ \mathbf{x}^T \mathbf{H}_n^{\text{indep}} \mathbf{x}
\end{aligned}$$

$$\geq C_0 \mathbf{x}^T \mathbf{H}_n^{\text{indep}} \mathbf{x},$$

using (II') and (E). Taking the expectation (with respect to P_{β_0}), we conclude that $\mathbf{x}^T \mathbf{M}_n \mathbf{x} \geq C_0 \mathbf{x}^T \mathbf{H}_n^{\text{indep}} \mathbf{x}$, i.e.

$$\mathbf{M}_n \geq C_0 \mathbf{H}_n^{\text{indep}}, \quad \forall n \geq 1.$$

The fact that (I) holds follows from our hypothesis (D').

We now prove that (S) holds. Part (ii) follows directly from Lemmas 3.3.3, 3.3.4 and 3.3.6. To prove that parts (i) and (iii) hold, note that condition (E) and our hypothesis (S'_δ) imply that for any $p \times 1$ vector $\mathbf{x}$ with $\|\mathbf{x}\| = 1$:

$$\begin{aligned} \mathbf{x}^T \mathbf{H}_n \mathbf{x} &= \sum_{i=1}^n \mathbf{x}^T \mathbf{X}_i^T \mathbf{A}_i^{1/2} \mathcal{R}_{i-1}^{-1} \mathbf{A}_i^{1/2} \mathbf{X}_i \mathbf{x} \geq \lambda_{\min}(\mathcal{R}_{i-1}^{-1}) \cdot \mathbf{x}^T \mathbf{H}_n^{\text{indep}} \mathbf{x} \\ &\geq C \quad \mathbf{x}^T \mathbf{H}_n^{\text{indep}} \mathbf{x} \geq C \alpha_n^{1/2+\delta}, \ \forall n \geq N. \end{aligned}$$

From Lemmas 3.3.3, 3.3.4 and 3.3.6, it follows that, with probability 1, there exist some random numbers $r_1 > 0, n_1 \geq 1$ such that

$$\alpha_n^{-1/2-\delta} |\mathbf{x}^T [\mathbf{H}_n(\beta) - \mathbf{H}_n] \mathbf{x} + \mathbf{x}^T [\mathbf{B}_n(\beta) + \mathcal{E}_n(\beta)] \mathbf{x}| < C/2, \quad \forall \beta \in B_{r_1}, n \geq n_1.$$

Recalling that $\mathcal{D}_n(\beta) = \mathbf{H}_n(\beta) - \mathbf{B}_n(\beta) - \mathcal{E}_n(\beta)$, we conclude that:

$$\begin{aligned} |\mathbf{x}^T \mathcal{D}_n(\beta) \mathbf{x}| &\geq |\mathbf{x}^T \mathbf{H}_n \mathbf{x}| - |\mathbf{x}^T [\mathbf{H}_n(\beta) - \mathbf{H}_n] \mathbf{x} + \mathbf{x}^T [\mathbf{B}_n(\beta) + \mathcal{E}_n(\beta)] \mathbf{x}| \\ &\geq C\alpha_n^{1/2+\delta} - C\alpha_n^{1/2+\delta}/2 > 0, \quad \text{for all } \beta \in B_{r_1}, n \geq n_1, \end{aligned}$$

This concludes the proof of parts (i) and (iii) of (S). □

Remark 3.4.2 The proof of Theorem 3.4.1 can be adapted to apply to the "working independence" GEE introduced in Example 2.2.2. More precisely, assume that (H') and (K') hold. Under conditions (D') and (S'_δ) of Theorem 3.4.1, one can prove that there exist a sequence $\{\widetilde{\beta}_n\}_n$ and a random integer n_0 such that

$$P(\mathbf{g}_n^{\text{indep}}(\widetilde{\beta}_n) = 0, \forall n \geq n_0) = 1 \quad \text{and} \quad \widetilde{\beta}_n \to \beta_0 \text{ a.s.}$$

Our set-up covers a more general situation than Theorem 2 of [13]; in our case, neither the joint distribution of the data nor the covariance matrices are known.

Chapter 4

The Error-in-Variables Model

In this chapter we introduce a multivariate structural linear error-in-variables model which is suitable for longitudinal data. We construct estimators of the regression parameters, which correspond to the modified least squares estimators used in the univariate case (see [8]). We show that these estimators are consistent. We prove a central limit theorem, which is completely data-based, under the assumption that the vector of latent variables belongs to the generalized domain of attraction of the normal law. Our results can be viewed as an extension of the results of [27] to include the longitudinal case.

This chapter is organized as follows. In Section 4.1 we state our model assumptions and obtain the estimators of the parameters of interest. In Section 4.2 we introduce the concept of GDAN and give preliminary results, which will be used in the next sections. The main result of this section is Theorem 4.2.4, which states that the inner product between a vector in GDAN and a vector whose components have moments of order four is in DAN. In Section 4.3 we give conditions under which the estimators are either weakly or strongly consistent. The CLTs are presented in Section 4.4 and Section 4.5. We remark that throughout the chapter we consider the euclidean norm.

4.1 Model assumptions and estimators of the parameters

4.1.1 The model

We consider a sample of n individuals, whose responses are recorded, at fixed moments of time (which are denoted for simplicity by $1, 2, \ldots, m$). For any $i \in \{1, 2, \ldots, n\}$, let $\mathbf{y}_i = (y_{i1}, \ldots, y_{im})^T$ be the collection of m responses supplied by the i-th individual. More precisely, y_{ij} represents the response of the i-th individual at time j. Alternatively, we can think of n as a sample of independent clusters of equal size m. Within each cluster i, the observations $(y_{i1}, \ldots, y_{im})$ might be correlated.

Each response y_{ij} depends on a covariate ξ_{ij}, which is *unobservable* (or latent): instead of the covariate ξ_{ij}, one observes a surrogate variable x_{ij}. This happens for any individual $i \in \{1, 2, \ldots, n\}$, and for any occasion $j \in \{1, 2, \ldots, m\}$. For any $i \in \{1, 2, \ldots, n\}$, we denote by $\boldsymbol{\xi}_i = (\xi_{i1}, \ldots, \xi_{im})^T$ the collection of m unobservable covariates which correspond to the i-th individual, and by $\mathbf{x}_i = (x_{i1}, \ldots, x_{im})^T$ the collection of m respective surrogate variables.

Both variables y_{ij} and x_{ij} are observed with errors. More precisely, we consider the following error-in-variables model: for any $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$

$$\begin{aligned} y_{ij} &= \alpha + \beta \xi_{ij} + \varepsilon_{ij}, \\ x_{ij} &= \xi_{ij} + \delta_{ij}, \end{aligned} \tag{4.1.1}$$

where α, β are unknown parameters (of dimension 1), and $\varepsilon_{ij}, \delta_{ij}$ are the random error terms. For any $i \in \{1, \ldots, n\}$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{im})^T$ and $\boldsymbol{\delta}_i = (\delta_{i1}, \ldots, \delta_{im})^T$ represent the error random vectors.

We assume that $\{\boldsymbol{\xi}_i\}_{1 \leq i \leq n}$ is a sequence of i.i.d. random vectors with mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)^T$, i.e. $\mu_j := E(\xi_{ij})$, for $j \in \{1, \ldots, m\}$ and $i \in \{1, \ldots, n\}$. We denote for any $i \in \{1, \ldots n\}$, and $j, k \in \{1, \ldots m\}$

$$\sigma_{\boldsymbol{\xi},jj} := \mathrm{Var}(\xi_{ij}), \text{ and } \sigma_{\boldsymbol{\xi},jk} := \mathrm{Cov}(\xi_{ij}, \xi_{ik}).$$

In matrix notation, we write

$$\boldsymbol{\Sigma}_{\boldsymbol{\xi}} = (\sigma_{\boldsymbol{\xi},jk})_{1 \leq j,k \leq m}.$$

We assume that the errors $\{(\boldsymbol{\varepsilon}_i, \boldsymbol{\delta}_i)\}_{1 \leq i \leq n}$ form a sequence of i.i.d. zero-mean random vectors, and denote

$$\sigma_{\boldsymbol{\varepsilon},jk} := E(\varepsilon_{ij}\varepsilon_{ik}), \quad \sigma_{\boldsymbol{\delta},jk} := E(\delta_{ij}\delta_{ik}), \quad \sigma_{\boldsymbol{\varepsilon\delta},jk} := E(\varepsilon_{ij}\delta_{ik}),$$

for any $i \in \{1, \ldots, n\}$ and $j, k \in \{1, \ldots, m\}$. In matrix notation,

$$\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} := (\sigma_{\boldsymbol{\varepsilon},jk})_{1 \leq j,k \leq m}, \quad \boldsymbol{\Sigma}_{\boldsymbol{\delta}} := (\sigma_{\boldsymbol{\delta},jk})_{1 \leq j,k \leq m}, \quad \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon\delta}} := (\sigma_{\boldsymbol{\varepsilon\delta},jk})_{1 \leq j,k \leq m}.$$

We remark here that when there is no risk of confusion we use a generic notation $\boldsymbol{\xi}$, $\boldsymbol{\delta}$, $\boldsymbol{\varepsilon}$ to identify random vectors which have the same distributions as $\boldsymbol{\xi}_i$, $\boldsymbol{\delta}_i$ and $\boldsymbol{\varepsilon}_i$ respectively, $i \leq n$.

4.1.2 Estimation

We are interested in the estimation of α and β. To obtain consistent estimators for α and β, we assume that the diagonal elements of $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon\delta}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\delta}}$ are known.

For any $i \leq n, j \leq m$, we have:

$$\begin{aligned}
E(x_{ij}) &= E(\xi_{ij}) + E(\delta_{ij}) = \mu_j, \\
E(y_{ij}) &= \alpha + \beta E(\xi_{ij}) + E(\varepsilon_{ij}) = \alpha + \beta\mu_j, \\
\mathrm{Var}(x_{ij}) &= \mathrm{Var}(\xi_{ij}) + \mathrm{Var}(\delta_{ij}) = \sigma_{\boldsymbol{\xi},jj} + \sigma_{\boldsymbol{\delta},jj}, \\
\mathrm{Var}(y_{ij}) &= \beta^2 \mathrm{Var}(\xi_{ij}) + \mathrm{Var}(\varepsilon_{ij}) = \beta^2 \sigma_{\boldsymbol{\xi},jj} + \sigma_{\boldsymbol{\varepsilon},jj}, \\
\mathrm{Cov}(x_{ij}, y_{ij}) &= \beta \mathrm{Var}(\xi_{ij}) + \mathrm{Cov}(\varepsilon_{ij}, \delta_{ij}) = \beta\sigma_{\boldsymbol{\xi},jj} + \sigma_{\boldsymbol{\varepsilon\delta},jj}.
\end{aligned}$$

Using the method of moments (see also [8], Section 1.3.1), we obtain the following system of equations, for any $j \in \{1, \dots, m\}$:

$$\widehat{\mu}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij} := \overline{x}_j, \tag{4.1.2}$$

$$\widehat{\alpha} + \widehat{\beta}\widehat{\mu}_j = \frac{1}{n}\sum_{i=1}^{n} y_{ij} := \overline{y}_j, \tag{4.1.3}$$

$$\widehat{\sigma}_{\xi,jj} + \sigma_{\delta,jj} = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2, \tag{4.1.4}$$

$$\widehat{\beta}^2\widehat{\sigma}_{\xi,jj} + \widehat{\sigma}_{\varepsilon,jj} = \frac{1}{n}\sum_{i=1}^{n}(y_{ij} - \overline{y}_j)^2, \tag{4.1.5}$$

$$\widehat{\beta}\widehat{\sigma}_{\xi,jj} + \sigma_{\varepsilon\delta,jj} = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)(y_{ij} - \overline{y}_j). \tag{4.1.6}$$

By taking the sum of the m equations in (4.1.2) and (4.1.3), we obtain an estimator of α:

$$\widehat{\alpha}_n = \overline{\overline{y}} - \widehat{\beta}_n \overline{\overline{x}}, \tag{4.1.7}$$

where

$$\overline{\overline{x}} = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} x_{ij} \quad \text{and} \quad \overline{\overline{y}} = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij}.$$

It is useful to develop a vector notation. Let

$$\overline{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i \quad \text{and} \quad \overline{\mathbf{y}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{y}_i,$$

be the sample mean of $(\mathbf{x}_1, \dots, \mathbf{x}_n)$, respectively $(\mathbf{y}_1, \dots, \mathbf{y}_n)$. Note that $\overline{\mathbf{x}}$ is an m-dimensional random vector whose components are $\overline{x}_j, 1 \le j \le m$. Similarly, $\overline{\mathbf{y}} = (\overline{y}_1, \dots, \overline{y}_m)^T$.

By taking the sum of the m equations in (4.1.4) and (4.1.6), we obtain

$$\widehat{\beta}_n = \frac{\sum_{i=1}^{n}[(\mathbf{x}_i - \overline{\mathbf{x}})^T(\mathbf{y}_i - \overline{\mathbf{y}}) - \mathrm{trace}(\boldsymbol{\Sigma}_{\varepsilon\delta})]}{\sum_{i=1}^{n}[\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \mathrm{trace}(\boldsymbol{\Sigma}_{\delta})]}, \tag{4.1.8}$$

where $\mathrm{trace}(\boldsymbol{\Sigma}_{\varepsilon\delta}) = \sum_{j=1}^{m}\sigma_{\varepsilon\delta,jj}$ and $\mathrm{trace}(\boldsymbol{\Sigma}_{\delta}) = \sum_{j=1}^{m}\sigma_{\delta,jj}$. The estimator $\widehat{\beta}_n$ is obtained under the additional assumption $\sum_{i=1}^{n}[\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \mathrm{trace}(\boldsymbol{\Sigma}_{\delta})] > 0$. We remark that a similar assumption is required in the univariate case to obtain the modified least square estimators (see [8]).

Remark 4.1.1 The errors could be taken to be independent, as this is the case which is more often encountered in the literature. The fact that we allow for some degree of correlation of the errors, makes the model more general, but all the results could also be carried out in the case when $\mathbf{\Sigma}_{\boldsymbol{\varepsilon\delta}} = \mathbf{0}_{m\times m}$.

Remark 4.1.2 It is known that the error-in-variables models are, in general, not identifiable and therefore, it becomes impossible to consistently estimate the parameters from data. In a multivariate context, the assumption that $\mathbf{\Sigma}_{\boldsymbol{\delta}}$ is known ($\mathbf{\Sigma}_{\boldsymbol{\varepsilon\delta}} = \mathbf{0}_{m\times m}$) is needed to guarantee identifiability of the model (see [19]). In the univariate case (i.e. $m = 1$), the same assumption is used to construct the modified least squares estimators of α and β (see [8]).

The information regarding the diagonal of the covariance matrix of the error $\boldsymbol{\delta}$ might be available from previous studies or, if additional replications are available, trace($\mathbf{\Sigma}_{\boldsymbol{\delta}}$) might be estimated. In the latter case, two or more replications are needed, at each occasion, for each individual.

Assume two replications are available, $x_{ij}^{(r)}$, with $r = 1,\ 2$ (the r-th replication taken on individual i, at time j):

$$\begin{aligned}\mathbf{x}_i^{(1)} &= \boldsymbol{\xi}_i + \boldsymbol{\delta}_i^{(1)},\\ \mathbf{x}_i^{(2)} &= \boldsymbol{\xi}_i + \boldsymbol{\delta}_i^{(2)}.\end{aligned}$$

Then a consistent estimator of trace($\mathbf{\Sigma}_{\boldsymbol{\delta}}$) is

$$\frac{1}{2(n-1)}\sum_{i=1}^{n}\|(\mathbf{x}_i^{(1)} - \overline{\mathbf{x}^{(1)}}) - (\mathbf{x}_i^{(2)} - \overline{\mathbf{x}^{(2)}})\|^2,$$

where $\overline{\mathbf{x}^{(\mathbf{r})}} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i^{(r)}$, for $r = 1,\ 2$.

Remark 4.1.3 A different estimator of β can be obtained using (4.1.5) and (4.1.6), when the diagonal elements of $\Sigma_{\varepsilon\delta}$ and Σ_{ε} (instead of Σ_{δ}) are assumed to be known. Its form is similar to (4.1.8) and the asymptotic results are similar to those derived in this chapter.

Remark 4.1.4 The error-in-variables model with replications is a particular case of model (4.1.1), obtained when $\xi_{i1} = \xi_{i2} = \ldots = \xi_{im}$. More precisely, if we consider the model:

$$\begin{aligned} y_{ij} &= \alpha + \beta\xi_i + \varepsilon_{ij}, \\ x_{ij} &= \xi_i + \delta_{ij}, \end{aligned}$$

using the same method, the estimators of α and β have the same form as in (4.1.7) and (4.1.8), respectively.

The goal is to evaluate the asymptotic properties of the estimators defined by (4.1.7) and (4.1.8). This purpose requires the introduction of additional model assumptions as follows.

The distribution of $\boldsymbol{\xi}_1$ is assumed to be *full*, i.e. for each m-dimensional vector of norm 1, $\mathbf{u}$, the random variable $\mathbf{u}^T\boldsymbol{\xi}_1$ is not a constant, almost surely. This is a standard assumption in the context of generalized domain of attraction and allows us to use results obtained in [26] and [17]. It also guarantees that $\Sigma_{\boldsymbol{\xi}} > 0$ whenever the components of $\boldsymbol{\xi}_1$ have finite variance.

Recall that $\boldsymbol{\xi}_1 - \boldsymbol{\mu}$ has a *symmetric* distribution if its distribution is equal to the distribution of $-(\boldsymbol{\xi}_1 - \boldsymbol{\mu})$.

Our model assumptions are the following:

$(A1)$ $\boldsymbol{\xi}_1$ lies in the generalized domain of attraction of the normal law (GDAN), $\boldsymbol{\xi}_1 - \boldsymbol{\mu}$ has a symmetric distribution, whenever the matrix $\Sigma_{\boldsymbol{\xi}}$ has at least one element on the diagonal which is ∞

$(A2)$ $E(\varepsilon_{1j}) = 0,\ E(\delta_{1j}) = 0,\ E(\varepsilon_{1j}^4) < \infty,\ E(\delta_{1j}^4) < \infty$ with $1 \le j \le m$,

and $\Sigma_{\text{error}} := \begin{pmatrix} \Sigma_\varepsilon & \Sigma_{\varepsilon\delta} \\ \Sigma_{\varepsilon\delta}^T & \Sigma_\delta \end{pmatrix}$ is positive definite

$(A3)$ $(\boldsymbol{\xi}_i)_{1\le i\le n}$ and $\{(\boldsymbol{\varepsilon}_i, \boldsymbol{\delta}_i)\}_{1\le i\le n}$ are independent.

We note that assumptions $(A1) - (A3)$ are the multidimensional versions of the assumptions used in [27].

We mention that, as in [27], our proofs cover two distinct cases: the case when all the components of the covariance matrix of the latent vector are finite, and the case when at least one of these components is infinite.

4.2 Preliminary results for GDAN

In this section we present the concept of the generalized domain of attraction of the normal law (GDAN), introduced in [20]. For additional properties, see [26] or [34]. We begin with a definition.

Definition 4.2.1 *Let $\boldsymbol{\xi}$ be an m-dimensional random vector. We say that $\boldsymbol{\xi}$ belongs to the generalized domain of attraction of the normal law (and we write $\boldsymbol{\xi} \in GDAN$) if there exists a sequence $(\mathbf{B}_n)_{n\ge1}$ of non-stochastic $m \times m$ matrices and a sequence $(\mathbf{A}_n)_{n\ge1}$of non-stochastic m-dimensional vectors, such that:*

$$\mathbf{B}_n \left(\sum_{i=1}^{n} \boldsymbol{\xi}_i - \mathbf{A}_n \right) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{I}), \tag{4.2.1}$$

where $(\boldsymbol{\xi}_i)_{1\le i\le n}$ is a sequence of i.i.d. random vectors with the same distribution as $\boldsymbol{\xi}$.

Remark 4.2.2 (i) If $m = 1$, the above definition coincides with the definition of the domain of attraction of the normal law (DAN). It is known that $\boldsymbol{\xi} \in GDAN$ implies that each component $\xi_j \in DAN$, for $j \le m$. In general, the converse of this statement is not true. However, Remark (ii) of [26] points out that in the case when $\boldsymbol{\xi}$ has a spherically symmetric distribution, the condition $\boldsymbol{\xi} \in GDAN$ becomes equivalent with the condition that each $\xi_j \in DAN$, for $j \leq m$ (or to $\|\boldsymbol{\xi}\| \in DAN$).

(ii) If $\boldsymbol{\xi} = (\xi_1, \dots, \xi_m)^T \in GDAN$, then $\mathrm{E}(\|\boldsymbol{\xi}\|^r) < \infty$, for $0 \leq r < 2$ (in particular, $\mu_j = \mathrm{E}(\xi_j) < \infty$, for all $j \leq m$) and the sequence $(\mathbf{A}_n)_{n\geq 1}$ can be taken as $\mathbf{A}_n = n\boldsymbol{\mu}$, where $\boldsymbol{\mu} = \mathrm{E}(\boldsymbol{\xi})$ (see also Remark (ii), p.193 in [26]).
Note that, if $\boldsymbol{\xi} \in GDAN$, since $\mathrm{E}(\|\boldsymbol{\xi}\|) < \infty$, the condition $\mathrm{Var}(\|\boldsymbol{\xi}\|) < \infty$ is equivalent to $\mathrm{E}(\|\boldsymbol{\xi}\|^2) < \infty$ (or $\mathrm{E}(|\xi_j|^2) < \infty$, for all $j \leq m$).

(iii) If $\mathrm{Var}(\|\boldsymbol{\xi}\|) < \infty$, the condition that the distribution of $\boldsymbol{\xi}$ is full is equivalent to $\boldsymbol{\Sigma}_{\boldsymbol{\xi}} > 0$.

(iv) If $\mathrm{Var}(\|\boldsymbol{\xi}\|) < \infty$, by the classical CLT, the sequence $(\mathbf{B}_n)_{n\geq 1}$ of (4.2.1) can be taken to be $\mathbf{B}_n = \sqrt{n}\ \boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{-1/2}$.

(v) The matrices $\mathbf{B}_n$ can be taken to be nonsingular and symmetric. For a complete discussion, see Remark (ii) in [26].

The next result is a consequence of Remarks (ii) and (iii), p. 193-194 in [26].

Lemma 4.2.3 *If $\boldsymbol{\xi} \in GDAN$ and $\mathrm{E}(\boldsymbol{\xi}) = \boldsymbol{\mu}$, then $\|\boldsymbol{\xi} - \boldsymbol{\mu}\| \in DAN$,*

$$\frac{1}{b_n}\left(\sum_{i=1}^{n} \|\boldsymbol{\xi}_i - \boldsymbol{\mu}\| - a_n\right) \xrightarrow{\mathcal{D}} N(0,1), \text{ and } \frac{\sum_{i=1}^{n} \|\boldsymbol{\xi}_i - \boldsymbol{\mu}\|^2}{b_n^2} \xrightarrow{P} 1, \tag{4.2.2}$$

where $(\boldsymbol{\xi}_i)_{1\leq i\leq n}$ are i.i.d copies of $\boldsymbol{\xi}$, $b_n^2 = \mathrm{trace}(\mathbf{B}_n^{-2})$, and $(\mathbf{B}_n)_{n\geq 1}$ is a sequence of matrices for which (4.2.1) *holds.*

The following theorem is an essential tool for the development of our asymptotic results. The result is important in its own right since it provides sufficient conditions for the inner product between a random vector in GDAN and a random vector with finite fourth order moments to be in the domain of attraction of the normal law. The case of $m = 1$ was fully resolved under more general conditions in [25] (see Theorem D.1.6). In the case of infinite variance, we assume that $\boldsymbol{\xi} - \boldsymbol{\mu}$ has a symmetric full distribution, as we rely on Theorem D.2.5 (see Theorem 3.4 in [16]). We use the method of the proof of Lemma 4 in [27], adapted to a multivariate setting.

Theorem 4.2.4 *Let $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_m)^T \in GDAN$ be an m-dimensional random vector with a full distribution. If $\mathrm{Var}(\|\boldsymbol{\xi}\|) = \infty$ we assume, in addition, that the distribution of $\boldsymbol{\xi} - \boldsymbol{\mu}$ is symmetric.*

Let $\boldsymbol{\varepsilon}$ be an m-dimensional random vector, such that $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ and $\mathrm{E}(|\varepsilon_j|^4) < \infty$, for all $j \leq m$. Assume that $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} > 0$, where $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ is the covariance matrix of $\boldsymbol{\varepsilon}$. If $\boldsymbol{\xi}$ and $\boldsymbol{\varepsilon}$ are independent, then,

$$\boldsymbol{\xi}^T \boldsymbol{\varepsilon} = \sum_{j=1}^{m} \xi_j \varepsilon_j \in \mathrm{DAN}.$$

Proof. *Case* 1. Assume $\mathrm{Var}(\|\boldsymbol{\xi}\|) < \infty$.

Since $\boldsymbol{\xi}$ is independent of $\boldsymbol{\varepsilon}$ and $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$,

$$\mathrm{Var}(\boldsymbol{\xi}^T \boldsymbol{\varepsilon}) = \mathrm{E}[(\boldsymbol{\xi}^T \boldsymbol{\varepsilon})^2] = \mathrm{E}[\mathrm{trace}(\boldsymbol{\xi}\boldsymbol{\xi}^T \boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T)] = \mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\xi}} \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}).$$

Hence, $\mathrm{Var}(\boldsymbol{\xi}^T \boldsymbol{\varepsilon}) < \infty$. since the covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}$ has only finite entries in this case.

Since $\boldsymbol{\xi}$ is full, $\boldsymbol{\Sigma}_{\boldsymbol{\xi}} > 0$ and so, using the fact that $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} > 0$, we obtain that $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{1/2} \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} \boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{1/2} > 0$. Therefore, $\mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\xi}} \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}) = \mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{1/2} \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} \boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{1/2}) > 0$ and so $\mathrm{Var}(\boldsymbol{\xi}^T \boldsymbol{\varepsilon}) > 0$. The conclusion follows by applying the CLT.

Case 2. Assume that $\mathrm{Var}(\|\boldsymbol{\xi}\|) = \infty$ (at least one component of $\boldsymbol{\xi}$ has infinite variance). In this case, we assume in addition that $\boldsymbol{\xi} - \boldsymbol{\mu}$ has a symmetric distribution.

Let $\{\boldsymbol{\xi}_i\}_{1 \leq i \leq n}$ and $\{\boldsymbol{\varepsilon}_i\}_{1 \leq i \leq n}$ be i.i.d copies of $\boldsymbol{\xi}$ and $\boldsymbol{\varepsilon}$, respectively. We denote $\boldsymbol{\xi}_i = (\xi_{i1}, \ldots, \xi_{im})^T$ and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \ldots, \varepsilon_{im})^T$.

(a) Assume that $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} = \mathbf{I}$. First, we consider the case when $\mathrm{E}(\boldsymbol{\xi}) = \mathbf{0}$.

Since $\boldsymbol{\xi} \in GDAN$, there exist a sequence of symmetric and non-singular non-stochastic matrices $\mathbf{B}_n$ such that $\mathbf{B}_n \sum_{i=1}^{n} \boldsymbol{\xi}_i \xrightarrow{\mathfrak{D}} N(\mathbf{0}, \mathbf{I})$. Denote $b_n^2 := \mathrm{trace}[(\mathbf{B}_n)^{-2}]$ and note that $b_n^2 \to \infty$. To show that $\boldsymbol{\xi}^T \boldsymbol{\varepsilon} \in DAN$, by Remark D.1.4, it suffices to prove that

$$b_n^{-2}\sum_{i=1}^{n}(\boldsymbol{\xi}_i^T\boldsymbol{\varepsilon}_i)^2 \xrightarrow{P} 1. \tag{4.2.3}$$

We have $b_n^{-2}\sum_{i=1}^{n}(\boldsymbol{\xi}_i^T\boldsymbol{\varepsilon}_i)^2 = \gamma_{n,1}\gamma_{n,2}$, where:

$$\gamma_{n,1} := \frac{\sum_{i=1}^{n}(\boldsymbol{\xi}_i^T\boldsymbol{\varepsilon}_i)^2}{\sum_{i=1}^{n}\boldsymbol{\xi}_i^T\boldsymbol{\xi}_i} \text{ and } \gamma_{n,2} := \frac{\sum_{i=1}^{n}\boldsymbol{\xi}_i^T\boldsymbol{\xi}_i}{b_n^2}.$$

By Lemma 4.2.3, $\gamma_{n,2} \xrightarrow{P} 1$.

In what follows we show that $\gamma_{n,1} \xrightarrow{P} 1$. Let $\epsilon > 0$ be fixed. Using Chebychev's inequality, we have the following:

$$\begin{aligned} P(|\gamma_{n,1} - 1| > \epsilon) &\leq \epsilon^{-2}\mathrm{E}\left[\left(\sum_{i=1}^{n}\frac{\boldsymbol{\xi}_i^T(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^T - \mathbf{I})\boldsymbol{\xi}_i}{\sum_{i=1}^{n}\boldsymbol{\xi}_i^T\boldsymbol{\xi}_i}\right)^2\right] \\ &= \epsilon^{-2}(I_1 + I_2), \end{aligned} \tag{4.2.4}$$

where:

$$\begin{aligned} I_1 &= \mathrm{E}\left[\sum_{i=1}^{n}\left(\frac{\boldsymbol{\xi}_i^T(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^T - \mathbf{I})\boldsymbol{\xi}_i}{\sum_{i=1}^{n}\boldsymbol{\xi}_i^T\boldsymbol{\xi}_i}\right)^2\right], \\ I_2 &= \mathrm{E}\left[\sum_{i\neq l}^{n}\frac{\boldsymbol{\xi}_i^T(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i^T - \mathbf{I})\boldsymbol{\xi}_i\boldsymbol{\xi}_l^T(\boldsymbol{\varepsilon}_l\boldsymbol{\varepsilon}_l^T - \mathbf{I})\boldsymbol{\xi}_l}{(\sum_{i=1}^{n}\boldsymbol{\xi}_i^T\boldsymbol{\xi}_i)^2}\right]. \end{aligned}$$

We treat I_1 first. Let $\mathbf{C}_n := \sum_{i=1}^{n}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T$ which can be assumed by Lemma D.2.4 to be non-singular. By Theorem A.1.3, we have the following evaluation:

$$\begin{aligned} [\boldsymbol{\xi}_1^T(\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1^T - \mathbf{I})\boldsymbol{\xi}_1]^2 &\leq \lambda_{\max}(\boldsymbol{\xi}_1\boldsymbol{\xi}_1^T)\lambda_{\max}[(\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1^T - \mathbf{I})^2]\boldsymbol{\xi}_1^T\boldsymbol{\xi}_1 \\ &\leq \mathrm{trace}(\boldsymbol{\xi}_1\boldsymbol{\xi}_1^T)\lambda_{\max}[(\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1^T - \mathbf{I})^2]\boldsymbol{\xi}_1^T\boldsymbol{\xi}_1 \\ &= \lambda_{\max}[(\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1^T - \mathbf{I})^2](\boldsymbol{\xi}_1^T\boldsymbol{\xi}_1)^2 \\ &\leq \lambda_{\max}[(\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1^T - \mathbf{I})^2]\lambda_{\max}^2(\mathbf{C}_n)(\boldsymbol{\xi}_1^T\mathbf{C}_n^{-1}\boldsymbol{\xi}_1)^2. \end{aligned}$$

By Remark A.1.11,

$$\lambda_{\max}(\mathbf{C}_n) = \lambda_{\max}(\sum_{i=1}^{n}\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T) \leq \sum_{i=1}^{n}\lambda_{\max}(\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T) \leq \sum_{i=1}^{n}\mathrm{trace}(\boldsymbol{\xi}_i\boldsymbol{\xi}_i^T) = \sum_{i=1}^{n}\boldsymbol{\xi}_i^T\boldsymbol{\xi}_i.$$

Using the independence between $\boldsymbol{\xi}$ and $\boldsymbol{\varepsilon}$ we obtain that

$$\begin{aligned} I_1 &= n\mathrm{E}\left[\left(\frac{\boldsymbol{\xi}_1^T(\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1^T - \mathbf{I})\boldsymbol{\xi}_1}{\sum_{i=1}^n \boldsymbol{\xi}_i^T\boldsymbol{\xi}_i}\right)^2\right] \\ &\leq n\mathrm{E}(\lambda_{\max}[(\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1^T - \mathbf{I})^2])\mathrm{E}[(\boldsymbol{\xi}_1^T\mathbf{C}_n^{-1}\boldsymbol{\xi}_1)^2]. \end{aligned}$$

Using Remark A.1.9, Remark A.1.10, the fact that $\mathrm{E}(\varepsilon_j^4) < \infty$, for all $j \leq m$ and the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned} \mathrm{E}\{\lambda_{\max}[(\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1^T - \mathbf{I})^2]\} &= \mathrm{E}(\|\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1^T - \mathbf{I}\|^2) \leq \sum_{j,k=1}^m \mathrm{E}[(\varepsilon_{1j}\varepsilon_{1k} - \delta_{jk})^2] \\ &\leq \sum_{j,k=1}^m \mathrm{E}(\varepsilon_{1j}^2\varepsilon_{1k}^2) \leq \sum_{j=1}^m \mathrm{E}(\varepsilon_{1j}^4) < \infty, \end{aligned}$$

Here we let $\delta_{jk} = 1$, if $j = k$ and $\delta_{jk} = 0$, if $j \neq k$.

By Theorem D.2.5, since $\boldsymbol{\xi} \in GDAN$, it follows that $n\mathrm{E}[(\boldsymbol{\xi}_1^T\mathbf{C}_n^{-1}\boldsymbol{\xi}_1)^2] \longrightarrow 0$. Hence,

$$I_1 \longrightarrow 0, \text{ as } n \to \infty. \tag{4.2.5}$$

We now treat I_2. Using the fact that $(\boldsymbol{\xi}_i)_{i\leq n}$ and $(\boldsymbol{\varepsilon}_i)_{i\leq n}$ are identical distributed, the independence between $(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2)$ and $(\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2)$ and the independence between $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\varepsilon}_2$, we obtain:

$$\begin{aligned} I_2 &= (n^2 - n)\mathrm{E}\left[\frac{\boldsymbol{\xi}_1^T(\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1^T - \mathbf{I})\boldsymbol{\xi}_1\boldsymbol{\xi}_2^T(\boldsymbol{\varepsilon}_2\boldsymbol{\varepsilon}_2^T - \mathbf{I})\boldsymbol{\xi}_2}{(\sum_{i=1}^n \boldsymbol{\xi}_i^T\boldsymbol{\xi}_i)^2}\right] \\ &= (n^2 - n)\sum_{j,k,l,p=1}^m \mathrm{E}(\varepsilon_{1j}\varepsilon_{1k} - \delta_{jk})\mathrm{E}(\varepsilon_{2l}\varepsilon_{2p} - \delta_{lp})\mathrm{E}\left[\frac{\xi_{1j}\xi_{1k}\xi_{2l}\xi_{2p}}{(\sum_{i=1}^n \boldsymbol{\xi}_i^T\boldsymbol{\xi}_i)^2}\right] \\ &= 0, \end{aligned} \tag{4.2.6}$$

since, $\mathrm{E}(\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1^T) = \mathbf{I}$. Note that by Cauchy-Schwarz inequality:

$$\begin{aligned} \left|\mathrm{E}\left[\frac{\xi_{1j}\xi_{1k}\xi_{2l}\xi_{2p}}{(\sum_{i=1}^n \boldsymbol{\xi}_i^T\boldsymbol{\xi}_i)^2}\right]\right| &\leq \left\{\mathrm{E}\left[\frac{\xi_{1j}}{(\sum_{i=1}^n \|\boldsymbol{\xi}_i\|^2)^{1/2}}\right]^4 \mathrm{E}\left[\frac{\xi_{1k}}{(\sum_{i=1}^n \|\boldsymbol{\xi}_i\|^2)^{1/2}}\right]^4 \cdot\right. \\ &\quad \left. \cdot \ \mathrm{E}\left[\frac{\xi_{2l}}{(\sum_{i=1}^n \|\boldsymbol{\xi}_i\|^2)^{1/2}}\right]^4 \mathrm{E}\left[\frac{\xi_{2p}}{(\sum_{i=1}^n \|\boldsymbol{\xi}_i\|^2)^{1/2}}\right]^4\right\}^{1/4} \end{aligned}$$

$$\leq \ 1.$$

From (4.2.4), (4.2.5) and (4.2.6), it follows that $\gamma_{n.1} \xrightarrow{P} 1$. This concludes the proof of (4.2.3), when $\mathrm{E}(\boldsymbol{\xi}) = \mathbf{0}$.

Now, consider the case when $\mathrm{E}(\boldsymbol{\xi}) = \boldsymbol{\mu}$. Since $\boldsymbol{\xi} \in GDAN$, it follows that $(\boldsymbol{\xi} - \boldsymbol{\mu}) \in GDAN$ and $\mathrm{E}(\boldsymbol{\xi} - \boldsymbol{\mu}) = \mathbf{0}$. Hence, by applying the first part of the proof, we obtain $(\boldsymbol{\xi} - \boldsymbol{\mu})^T\boldsymbol{\varepsilon} \in DAN$. We write:

$$\boldsymbol{\xi}^T\boldsymbol{\varepsilon} = (\boldsymbol{\xi} - \boldsymbol{\mu})^T\boldsymbol{\varepsilon} + \boldsymbol{\mu}^T\boldsymbol{\varepsilon},$$

and note that $0 < \mathrm{Var}(\boldsymbol{\mu}^T\boldsymbol{\varepsilon}) < \infty$. Hence, by CLT. $\boldsymbol{\mu}^T\boldsymbol{\varepsilon} \in DAN$ and since $\mathrm{E}((\boldsymbol{\xi} - \boldsymbol{\mu})^T\boldsymbol{\varepsilon} \cdot \boldsymbol{\mu}^T\boldsymbol{\varepsilon}) < \infty$, we apply Lemma D.1.9 to obtain $\boldsymbol{\xi}^T\boldsymbol{\varepsilon} \in DAN$.

(b) Assume $\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}$ is a positive definite covariance matrix. It follows that $\boldsymbol{\varepsilon}' := \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}^{-1/2}\boldsymbol{\varepsilon}$ is such that $\mathrm{E}(\boldsymbol{\varepsilon}') = \mathbf{0}$ and $\mathrm{E}[\boldsymbol{\varepsilon}'(\boldsymbol{\varepsilon}')^T] = \mathbf{I}$. If $\boldsymbol{\xi} \in GDAN$, there exist a sequence of matrices $\mathbf{B}_n$ such that $\mathbf{B}_n(\sum_{i=1}^n \boldsymbol{\xi}_i - n\mu) \xrightarrow{\mathfrak{D}} N(0, \mathbf{I})$. It follows that $\boldsymbol{\xi}' := \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}{}^{1/2}\boldsymbol{\xi} \in GDAN$, with $\mathbf{B}_n' := \mathbf{B}_n\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}}{}^{-1/2}$ as the sequence of normalizing matrices. We apply part (a) of the proof and obtain $(\boldsymbol{\xi}')^T\boldsymbol{\varepsilon}' = \boldsymbol{\xi}^T\boldsymbol{\varepsilon} \in DAN$. □

Lemma 4.2.5 *Assume that* $\boldsymbol{\xi} \in GDAN$, $\mathrm{E}(\boldsymbol{\xi}) = 0$ *and* $\mathbf{B}_n \sum_{i=1}^n \boldsymbol{\xi}_i \xrightarrow{\mathfrak{D}} N(\mathbf{0}, \mathbf{I})$, *where* $(\boldsymbol{\xi})_{1\leq i\leq n}$ *are i.i.d copies of* $\boldsymbol{\xi}$. *Let* $\boldsymbol{\Sigma}$ *be a positive definite matrix and denote* $\boldsymbol{\xi}' = \boldsymbol{\Sigma}^{1/2}\boldsymbol{\xi}$. *Then* $\boldsymbol{\xi}' \in GDAN$, *and* $\mathbf{B}_n' \sum_{i=1}^n \boldsymbol{\xi}_i' \xrightarrow{\mathfrak{D}} N(\mathbf{0}, \mathbf{I})$, *where* $\mathbf{B}_n' = (\boldsymbol{\Sigma}^{-1/2}\mathbf{B}_n^T\mathbf{B}_n\boldsymbol{\Sigma}^{-1/2})^{1/2}$.

Proof. We denote by $\mathbf{C}_n := \mathbf{B}_n\boldsymbol{\Sigma}^{-1/2}$ and so $\mathbf{C}_n \sum_{i=1}^n \boldsymbol{\xi}_i' \xrightarrow{\mathfrak{D}} N(0, \mathbf{I})$. We apply Lemma D.2.1 to conclude that:

$$(\mathbf{C}_n^T\mathbf{C}_n)^{1/2}\sum_{i=1}^n \boldsymbol{\xi}_i' \xrightarrow{\mathfrak{D}} N(0, \mathbf{I}). \ \square$$

The next result gives the rate of convergence for the trace of the sequence $(\mathbf{B}_n)_{n\geq 1}$ of normalizing matrices. It will be used frequently in the proofs of the main results.

Lemma 4.2.6 *Let $\boldsymbol{\xi} \in GDAN$, $\mathrm{E}(\boldsymbol{\xi}) = \boldsymbol{\mu}$ and $(\mathbf{B}_n)_{n\geq 1}$ be a sequence for which* (4.2.1) *holds, with $\mathbf{A}_n = n\boldsymbol{\mu}$.*

(a) (i) *If* $\mathrm{Var}(\|\boldsymbol{\xi}\|) < \infty$, $\dfrac{n}{\mathrm{trace}(\mathbf{B}_n^{-2})} = \dfrac{1}{\mathrm{Var}(\|\boldsymbol{\xi} - \boldsymbol{\mu}\|)}$, *for any $n \geq 1$.*

(ii) *If* $\mathrm{Var}(\|\boldsymbol{\xi}\|) = \infty$, $\dfrac{n}{\mathrm{trace}(\mathbf{B}_n^{-2})} \to 0$.

(b) $\dfrac{\mathrm{trace}(\mathbf{B}_n^{-2})}{n^2} \to 0$, *as $n \to \infty$.*

Proof. By Lemma 4.2.3, $\|\boldsymbol{\xi} - \boldsymbol{\mu}\| \in DAN$ and $\dfrac{1}{b_n}\left(\sum_{i=1}^{n} \|\boldsymbol{\xi}_i - \boldsymbol{\mu}\| - a_n\right) \xrightarrow{\mathcal{D}} N(0,1)$, where $(\boldsymbol{\xi}_i)_{i\leq n}$ are i.i.d copies of $\boldsymbol{\xi}$ and $a_n = n\mathrm{E}(\|\boldsymbol{\xi}_1 - \boldsymbol{\mu}\|)$.

(i) If $\mathrm{Var}(\|\boldsymbol{\xi}\|) < \infty$, it follows that $\sigma^2 := \mathrm{Var}(\|\boldsymbol{\xi} - \boldsymbol{\mu}\|) < \infty$. Then $b_n^2 = n\sigma^2$ and so $\dfrac{n}{b_n^2} = \dfrac{1}{\sigma^2}$ and $\dfrac{b_n^2}{n^2} = \dfrac{\sigma^2}{n} \to 0$.

(ii) If $\mathrm{Var}(\|\boldsymbol{\xi}\|) = \infty$, then $\mathrm{Var}(\|\boldsymbol{\xi} - \boldsymbol{\mu}\|) = \infty$ and $b_n^2 = nl^2(n)$, where $l(n)$ is a slowly varying function at infinity. Hence $\dfrac{n}{b_n^2} = \dfrac{1}{l^2(n)} \to 0$ and $\dfrac{b_n^2}{n^2} = \dfrac{l^2(n)}{n} \to 0$. □

4.3 Consistency

In this section, we examine the consistency property of our estimators, which is relatively straight-forward. Lemma 4.3.2 shows that in the case when all components of $\boldsymbol{\xi}$ have finite variance, the estimators given by (4.1.7) and (4.1.8) are strongly consistent. The assumptions required for weak consistency are slightly weaker than the assumptions $(A1) - (A3)$, which are required for the CLTs proved in the next sections. More precisely, here we replace $(A1)$ by

$$(A1') \qquad \boldsymbol{\xi}_1 \in GDAN.$$

Theorem 4.3.1 *Assume that $(A1')$ and $(A3)$ are satisfied. Then, the estimators $\widehat{\beta}_n$ and $\widehat{\alpha}_n$, given by* (4.1.8) *and* (4.1.7) *are weakly consistent, i.e. $\widehat{\beta}_n \xrightarrow{P} \beta$ and $\widehat{\alpha}_n \xrightarrow{P} \alpha$.*

Proof. Let $b_n^2 = \text{trace}(\mathbf{B}_n^{-2})$, where $(\mathbf{B}_n)_{n\geq 1}$ is a sequence of normalizing matrices for which (4.2.1) holds, with $\mathbf{A}_n = n\boldsymbol{\mu}$. From (4.1.8), the consistency of $\widehat{\beta}_n$ follows once we proved that:

$$\frac{\sum_{i=1}^n \left[(\mathbf{x}_i - \overline{\mathbf{x}})^T(\mathbf{y}_i - \overline{\mathbf{y}}) - \text{trace}(\boldsymbol{\Sigma}_{\varepsilon\boldsymbol{\delta}})\right]}{b_n^2} \xrightarrow{P} \beta, \tag{4.3.1}$$

$$\frac{\sum_{i=1}^n \left[\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \text{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\delta}})\right]}{b_n^2} \xrightarrow{P} 1, \tag{4.3.2}$$

We denote $\overline{\boldsymbol{\delta}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{\delta}_i$, $\overline{\varepsilon} = \frac{1}{n}\sum_{i=1}^n \varepsilon_i$, $\overline{\boldsymbol{\xi}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{\xi}_i$. Note that $\overline{\boldsymbol{\delta}}$ is an m-dimensional random vector with components $\overline{\delta}_j = \frac{1}{n}\sum_{i=1}^n \delta_{ij}$, where $j \leq m$.

Writing $\mathbf{x}_i$ and $\mathbf{y}_i$ in terms of $\boldsymbol{\xi}_i$, ε_i and $\boldsymbol{\delta}_i$ we obtain:

$$\begin{aligned}
&\frac{\sum_{i=1}^n \left[(\mathbf{x}_i - \overline{\mathbf{x}})^T(\mathbf{y}_i - \overline{\mathbf{y}}) - \text{trace}(\boldsymbol{\Sigma}_{\varepsilon\boldsymbol{\delta}})\right]}{b_n^2} = \beta\frac{\sum_{i=1}^n \|\boldsymbol{\xi}_i - \overline{\boldsymbol{\xi}}\|^2}{b_n^2} \\
&+\beta\frac{\sum_{i=1}^n (\boldsymbol{\xi}_i - \overline{\boldsymbol{\xi}})^T(\boldsymbol{\delta}_i - \overline{\boldsymbol{\delta}})}{b_n^2} + \frac{\sum_{i=1}^n (\boldsymbol{\xi}_i - \overline{\boldsymbol{\xi}})^T(\varepsilon_i - \overline{\varepsilon})}{b_n^2} \\
&+\frac{\sum_{i=1}^n \left[(\boldsymbol{\delta}_i - \overline{\boldsymbol{\delta}})^T(\varepsilon_i - \overline{\varepsilon}) - \text{trace}(\boldsymbol{\Sigma}_{\varepsilon\boldsymbol{\delta}})\right]}{b_n^2} \\
&:= \beta I_1 + \beta I_2 + I_3 + I_4,
\end{aligned}$$

and

$$\begin{aligned}
&\frac{\sum_{i=1}^n \left[\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \text{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\delta}})\right]}{b_n^2} = \frac{\sum_{i=1}^n \|\boldsymbol{\xi}_i - \overline{\boldsymbol{\xi}}\|^2}{b_n^2} \\
&+2\frac{\sum_{i=1}^n (\boldsymbol{\xi}_i - \overline{\boldsymbol{\xi}})^T(\boldsymbol{\delta}_i - \overline{\boldsymbol{\delta}})}{b_n^2} + \frac{\sum_{i=1}^n \left[\|\boldsymbol{\delta}_i - \overline{\boldsymbol{\delta}}\|^2 - \text{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\delta}})\right]}{b_n^2} \\
&:= I_1 + 2I_2 + I_5.
\end{aligned}$$

Note that:

$$I_1 = \frac{\sum_{i=1}^n \|\boldsymbol{\xi}_i - \boldsymbol{\mu}\|^2}{b_n^2} - \frac{n\|\overline{\boldsymbol{\xi}} - \boldsymbol{\mu}\|^2}{b_n^2},$$

where the first term converges to 1 in probability, by Lemma 4.2.3, and the second term converges to 0 in probability, by the weak law of large numbers (WLLN) and Lemma 4.2.6. Hence, $I_1 \xrightarrow{P} 1$.

By WLLN, $\frac{1}{n}\sum_{i=1}^n \|\boldsymbol{\delta}_i - \overline{\boldsymbol{\delta}}\|^2 \xrightarrow{P} \text{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\delta}})$, and using Lemma 4.2.6, it follows that $I_5 \xrightarrow{P} 0$. Similarly, $I_4 \xrightarrow{P} 0$.

Since $(\boldsymbol{\xi}_i)_{i\leq n}$ and $(\boldsymbol{\delta}_i)_{i\leq n}$ are independent, by WLLN, we obtain:

$$\frac{\sum_{i=1}^{n}(\boldsymbol{\xi}_i-\overline{\boldsymbol{\xi}})^T(\boldsymbol{\delta}_i-\overline{\boldsymbol{\delta}})}{n}\xrightarrow{P}0.$$

This, together with Lemma 4.2.6, implies that $I_2\xrightarrow{P}0$. Similarly, $I_3\xrightarrow{P}0$. Hence, (4.3.1) and (4.3.2) hold and therefore $\widehat{\beta}_n\xrightarrow{P}\beta$.

By WLLN, we have $\overline{\overline{x}}\xrightarrow{P}\frac{1}{m}\sum_{j=1}^{m}\mu_j$ and $\overline{\overline{y}}\xrightarrow{P}\alpha+\beta\frac{1}{m}\sum_{j=1}^{m}\mu_j$. Since $\widehat{\beta}_n\xrightarrow{P}\beta$, we obtain $\widehat{\alpha}_n=\overline{\overline{y}}-\widehat{\beta}_n\overline{\overline{x}}\xrightarrow{P}\alpha$. $\square$

Lemma 4.3.2 *If* $\mathrm{Var}(\|\boldsymbol{\xi}\|)<\infty$, *then* $\widehat{\alpha}_n$ *and* $\widehat{\beta}_n$ *are strongly consistent, i.e.* $\widehat{\beta}_n\xrightarrow{a.s.}\beta$ *and* $\widehat{\alpha}_n\xrightarrow{a.s.}\alpha$.

Proof. The proof follows by applying the strong law of large numbers (SLLN).

First, we consider the denominator and write:

$$\frac{1}{n}\sum_{i=1}^{n}\|\mathbf{x}_i-\overline{\mathbf{x}}\|^2=\frac{1}{n}\sum_{i=1}^{n}\|\mathbf{x}_i-\boldsymbol{\mu}\|^2-\|\overline{\mathbf{x}}-\boldsymbol{\mu}\|^2:=T_1-T_2,$$

By SLLN, we have:

$$\overline{\mathbf{x}}\xrightarrow{a.s.}\boldsymbol{\mu},\quad\text{and hence }\|\overline{\mathbf{x}}-\boldsymbol{\mu}\|^2\xrightarrow{a.s.}0,\text{ or }T_2\xrightarrow{a.s.}0.$$

Also,

$$\frac{1}{n}\sum_{i=1}^{n}\|\mathbf{x}_i-\boldsymbol{\mu}\|^2\xrightarrow{a.s.}\mathrm{E}(\|\mathbf{x}_1-\boldsymbol{\mu}\|^2)=\sum_{j=1}^{m}\mathrm{E}[(\mathbf{x}_{1j}-\boldsymbol{\mu}_j)^2]=\sum_{j=1}^{m}\mathrm{Var}(\mathbf{x_{1j}}),$$

and so $T_1\xrightarrow{a.s.}\mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\xi}})+\mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\delta}})$. Hence, the denominator of $\widehat{\beta}_n$ converges to $\mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\xi}})$, a.s.

We consider the numerator and write

$$\frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i-\overline{\mathbf{x}})^T(\mathbf{y}_i-\overline{\mathbf{y}})\quad=\quad\frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu})^T[\mathbf{y}_i-(\boldsymbol{\alpha}+\beta\boldsymbol{\mu})]$$

$$
\begin{aligned}
&- \quad (\overline{\mathbf{x}}-\boldsymbol{\mu})^T[\overline{\mathbf{y}}-(\boldsymbol{\alpha}+\beta\boldsymbol{\mu})] \\
&:= \quad S_1 - S_2.
\end{aligned}
$$

By SLLN,

$$\overline{\mathbf{x}} \xrightarrow{a.s.} \boldsymbol{\mu} \text{ and } \overline{\mathbf{y}} \xrightarrow{a.s.} \alpha+\beta\boldsymbol{\mu}, \text{ and so } S_2 \xrightarrow{a.s.} 0.$$

We apply SLLN again to obtain

$$
\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu})^T[\mathbf{y}_i-(\boldsymbol{\alpha}+\beta\boldsymbol{\mu})] \xrightarrow{a.s.} \mathrm{E}\{(\mathbf{x}_1-\boldsymbol{\mu})^T[\mathbf{y}_1-(\boldsymbol{\alpha}+\beta\boldsymbol{\mu})]\} \\
& = \mathrm{trace}\{\mathrm{E}\{(\mathbf{x}_1-\boldsymbol{\mu})[\mathbf{y}_1-(\boldsymbol{\alpha}+\beta\boldsymbol{\mu})]^T\}\} = \mathrm{trace}[\mathrm{Cov}(\mathbf{x}_1,\mathbf{y}_1)] \\
= \quad & \beta\mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\xi}}) + \mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon\delta}}),
\end{aligned}
$$

and so $S_1 \xrightarrow{a.s.} \beta\mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\xi}}) + \mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon\delta}})$. Hence, the nominator of $\widehat{\beta}_n$ converges to $\beta\mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\xi}})$, a.s.

We conclude that the estimator $\widehat{\beta}_n$, given by (4.1.8) is a.s. consistent.

By SLLN, we have $\overline{\overline{x}} \xrightarrow{a.s.} \frac{1}{m}\sum_{j=1}^{m}\mu_j$ and $\overline{\overline{y}} \xrightarrow{a.s.} \alpha+\beta\frac{1}{m}\sum_{j=1}^{m}\mu_j$. Since $\widehat{\beta}_n$ is a.s. consistent, it follows that the estimator $\widehat{\alpha}_n$, given by (4.1.7) is also a.s. consistent. □

4.4 Central Limit Theorems

In this section we state and prove CLTs for our estimators. The CLT requires more work and we divide the demonstration into several steps. First, we consider the case $\alpha = 0$ and obtain CLTs for the estimator of β (Theorem 4.4.2 and Theorem 4.4.3). Next, we remove the condition $\alpha = 0$ and prove a CLT for $\widehat{\beta}_n$ (Theorem 4.4.5) and a CLT for $\widehat{\alpha}_n$ (Theorem 4.4.9).

4.4.1 CLTs in the case $\alpha = 0$

When $\alpha = 0$, the estimator of β has the following form

$$\widehat{\beta}'_n = \frac{\sum_{i=1}^{n}[\mathbf{x}_i^T\mathbf{y}_i - \mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\delta\varepsilon}})]}{\sum_{i=1}^{n}[\|\mathbf{x}_i\|^2 - \mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\delta}})]}. \tag{4.4.1}$$

It follows that:

$$\left[\sum_{i=1}^{n}[\|\mathbf{x}_i\|^2 - \mathrm{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}})]\right](\widehat{\beta}'_n - \beta) = \sum_{i=1}^{n} u_i := n\overline{u}, \tag{4.4.2}$$

where

$$u_i = \mathbf{x}_i^T\mathbf{y}_i - \mathrm{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta\varepsilon}}) - \beta[\|\mathbf{x}_i\|^2 - \mathrm{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}})],\ 1 \le i \le n. \tag{4.4.3}$$

In terms of $\boldsymbol{\xi}_i$, $\boldsymbol{\varepsilon}_i$, $\boldsymbol{\delta}_i$, we obtain, for $1 \le i \le n$

$$u_i \quad = \quad \boldsymbol{\xi}_i^T\boldsymbol{\varepsilon}_i - \beta\boldsymbol{\xi}_i^T\boldsymbol{\delta}_i + \boldsymbol{\delta}_i^T\boldsymbol{\varepsilon}_i - \mathrm{trace}(\mathbf{\Sigma}_{\boldsymbol{\varepsilon\delta}}) - \beta[\|\boldsymbol{\delta}_i\|^2 - \mathrm{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}})].$$

Our goal is to prove that $u_1 \in DAN$, which by (4.4.2) and Lemma D.1.7 gives a CLT for the estimator of β.

We first present the general idea of the proof. By Theorem 4.2.4, the first two terms of u_1 are in DAN. The remaining terms are clearly also in DAN. Then, the fact that $u_1 \in DAN$ follows by Lemma D.1.9.

To obtain CLTs in the case when α is arbitrary, we need a general result. We write, for $1 \le i \le n$

$$u_i = \mathbf{b}^T\boldsymbol{\zeta}_i. \tag{4.4.4}$$

where $\mathbf{b} = (1, -\beta, 0, 0, 1, -\beta) \in \mathbb{R}^6$ and

$$\begin{aligned}\boldsymbol{\zeta}_i \quad = \quad & \Bigg(\boldsymbol{\xi}_i^T\boldsymbol{\varepsilon}_i, \boldsymbol{\xi}_i^T\boldsymbol{\delta}_i, \frac{1}{m}\sum_{j=1}^{m}\varepsilon_{ij}, \frac{1}{m}\sum_{j=1}^{m}\delta_{ij}, \boldsymbol{\delta}_i^T\boldsymbol{\varepsilon}_i - \mathrm{trace}(\mathbf{\Sigma}_{\boldsymbol{\varepsilon\delta}}), \\ & \|\boldsymbol{\delta}_i\|^2 - \mathrm{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}})\Bigg)^T.\end{aligned} \tag{4.4.5}$$

Note that u_i are i.i.d. random variables and that $\mathrm{E}(\boldsymbol{\zeta}_i) = 0$, for any $i \le n$.

Let

$$\begin{aligned}\boldsymbol{\zeta}_i' \quad = \quad & \Bigg((\boldsymbol{\xi}_i - \boldsymbol{\mu})^T\boldsymbol{\varepsilon}_i, (\boldsymbol{\xi}_i - \boldsymbol{\mu})^T\boldsymbol{\delta}_i, \frac{1}{m}\sum_{j=1}^{m}\varepsilon_{ij}, \frac{1}{m}\sum_{j=1}^{m}\delta_{ij}, \boldsymbol{\delta}_i^T\boldsymbol{\varepsilon}_i - \mathrm{trace}(\mathbf{\Sigma}_{\boldsymbol{\varepsilon\delta}}), \\ & \|\boldsymbol{\delta}_i\|^2 - \mathrm{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}})\Bigg)^T.\end{aligned} \tag{4.4.6}$$

Note that this result is a generalization of Lemma 6, in [27] to the case when $\boldsymbol{\xi} \in GDAN$.

Lemma 4.4.1 *Assume that conditions* $(A1)$, $(A2)$ *and* $(A3)$ *are satisfied.*

Let $\mathbf{b} \in \mathbb{R}^6$ *and* $\boldsymbol{\zeta}_i$ *be given by* (4.4.5). *If* $|b_1|+|b_2| = 0$, *assume that* $\mathrm{Var}(\mathbf{b}^T\boldsymbol{\zeta}_1) > 0$. *Then,*

$$(a)\ \mathbf{b}^T\boldsymbol{\zeta}_1 \in DAN \text{ and } (b)\ \mathbf{b}^T\boldsymbol{\zeta}_1' \in DAN.$$

In particular, $u_1 \in DAN$, *where* u_1 *is defined by* (4.4.3).

Proof. (a) Since in this proof there is no risk of confusion, we suppress the index i of the random vectors $\boldsymbol{\zeta}_i$, $\boldsymbol{\xi}_i$, $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\delta}_i$.

We write $\mathbf{b}^T\boldsymbol{\zeta} = \boldsymbol{\xi}^T(b_1\boldsymbol{\varepsilon} + b_2\boldsymbol{\delta}) + f_{\mathbf{b}}(\boldsymbol{\varepsilon}, \boldsymbol{\delta})$, where

$$\begin{aligned} f_{\mathbf{b}}(\boldsymbol{\varepsilon}, \boldsymbol{\delta}) &= b_3\left(\frac{1}{m}\sum_{j=1}^{m}\varepsilon_j\right) + b_4\left(\frac{1}{m}\sum_{j=1}^{m}\delta_j\right) + b_5[\boldsymbol{\delta}^T\boldsymbol{\varepsilon} - \mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}\boldsymbol{\delta}})] \\ &+ b_6[\|\boldsymbol{\delta}\|^2 - \mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\delta}})]. \end{aligned}$$

Case 1. $|b_1|+|b_2| = 0$. By assumption (A2), $0 < \mathrm{Var}(\mathbf{b}^T\boldsymbol{\zeta}) = \mathrm{Var}(f_{\mathbf{b}}(\boldsymbol{\varepsilon}, \boldsymbol{\delta})) < \infty$, since $f_{\mathbf{b}}(\boldsymbol{\varepsilon}, \boldsymbol{\delta})$ is a function containing powers of maximum order 2 of $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$. Hence, we apply the CLT and obtain $\mathbf{b}^T\boldsymbol{\zeta} \in DAN$.

Case 2. $|b_1| + |b_2| > 0$.

I. Assume $\mathrm{Var}(\|\boldsymbol{\xi}\|) < \infty$. Then, $\mathrm{Var}(\mathbf{b}^T\boldsymbol{\zeta}) < \infty$ and if $\mathrm{Var}(\mathbf{b}^T\boldsymbol{\zeta}) > 0$, we apply the CLT to reach the conclusion.

Assume, by contradiction, that $\mathrm{Var}(\mathbf{b}^T\boldsymbol{\zeta}) = 0$. Then, $\mathbf{b}^T\boldsymbol{\zeta} = C$ a.s., where C is a constant. Since $\mathrm{E}(\mathbf{b}^T\boldsymbol{\zeta}) = 0$, it follows that $C = 0$, (i.e. $\mathbf{b}^T\boldsymbol{\zeta} = 0$ a.s.).

Now $0 = \mathrm{E}[\mathbf{b}^T\boldsymbol{\zeta}|\boldsymbol{\varepsilon}, \boldsymbol{\delta}] = \mathrm{E}[\boldsymbol{\xi}^T(b_1\boldsymbol{\varepsilon}+b_2\boldsymbol{\delta})+f_{\mathbf{b}}(\boldsymbol{\varepsilon}.\boldsymbol{\delta})|\boldsymbol{\varepsilon}, \boldsymbol{\delta}]$ a.s. Since $\boldsymbol{\xi}$ is independent of $(\boldsymbol{\varepsilon}, \boldsymbol{\delta})$ and the other terms and factors are $(\boldsymbol{\varepsilon}, \boldsymbol{\delta})$- measurable functions, we have $\boldsymbol{\mu}^T(b_1\boldsymbol{\varepsilon} + b_2\boldsymbol{\delta}) + f_{\mathbf{b}}(\boldsymbol{\varepsilon}.\boldsymbol{\delta}) = 0$ a.s., so $(\boldsymbol{\xi} - \boldsymbol{\mu})^T(b_1\boldsymbol{\varepsilon} + b_2\boldsymbol{\delta}) = 0$ a.s.

Using (A_3), $\mathrm{Var}[(\boldsymbol{\xi} - \boldsymbol{\mu})^T(b_1\boldsymbol{\varepsilon} + b_2\boldsymbol{\delta})] = \mathrm{trace}[\boldsymbol{\Sigma}_{\boldsymbol{\xi}}\mathbf{V}_{\mathbf{b}}(\boldsymbol{\varepsilon}, \boldsymbol{\delta})]$, where $\mathbf{V}_{\mathbf{b}}(\boldsymbol{\varepsilon}, \boldsymbol{\delta}) := \mathrm{Var}(b_1\boldsymbol{\varepsilon} + b_2\boldsymbol{\delta}) = b_1^2\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}} + b_1b_2\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}\boldsymbol{\delta}} + b_1b_2\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}\boldsymbol{\delta}}^T + b_2^2\boldsymbol{\Sigma}_{\boldsymbol{\delta}}$. Note that, by $(A2)$, $\mathbf{V}_{\mathbf{b}}(\boldsymbol{\varepsilon}, \boldsymbol{\delta}) > 0$

which together with Remark 4.2.2, (iii) implies that $\mathrm{Var}[(\boldsymbol{\xi}-\boldsymbol{\mu})^T(b_1\boldsymbol{\varepsilon}+b_2\boldsymbol{\delta})] > 0$. We reached a contradiction and therefore, $\mathrm{Var}(\mathbf{b}^T\boldsymbol{\zeta}) > 0$.

II. Assume $\mathrm{Var}(\|\boldsymbol{\xi}\|) = \infty$. By Theorem 4.2.4 we have $\boldsymbol{\xi}^T(b_1\boldsymbol{\varepsilon}+b_2\boldsymbol{\delta}) \in DAN$. Since $f_{\mathbf{b}}(\boldsymbol{\varepsilon},\boldsymbol{\delta}) \in DAN$, we apply Lemma D.1.9 to obtain $\mathbf{b}^T\boldsymbol{\zeta} \in DAN$, provided we checked two conditions:

(i) $\mathrm{E}[\boldsymbol{\xi}^T(b_1\boldsymbol{\varepsilon}+b_2\boldsymbol{\delta})f_{\mathbf{b}}(\boldsymbol{\varepsilon},\boldsymbol{\delta})] < \infty$ and

(ii) $\mathrm{Var}(\mathbf{b}^T\boldsymbol{\zeta}) > 0$. Denote $\mathbf{V}_{\boldsymbol{\xi}} = \mathrm{E}(\boldsymbol{\xi}\boldsymbol{\xi}^T)$. We have:

$$\begin{aligned}
\mathrm{Var}(\mathbf{b}^T\boldsymbol{\zeta}) &= \mathrm{Var}[\boldsymbol{\xi}^T(b_1\boldsymbol{\varepsilon}+b_2\boldsymbol{\delta})] + \mathrm{Var}[f_{\mathbf{b}}(\boldsymbol{\varepsilon},\boldsymbol{\delta})] + 2\mathrm{Cov}[\boldsymbol{\xi}^T(b_1\boldsymbol{\varepsilon}+b_2\boldsymbol{\delta}), f_{\mathbf{b}}(\boldsymbol{\varepsilon},\boldsymbol{\delta})] \\
&= \mathrm{trace}[\mathbf{V}_{\boldsymbol{\xi}}\mathbf{V}_{\mathbf{b}}(\boldsymbol{\varepsilon},\boldsymbol{\delta})] + \mathrm{Var}[f_{\mathbf{b}}(\boldsymbol{\varepsilon},\boldsymbol{\delta})] + 2\boldsymbol{\mu}^T\mathrm{E}[(b_1\boldsymbol{\varepsilon}+b_2\boldsymbol{\delta})f_{\mathbf{b}}(\boldsymbol{\varepsilon},\boldsymbol{\delta})] \\
&= \infty,
\end{aligned}$$

since the last two terms are finite and, by Lemma A.1.17,

$$\mathrm{trace}[\mathbf{V}_{\boldsymbol{\xi}}\mathbf{V}_{\mathbf{b}}(\boldsymbol{\varepsilon},\boldsymbol{\delta})] \geq \frac{\mathrm{trace}(\mathbf{V}_{\boldsymbol{\xi}})}{\mathrm{trace}[\mathbf{V}_{\mathbf{b}}^{-1}(\boldsymbol{\varepsilon},\boldsymbol{\delta})]} = \infty.$$

(b) Since $\boldsymbol{\xi} \in GDAN$ is equivalent to $\boldsymbol{\xi}-\boldsymbol{\mu} \in GDAN$, we apply part (a) to $\boldsymbol{\xi}' = \boldsymbol{\xi}-\boldsymbol{\mu}$ to obtain the conclusion. □

Using Lemma 4.4.1 and the self-normalized and Studentized versions of the classical CLT (Lemma D.1.7), we obtain the following CLTs for $\widehat{\beta}_n$, when $\alpha = 0$.

Theorem 4.4.2 *Assume that* $(A1)$, $(A2)$ *and* $(A3)$ *hold and* $\alpha = 0$.

Let $\widehat{\beta}'_n$ *be the estimator given by* (4.4.1). *Then, as* $n \to \infty$, *we have:*

$$(a)\ \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n[\|\mathbf{x}_i\|^2 - \mathrm{trace}(\boldsymbol{\Sigma}_{\delta})](\widehat{\beta}'_n - \beta)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n(u_i-\overline{u})^2}} \xrightarrow{\mathcal{D}} N(0,1),$$

$$(b)\ \frac{\sum_{i=1}^n\left[\|\mathbf{x}_i\|^2 - \mathrm{trace}(\boldsymbol{\Sigma}_{\delta})\right](\widehat{\beta}'_n - \beta)}{\sqrt{\sum_{i=1}^n u_i^2}} \xrightarrow{\mathcal{D}} N(0,1),$$

where u_i *is given by* (4.4.3) *and* $\overline{u} = \dfrac{1}{n}\displaystyle\sum_{i=1}^n u_i$.

Proof. By (4.4.4) and Lemma 4.4.1, $u_1 \in DAN$. Note that $\mathrm{E}(u_1) = 0$. Hence,

$$\frac{\sqrt{n}\overline{u}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n (u_i - \overline{u})^2}} \xrightarrow{\mathcal{D}} N(0,1),$$
$$\frac{\sum_{i=1}^n u_i}{\sqrt{\sum_{i=1}^n u_i^2}} \xrightarrow{\mathcal{D}} N(0,1),$$

by Lemma D.1.7. The conclusion follows using (4.4.2). □

The next result shows that Theorem 4.4.2 continues to hold if we replace β by $\widehat{\beta}'_n$ in the definition of u_i. Its proof is omitted, as it is similar to the proof of Theorem 4.5.1 (below).

Theorem 4.4.3 *Assume that (A1), (A2) and (A3) hold and $\alpha = 0$.*

Let $\widehat{\beta}'_n$ be the estimator given by (4.4.1). Then, as $n \to \infty$, we have:

$$\frac{\sum_{i=1}^n [\|\mathbf{x}_i\|^2 - \mathrm{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}})](\widehat{\beta}'_n - \beta)}{\sqrt{\sum_{i=1}^n \tilde{u}_i(n)^2}} \xrightarrow{\mathcal{D}} N(0,1),$$

where $\tilde{u}_i(n) = \mathbf{x}_i^T\mathbf{y}_i - \mathrm{trace}(\mathbf{\Sigma}_{\boldsymbol{\varepsilon\delta}}) - \widehat{\beta}_n[\|\mathbf{x}_i\|^2 - \mathrm{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}})]$.

4.4.2 CLTs in the case of α arbitrary

We write:

$$\sum_{i=1}^n \left[\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \mathrm{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}})\right](\widehat{\beta}_n - \beta) = \sum_{i=1}^n u_i(n), \tag{4.4.7}$$

with

$$u_i(n) = (\mathbf{x}_i - \overline{\mathbf{x}})^T(\mathbf{y}_i - \overline{\mathbf{y}}) - \mathrm{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta\varepsilon}}) - \beta[\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \mathrm{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}})]. \tag{4.4.8}$$

Note that $u_i(n)$ depends on n and β.

Relation (4.4.8) is of crucial importance and lies at the core of our developments. The goal is to obtain CLTs for the right hand side of (4.4.7), which will give a CLT for $\widehat{\beta}_n$.

We express $u_i(n)$ in terms of $\boldsymbol{\xi}_i$, $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\delta}_i$, for every $1 \leq i \leq n$.

$$\begin{aligned} u_i(n) &= (\boldsymbol{\xi}_i - \overline{\boldsymbol{\xi}})^T(\boldsymbol{\varepsilon}_i - \overline{\boldsymbol{\varepsilon}}) - \beta(\boldsymbol{\xi}_i - \overline{\boldsymbol{\xi}})^T(\boldsymbol{\delta}_i - \overline{\boldsymbol{\delta}}) + (\boldsymbol{\delta}_i - \overline{\boldsymbol{\delta}})^T(\boldsymbol{\varepsilon}_i - \overline{\boldsymbol{\varepsilon}}) - \operatorname{trace}(\boldsymbol{\Sigma}_{\varepsilon\delta}) \\ &- \beta[\|\boldsymbol{\delta}_i - \overline{\boldsymbol{\delta}}\|^2 - \operatorname{trace}(\boldsymbol{\Sigma}_{\delta})] \end{aligned}$$

Note that

$$u_i(n) = \mathbf{d}^T \boldsymbol{\eta}_i(n). \tag{4.4.9}$$

where $\mathbf{d} = (0, 0, 1. -\beta)^T$ and

$$\begin{aligned} \boldsymbol{\eta}_i(n) &= \Bigg(\frac{1}{m}\sum_{j=1}^m y_{ij} - \alpha, \frac{1}{m}\sum_{j=1}^m x_{ij}, (\mathbf{x_i} - \overline{\mathbf{x}})^T(\mathbf{y_i} - \overline{\mathbf{y}}) - \operatorname{trace}(\boldsymbol{\Sigma}_{\varepsilon\delta}). \\ &\qquad \|\mathbf{x_i} - \overline{\mathbf{x}}\|^2 - \operatorname{trace}(\boldsymbol{\Sigma}_{\delta})\Bigg)^T \qquad (4.4.10) \\ &= \Bigg(\beta\frac{1}{m}\sum_{j=1}^m \xi_{ij} + \frac{1}{m}\sum_{j=1}^m \varepsilon_{ij}. \frac{1}{m}\sum_{j=1}^m \xi_{ij} + \frac{1}{m}\sum_{j=1}^m \delta_{ij}, \\ &\qquad \beta\|\boldsymbol{\xi}_i - \overline{\boldsymbol{\xi}}\|^2 + (\boldsymbol{\xi}_i - \overline{\boldsymbol{\xi}})^T(\boldsymbol{\varepsilon}_i - \overline{\boldsymbol{\varepsilon}}) + \beta(\boldsymbol{\xi}_i - \overline{\boldsymbol{\xi}})^T(\boldsymbol{\delta}_i - \overline{\boldsymbol{\delta}}) \\ &\qquad +(\boldsymbol{\delta}_i - \overline{\boldsymbol{\delta}})^T(\boldsymbol{\varepsilon}_i - \overline{\boldsymbol{\varepsilon}}) - \operatorname{trace}(\boldsymbol{\Sigma}_{\varepsilon\delta}), \\ &\qquad \|\boldsymbol{\xi}_i - \overline{\boldsymbol{\xi}}\|^2 + \|\boldsymbol{\delta}_i - \overline{\boldsymbol{\delta}}\|^2 + 2(\boldsymbol{\xi}_i - \overline{\boldsymbol{\xi}})^T(\boldsymbol{\delta}_i - \overline{\boldsymbol{\delta}}) - \operatorname{trace}(\boldsymbol{\Sigma}_{\delta})\Bigg). \end{aligned}$$

The first two components in the definition of $\boldsymbol{\eta}_i(n)$ are introduced artificially at this point, but they will be used in the proofs of the asymptotic properties of the estimator $\widehat{\alpha}_n$.

The idea is to replace $\boldsymbol{\xi}_i - \overline{\boldsymbol{\xi}}$ by $\boldsymbol{\xi}_i - \boldsymbol{\mu}$ in the expression of $u_i(n)$. i.e. to obtain an expression for $u_i(n)$ which contains an inner product of the form $\mathbf{e}^T\boldsymbol{\zeta}_i'$ (for some $\mathbf{e} \in \mathbb{R}^6$), plus a remainder term. Then, we use the fact that $\mathbf{e}^T\boldsymbol{\zeta}_1' \in DAN$. which was shown in Lemma 4.4.1, (b), and show that the normalized remainder term converges to 0.

For a sake of a generalization needed in the sequel, we consider $\mathbf{d} \in \mathbb{R}^4$ such that $\beta d_1 + d_2 = 0$ and $\beta d_3 + d_4 = 0$. We have the following decomposition:

$$\mathbf{d}^T\boldsymbol{\eta}_i(n) = e_1[(\boldsymbol{\xi}_i - \boldsymbol{\mu})^T\boldsymbol{\varepsilon}_i - \boldsymbol{\varepsilon}_i^T(\overline{\boldsymbol{\xi}} - \boldsymbol{\mu}) - \overline{\boldsymbol{\varepsilon}}^T(\boldsymbol{\xi}_i - \boldsymbol{\mu}) + \overline{\boldsymbol{\varepsilon}}^T(\overline{\boldsymbol{\xi}} - \boldsymbol{\mu})]$$

$$
\begin{aligned}
&+ \quad e_2[(\boldsymbol{\xi}_i - \boldsymbol{\mu})^T\boldsymbol{\delta}_i - \boldsymbol{\delta}_i^T(\overline{\boldsymbol{\xi}} - \boldsymbol{\mu}) - \overline{\boldsymbol{\delta}}^T(\boldsymbol{\xi}_i - \boldsymbol{\mu}) + \overline{\boldsymbol{\delta}}^T(\overline{\boldsymbol{\xi}} - \boldsymbol{\mu})] \\
&+ \quad e_3\frac{1}{m}\sum_{j=1}^{m}\varepsilon_{ij} + e_4\frac{1}{m}\sum_{j=1}^{m}\delta_{ij} + e_5[\boldsymbol{\delta}_i^T\boldsymbol{\varepsilon}_i - \boldsymbol{\delta}_i^T\overline{\boldsymbol{\varepsilon}} - \overline{\boldsymbol{\delta}}^T\boldsymbol{\varepsilon}_i + \overline{\boldsymbol{\delta}}^T\overline{\boldsymbol{\varepsilon}} - \mathrm{trace}(\boldsymbol{\Sigma}_{\varepsilon\delta})] \\
&+ \quad e_6[\|\boldsymbol{\delta}_i\|^2 - 2\boldsymbol{\delta}_i^T\overline{\boldsymbol{\delta}} + \|\overline{\boldsymbol{\delta}}\|^2 - \mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\delta}})], \qquad (4.4.11)
\end{aligned}
$$

with

$$\mathbf{e} = (d_3, \beta d_3 + 2d_4, d_1, d_2, d_3, d_4)^T. \qquad (4.4.12)$$

Therefore, for any $\mathbf{d} \in \mathbb{R}^4$ which satisfies $\beta d_1 + d_2 = 0$ and $\beta d_3 + d_4 = 0$, we have the decomposition:

$$\mathbf{d}^T\boldsymbol{\eta}_i(n) = \mathbf{e}^T\boldsymbol{\zeta}_i' + R_i(n), \qquad (4.4.13)$$

where $\boldsymbol{\zeta}_i'$ is given by (4.4.6) and

$$
\begin{aligned}
R_i(n) \quad := \quad & e_1[-\boldsymbol{\varepsilon}_i^T(\overline{\boldsymbol{\xi}} - \boldsymbol{\mu}) - \overline{\boldsymbol{\varepsilon}}^T(\boldsymbol{\xi}_i - \boldsymbol{\mu}) + \overline{\boldsymbol{\varepsilon}}^T(\overline{\boldsymbol{\xi}} - \boldsymbol{\mu})] \\
+ \quad & e_2[-\boldsymbol{\delta}_i^T(\overline{\boldsymbol{\xi}} - \boldsymbol{\mu}) - \overline{\boldsymbol{\delta}}^T(\boldsymbol{\xi}_i - \boldsymbol{\mu}) + \overline{\boldsymbol{\delta}}^T(\overline{\boldsymbol{\xi}} - \boldsymbol{\mu})] \\
+ \quad & e_5(-\boldsymbol{\delta}_i^T\overline{\boldsymbol{\varepsilon}} - \overline{\boldsymbol{\delta}}^T\boldsymbol{\varepsilon}_i + \overline{\boldsymbol{\delta}}^T\overline{\boldsymbol{\varepsilon}}) \\
+ \quad & e_6(-2\boldsymbol{\delta}_i^T\overline{\boldsymbol{\delta}} + \|\overline{\boldsymbol{\delta}}\|^2).
\end{aligned}
$$

In particular, (4.4.13) holds for $d = (0, 0, 1, -\beta)^T$, in which case (4.4.13) gives a representation of $u_i(n)$, definied by (4.4.8).

The next result contains the Studentized and self-normalized forms of the CLT for the sequence $\{\mathbf{d}^T\eta_i(n)\}_{i\leq n}$, (and in particular, for $\{u_i(n)\}_{i\leq n}$).

Lemma 4.4.4 *Assume that conditions (A1), (A2) and (A3) are satisfied.*

Let $\mathbf{d} \in \mathbb{R}^4$, which satisfies $\beta d_1 + d_2 = 0$, $\beta d_3 + d_4 = 0$, and $\mathbf{e}$ be given by (4.4.12). If $|e_1| + |e_2| = 0$, assume that $\mathrm{Var}(\mathbf{e}^{\mathrm{T}}\boldsymbol{\zeta}_1') > 0$, where $\boldsymbol{\zeta}_i'$ is given by (4.4.6). Then, as $n \to \infty$:

$$(a)\ \frac{\sqrt{n}\ \overline{\mathbf{d}^T\boldsymbol{\eta}(n)}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}[\mathbf{d}^T(\boldsymbol{\eta}_i(n) - \overline{\boldsymbol{\eta}(n)})]^2}} \xrightarrow{\mathcal{D}} N(0, 1),$$

$$(b)\frac{n\overline{\mathbf{d}^T\boldsymbol{\eta}(n)}}{\sqrt{\sum_{i=1}^n(\mathbf{d}^T\boldsymbol{\eta}_i(n))^2}} \xrightarrow{\mathcal{D}} N(0,1),$$

where $\eta_i(n)$ *is given by* (4.4.10), $\overline{\mathbf{d}^T\boldsymbol{\eta}(n)} = \frac{1}{n}\sum_{i=1}^n \mathbf{d}^T\boldsymbol{\eta}_i(n)$ *and* $\overline{\boldsymbol{\eta}(n)} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{\eta}_i(n)$.

Proof. By (4.4.13), it follows that $\overline{\mathbf{d}^T\boldsymbol{\eta}(n)} = \overline{\mathbf{e}^T\boldsymbol{\zeta}'} + \overline{R(n)}$, where $\overline{\mathbf{e}^T\boldsymbol{\zeta}'} = \frac{1}{n}\sum_{i=1}^n \mathbf{e}^T\boldsymbol{\zeta}_i'$, $\overline{R(n)} = \frac{1}{n}\sum_{i=1}^n R_i(n)$ and therefore:

$$\frac{\sqrt{n}\ \overline{\mathbf{d}^T\boldsymbol{\eta}(n)}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n[\mathbf{d}^T(\boldsymbol{\eta}_i(n)-\overline{\boldsymbol{\eta}(n)})]^2}} = I_1 + I_2.$$

where:

$$I_1 := \frac{\sqrt{n}\ \overline{\mathbf{e}^T\boldsymbol{\zeta}'}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n[\mathbf{d}^T(\boldsymbol{\eta}_i(n)-\overline{\boldsymbol{\eta}(n)})]^2}},$$
$$I_2 := \frac{\sqrt{n}\ \overline{R(n)}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n[\mathbf{d}^T(\boldsymbol{\eta}_i(n)-\overline{\boldsymbol{\eta}(n)})]^2}}.$$

We write:

$$\begin{aligned} I_1 &= \frac{\sqrt{n}\ \overline{\mathbf{e}^T\boldsymbol{\zeta}'}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n[\mathbf{e}^T(\boldsymbol{\zeta}_i'-\overline{\boldsymbol{\zeta}'})]^2}}\cdot\sqrt{\frac{\sum_{i=1}^n[\mathbf{e}^T(\boldsymbol{\zeta}_i'-\overline{\boldsymbol{\zeta}'})]^2}{\sum_{i=1}^n[\mathbf{d}^T(\boldsymbol{\eta}_i(n)-\overline{\boldsymbol{\eta}(n)})]^2}}, \\ I_2 &= \frac{\sqrt{n}\ \overline{R(n)}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n[\mathbf{e}^T(\boldsymbol{\zeta}_i'-\overline{\boldsymbol{\zeta}'})]^2}}\cdot\sqrt{\frac{\sum_{i=1}^n[\mathbf{e}^T(\boldsymbol{\zeta}_i'-\overline{\boldsymbol{\zeta}'})]^2}{\sum_{i=1}^n[\mathbf{d}^T(\boldsymbol{\eta}_i(n)-\overline{\boldsymbol{\eta}(n)})]^2}}, \end{aligned}$$

where $\overline{\boldsymbol{\zeta}'} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{\zeta}_i'$. We prove that, as $n \to \infty$:

$$\frac{\sqrt{n}\ \overline{R(n)}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n[\mathbf{e}^T(\boldsymbol{\zeta}_i'-\overline{\boldsymbol{\zeta}'})]^2}} \xrightarrow{P} 0, \tag{4.4.14}$$
$$\frac{\sum_{i=1}^n[\mathbf{d}^T(\boldsymbol{\eta}_i(n)-\overline{\boldsymbol{\eta}(n)})]^2}{\sum_{i=1}^n[\mathbf{e}^T(\boldsymbol{\zeta}_i'-\overline{\boldsymbol{\zeta}'})]^2} \xrightarrow{P} 1. \tag{4.4.15}$$

By Lemma 4.4.1. (b) with $\mathbf{b} = \mathbf{e}$, $\mathbf{e}^T\boldsymbol{\zeta}'_1 \in DAN$ and so:

$$\frac{\sqrt{n}\ \overline{\mathbf{e}^T\boldsymbol{\zeta}'}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n[\mathbf{e}^T(\boldsymbol{\zeta}'_i - \overline{\boldsymbol{\zeta}'})]^2}} \xrightarrow{\mathfrak{D}} N(0, 1).$$

This result and (4.4.15) imply $I_1 \xrightarrow{\mathfrak{D}} N(0,1)$, by Slutsky's Theorem. Furthermore, from (4.4.14) and (4.4.15), we obtain $I_2 \xrightarrow{P} 0$. Therefore, the convergence in (a) would follow by applying Slutsky's Theorem.

Recall that $\overline{\boldsymbol{\delta}} = (\overline{\delta}_1, \ldots, \overline{\delta}_m)^T$, where $\overline{\delta}_j = \dfrac{1}{n}\sum_{i=1}^n \delta_{ij}$, $1 \leq j \leq m$. Note that:

$$\begin{aligned}\overline{R(n)} &= e_1[-(\overline{\boldsymbol{\xi}} - \boldsymbol{\mu})^T\overline{\boldsymbol{\varepsilon}}] + e_2[-(\overline{\boldsymbol{\xi}} - \boldsymbol{\mu})^T\overline{\boldsymbol{\delta}}] \\ &+ e_5(-\overline{\boldsymbol{\delta}}^T\overline{\boldsymbol{\varepsilon}}) + e_6(-\|\overline{\boldsymbol{\delta}}\|^2).\end{aligned} \tag{4.4.16}$$

To prove (4.4.14), we consider two cases.

Case 1. Assume $\mathrm{Var}(\|\boldsymbol{\xi}\|) < \infty$ or $|e_1| + |e_2| = 0$.

In this case, we first prove:

$$\sqrt{n}\ \overline{R(n)} \xrightarrow{P} 0. \tag{4.4.17}$$

For each $1 \leq j \leq m$, by applying the CLT, we obtain $\sqrt{n}\overline{\delta}_j = O_P(1)$ and so $\sqrt{n}\overline{\delta}_j^2 = o_P(1)$. Hence, $\sqrt{n}\|\overline{\boldsymbol{\delta}}\|^2 = o_P(1)$. Similarly, we have $\sqrt{n}\|\overline{\boldsymbol{\varepsilon}}\|^2 = o_P(1)$. By Cauchy-Schwarz inequality, $n|\overline{\boldsymbol{\delta}}^T\overline{\boldsymbol{\varepsilon}}|^2 \leq n\|\overline{\boldsymbol{\varepsilon}}\|^2 \cdot \|\overline{\boldsymbol{\delta}}\|^2$ and therefore, we obtain $\sqrt{n}|\overline{\boldsymbol{\delta}}^T\overline{\boldsymbol{\varepsilon}}| = o_P(1)$.

Using the same arguments as above, $\sqrt{n}|(\overline{\boldsymbol{\xi}}-\boldsymbol{\mu})^T\overline{\boldsymbol{\varepsilon}}| = o_P(1)$ and $\sqrt{n}|(\overline{\boldsymbol{\xi}}-\boldsymbol{\mu})^T\overline{\boldsymbol{\delta}}| = o_P(1)$, so (4.4.17) follows.

To prove (4.4.14) in this case, we write:

$$\frac{\sqrt{n}\overline{R(n)}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n[\mathbf{e}^T(\boldsymbol{\zeta}'_i - \overline{\boldsymbol{\zeta}'})]^2}} = \frac{\sqrt{n}\overline{R(n)}}{\sqrt{\mathrm{Var}(\mathbf{e}^T\boldsymbol{\zeta}')}} \cdot \sqrt{\frac{\mathrm{Var}(\mathbf{e}^T\boldsymbol{\zeta}')}{\frac{1}{n-1}\sum_{i=1}^n[\mathbf{e}^T(\boldsymbol{\zeta}'_i - \overline{\boldsymbol{\zeta}'})]^2}}.$$

By WLLN (see also (4.4.24), with $\mathbf{b} = \mathbf{e}$ and $\boldsymbol{\zeta}_i$ replaced by $\boldsymbol{\zeta}'_i$), the second factor of the product converges in probability to 1. Since $0 < \mathrm{Var}(\mathbf{e}^T\boldsymbol{\zeta}') < \infty$, (4.4.17) implies (4.4.14).

Case 2. Assume that $\mathrm{Var}(\|\boldsymbol{\xi}\|) = \infty$ and $|e_1| + |e_2| > 0$.

Let $(\mathbf{B}_n)_{n\geq 1}$ be a sequence of (nonsingular, symmetric) non-stochastic matrices such that $\mathbf{B}_n \sum_{i=1}^{n}(\boldsymbol{\xi}_i - \boldsymbol{\mu}) \xrightarrow{\mathcal{D}} N(\mathbf{0},\mathbf{I})$. Let $b_n^2 = \text{trace}(\mathbf{B}_n^{-2})$. Define $c_n^2 = \text{trace}(\boldsymbol{\Sigma}_\gamma \mathbf{B}_n^{-2})$, where $\boldsymbol{\gamma}_i = e_1\boldsymbol{\varepsilon}_i + e_2\boldsymbol{\delta}_i$ and $\boldsymbol{\Sigma}_\gamma = \text{Var}(\boldsymbol{\gamma}_i)$. In this case, in lieu of (4.4.17), we prove:

$$\frac{n}{\sqrt{c_n^2}}\overline{R(n)} \xrightarrow{P} 0. \tag{4.4.18}$$

Using Lemma A.1.17 and Lemma 4.2.6, $\frac{n}{c_n^2} \to 0$. Therefore, the last two terms in (4.4.16), multiplied by $\frac{n}{\sqrt{c_n^2}}$ converge in probability to 0.

As for the first term, by applying Cauchy-Schwarz inequality, we have:

$$\frac{n}{\sqrt{c_n^2}}|(\overline{\boldsymbol{\xi}} - \boldsymbol{\mu})^T\overline{\boldsymbol{\varepsilon}}| \leq \frac{n}{\sqrt{c_n^2}}\|\overline{\boldsymbol{\xi}} - \boldsymbol{\mu}\| \cdot \|\overline{\boldsymbol{\varepsilon}}\|.$$

Recall that $\|\overline{\boldsymbol{\varepsilon}}\| = \frac{O_P(1)}{\sqrt{n}}$.

Since $\|\mathbf{B}_n \sum_{i=1}^{n}(\boldsymbol{\xi}_i - \boldsymbol{\mu})\| = O_P(1)$, we have the following relations:

$$\begin{aligned}
\|\overline{\boldsymbol{\xi}} - \boldsymbol{\mu}\|^2 &= \frac{1}{n^2}\left\|\sum_{i=1}^{n}(\boldsymbol{\xi}_i - \boldsymbol{\mu})\right\|^2 = \frac{1}{n^2}\text{trace}\left[\left(\sum_{i=1}^{n}(\boldsymbol{\xi}_i - \boldsymbol{\mu})\right)\cdot\left(\sum_{i=1}^{n}(\boldsymbol{\xi}_i - \boldsymbol{\mu})\right)^T\right] \\
&= \frac{1}{n^2}\text{trace}\left[\mathbf{B}_n^{-1}\mathbf{B}_n\left(\sum_{i=1}^{n}(\boldsymbol{\xi}_i - \boldsymbol{\mu})\right)\cdot\left(\sum_{i=1}^{n}(\boldsymbol{\xi}_i - \boldsymbol{\mu})\right)^T\mathbf{B}_n\mathbf{B}_n^{-1}\right] \\
&= \frac{1}{n^2}\text{trace}\left[\mathbf{B}_n^{-2}\mathbf{B}_n\left(\sum_{i=1}^{n}(\boldsymbol{\xi}_i - \boldsymbol{\mu})\right)\cdot\left(\sum_{i=1}^{n}(\boldsymbol{\xi}_i - \boldsymbol{\mu})\right)^T\mathbf{B}_n\right] \\
&\leq \frac{1}{n^2}\text{trace}(\mathbf{B}_n^{-2})\text{trace}\left[\mathbf{B}_n\left(\sum_{i=1}^{n}(\boldsymbol{\xi}_i - \boldsymbol{\mu})\right)\cdot\left(\sum_{i=1}^{n}(\boldsymbol{\xi}_i - \boldsymbol{\mu})\right)^T\mathbf{B}_n\right] \\
&= \frac{1}{n^2}b_n^2\cdot\left\|\mathbf{B}_n\sum_{i=1}^{n}(\boldsymbol{\xi}_i - \boldsymbol{\mu})\right\|^2 = \frac{1}{n^2}b_n^2 O_P(1),
\end{aligned}$$

where we have used Lemma A.1.17 to obtain the inequality.

Therefore $\frac{n}{\sqrt{c_n^2}}|(\overline{\boldsymbol{\xi}} - \boldsymbol{\mu})^T\overline{\boldsymbol{\varepsilon}}| \leq \frac{n}{\sqrt{c_n^2}}\frac{\sqrt{b_n^2}}{n}\frac{O_P(1)}{\sqrt{n}} = \frac{O_P(1)}{\sqrt{n}}\sqrt{\frac{b_n^2}{c_n^2}}$. By Lemma A.1.17 we obtain $\frac{b_n^2}{c_n^2} \leq \text{trace}(\boldsymbol{\Sigma}_\gamma^{-1})$ and so $\frac{n}{\sqrt{c_n^2}}|(\overline{\boldsymbol{\xi}} - \boldsymbol{\mu})^T\overline{\boldsymbol{\varepsilon}}| = o_P(1)$.

Similarly, $\frac{n}{\sqrt{c_n^2}}|(\overline{\boldsymbol{\xi}}-\boldsymbol{\mu})^T\overline{\boldsymbol{\delta}}| = o_P(1)$ and so the convergence in (4.4.18) is proved. Therefore, if we write:

$$\frac{\sqrt{n}\overline{R(n)}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n[\mathbf{e}^T(\boldsymbol{\zeta}_i'-\overline{\boldsymbol{\zeta}'})]^2}} = \frac{n\overline{R(n)}}{\sqrt{c_n^2}}\cdot\sqrt{\frac{n-1}{n}}\cdot\sqrt{\frac{1}{\frac{1}{c_n^2}\sum_{i=1}^n[\mathbf{e}^T(\boldsymbol{\zeta}_i'-\overline{\boldsymbol{\zeta}'})]^2}},$$

the last factor converges in probability to 1, by (4.4.25), with $\mathbf{b}=\mathbf{e}$ and $\boldsymbol{\zeta}_i$ replaced by $\boldsymbol{\zeta}_i'$. This, together with (4.4.18), concludes the proof of (4.4.14).

Now we prove (4.4.15). Using (4.4.11), we write:

$$\begin{aligned}
\mathbf{d}^T(\boldsymbol{\eta}_i(n)-\overline{\boldsymbol{\eta}(n)}) &= e_1[(\boldsymbol{\xi}_i-\boldsymbol{\mu})^T\boldsymbol{\varepsilon}_i-(\overline{\boldsymbol{\xi}}-\boldsymbol{\mu})^T\boldsymbol{\varepsilon}_i-(\boldsymbol{\xi}_i-\boldsymbol{\mu})^T\overline{\boldsymbol{\varepsilon}}-\overline{(\boldsymbol{\xi}-\boldsymbol{\mu})^T\boldsymbol{\varepsilon}}\\
&+ 2(\overline{\boldsymbol{\xi}}-\boldsymbol{\mu})^T\overline{\boldsymbol{\varepsilon}}]\\
&+ e_2[(\boldsymbol{\xi}_i-\boldsymbol{\mu})^T\boldsymbol{\delta}_i-(\overline{\boldsymbol{\xi}}-\boldsymbol{\mu})^T\boldsymbol{\delta}_i-(\boldsymbol{\xi}_i-\boldsymbol{\mu})^T\overline{\boldsymbol{\delta}}-\overline{(\boldsymbol{\xi}-\boldsymbol{\mu})^T\boldsymbol{\delta}}\\
&+ 2(\overline{\boldsymbol{\xi}}-\boldsymbol{\mu})^T\overline{\boldsymbol{\delta}}]\\
&+ e_3\left(\frac{1}{m}\sum_{j=1}^m\varepsilon_{ij}-\overline{\overline{\varepsilon}}\right)+e_4\left(\frac{1}{m}\sum_{j=1}^m\delta_{ij}-\overline{\overline{\delta}}\right)\\
&+ e_5(\boldsymbol{\delta}_i^T\boldsymbol{\varepsilon}_i-\boldsymbol{\delta}_i^T\overline{\boldsymbol{\varepsilon}}-\overline{\boldsymbol{\delta}}^T\boldsymbol{\varepsilon}_i-\overline{\boldsymbol{\delta}^T\boldsymbol{\varepsilon}}+2\overline{\boldsymbol{\delta}}^T\overline{\boldsymbol{\varepsilon}})\\
&+ e_6(\|\boldsymbol{\delta}_i\|^2-2\boldsymbol{\delta}_i^T\overline{\boldsymbol{\delta}}-\overline{\|\boldsymbol{\delta}\|^2}+2\|\overline{\boldsymbol{\delta}}\|^2).
\end{aligned}$$

where $\overline{\overline{\varepsilon}}=\frac{1}{nm}\sum_{i=1}^n\sum_{j=1}^m\varepsilon_{ij}$, $\overline{\overline{\delta}}=\frac{1}{nm}\sum_{i=1}^n\sum_{j=1}^m\delta_{ij}$, $\overline{(\boldsymbol{\xi}-\boldsymbol{\mu})^T\boldsymbol{\varepsilon}}=\frac{1}{n}\sum_{i=1}^n(\boldsymbol{\xi}_i-\boldsymbol{\mu})^T\boldsymbol{\varepsilon}_i$, $\overline{(\boldsymbol{\xi}-\boldsymbol{\mu})^T\boldsymbol{\delta}}=\frac{1}{n}\sum_{i=1}^n(\boldsymbol{\xi}_i-\boldsymbol{\mu})^T\boldsymbol{\delta}_i$, $\overline{\boldsymbol{\delta}^T\boldsymbol{\varepsilon}}=\frac{1}{n}\sum_{i=1}^n\boldsymbol{\delta}_i^T\boldsymbol{\varepsilon}_i$ and $\overline{\|\boldsymbol{\delta}\|^2}=\frac{1}{n}\sum_{i=1}^n\|\boldsymbol{\delta}_i\|^2$. Also, by (4.4.6)

$$\begin{aligned}
\mathbf{e}^T(\boldsymbol{\zeta}_i'-\overline{\boldsymbol{\zeta}'}) &= e_1[(\boldsymbol{\xi}_i-\boldsymbol{\mu})^T\boldsymbol{\varepsilon}_i-\overline{(\boldsymbol{\xi}-\boldsymbol{\mu})^T\boldsymbol{\varepsilon}}]\\
&+ e_2[(\boldsymbol{\xi}_i-\boldsymbol{\mu})^T\boldsymbol{\delta}_i^T-\overline{(\boldsymbol{\xi}-\boldsymbol{\mu})^T\boldsymbol{\delta}}]\\
&+ e_3\left(\frac{1}{m}\sum_{j=1}^m\varepsilon_{ij}-\overline{\overline{\varepsilon}}\right)+e_4\left(\frac{1}{m}\sum_{j=1}^m\delta_{ij}-\overline{\overline{\delta}}\right)+e_5(\boldsymbol{\delta}_i^T\boldsymbol{\varepsilon}_i-\overline{\boldsymbol{\delta}^T\boldsymbol{\varepsilon}})\\
&+ e_6(\|\boldsymbol{\delta}_i\|^2-\overline{\|\boldsymbol{\delta}\|^2}),
\end{aligned}$$

and therefore:

$$\mathbf{d}^T(\boldsymbol{\eta}_i(n)-\overline{\boldsymbol{\eta}(n)})-\mathbf{e}^T(\boldsymbol{\zeta}_i'-\overline{\boldsymbol{\zeta}'}) = e_1[-(\overline{\boldsymbol{\xi}}-\boldsymbol{\mu})^T\boldsymbol{\varepsilon}_i-(\boldsymbol{\xi}_i-\boldsymbol{\mu})^T\overline{\boldsymbol{\varepsilon}}+2(\overline{\boldsymbol{\xi}}-\boldsymbol{\mu})^T\overline{\boldsymbol{\varepsilon}}]$$

$$
\begin{aligned}
&+ \quad e_2[-(\overline{\boldsymbol{\xi}}-\boldsymbol{\mu})^T\boldsymbol{\delta}_i - (\boldsymbol{\xi}_i-\boldsymbol{\mu})^T\overline{\boldsymbol{\delta}} + 2(\overline{\boldsymbol{\xi}}-\boldsymbol{\mu})^T\overline{\boldsymbol{\delta}}] \\
&+ \quad e_5(-\boldsymbol{\delta}_i^T\overline{\boldsymbol{\varepsilon}} - \overline{\boldsymbol{\delta}}^T\boldsymbol{\varepsilon}_i + 2\overline{\boldsymbol{\delta}}^T\overline{\boldsymbol{\varepsilon}}) \qquad (4.4.19) \\
&+ \quad e_6(-2\boldsymbol{\delta}_i^T\overline{\boldsymbol{\delta}} + 2\|\overline{\boldsymbol{\delta}}\|^2).
\end{aligned}
$$

By Lemma C.0.12, to prove (4.4.15), it is enough to prove

$$\frac{\sum_{i=1}^n[\mathbf{d}^T(\boldsymbol{\eta}_i(n)-\overline{\boldsymbol{\eta}(n)}) - \mathbf{e}^T(\boldsymbol{\zeta}_i'-\overline{\boldsymbol{\zeta}'})]^2}{\sum_{i=1}^n[\mathbf{e}^T(\boldsymbol{\zeta}_i'-\overline{\boldsymbol{\zeta}'})]^2} \xrightarrow{P} 0.$$

If $\mathrm{Var}(\|\boldsymbol{\xi}\|)<\infty$ or $|e_1|+|e_2|=0$, $\frac{1}{n}\sum_{i=1}^n[\mathbf{e}^T(\boldsymbol{\zeta}_i'-\overline{\boldsymbol{\zeta}'})]^2 = O_P(1)$, by the WLLN (see also (4.4.24)). Therefore, it is enough to prove:

$$\frac{1}{n}\sum_{i=1}^n[\mathbf{d}^T(\boldsymbol{\eta}_i(n)-\overline{\boldsymbol{\eta}(n)}) - \mathbf{e}^T(\boldsymbol{\zeta}_i'-\overline{\boldsymbol{\zeta}'})]^2 \xrightarrow{P} 0. \qquad (4.4.20)$$

If $\mathrm{Var}(\|\boldsymbol{\xi}\|)=\infty$ and $|e_1|+|e_2|>0$, by (4.4.25), with $\mathbf{b}=\mathbf{e}$, we have

$$\frac{1}{c_n^2}\sum_{i=1}^n[\mathbf{e}^T(\boldsymbol{\zeta}_i'-\overline{\boldsymbol{\zeta}'})]^2 = O_P(1).$$

Therefore, if we prove (4.4.20), the convergence (4.4.15) follows since:

$$\frac{1}{c_n^2}\sum_{i=1}^n[\mathbf{d}^T(\boldsymbol{\eta}_i(n)-\overline{\boldsymbol{\eta}(n)}) - \mathbf{e}^T(\boldsymbol{\zeta}_i'-\overline{\boldsymbol{\zeta}'})]^2 = \frac{n}{c_n^2}\frac{1}{n}\sum_{i=1}^n[\mathbf{d}^T(\boldsymbol{\eta}_i(n)-\overline{\boldsymbol{\eta}(n)}) - \mathbf{e}^T(\boldsymbol{\zeta}_i'-\overline{\boldsymbol{\zeta}'})]^2,$$

where the first ratio converges to 0.

Hence, it remains to prove (4.4.20). By (4.4.19), we have the following

$$\frac{1}{n}\sum_{i=1}^n[\mathbf{d}^T(\boldsymbol{\eta}_i(n)-\overline{\boldsymbol{\eta}(n)}) - \mathbf{e}^T(\boldsymbol{\zeta}_i'-\overline{\boldsymbol{\zeta}'})]^2 \le 4(e_1)^2T_1 + 4(e_2)^2T_2 + 4(e_5)^2T_3 + 4(e_6)^2T_4,$$

where:

$$
\begin{aligned}
T_1 &:= \frac{1}{n}\sum_{i=1}^n[-(\overline{\boldsymbol{\xi}}-\boldsymbol{\mu})^T\boldsymbol{\varepsilon}_i - (\boldsymbol{\xi}_i-\boldsymbol{\mu})^T\overline{\boldsymbol{\varepsilon}} + 2(\overline{\boldsymbol{\xi}}-\boldsymbol{\mu})^T\overline{\boldsymbol{\varepsilon}}]^2, \\
T_2 &:= \frac{1}{n}\sum_{i=1}^n[-(\overline{\boldsymbol{\xi}}-\boldsymbol{\mu})^T\boldsymbol{\delta}_i - (\boldsymbol{\xi}_i-\boldsymbol{\mu})^T\overline{\boldsymbol{\delta}} + 2(\overline{\boldsymbol{\xi}}-\boldsymbol{\mu})^T\overline{\boldsymbol{\delta}}]^2,
\end{aligned}
$$

$$\begin{aligned}
T_3 &:= \frac{1}{n}\sum_{i=1}^{n}(-\boldsymbol{\delta}_i^T\overline{\boldsymbol{\varepsilon}} - \overline{\boldsymbol{\delta}}^T\boldsymbol{\varepsilon}_i + 2\overline{\boldsymbol{\delta}}^T\overline{\boldsymbol{\varepsilon}})^2,\\
T_4 &:= \frac{1}{n}\sum_{i=1}^{n}(-2\boldsymbol{\delta}_i^T\overline{\boldsymbol{\delta}} + 2\|\overline{\boldsymbol{\delta}}\|^2)^2.
\end{aligned}$$

Using the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
T_1 &\leq 4\frac{1}{n}\sum_{i=1}^{n}[(\overline{\boldsymbol{\xi}}-\boldsymbol{\mu})^T\boldsymbol{\varepsilon}_i]^2 + 4\frac{1}{n}\sum_{i=1}^{n}[(\boldsymbol{\xi}_i-\boldsymbol{\mu})^T\overline{\boldsymbol{\varepsilon}}]^2 + 8\frac{1}{n}\sum_{i=1}^{n}[(\overline{\boldsymbol{\xi}}-\boldsymbol{\mu})^T\overline{\boldsymbol{\varepsilon}}]^2\\
&\leq 4\frac{1}{n}\sum_{i=1}^{n}\left[\|\overline{\boldsymbol{\xi}}-\boldsymbol{\mu}\|^2\|\boldsymbol{\varepsilon}_i\|^2 + \|\overline{\boldsymbol{\varepsilon}}\|^2\|\boldsymbol{\xi}_i-\boldsymbol{\mu}\|^2\right] + 8\|\overline{\boldsymbol{\xi}}-\boldsymbol{\mu}\|^2\|\overline{\boldsymbol{\varepsilon}}\|^2. \qquad (4.4.21)
\end{aligned}$$

By WLLN, $\|\overline{\boldsymbol{\xi}}-\boldsymbol{\mu}\|^2 \xrightarrow{P} 0$ and $\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{\varepsilon}_i\|^2 \xrightarrow{P} \text{trace}(\boldsymbol{\Sigma}_{\varepsilon})$. It follows that the first term of (4.4.21) converges in probability to 0.

Recall that $n\|\overline{\boldsymbol{\varepsilon}}\|^2 = O_P(1)$, and $\frac{\sum_{i=1}^{n}\|\boldsymbol{\xi}_i-\boldsymbol{\mu}\|^2}{b_n^2} \xrightarrow{P} 1$, by Lemma 4.2.3. Hence, we have $\|\overline{\boldsymbol{\varepsilon}}\|^2\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{\xi}_i-\boldsymbol{\mu}\|^2 = O_P(1)\frac{b_n^2}{n^2} = o_P(1)$, by Lemma 4.2.6, (b).

Since $\|\overline{\boldsymbol{\xi}}-\boldsymbol{\mu}\|^2\|\overline{\boldsymbol{\varepsilon}}\|^2 = o_P(1)\frac{O_P(1)}{n} = o_P(1)$, we conclude that $T_1 \to 0$, in probability.

By replacing $\boldsymbol{\varepsilon}$ with $\boldsymbol{\delta}$ we obtain $T_2 \to 0$, in probability. Using the same technique as above, one can prove $T_i \to 0$, in probability, for $i = 3, 4$. This concludes the proof of (4.4.20).

Part (a) can be used to prove part (b). We have:

$$\begin{aligned}
\frac{n\overline{\mathbf{d}^T\boldsymbol{\eta}(n)}}{\sqrt{\sum_{i=1}^{n}(\mathbf{d}^T\boldsymbol{\eta}_i(n))^2}} &= \frac{\sqrt{n}\;\overline{\mathbf{d}^T\boldsymbol{\eta}(n)}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}[\mathbf{d}^T(\boldsymbol{\eta}_i(n)-\overline{\boldsymbol{\eta}(n)})]^2}}\cdot\sqrt{\frac{n}{n-1}}\\
&\quad\cdot\sqrt{\frac{\sum_{i=1}^{n}[\mathbf{d}^T(\boldsymbol{\eta}_i(n)-\overline{\boldsymbol{\eta}(n)})]^2}{\sum_{i=1}^{n}(\mathbf{d}^T\boldsymbol{\eta}_i(n))^2}}.
\end{aligned}$$

By (a), the first factor converges in distribution to $N(0,1)$. Hence, by Slutsky's Theorem, to obtain the convergence in part (b), it suffices to prove:

$$\frac{\sum_{i=1}^{n}(\mathbf{d}^T\boldsymbol{\eta}_i(n))^2}{\sum_{i=1}^{n}[\mathbf{d}^T(\boldsymbol{\eta}_i(n)-\overline{\boldsymbol{\eta}(n)})]^2} \xrightarrow{P} 1. \qquad (4.4.22)$$

Using the result in part (a), we obtain:

$$\frac{n(\overline{\mathbf{d}^T\boldsymbol{\eta}(n)})^2}{\sum_{i=1}^n[\mathbf{d}^T(\boldsymbol{\eta}_i(n)-\overline{\boldsymbol{\eta}(n)})]^2} = \frac{1}{n-1}\left\{\frac{\sqrt{n}\,\overline{\mathbf{d}^T\boldsymbol{\eta}(n)}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n[\mathbf{d}^T(\boldsymbol{\eta}_i(n)-\overline{\boldsymbol{\eta}(n)})]^2}}\right\}^2$$
$$= \frac{O_P(1)}{n-1} = o_p(1).$$

To conclude (4.4.22), we apply Lemma C.0.12, with $s_i = \mathbf{d}^T\boldsymbol{\eta}_i(n)$, $t_i = \mathbf{d}^T[\boldsymbol{\eta}_i(n) - \overline{\boldsymbol{\eta}(n)}]$. □

The next theorem gives CLTs for $\widehat{\beta}_n$, as a direct consequence of Lemma 4.4.4. Note that the normalizing factors in the CLTs depend on the parameter β. In Theorem 4.5.1 we show that the result is still valid with β replaced by $\widehat{\beta}_n$.

Theorem 4.4.5 *Assume that (A1), (A2) and (A3) hold.*

Let $\widehat{\beta}_n$ be the estimator given by (4.1.8). Then, as $n \to \infty$, we have:

$$(a)\quad \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\|\mathbf{x}_i-\overline{\mathbf{x}}\|^2-\mathrm{trace}(\boldsymbol{\Sigma}_\delta)\right](\widehat{\beta}_n-\beta)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n[u_i(n)-\overline{u(n)}]^2}} \xrightarrow{\mathcal{D}} N(0,1),$$

$$(b)\quad \frac{\sum_{i=1}^n\left[\|\mathbf{x}_i-\overline{\mathbf{x}}\|^2-\mathrm{trace}(\boldsymbol{\Sigma}_\delta)\right](\widehat{\beta}_n-\beta)}{\sqrt{\sum_{i=1}^n u_i(n)^2}} \xrightarrow{\mathcal{D}} N(0,1),$$

where $u_i(n)$ is given by (4.4.8) and $\overline{u(n)} = \dfrac{1}{n}\displaystyle\sum_{i=1}^n u_i(n)$.

Proof. The results follow by applying Lemma 4.4.4, with $\mathbf{d} = (0,0,1,-\beta)^T$ (and $\mathbf{e} = (1,-\beta,0,0,1,-\beta)^T$), due to (4.4.9) and (4.4.7).□

To examine the asymptotic behavior of $\widehat{\alpha}_n$ we need three auxiliary results.

The first result gives the convergence rate of $\dfrac{1}{n}\displaystyle\sum_{i=1}^n \mathbf{b}^T\boldsymbol{\zeta}_i$, with $\boldsymbol{\zeta}_i$ given by (4.4.5), $1 \le i \le n$.

Lemma 4.4.6 *Under the same hypotheses as in Lemma 4.4.1, as $n \to \infty$:*

$$\frac{1}{n}\sum_{i=1}^n \mathbf{b}^T\boldsymbol{\zeta}_i = \begin{cases} \frac{1}{\sqrt{n}}O_P(1), & \text{if } \mathrm{Var}(\|\boldsymbol{\xi}\|) < \infty, \text{ or } b_1 = b_2 = 0 \\ \sqrt{\frac{c_n^2}{n^2}}O_P(1), & \text{if } \mathrm{Var}(\|\boldsymbol{\xi}\|) = \infty, \text{ and } |b_1|+|b_2| > 0, \end{cases}$$

where $c_n^2 = \operatorname{trace}(\boldsymbol{\Sigma}_\gamma \mathbf{B}_n^{-2})$, *with* $\boldsymbol{\Sigma}_\gamma = \mathrm{E}(\gamma_i\gamma_i^T)$, $\gamma_i = b_1\boldsymbol{\varepsilon}_i + b_2\boldsymbol{\delta}_i$, *for* $i \le n$ *and* $(\mathbf{B}_n)_{n\ge1}$ *is a sequence of symmetric matrices such that* $\mathbf{B}_n\sum_{i=1}^n(\boldsymbol{\xi}_i - \boldsymbol{\mu}) \xrightarrow{\mathcal{D}} N(\mathbf{0},\mathbf{I})$. *The sequence* $(c_n)_{n\ge1}$ *satisfies* $\frac{c_n^2}{n^2} \to 0$.

Proof. We apply Lemma 4.4.1 to obtain $\mathbf{b}^T\boldsymbol{\zeta}_1 \in DAN$, which by Lemma D.1.7 implies

$$\frac{\sqrt{n}\,\overline{\mathbf{b}^T\boldsymbol{\zeta}}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n[\mathbf{b}^T(\boldsymbol{\zeta}_i - \overline{\boldsymbol{\zeta}})]^2}} \xrightarrow{\mathcal{D}} N(0,1), \tag{4.4.23}$$

where $\overline{\mathbf{b}^T\boldsymbol{\zeta}} = \frac{1}{n}\sum_{i=1}^n \mathbf{b}^T\boldsymbol{\zeta}_i$ and $\overline{\boldsymbol{\zeta}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{\zeta}_i$.

Case 1. Assume that $\mathrm{Var}(\|\boldsymbol{\xi}\|) < \infty$ or $b_1 = b_2 = 0$. Since $\mathrm{Var}(\mathbf{b}^T\boldsymbol{\zeta}_1) < \infty$, we apply the WLLN to obtain:

$$\frac{1}{n}\sum_{i=1}^n[\mathbf{b}^T(\boldsymbol{\zeta}_i - \overline{\boldsymbol{\zeta}})]^2 \xrightarrow{P} \mathrm{Var}(\mathbf{b}^{\mathbf{T}}\boldsymbol{\zeta}_1) > \mathbf{0}. \tag{4.4.24}$$

(The fact that $\mathrm{Var}(\mathbf{b}^{\mathbf{T}}\boldsymbol{\zeta}_1) > \mathbf{0}$ follows either by hypothesis, if $b_1 = b_2 = 0$, or as it was shown in the proof of Lemma 4.4.1, if $|b_1| + |b_2| > 0$.) Using Slutsky's Theorem, (4.4.23) and (4.4.24) we obtain $\overline{\mathbf{b}^T\boldsymbol{\zeta}} = \frac{O_P(1)}{\sqrt{n}}$.

Case 2. Assume $\mathrm{Var}(\|\boldsymbol{\xi}\|) = \infty$ and $|b_1| + |b_2| > 0$.

We write

$$\sqrt{\frac{n(n-1)}{c_n^2}}\,\overline{\mathbf{b}^T\boldsymbol{\zeta}} = \frac{\sqrt{n}\,\overline{\mathbf{b}^T\boldsymbol{\zeta}}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n[\mathbf{b}^T(\boldsymbol{\zeta}_i - \overline{\boldsymbol{\zeta}})]^2}}\sqrt{\frac{\sum_{i=1}^n[\mathbf{b}^T(\boldsymbol{\zeta}_i - \overline{\boldsymbol{\zeta}})]^2}{c_n^2}}.$$

By (4.4.23), the first factor of the product converges in distribution to $N(0,1)$ and we show

$$\sum_{i=1}^n \frac{[\mathbf{b}^T(\boldsymbol{\zeta}_i - \overline{\boldsymbol{\zeta}})]^2}{c_n^2} \xrightarrow{P} 1. \tag{4.4.25}$$

From the proof of Lemma 4.4.1, we recall that, for each $i \le n$

$$\mathbf{b}^T\boldsymbol{\zeta}_i = \boldsymbol{\xi}_i^T\boldsymbol{\gamma}_i + f_b(\boldsymbol{\varepsilon}_i, \boldsymbol{\delta}_i), \tag{4.4.26}$$

where $\boldsymbol{\gamma}_i = b_1\boldsymbol{\varepsilon}_i + b_2\boldsymbol{\delta}_i$ and

$$
\begin{aligned}
f_{\mathbf{b}}(\boldsymbol{\varepsilon}_i, \boldsymbol{\delta}_i) &= b_3\left(\frac{1}{m}\sum_{j=1}^{m}\varepsilon_{ij}\right) + b_4\left(\frac{1}{m}\sum_{j=1}^{m}\delta_{ij}\right) + b_5[\boldsymbol{\delta}_i^T\boldsymbol{\varepsilon}_i - \mathrm{trace}(\boldsymbol{\Sigma}_{\varepsilon\delta})] \\
&+ b_6[\|\boldsymbol{\delta}_i\|^2 - \mathrm{trace}(\boldsymbol{\Sigma}_{\delta})].
\end{aligned}
$$

Note that $\boldsymbol{\Sigma}_\gamma = \mathrm{E}(\boldsymbol{\gamma}_1\boldsymbol{\gamma}_1^T) = b_1^2\boldsymbol{\Sigma}_\varepsilon + b_1b_2\boldsymbol{\Sigma}_{\varepsilon\delta} + b_1b_2\boldsymbol{\Sigma}_{\varepsilon\delta}^T + b_2^2\boldsymbol{\Sigma}_\delta > 0$, since for any $\mathbf{z} \in \mathbb{R}^m$, we have $\mathbf{z}^T\boldsymbol{\Sigma}_\gamma\mathbf{z} = \mathbf{s}^T\boldsymbol{\Sigma}_{\mathrm{error}}\mathbf{s} > 0$, with $\mathbf{s}^T = (b_1\mathbf{z}^T\ b_2\mathbf{z}^T) \in \mathbb{R}^{2m}$. We denote $\boldsymbol{\gamma}_i' = \boldsymbol{\Sigma}_\gamma^{-1/2}\boldsymbol{\gamma}_i$, and $\boldsymbol{\xi}_i' := \boldsymbol{\Sigma}_\gamma^{1/2}(\boldsymbol{\xi}_i - \boldsymbol{\mu})$ for all $i \leq n$. It follows that $\mathrm{Var}(\boldsymbol{\gamma}_i') = \mathbf{I}$ and $\mathrm{E}(\boldsymbol{\xi}_i') = \mathbf{0}$, for all $i \leq n$.

We first prove that

$$\frac{1}{c_n^2}\sum_{i=1}^{n}(\boldsymbol{\xi}_i^T\boldsymbol{\gamma}_i)^2 \xrightarrow{P} 1.$$

Let $(\mathbf{B}_n)_{n\geq 1}$ be a sequence of symmetric matrices such that $\mathbf{B}_n\sum_{i=1}^{n}(\boldsymbol{\xi}_i - \boldsymbol{\mu}) \xrightarrow{\mathfrak{D}} N(\mathbf{0}, \mathbf{I})$.

By Lemma 4.2.5, $\boldsymbol{\xi}_1' \in GDAN$, and $\mathbf{B}_n'\sum_{i=1}^{n}\boldsymbol{\xi}_i' \xrightarrow{\mathfrak{D}} N(\mathbf{0}, \mathbf{I})$, where $\mathbf{B}_n' = (\mathbf{C}_n^T\mathbf{C}_n)^{1/2}$, and $\mathbf{C}_n = \mathbf{B}_n\boldsymbol{\Sigma}_\gamma^{-1/2}$. From the proof of Theorem 4.2.4, Case 2, (a) (applied to $\boldsymbol{\xi}_1'$ and $\boldsymbol{\gamma}_1'$) it follows that:

$$\frac{1}{c_n^2}\sum_{i=1}^{n}[(\boldsymbol{\xi}_i')^T\boldsymbol{\gamma}_i']^2 \xrightarrow{P} 1, \tag{4.4.27}$$

where

$$c_n^2 = \mathrm{trace}[(\mathbf{B}_n')^{-2}] = \mathrm{trace}[(\mathbf{C}_n^T\mathbf{C}_n)^{-1}] = \mathrm{trace}[\boldsymbol{\Sigma}_\gamma^{1/2}\mathbf{B}_n^{-2}\boldsymbol{\Sigma}_\gamma^{1/2}] = \mathrm{trace}[\boldsymbol{\Sigma}_\gamma(\mathbf{B}_n)^{-2}].$$

Since $(\boldsymbol{\xi}_i - \boldsymbol{\mu})^T\boldsymbol{\gamma}_i = (\boldsymbol{\xi}_i')^T\boldsymbol{\gamma}_i'$, (4.4.27) becomes:

$$\frac{1}{c_n^2}\sum_{i=1}^{n}[(\boldsymbol{\xi}_i - \boldsymbol{\mu})^T\boldsymbol{\gamma}_i]^2 \xrightarrow{P} 1.$$

We write:

$$\frac{\sum_{i=1}^{n}(\boldsymbol{\xi}_i^T\boldsymbol{\gamma}_i)^2}{c_n^2} = \frac{\sum_{i=1}^{n}[(\boldsymbol{\xi}_i - \boldsymbol{\mu})^T\boldsymbol{\gamma}_i]^2}{c_n^2} + \frac{\sum_{i=1}^{n}(\boldsymbol{\mu}^T\boldsymbol{\gamma}_i)^2}{c_n^2} + 2\frac{\sum_{i=1}^{n}[(\boldsymbol{\xi}_i - \boldsymbol{\mu})^T\boldsymbol{\gamma}_i](\boldsymbol{\mu}^T\boldsymbol{\gamma}_i)}{c_n^2},$$

and prove that the last two terms of the sum converge to 0 in probability.

Using Remark 4.2.2, (iv), Lemma 4.2.3 and ($A2$), we have $\dfrac{\sum_{i=1}^n \|\boldsymbol{\gamma}_i\|^2}{d_n^2} \xrightarrow{P} 1$, where $d_n^2 := n\text{trace}(\boldsymbol{\Sigma}_\gamma) > 0$. By the Cauchy-Scwharz inequality,

$$\frac{\sum_{i=1}^n (\boldsymbol{\mu}^T\boldsymbol{\gamma}_i)^2}{c_n^2} \leq \|\boldsymbol{\mu}\|^2 \frac{\sum_{i=1}^n \|\boldsymbol{\gamma}_i\|^2}{d_n^2}\frac{d_n^2}{c_n^2}.$$

We have to show that $\lim_{n\to\infty} \dfrac{d_n^2}{c_n^2} = 0$. By Lemma A.1.17, $c_n^2 \geq \dfrac{\text{trace}[(\mathbf{B}_n)^{-2}]}{\text{trace}[(\boldsymbol{\Sigma}_\gamma)^{-1}]}$ and hence $\dfrac{d_n^2}{c_n^2} \leq \text{trace}(\boldsymbol{\Sigma}_\gamma)\text{trace}[(\boldsymbol{\Sigma}_\gamma)^{-1}]\dfrac{n}{\text{trace}[(\mathbf{B}_n)^{-2}]} \to 0$, where we used Lemma 4.2.6, part (a).

By the Cauchy-Schwarz inequality, it also follows that:

$$\left|\frac{\sum_{i=1}^n [(\boldsymbol{\xi}_i - \boldsymbol{\mu})^T\boldsymbol{\gamma}_i](\boldsymbol{\mu}^T\boldsymbol{\gamma}_i)}{c_n^2}\right| \leq \left[\frac{\sum_{i=1}^n [(\boldsymbol{\xi}_i - \boldsymbol{\mu})^T\boldsymbol{\gamma}_i]^2}{c_n^2}\right]^{1/2}\left[\frac{\sum_{i=1}^n (\boldsymbol{\mu}^T\boldsymbol{\gamma}_i)^2}{c_n^2}\right]^{1/2} \xrightarrow{P} 0,$$

and so $\dfrac{\sum_{i=1}^n (\boldsymbol{\xi}_i^T\boldsymbol{\gamma}_i)^2}{c_n^2} \xrightarrow{P} 1$.

We now return to the study of $\overline{\mathbf{b}^T\boldsymbol{\zeta}}$. Using (4.4.26), we obtain:

$$\begin{aligned}
\sum_{i=1}^n \frac{[\mathbf{b}^T(\boldsymbol{\zeta}_i - \overline{\boldsymbol{\zeta}})]^2}{c_n^2} &= \sum_{i=1}^n \frac{(\boldsymbol{\xi}_i^T\boldsymbol{\gamma}_i - \overline{\boldsymbol{\xi}^T\boldsymbol{\gamma}})^2}{c_n^2} + \sum_{i=1}^n \frac{[f_{\mathbf{b}}(\boldsymbol{\varepsilon}_i, \boldsymbol{\delta}_i) - \overline{f_{\mathbf{b}}(\boldsymbol{\varepsilon}, \boldsymbol{\delta})}]^2}{c_n^2} \\
&+ 2\sum_{i=1}^n \frac{(\boldsymbol{\xi}_i^T\boldsymbol{\gamma}_i - \overline{\boldsymbol{\xi}^T\boldsymbol{\gamma}})[f_{\mathbf{b}}(\boldsymbol{\varepsilon}_i, \boldsymbol{\delta}_i) - \overline{f_{\mathbf{b}}(\boldsymbol{\varepsilon}, \boldsymbol{\delta})}]}{c_n^2} \\
&:= I_1 + I_2 + I_3,
\end{aligned}$$

where $\overline{\boldsymbol{\xi}^T\boldsymbol{\gamma}} = \dfrac{1}{n}\sum_{i=1}^n \boldsymbol{\xi}_i^T\boldsymbol{\gamma}_i$ and $\overline{f_{\mathbf{b}}(\boldsymbol{\varepsilon}, \boldsymbol{\delta})} = \dfrac{1}{n}\sum_{i=1}^n f_{\mathbf{b}}(\boldsymbol{\varepsilon}_i, \boldsymbol{\delta}_i)$.

We have $I_1 = \sum_{i=1}^n \dfrac{(\boldsymbol{\xi}_i^T\boldsymbol{\gamma}_i - \overline{\boldsymbol{\xi}^T\boldsymbol{\gamma}})^2}{c_n^2} = \sum_{i=1}^n \dfrac{(\boldsymbol{\xi}_i^T\boldsymbol{\gamma}_i)^2}{c_n^2} - \dfrac{n}{c_n^2}(\overline{\boldsymbol{\xi}^T\boldsymbol{\gamma}})^2 \xrightarrow{P} 1$, since by applying the WLLN and using the fact that $\lim_{n\to\infty} \dfrac{n}{c_n^2} = 0$, the second term converges in probability to 0.

We write:

$$I_2 = \sum_{i=1}^n \frac{[f_{\mathbf{b}}(\boldsymbol{\varepsilon}_i, \boldsymbol{\delta}_i) - \overline{f_{\mathbf{b}}(\boldsymbol{\varepsilon}, \boldsymbol{\delta})}]^2}{c_n^2} = \frac{\sum_{i=1}^n [f_{\mathbf{b}}(\boldsymbol{\varepsilon}_i, \boldsymbol{\delta}_i) - \overline{f_{\mathbf{b}}(\boldsymbol{\varepsilon}, \boldsymbol{\delta})}]^2}{(n-1)\text{Var}[f_{\mathbf{b}}(\boldsymbol{\varepsilon}_1, \boldsymbol{\delta}_1)]}\frac{(n-1)\text{Var}[f_{\mathbf{b}}(\boldsymbol{\varepsilon}_1, \boldsymbol{\delta}_1)]}{c_n^2}.$$

By the WLLN, the first factor converges in probability to 1 and since $\lim_{n\to\infty} \frac{n}{c_n^2} = 0$, the second factor of the product converges in probability to 0.

By the Cauchy-Schwarz inequality, $I_3 \le \sqrt{I_1 I_2}$ and hence $I_3 \xrightarrow{P} 0$. Therefore, (4.4.25) holds.

It remains to prove that $\lim_{n\to\infty} \frac{c_n^2}{n^2} = 0$. Using Lemma A.1.17 and Lemma 4.2.6, (b) we obtain:

$$\frac{c_n^2}{n^2} = \frac{\mathrm{trace}(\mathbf{\Sigma}_\gamma \mathbf{B}_n^{-2})}{n^2} \le \mathrm{trace}(\mathbf{\Sigma}_\gamma)\frac{\mathrm{trace}(\mathbf{B}_n^{-2})}{n^2} \to 0.\square$$

The next result gives the convergence rate of $\overline{\mathbf{d}^T\boldsymbol{\eta}(n)}$. Its proof essentially follows from Lemma 4.4.6, with $\mathbf{b} = \mathbf{e}$, and (4.4.15).

Lemma 4.4.7 *Under the same hypotheses as in Lemma 4.4.4, as $n \to \infty$:*

$$\frac{1}{n}\sum_{i=1}^n \mathbf{d}^T\boldsymbol{\eta}_i(n) = \overline{\mathbf{d}^T\boldsymbol{\eta}(n)} = \begin{cases} \frac{1}{\sqrt{n}} O_P(1), & \textit{if } \mathrm{Var}(\|\boldsymbol{\xi}\|) < \infty \textit{ or } e_1 = e_2 = 0 \\ \sqrt{\frac{c_n^2}{n^2}} O_P(1), & \textit{if } \mathrm{Var}(\|\boldsymbol{\xi}\|) = \infty \textit{ and } |e_1| + |e_2| > 0. \end{cases}$$

where c_n is a sequence of constants, given by Lemma 4.4.6, with $\boldsymbol{\gamma}_i = e_1\boldsymbol{\varepsilon}_i + e_2\boldsymbol{\delta}_i$, $1 \le i \le n$.

Proof. By Lemma 4.4.4, (a)

$$\frac{\sqrt{n}\,\overline{\mathbf{d}^T\boldsymbol{\eta}(n)}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n [\mathbf{d}^T(\boldsymbol{\eta}_i(n) - \overline{\boldsymbol{\eta}(n)})]^2}} \xrightarrow{\mathcal{D}} N(0,1).$$

Case 1. Assume $\mathrm{Var}(\|\boldsymbol{\xi}\|) < \infty$ or $e_1 = e_2 = 0$. We write:

$$\overline{\mathbf{d}^T\boldsymbol{\eta}(n)} = \frac{\sqrt{n}\,\overline{\mathbf{d}^T\boldsymbol{\eta}(n)}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n [\mathbf{d}^T(\boldsymbol{\eta}_i(n) - \overline{\boldsymbol{\eta}(n)})]^2}} \sqrt{\frac{1}{n-1}} \sqrt{\frac{1}{n}\sum_{i=1}^n [\mathbf{d}^T(\boldsymbol{\eta}_i(n) - \overline{\boldsymbol{\eta}(n)})]^2}.$$

From the proof of Lemma 4.4.6, with $\mathbf{b} = \mathbf{e}$ and $\boldsymbol{\zeta}_i$ replaced by $\boldsymbol{\zeta}_i'$ (see relation (4.4.24)) we have $\frac{1}{n}\sum_{i=1}^n [\mathbf{e}^T(\boldsymbol{\zeta}_i' - \overline{\boldsymbol{\zeta}'})]^2 \xrightarrow{P} \mathrm{Var}(\mathbf{e}^T\boldsymbol{\zeta}_1') > 0$. This, together with (4.4.15), implies:

$$\frac{1}{n}\sum_{i=1}^n [\mathbf{d}^T(\boldsymbol{\eta}_i(n) - \overline{\boldsymbol{\eta}(n)})]^2 \xrightarrow{P} \mathrm{Var}(\mathbf{e}^T\boldsymbol{\zeta}_1') > 0, \qquad (4.4.28)$$

and so $\overline{\mathbf{d}^T\boldsymbol{\eta}(n)} = \dfrac{O_P(1)}{\sqrt{n}}$.

Case 2. Assume $\mathrm{Var}(\|\boldsymbol{\xi}\|) = \infty$ and $|e_1| + |e_2| > 0$. We write:

$$\overline{\mathbf{d}^T\boldsymbol{\eta}(n)} = \frac{\sqrt{n}\overline{\mathbf{d}^T\boldsymbol{\eta}(n)}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n[\mathbf{d}^T(\boldsymbol{\eta}_i(n) - \overline{\boldsymbol{\eta}(n)})]^2}}\sqrt{\frac{c_n^2}{n(n-1)}}\sqrt{\frac{1}{c_n^2}\sum_{i=1}^n[\mathbf{d}^T(\boldsymbol{\eta}_i(n) - \overline{\boldsymbol{\eta}(n)})]^2}.$$

From the proof of Lemma 4.4.6, with $\mathbf{b} = \mathbf{e}$ and $\boldsymbol{\zeta}_i$ replaced by $\boldsymbol{\zeta}_i'$ (see relation (4.4.25)), it follows that $\dfrac{1}{c_n^2}\sum_{i=1}^n[\mathbf{e}^T(\boldsymbol{\zeta}_i' - \overline{\boldsymbol{\zeta}'})]^2 \xrightarrow{P} 1$. Hence, we use (4.4.15) to obtain:

$$\frac{1}{c_n^2}\sum_{i=1}^n[\mathbf{d}^T(\boldsymbol{\eta}_i(n) - \overline{\boldsymbol{\eta}(n)})]^2 \xrightarrow{P} 1. \tag{4.4.29}$$

and so $\overline{\mathbf{d}^T\boldsymbol{\eta}(n)} = \sqrt{\dfrac{c_n^2}{n^2}}O_P(1)$.

Here, the convergence rate depends on the sequence $c_n^2 = \mathrm{trace}(\boldsymbol{\Sigma}_\gamma \mathbf{B}_n^{-2})$, where $\gamma_i = e_1\boldsymbol{\varepsilon}_i + e_2\boldsymbol{\delta}_i$, $1 \le i \le n$ and $(\mathbf{B}_n)_{n\ge1}$ is a sequence of symmetric matrices such that $\mathbf{B}_n\sum_{i=1}^n(\boldsymbol{\xi}_i - \boldsymbol{\mu}) \xrightarrow{\mathfrak{D}} N(\mathbf{0}, \mathbf{I})$. □

The next result gives the asymptotic convergence rate of the estimator $\widehat{\beta}_n$ to β.

Lemma 4.4.8 *Assume that* (*A*1), (*A*2) *and* (*A*3) *hold.*

Let $\widehat{\beta}_n$ *be the estimator given by* (4.1.8). *Then, as* $n \to \infty$:

$$(a)\ \frac{b_n^2}{\sqrt{n}}(\widehat{\beta}_n - \beta) \xrightarrow{\mathfrak{D}} N(0, \lambda^2),\ \textit{if}\ \mathrm{Var}(\|\boldsymbol{\xi}\|) < \infty,$$

$$(b)\ \frac{b_n^2}{\sqrt{c_n^2}}(\widehat{\beta}_n - \beta) \xrightarrow{\mathfrak{D}} N(0, 1),\ \textit{if}\ \mathrm{Var}(\|\boldsymbol{\xi}\|) = \infty,$$

where $b_n^2 = \mathrm{trace}(\mathbf{B}_n^{-2})$, $(\mathbf{B}_n)_{n\ge1}$ *is a sequence of* $m \times m$ *non-stochastic matrices such that* $\mathbf{B}_n\sum_{i=1}^n(\boldsymbol{\xi}_i - \boldsymbol{\mu}) \xrightarrow{\mathfrak{D}} N(\mathbf{0}, \mathbf{I})$, $c_n^2 = \mathrm{trace}(\boldsymbol{\Sigma}_\gamma\mathbf{B}_n^{-2})$, $\boldsymbol{\Sigma}_\gamma = \mathrm{E}(\gamma_i\gamma_i^T)$, *with* $\gamma_i = \boldsymbol{\varepsilon}_i - \beta\boldsymbol{\delta}_i$, $1 \le i \le n$ *and* $\lambda > 0$ *is a constant.*

Proof. Recall that $u_i(n) = \mathbf{d}^T\boldsymbol{\eta}_i(n)$, where $\mathbf{d} = (0, 0, 1, -\beta)^T$ (see (4.4.9)). By Lemma 4.4.4, (a):

$$\frac{\sqrt{n}\ \overline{u(n)}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n[u_i(n) - \overline{u(n)}]^2}} \xrightarrow{\mathfrak{D}} N(0, 1), \tag{4.4.30}$$

where $\overline{u(n)} = \frac{1}{n}\sum_{i=1}^{n} u_i(n)$.

Case 1. Assume that $\mathrm{Var}(\|\boldsymbol{\xi}\|) < \infty$. Then, by (4.4.28), with $\mathbf{d}^T = (0, 0, 1, -\beta)$, it follows that:

$$\frac{1}{n-1}\sum_{i=1}^{n}[u_i(n) - \overline{u(n)}]^2 \xrightarrow{P} \lambda = \mathrm{Var}(\mathbf{e}^T\boldsymbol{\zeta}_1') > 0,$$

where $\mathbf{e} = (1, -\beta, 0, 0, 1, -\beta)^T$. We use (4.4.7) and write:

$$\begin{aligned}
\frac{b_n^2}{\sqrt{n}}(\widehat{\beta}_n - \beta) &= \frac{b_n^2}{\sqrt{n}} \cdot \frac{n\overline{u(n)}}{\sum_{i=1}^{n}[\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\delta}})]} \\
&= \frac{\sqrt{n}\overline{u(n)}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}[u_i(n) - \overline{u(n)}]^2}} \cdot b_n^2 \frac{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}[u_i(n) - \overline{u(n)}]^2}}{\sum_{i=1}^{n}[\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\delta}})]}.
\end{aligned}$$

We apply Slutsky's theorem, (4.4.30) and (4.3.2) to obtain the conclusion.

Case 2. Assume that $\mathrm{Var}(\|\boldsymbol{\xi}\|) = \infty$. By (4.4.29), with $\mathbf{d} = (0, 0, 1, -\beta)^T$, it follows that:

$$\frac{1}{c_n^2}\sum_{i=1}^{n}[u_i(n) - \overline{u(n)}]^2 \xrightarrow{P} 1,$$

with $c_n^2 = \mathrm{trace}(\boldsymbol{\Sigma}_{\gamma}\mathbf{B}_n^{-2})$, and $\gamma_i = e_1\boldsymbol{\varepsilon}_i + e_2\boldsymbol{\delta}_i = \boldsymbol{\varepsilon}_i - \beta\boldsymbol{\delta}_i$, since $\mathbf{e} = (1, -\beta, 0, 0, 1, -\beta)^T$. Using (4.4.7) again, we obtain:

$$\begin{aligned}
\frac{b_n^2}{\sqrt{c_n^2}}(\widehat{\beta}_n - \beta) &= \frac{b_n^2}{\sqrt{c_n^2}} \cdot \frac{n\overline{u(n)}}{\sum_{i=1}^{n}[\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\delta}})]} \\
&= \frac{\sqrt{n}\overline{u(n)}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}[u_i(n) - \overline{u(n)}]^2}} \cdot \sqrt{\frac{1}{c_n^2}\sum_{i=1}^{n}[u_i(n) - \overline{u(n)}]^2} \\
&\quad \cdot \sqrt{\frac{n}{n-1}} \cdot \frac{b_n^2}{\sum_{i=1}^{n}[\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\delta}})]}.
\end{aligned}$$

The conclusion follows by Slutsky's theorem, (4.4.30) and (4.3.2).□

The next theorem gives a CLT for $\widehat{\alpha}_n$.

Theorem 4.4.9 *Assume that (A1), (A2) and (A3) hold.*

Let $\widehat{\alpha}_n$ be the estimator given by (4.1.7). Then, as $n \to \infty$, we have:

$$\frac{\sqrt{n}(\widehat{\alpha}_n - \alpha)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^n (v_i(n) - \overline{v(n)})^2}} \xrightarrow{\mathcal{D}} N(0,1),$$

where $v_i(n) = \frac{1}{m}\sum_{j=1}^m y_{ij} - \alpha - \beta\frac{1}{m}\sum_{j=1}^m x_{ij} - \frac{\overline{\overline{x}}}{\frac{1}{n}\sum_{i=1}^n \|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \text{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}})} u_i(n)$,

$u_i(n)$ *is given by* (4.4.8), $\overline{\overline{x}} = \frac{1}{nm}\sum_{i=1}^n\sum_{j=1}^m x_{ij}$ *and* $\overline{v(n)} = \frac{1}{n}\sum_{i=1}^n v_i(n)$.

Proof. We divide the proof into two cases and use the notation $\overline{\mu} = \frac{1}{m}\sum_{j=1}^m \mu_j$.

Case 1. Assume that $\text{Var}(\|\boldsymbol{\xi}\|) < \infty$. We use (4.4.7) in the following decomposition of $\widehat{\alpha}_n - \alpha$:

$$\begin{aligned}
\widehat{\alpha}_n - \alpha &= \overline{\overline{y}} - \alpha - \widehat{\beta}_n \overline{\overline{x}} \\
&= \overline{\overline{y}} - \alpha - \overline{\overline{x}}\frac{\sum_{i=1}^n [u_i(n) + \beta(\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \text{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}}))]}{\sum_{i=1}^n [\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \text{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}})]} \\
&= \overline{\overline{y}} - \alpha - \beta\overline{\overline{x}} - \overline{u(n)}\frac{n\overline{\overline{x}}}{\sum_{i=1}^n [\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \text{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}})]},
\end{aligned}$$

where $\overline{\overline{y}} = \frac{1}{nm}\sum_{i=1}^n\sum_{j=1}^m y_{ij}$.

We introduce $v_i'(n) = \frac{1}{m}\sum_{j=1}^m y_{ij} - \alpha - \beta\frac{1}{m}\sum_{j=1}^m x_{ij} - \frac{\overline{\mu}}{\text{trace}(\mathbf{\Sigma}_{\boldsymbol{\xi}})} u_i(n)$ and note that

$$\overline{v'(n)} = \frac{1}{n}\sum_{i=1}^n v_i'(n) = \overline{\overline{y}} - \alpha - \beta\overline{\overline{x}} - \overline{u(n)}\frac{\overline{\mu}}{\text{trace}(\mathbf{\Sigma}_{\boldsymbol{\xi}})}.$$

Hence

$$\widehat{\alpha}_n - \alpha = \overline{v'(n)} + \overline{u(n)}\left(\frac{\overline{\mu}}{\text{trace}(\mathbf{\Sigma}_{\boldsymbol{\xi}})} - \frac{n\overline{\overline{x}}}{\sum_{i=1}^n [\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \text{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}})]}\right).$$

Using (4.4.9) and Lemma 4.4.7, with $\mathbf{d} = (0, 0.1, -\beta)^T$, we obtain $\sqrt{n}\overline{u(n)} = O_P(1)$.

By (4.3.2), we have $\frac{\sum_{i=1}^n [\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \text{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}})]}{b_n^2} \xrightarrow{P} 1$, where $b_n^2 = n\text{trace}(\mathbf{\Sigma}_{\boldsymbol{\xi}})$ and so

$$\frac{1}{n}\sum_{i=1}^n \|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \text{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}}) \xrightarrow{P} \text{trace}(\mathbf{\Sigma}_{\boldsymbol{\xi}}). \qquad (4.4.31)$$

Using the WLLN, $\overline{\overline{\mathbf{x}}} \xrightarrow{P} \overline{\mu}$ and it follows that

$$\sqrt{n u(n)}\left(\frac{\overline{\mu}}{\operatorname{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\xi}})} - \frac{\overline{\overline{x}}}{\frac{1}{n}\sum_{i=1}^{n}\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \operatorname{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\delta}})}\right) = o_P(1).$$

Therefore, we have

$$\sqrt{n}(\widehat{\alpha}_n - \alpha) = \sqrt{n}\,\overline{v'(n)} + o_P(1). \tag{4.4.32}$$

Note that $v_i'(n) = \mathbf{d}^T\boldsymbol{\eta}_i(n)$, where $\mathbf{d} = \left(1, -\beta, -\frac{\overline{\mu}}{\operatorname{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\xi}})}, \frac{\beta\overline{\mu}}{\operatorname{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\xi}})}\right)^T$, and so using (4.4.28), it follows that

$$\frac{1}{n}\sum_{i=1}^{n}[v_i'(n) - \overline{v_i'(n)}]^2 \xrightarrow{P} \operatorname{Var}(\mathbf{e}^T\boldsymbol{\zeta}_1') > 0, \tag{4.4.33}$$

where $\mathbf{e} = \left(-\frac{\overline{\mu}}{\operatorname{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\xi}})}, \frac{\beta\overline{\mu}}{\operatorname{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\xi}})}, 1, -\beta, -\frac{\overline{\mu}}{\operatorname{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\xi}})}, \frac{\beta\overline{\mu}}{\operatorname{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\xi}})}\right)^T$. Therefore, we apply Slutsky's Theorem, to obtain

$$\frac{\sqrt{n}(\widehat{\alpha}_n - \alpha)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}[v_i'(n) - \overline{v'(n)}]^2}} \xrightarrow{\mathcal{D}} N(0,1). \tag{4.4.34}$$

Slutsky's Theorem applied again completes the proof once we proved that:

$$\frac{\sum_{i=1}^{n}[v_i(n) - \overline{v(n)}]^2}{\sum_{i=1}^{n}[v_i'(n) - \overline{v'(n)}]^2} \xrightarrow{P} 1. \tag{4.4.35}$$

By Lemma C.0.12, to prove (4.4.35), it is enough to show

$$\left[\frac{n\overline{\overline{x}}}{\sum_{i=1}^{n}[\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \operatorname{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\delta}})]} - \frac{\overline{\mu}}{\operatorname{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\xi}})}\right]^2 \cdot \frac{\sum_{i=1}^{n}[u_i(n) - \overline{u(n)}]^2}{\sum_{i=1}^{n}[v_i'(n) - \overline{v'(n)}]^2} = o_P(1).$$

By (4.4.31) the first factor converges in probability to 0.

By (4.4.28), with $\mathbf{d}^* = (0,0,1,-\beta)^T$ and $\mathbf{e}^* = (1,-\beta,0,0,1,-\beta)^T$, we have:

$$\frac{1}{n}\sum_{i=1}^{n}[u_i(n) - \overline{u(n)}]^2 \xrightarrow{P} \operatorname{Var}(\mathbf{e}^{*T}\boldsymbol{\zeta}_1') > 0. \tag{4.4.36}$$

Hence, using (4.4.33), we complete the proof of (4.4.35).

Case 2. Assume that $\mathrm{Var}(\|\boldsymbol{\xi}\|) = \infty$. In this case we use a different decomposition of $\widehat{\alpha}_n - \alpha$. We write:

$$\widehat{\alpha}_n - \alpha = \overline{\overline{y}} - \alpha - \widehat{\beta}_n \overline{\overline{x}} = \overline{\overline{y}} - \alpha - \beta\overline{\overline{x}} - (\widehat{\beta}_n - \beta)\overline{\overline{x}} = \overline{v''} - (\widehat{\beta}_n - \beta)\overline{\overline{x}},$$

where $\overline{v''} = \frac{1}{n}\sum_{i=1}^{n} v_i''$ and $v_i'' = \frac{1}{m}\sum_{j=1}^{m} y_{ij} - \alpha - \beta\frac{1}{m}\sum_{j=1}^{m} x_{ij} = \frac{1}{m}\sum_{j=1}^{m}(\varepsilon_{ij} - \beta\delta_{ij})$. Note that $v_i'' = \mathbf{d}^T\boldsymbol{\eta}_i(n)$, with $\mathbf{d} = (1, -\beta, 0, 0)^T$.

We have:

$$\frac{\sqrt{n}(\widehat{\alpha}_n - \alpha)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}[v_i(n) - \overline{v(n)}]^2}} = \frac{\sqrt{n}\,\overline{v''} - \sqrt{n}\,\overline{\overline{x}}(\widehat{\beta}_n - \beta)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}[v_i'' - \overline{v''}]^2}} \cdot \sqrt{\frac{\sum_{i=1}^{n}[v_i'' - \overline{v''}]^2}{\sum_{i=1}^{n}[v_i(n) - \overline{v(n)}]^2}}. \tag{4.4.37}$$

First we prove $\sqrt{n}(\widehat{\beta}_n - \beta) = o_P(1)$. We apply Lemma 4.4.8 (b) to obtain:

$$\frac{b_n^2}{\sqrt{c_n^2}}(\widehat{\beta}_n - \beta) = O_P(1),$$

where $c_n^2 = \mathrm{trace}(\boldsymbol{\Sigma}_\gamma \mathbf{B}_\mathbf{n}^{-2})$ with $\boldsymbol{\gamma} = \boldsymbol{\varepsilon} - \beta\boldsymbol{\delta}$.

By Lemma A.1.17 we have

$$\begin{aligned} c_n^2 &\leq \mathrm{trace}(\boldsymbol{\Sigma}_\gamma)\mathrm{trace}(\mathbf{B}_\mathbf{n}^{-2}) = (\mathrm{const}_1)b_n^2, \text{ with } \mathrm{const}_1 = \mathrm{trace}(\boldsymbol{\Sigma}_\gamma), \\ c_n^2 &\geq \frac{\mathrm{trace}(\mathbf{B}_\mathbf{n}^{-2})}{\mathrm{trace}(\boldsymbol{\Sigma}_\gamma^{-1})} = (\mathrm{const}_2)b_n^2. \text{ with } \mathrm{const}_2 = [\mathrm{trace}(\boldsymbol{\Sigma}_\gamma^{-1})]^{-1}, \end{aligned}$$

and therefore, the sequence $\left\{\frac{c_n^2}{b_n^2}\right\}_{n\geq 1}$ is bounded. Using Lemma 4.2.6 (a) (ii), we have the following evaluation

$$\begin{aligned} \sqrt{n}(\widehat{\beta}_n - \beta) &= \sqrt{n}\frac{\sqrt{c_n^2}}{b_n^2}\left[\frac{b_n^2}{\sqrt{c_n^2}}(\widehat{\beta}_n - \beta)\right] = \sqrt{n}\frac{\sqrt{c_n^2}}{b_n^2}O_P(1) \\ &= \sqrt{\frac{n}{b_n^2}}\sqrt{\frac{c_n^2}{b_n^2}}O_P(1) = o_P(1). \end{aligned}$$

From here, we obtain $\sqrt{n}(\widehat{\beta}_n - \beta) = o_P(1)$.

By WLLN we have $\overline{\overline{\mathbf{x}}} \xrightarrow{P} \overline{\mu}$ and

$$\frac{1}{n-1}\sum_{i=1}^{n}[v_i'' - \overline{v''}]^2 = \frac{1}{n-1}\sum_{i=1}^{n}[v_i'' - \overline{v''}]^2 \xrightarrow{P} \mathrm{Var}(v_1'') > 0, \tag{4.4.38}$$

where the inequality follows from assumption $(A2)$. To see this, write

$$\begin{aligned}
\mathrm{Var}(v_1'') &= \mathrm{Var}\left[\frac{1}{m}\sum_{j=1}^{m}(\varepsilon_{1j} - \beta\delta_{1j})\right] = \mathrm{Var}\left(\frac{1}{m}\sum_{j=1}^{m}\varepsilon_{1j}\right) + \beta^2\mathrm{Var}\left(\frac{1}{m}\sum_{j=1}^{m}\delta_{1j}\right) \\
&- 2\beta\mathrm{Cov}\left(\frac{1}{m}\sum_{j=1}^{m}\varepsilon_{1j}, \frac{1}{m}\sum_{j=1}^{m}\delta_{1j}\right) \\
&= \frac{1}{m^2}\left(\beta^2\sum_{j,k=1}^{m}\sigma_{\delta,jk} - 2\beta\sum_{j,k=1}^{m}\sigma_{\varepsilon\delta,jk} + \sum_{j,k=1}^{m}\sigma_{\varepsilon,jk}\right) \\
&= \frac{1}{m^2}\mathbf{z}^T\mathbf{\Sigma}_{\text{error}}\mathbf{z} > 0,
\end{aligned}$$

where $\mathbf{z}^T = ((\mathbf{z}^1)^T\ (\mathbf{z}^2)^T)$, with $(\mathbf{z}^1)^T = (1,\ldots,1)$, and $\mathbf{z}^2 = -\beta\mathbf{z}^1$.

Hence, $\dfrac{\sqrt{n}\ \overline{\overline{x}}(\widehat{\beta}_n - \beta)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}[v_i'' - \overline{v''}]^2}} = o_P(1)$, which implies that the first factor in (4.4.37) converges in distribution to $N(0,1)$, by Lemma 4.4.4, with $\mathbf{d} = (1,-\beta,0,0)^T$ and Slutsky's Theorem. To finish the proof, it remains to show:

$$\frac{\sum_{i=1}^{n}[v_i(n) - \overline{v(n)}]^2}{\sum_{i=1}^{n}[v_i'' - \overline{v''}]^2} \xrightarrow{P} 1. \tag{4.4.39}$$

Using Lemma C.0.12, it is enough to prove:

$$\left[\frac{n\overline{\overline{\mathbf{x}}}}{\sum_{i=1}^{n}[\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \mathrm{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}})]}\right]^2 \cdot \frac{\sum_{i=1}^{n}[u_i(n) - \overline{u(n)}]^2}{\sum_{i=1}^{n}[v_i'' - \overline{v''}]^2} = o_P(1). \tag{4.4.40}$$

By (4.3.2), $\sum_{i=1}^{n}[\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \mathrm{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}})] = b_n^2 O_P(1)$, where $b_n^2 = \mathrm{trace}(\mathbf{B}_n^{-2})$ and $(\mathbf{B}_n)_{n\geq 1}$ is a sequence of matrices such that $\mathbf{B}_n\sum_{i=1}^{n}(\boldsymbol{\xi}_i - \boldsymbol{\mu}) \xrightarrow{\mathcal{D}} N(\mathbf{0},\mathbf{I})$. Since, by WLLN, $\overline{\overline{x}} \to \overline{\mu}$, in probability, we conclude that

$$\frac{n\overline{\overline{\mathbf{x}}}}{\sum_{i=1}^{n}[\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \mathrm{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}})]} = \frac{n}{b_n^2}O_P(1). \tag{4.4.41}$$

By (4.4.29), with $\mathbf{d} = (0, 0, 1, -\beta)^T$,

$$\frac{1}{c_n^2} \sum_{i=1}^{n} [u_i(n) - \overline{u(n)}]^2 = O_P(1), \tag{4.4.42}$$

where $c_n^2 = \text{trace}(\mathbf{\Sigma}_\gamma \mathbf{B}_n^{-2})$ and $\boldsymbol{\gamma}_i = \boldsymbol{\varepsilon}_i - \beta \boldsymbol{\delta}_i$, for $i \le n$.

Also, (4.4.38) implies $\dfrac{\sum_{i=1}^n [u_i(n) - \overline{u(n)}]^2}{\sum_{i=1}^n [v_i'' - \overline{v''}]^2} = \dfrac{c_n^2}{n} O_P(1)$.

Finally, using Lemma A.1.17:

$$\left[\frac{n\overline{\overline{\mathbf{x}}}}{\sum_{i=1}^n [\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \text{trace}(\mathbf{\Sigma}_\delta)]} \right]^2 \cdot \frac{\sum_{i=1}^n [u_i(n) - \overline{u(n)}]^2}{\sum_{i=1}^n [v_i'' - \overline{v''}]^2}$$
$$= \left(\frac{n}{b_n^2} \right)^2 \frac{c_n^2}{n} O_P(1) = n \frac{c_n^2}{(b_n^2)^2} O_P(1) \le \frac{n}{b_n^2} \text{trace}(\mathbf{\Sigma}_\gamma) O_P(1) = o_P(1),$$

where we used Lemma 4.2.6, (a) (ii). This proves (4.4.40).□

4.5 General CLTs

The normalizers of the CLTs in Theorem 4.4.5 and Theorem 4.4.9 depend on β. The next result proves that, under the same assumptions, we can substitute the unknown parameter β by its estimator, $\widehat{\beta}_n$, for large values of n. The idea is to show that the normalizer $\sum_{i=1}^n u_i(n)^2$ has the same asymptotic behavior as $\sum_{i=1}^n \tilde{u}_i(n)^2$, where $\tilde{u}_i(n)$ has the same expression as $u_i(n)$, with β replaced by $\widehat{\beta}_n$.

Theorem 4.5.1 *Assume that (A1), (A2) and (A3) hold.*

Then, as $n \to \infty$, we have:

$$(a) \quad \frac{\sum_{i=1}^n [\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \text{trace}(\mathbf{\Sigma}_\delta)](\widehat{\beta}_n - \beta)}{\sqrt{\sum_{i=1}^n \tilde{u}_i(n)^2}} \xrightarrow{\mathcal{D}} N(0,1).$$

$$(b) \quad \frac{\sqrt{n}(\widehat{\alpha}_n - \alpha)}{\sqrt{\frac{1}{n-1} \sum_{i=1}^n (\tilde{v}_i(n) - \overline{\tilde{v}(n)})^2}} \xrightarrow{\mathcal{D}} N(0,1),$$

where $\tilde{u}_i(n) = (\mathbf{x}_i - \overline{\mathbf{x}})^T (\mathbf{y}_i - \overline{\mathbf{y}}) - \text{trace}(\mathbf{\Sigma}_{\varepsilon\delta}) - \widehat{\beta}_n [\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \text{trace}(\mathbf{\Sigma}_\delta)]$, $\widehat{\beta}_n$ *is given by* (4.1.8), $\widehat{\alpha}_n$ *is given by* (4.1.7),

$$\tilde{v}_i(n) = \frac{1}{m} \sum_{j=1}^m y_{ij} - \alpha - \widehat{\beta}_n \frac{1}{m} \sum_{j=1}^m x_{ij} - \frac{n\overline{\overline{x}}}{\sum_{i=1}^n [\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \text{trace}(\mathbf{\Sigma}_\delta)]} \tilde{u}_i(n),$$

and $\overline{\tilde{v}(n)} = \frac{1}{n}\sum_{i=1}^{n}\tilde{v}_i(n)$.

Proof. (a) Note that the numerator of the ratio is equal to $\sum_{i=1}^{n} u_i(n)$, by (4.4.7). We write:

$$\frac{\sum_{i=1}^{n} u_i(n)}{\sqrt{\sum_{i=1}^{n}\tilde{u}_i(n)^2}} = \frac{\sum_{i=1}^{n} u_i(n)}{\sqrt{\sum_{i=1}^{n} u_i(n)^2}} \cdot \sqrt{\frac{\sum_{i=1}^{n} u_i(n)^2}{\sum_{i=1}^{n}\tilde{u}_i(n)^2}}. \tag{4.5.1}$$

The first factor in (4.5.1) converges in distribution to $N(0,1)$, by Theorem 4.4.5 (b) and (4.4.7). The proof is complete once we showed that:

$$\frac{\sum_{i=1}^{n}\tilde{u}_i(n)^2}{\sum_{i=1}^{n} u_i(n)^2} \xrightarrow{P} 1. \tag{4.5.2}$$

By Lemma C.0.12, it is enough to prove:

$$\frac{\sum_{i=1}^{n}[\tilde{u}_i(n) - u_i(n)]^2}{\sum_{i=1}^{n} u_i(n)^2} \xrightarrow{P} 0. \tag{4.5.3}$$

Using (4.4.7) and the definitions of $u_i(n)$ and $\tilde{u}_i(n)$, we have

$$\sum_{i=1}^{n}[\tilde{u}_i(n) - u_i(n)]^2 = \left[\sum_{i=1}^{n} u_i(n)\right]^2 \frac{\sum_{i=1}^{n}[\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \mathrm{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}})]^2}{\{\sum_{i=1}^{n}[\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \mathrm{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}})]\}^2}.$$

Since $\frac{\sum_{i=1}^{n} u_i(n)}{\sqrt{\sum_{i=1}^{n} u_i(n)^2}} \xrightarrow{\mathcal{D}} N(0,1)$, by Theorem 4.4.5 (b), to prove (4.5.3), it suffices to show:

$$\frac{\sum_{i=1}^{n}\left[\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \mathrm{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}})\right]^2}{\left\{\sum_{i=1}^{n}\left[\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \mathrm{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}})\right]\right\}^2} \xrightarrow{P} 0. \tag{4.5.4}$$

The convergence in (4.5.4) follows from (4.3.2) and:

$$\frac{\sum_{i=1}^{n}\left[\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \mathrm{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}})\right]^2}{(b_n^2)^2} \xrightarrow{P} 0, \tag{4.5.5}$$

where $b_n^2 = \mathrm{trace}(\mathbf{B}_n^{-2})$. Here the sequence $(\mathbf{B}_n)_{n\geq 1}$ is such that

$$\mathbf{B}_n\sum_{i=1}^{n}(\boldsymbol{\xi}_i - \boldsymbol{\mu}) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{I}).$$

We write:

$$\frac{\sum_{i=1}^{n}\left[\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \mathrm{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}})\right]^2}{(b_n^2)^2} = \frac{\sum_{i=1}^{n}\|\mathbf{x}_i - \overline{\mathbf{x}}\|^4}{(b_n^2)^2} - 2\mathrm{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}})\frac{\sum_{i=1}^{n}\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2}{(b_n^2)^2}$$

$$
\begin{aligned}
&+ \quad [\mathrm{trace}(\boldsymbol{\Sigma}_\delta)]^2 \frac{n}{(b_n^2)^2} \\
&:= \quad J_1 + J_2 + J_3.
\end{aligned}
$$

We use (4.3.2) to obtain $J_2 \xrightarrow{P} 0$ and Lemma 4.2.6, (a) to obtain $J_3 \to 0$. It remains to show that $J_1 \xrightarrow{P} 0$. We have:

$$
\begin{aligned}
\frac{\sum_{i=1}^n \|\mathbf{x}_i - \overline{\mathbf{x}}\|^4}{(b_n^2)^2} &\le 4\frac{\sum_{i=1}^n \|\boldsymbol{\xi}_i - \overline{\boldsymbol{\xi}}\|^4}{(b_n^2)^2} + 4\frac{\sum_{i=1}^n \|\boldsymbol{\delta}_i - \overline{\boldsymbol{\delta}}\|^4}{(b_n^2)^2} \\
&\le 16\left(\frac{\sum_{i=1}^n \|\boldsymbol{\xi}_i - \boldsymbol{\mu}\|^4}{(b_n^2)^2} + \frac{\sum_{i=1}^n \|\boldsymbol{\delta}_i\|^4}{(b_n^2)^2}\right. \\
&+ \left. n\frac{\|\overline{\boldsymbol{\xi}} - \boldsymbol{\mu}\|^4}{(b_n^2)^2} + n\frac{\|\overline{\boldsymbol{\delta}}\|^4}{(b_n^2)^2}\right) \\
&:= 16(T_1 + T_2 + T_3 + T_4).
\end{aligned}
$$

We that $T_i \xrightarrow{P} 0$, for all $i = 1, 2, 3, 4$. We first treat T_1 and write:

$$
T_1 = \frac{\sum_{i=1}^n \|\boldsymbol{\xi}_i - \boldsymbol{\mu}\|^4}{(b_n^2)^2} = \frac{\sum_{i=1}^n \|\boldsymbol{\xi}_i - \boldsymbol{\mu}\|^4}{\left(\sum_{i=1}^n \|\boldsymbol{\xi}_i - \boldsymbol{\mu}\|^2\right)^2} \cdot \left(\frac{\sum_{i=1}^n \|\boldsymbol{\xi}_i - \boldsymbol{\mu}\|^2}{b_n^2}\right)^2,
$$

where the last factor converges in probability to 1, by Lemma 4.2.3 and therefore is bounded, in probability. Hence, $T_1 \xrightarrow{P} 0$, if we prove that $\frac{\sum_{i=1}^n \|\boldsymbol{\xi}_i - \boldsymbol{\mu}\|^4}{\left(\sum_{i=1}^n \|\boldsymbol{\xi}_i - \boldsymbol{\mu}\|^2\right)^2} \xrightarrow{P} 0$.

Let $\epsilon > 0$ be arbitrary. By Markov's inequality and the fact that $(\boldsymbol{\xi}_i)_{1 \le i \le n}$ are i.i.d. random vectors:

$$
\begin{aligned}
P\left(\left|\frac{\sum_{i=1}^n \|\boldsymbol{\xi}_i - \boldsymbol{\mu}\|^4}{\left(\sum_{i=1}^n \|\boldsymbol{\xi}_i - \boldsymbol{\mu}\|^2\right)^2}\right| > \epsilon\right) &\le \epsilon^{-1} n \mathrm{E}\left[\frac{\|\boldsymbol{\xi}_1 - \boldsymbol{\mu}\|^4}{\left(\sum_{i=1}^n \|\boldsymbol{\xi}_i - \boldsymbol{\mu}\|^2\right)^2}\right] \\
&= \epsilon^{-1} n \mathrm{E}\left[\left(\frac{\mathbf{Z}_1^T \mathbf{Z}_1}{\sum_{i=1}^n \mathbf{Z}_i^T \mathbf{Z}_i}\right)^2\right], \qquad (4.5.6)
\end{aligned}
$$

where we have denoted $\mathbf{Z}_i = \boldsymbol{\xi}_i - \boldsymbol{\mu}$. We have the following inequalities:

$$
\begin{aligned}
\frac{\mathbf{Z}_1^T \mathbf{Z}_1}{\sum_{i=1}^n \mathbf{Z}_i^T \mathbf{Z}_i} &\le \frac{1}{\sum_{i=1}^n \lambda_{\max}(\mathbf{Z}_i \mathbf{Z}_i^T)} \mathbf{Z}_1^T \mathbf{Z}_1 \le \lambda_{\min}\left[\left(\sum_{i=1}^n \mathbf{Z}_i \mathbf{Z}_i^T\right)^{-1}\right] \mathbf{Z}_1^T \mathbf{Z}_1 \\
&\le \mathbf{Z}_1^T \left(\sum_{i=1}^n \mathbf{Z}_i \mathbf{Z}_i^T\right)^{-1} \mathbf{Z}_1. \qquad (4.5.7)
\end{aligned}
$$

From Theorem D.2.5, it follows that $n\mathrm{E}\left\{\left[\mathbf{Z}_1^T\left(\sum_{i=1}^{n}\mathbf{Z}_i\mathbf{Z}_i^T\right)^{-1}\mathbf{Z}_1\right]^2\right\} \to 0$, since $\mathbf{Z}_1 \in GDAN$ and hence by (4.5.7) and (4.5.6), we obtain $\frac{\sum_{i=1}^n \|\boldsymbol{\xi}_i - \boldsymbol{\mu}\|^4}{\left(\sum_{i=1}^n \|\boldsymbol{\xi}_i - \boldsymbol{\mu}\|^2\right)^2} = o_P(1)$.

Similarly, $\frac{\sum_{i=1}^n \|\boldsymbol{\delta}_i\|^4}{\left(\sum_{i=1}^n \|\boldsymbol{\delta}_i\|^2\right)^2} = o_P(1)$, since $\boldsymbol{\delta}_1 \in GDAN$. We write:

$$T_2 = \frac{\sum_{i=1}^n \|\boldsymbol{\delta}_i\|^4}{(b_n^2)^2} = o_P(1)\cdot\left[\frac{\sum_{i=1}^n \|\boldsymbol{\delta}_i\|^2}{n\mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\delta}})}\right]^2\cdot\left[\frac{n\mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\delta}})}{b_n^2}\right]^2.$$

By WLLN, $\frac{\sum_{i=1}^n \|\boldsymbol{\delta}_i\|^2}{n\mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\delta}})} \to 1$, in probability. In addition, $\frac{n\mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\delta}})}{b_n^2}$ is bounded, by Lemma 4 2.6, (a) and hence, $T_2 \to 0$, in probability.

Also, $T_i \to 0$, in probability, where $i = 3, 4$, by WLLN and Lemma 4.2.6, (a). This finishes the proof of $J_1 \xrightarrow{P} 0$.

For the proof of (b), we use Theorem 4.4.9 and Slutsky's Theorem. Hence, it suffices to show:

$$\frac{\sum_{i=1}^n[\tilde{v}_i(n) - \overline{\tilde{v}(n)}]^2}{\sum_{i=1}^n[v_i(n) - \overline{v(n)}]^2} \xrightarrow{P} 1 \tag{4 5.8}$$

By Lemma C.0.12, it suffices to prove.

$$\frac{\sum_{i=1}^n\{\tilde{v}_i(n) - v_i(n) - [\overline{\tilde{v}(n)} - \overline{v(n)}]\}^2}{\sum_{i=1}^n[v_i(n) - \overline{v(n)}]^2} \xrightarrow{P} 0 \tag{4.5.9}$$

We have:

$$\begin{aligned}
&\sum_{i=1}^n\{\tilde{v}_i(n) - v_i(n) - [\overline{\tilde{v}(n)} - \overline{v(n)}]\}^2 \\
&= \sum_{i=1}^n\left\{(\beta - \widehat{\beta}_n)\left(\frac{1}{m}\sum_{j=1}^m x_{ij} - \overline{\overline{x}}\right) - \frac{n\overline{\overline{x}}}{\sum_{i=1}^n[\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\delta}})]}\right. \\
&\quad \left.\cdot\{\tilde{u}_i(n) - u_i(n) - [\overline{\tilde{u}(n)} - \overline{u(n)}]\}\right\}^2 \\
&\le 2(\beta - \widehat{\beta}_n)^2\sum_{i=1}^n\left(\frac{1}{m}\sum_{j=1}^m x_{ij} - \overline{\overline{x}}\right)^2 + 2\left(\frac{n\overline{\overline{x}}}{\sum_{i=1}^n[\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \mathrm{trace}(\boldsymbol{\Sigma}_{\boldsymbol{\delta}})]}\right)^2 \\
&\quad \cdot\sum_{i=1}^n\{\tilde{u}_i(n) - u_i(n) - [\overline{\tilde{u}(n)} - \overline{u(n)}]\}^2
\end{aligned}$$

$$= 2I_1 + 2I_2 \cdot I_3,$$

where we have denoted by:

$$\begin{aligned}
I_1 &:= (\widehat{\beta}_n - \beta)^2 \sum_{i=1}^{n} \left(\frac{1}{m} \sum_{j=1}^{m} x_{ij} - \overline{\overline{x}} \right)^2, \\
I_2 &:= \left(\frac{n\overline{\overline{x}}}{\sum_{i=1}^{n} [\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \mathrm{trace}(\mathbf{\Sigma}_\delta)]} \right)^2, \\
I_3 &:= \sum_{i=1}^{n} \{\tilde{u}_i(n) - u_i(n) - [\overline{\tilde{u}(n)} - \overline{u(n)}]\}^2.
\end{aligned}$$

By (4.4.41),

$$I_2 = \frac{n^2}{(b_n^2)^2} O_P(1). \tag{4.5.10}$$

Since $\tilde{u}_i(n) - u_i(n) = (\beta - \widehat{\beta}_n)[\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \mathrm{trace}(\mathbf{\Sigma}_\delta)]$, it follows that:

$$\begin{aligned}
I_3 &\leq 2(\widehat{\beta}_n - \beta)^2 \left\{ \sum_{i=1}^{n} [\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \mathrm{trace}(\mathbf{\Sigma}_\delta)]^2 + n \left[\frac{1}{n} \sum_{i=1}^{n} \|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \mathrm{trace}(\mathbf{\Sigma}_\delta) \right]^2 \right\} \\
&= 2(\widehat{\beta}_n - \beta)^2 \left\{ \sum_{i=1}^{n} [\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \mathrm{trace}(\mathbf{\Sigma}_\delta)]^2 + \frac{1}{n} \left[\sum_{i=1}^{n} [\|\mathbf{x}_i - \overline{\mathbf{x}}\|^2 - \mathrm{trace}(\mathbf{\Sigma}_\delta)] \right]^2 \right\}, \\
&= (\widehat{\beta}_n - \beta)^2 \left[(b_n^2)^2 o_P(1) + \frac{1}{n} (b_n^2)^2 O_P(1) \right] = 2(\widehat{\beta}_n - \beta)^2 (b_n^2)^2 o_P(1),
\end{aligned} \tag{4.5.11}$$

by using (4.5.5) and (4.3.2).

From (4.5.10) and (4.5.11) we get $I_2 \cdot I_3 = n^2 (\widehat{\beta}_n - \beta)^2 o_P(1)$.

We have the following inequality:

$$I_1 \leq 2(\widehat{\beta}_n - \beta)^2 \left[\sum_{i=1}^{n} \left(\frac{1}{m} \sum_{j=1}^{m} \xi_{ij} - \overline{\overline{\boldsymbol{\xi}}} \right)^2 + \sum_{i=1}^{n} \left(\frac{1}{m} \sum_{j=1}^{m} \delta_{ij} - \overline{\overline{\boldsymbol{\delta}}} \right)^2 \right],$$

where $\overline{\overline{\boldsymbol{\xi}}} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \xi_{ij}$ and $\overline{\overline{\boldsymbol{\delta}}} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} \delta_{ij}$.

Since the asymptotic behavior of $\widehat{\beta}_n - \beta$ is different in the cases $\mathrm{Var}(\|\boldsymbol{\xi}\|) < \infty$ and $\mathrm{Var}(\|\boldsymbol{\xi}\|) = \infty$, we have to consider these cases separately.

Case 1. Assume $\mathrm{Var}(\|\boldsymbol{\xi}\|) < \infty$.

We apply WLLN and so:

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{m}\sum_{j=1}^{m}\xi_{ij} - \overline{\overline{\boldsymbol{\xi}}}\right)^2 \xrightarrow{P} \frac{1}{m^2}\sum_{j,k=1}^{m}\sigma_{\boldsymbol{\xi},jk},$$
$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{m}\sum_{j=1}^{m}\delta_{ij} - \overline{\overline{\boldsymbol{\delta}}}\right)^2 \xrightarrow{P} \frac{1}{m^2}\sum_{j,k=1}^{m}\sigma_{\boldsymbol{\delta},jk}.$$

By Lemma 4.4.8, (a) and Lemma 4.2.6, (a), we obtain $\frac{1}{n}I_1 \le \frac{n}{(b_n^2)^2}O_P(1) = o_P(1)$ and $\frac{1}{n}I_2 \cdot I_3 = \frac{1}{n}n^2 O_P(1)\frac{n}{(b_n^2)^2}o_P(1) = o_p(1)$.

Hence, $\frac{1}{n}\sum_{i=1}^{n}\{\tilde{v}_i(n) - v_i(n) - [\overline{\tilde{v}(n)} - \overline{v(n)}]\}^2 = o_P(1)$.

By (4.4.35) and (4.4.33), the sequence $\left\{\frac{1}{n}\sum_{i=1}^{n}[v_i(n) - \overline{v(n)}]^2\right\}_{n\ge1}$ has a positive finite limit, in probability. Hence, (4.5.9) follows in this case.

Case 2. Assume $\mathrm{Var}(\|\boldsymbol{\xi}\|) = \infty$.

By Assumption $(A1)$, $\boldsymbol{\xi}_1 \in GDAN$ and therefore, by Remark D.2.3, we have $\frac{1}{m}\sum_{j=1}^{m}\xi_{1j} \in DAN$. Let $(\overline{b}_n)_{n\ge1}$ be a sequence of constants such that

$$\frac{1}{\overline{b}_n}\sum_{i=1}^{n}\left(\frac{1}{m}\sum_{j=1}^{m}\xi_{ij} - \overline{\mu}\right) \xrightarrow{\mathcal{D}} N(0,1).$$

Hence, $\frac{1}{\overline{b}_n^2}\sum_{i=1}^{n}\left(\frac{1}{m}\sum_{j=1}^{m}\xi_{ij} - \overline{\mu}\right)^2 \xrightarrow{P} 1$. By WLLN, we have:

$$\frac{1}{\overline{b}_n^2}\sum_{i=1}^{n}\left(\frac{1}{m}\sum_{j=1}^{m}\xi_{ij} - \overline{\overline{\boldsymbol{\xi}}}\right)^2 = \frac{1}{\overline{b}_n^2}\left[\sum_{i=1}^{n}\left(\frac{1}{m}\sum_{j=1}^{m}\xi_{ij} - \overline{\mu}\right)^2 - n(\overline{\overline{\boldsymbol{\xi}}} - \overline{\mu})^2\right] \xrightarrow{P} 1,$$
$$\frac{1}{\overline{b}_n^2}\sum_{i=1}^{n}\left(\frac{1}{m}\sum_{j=1}^{m}\delta_{ij} - \overline{\overline{\boldsymbol{\delta}}}\right)^2 = \frac{n}{\overline{b}_n^2}O_P(1) = o_P(1),$$

since $\overline{b}_n^2 = nl^2(n)$, where $l(n)$ it is a slowly varying function at ∞. Hence

$$\sum_{i=1}^{n}\left(\frac{1}{m}\sum_{j=1}^{m}\xi_{ij}-\overline{\overline{\xi}}\right)^2+\sum_{i=1}^{n}\left(\frac{1}{m}\sum_{j=1}^{m}\delta_{ij}-\overline{\overline{\delta}}\right)^2=\overline{b}_n^2[O_P(1)+o_P(1)].$$

By Lemma 4.4.8, $\widehat{\beta}_n-\beta=\frac{\sqrt{c_n^2}}{b_n^2}O_P(1)$, with $c_n^2=\text{trace}(\mathbf{\Sigma}_\gamma\mathbf{B}_n^{-2})$ and $\gamma_i=\varepsilon_i-\beta\delta_i$, for $i\le n$. Hence, $(\widehat{\beta}_n-\beta)^2=\frac{c_n^2}{(b_n^2)^2}O_P(1)$. By Lemma A.1.17, we conclude that the sequence $\left\{\frac{c_n^2}{b_n^2}\right\}_{n\ge 1}$ is bounded, whereas Lemma 4.2.6 implies that $\frac{\overline{b}_n^2c_n^2}{(n-1)b_n^2}\to 0$ and $\frac{n}{b_n^2}\to 0$. Therefore,

$$\begin{aligned}\frac{1}{n-1}I_1 &= \frac{\overline{b}_n^2c_n^2}{(n-1)(b_n^2)^2}O_P(1)[O_P(1)+o_P(1)]=o_P(1).\\ \frac{1}{n-1}I_2\cdot I_3 &= \frac{n^2c_n^2}{(n-1)(b_n^2)^2}O_P(1)o_P(1)=o_P(1).\end{aligned}$$

and so,

$$\frac{1}{n-1}\sum_{i=1}^{n}\{\tilde{v}_i(n)-v_i(n)-[\overline{\tilde{v}(n)}-\overline{v(n)}]\}^2=o_P(1).$$

From (4.4.39) and (4.4.38), it follows that $\left\{\frac{1}{n-1}\sum_{i=1}^{n}[v_i(n)-\overline{v(n)}]^2\right\}_{n\ge 1}$ has a positive finite limit, in probability, which concludes the proof of (4.5.9).□

Remark 4.5.2 (see also Remark 4.1.4)

In the case of replications, the assumption that $\boldsymbol{\xi}$ has a full distribution does not hold. Nevertheless, the results remain valid, under slightly different conditions:

$(A1^*)$ ξ_1 lies in the domain of attraction of the normal law (DAN)

$(A2^*)$ $E(\varepsilon_{1j})=0$, $E(\delta_{1j})=0$, $E(\varepsilon_{1j}^4)<\infty$, $E(\delta_{1j}^4)<\infty$, with $1\le j\le m$, and $\mathbf{\Sigma}^*_{\text{error}}:=\begin{pmatrix}\sigma^*_{\varepsilon} & \sigma^*_{\varepsilon\delta}\\ \sigma^*_{\varepsilon\delta} & \sigma^*_{\delta}\end{pmatrix}$ is positive definite

$(A3^*)$ $(\xi_i)_{1\le i\le n}$ and $\{(\varepsilon_i,\delta_i)\}_{1\le i\le n}$ are independent.

where $\sigma_{\varepsilon}^{*} = \sum_{j,k=1}^{m} \sigma_{\varepsilon,jk}$, $\sigma_{\delta}^{*} = \sum_{j,k=1}^{m} \sigma_{\delta,jk}$. $\sigma_{\varepsilon\delta}^{*} = \sum_{j,k=1}^{m} \sigma_{\varepsilon\delta,jk}$.

Note that assumption (A_2^*) is weaker than assumption (A_2) since $\mathbf{\Sigma}_{\text{error}} > 0$ implies $\mathbf{\Sigma}_{\text{error}}^{*} > 0$.

In this case,

$$\begin{aligned}\zeta_1 \;=\; & \Bigg(\xi_1 \sum_{j=1}^{m} \varepsilon_{1j}, \xi_1 \sum_{j=1}^{m} \delta_{1j}, \frac{1}{m} \sum_{j=1}^{m} \varepsilon_{1j}, \frac{1}{m} \sum_{j=1}^{m} \delta_{1j}, \boldsymbol{\delta}_1^T \boldsymbol{\varepsilon}_1 - \text{trace}(\mathbf{\Sigma}_{\boldsymbol{\varepsilon}\boldsymbol{\delta}}), \\ & \|\boldsymbol{\delta}_1\|^2 - \text{trace}(\mathbf{\Sigma}_{\boldsymbol{\delta}})\Bigg)^T .\end{aligned}$$

One can follow the steps in the proof of Lemma 4.4.1 and use Lemma D.1.8 in lieu of Theorem 4.2.4 to obtain the result. Furthermore, one can prove Theorem 4.4.2, Theorem 4.4.5 and Theorem 4.5.1, using similar arguments.

Note that in this special case of repeated measurements, the converse of Lemma 4.4.1 also holds, by Theorem D.1.6. Hence, in this case the assumption $(A1^*)$ becomes necessary for the CLTs in Theorem 4.4.2 and Theorem 4.4.5, by Lemma 7 in [27].

Appendix A

Matrix Analysis Results

A.1 Norms and Inequalities

This appendix section contains some auxiliary results which are used in the present work. Our major reference is [33].

Definition A.1.1 *We say that $\lambda \in \mathbb{C}$ is an eigenvalue of an $n \times m$ matrix $\boldsymbol{A}$ if there exists an $m \times 1$ vector $\boldsymbol{x} \neq 0$ such that: $\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}$.*

If $\mathbf{A}$ is a symmetric $m \times m$ matrix, then all its eigenvalues are real numbers. We denote these eigenvalues by $\lambda_1(\mathbf{A}), \ldots, \lambda_m(\mathbf{A})$. We use $\lambda_{\min}(\mathbf{A})$ and $\lambda_{\max}(\mathbf{A})$ to denote the minimum and the maximum eigenvalues.

Theorem A.1.2 (Theorem of 3.4 [33]) *If $\boldsymbol{A}$ is a is a symmetric, non-singular $m \times m$ matrix, then for any $i \leq m$:*

$$\lambda_i(\boldsymbol{A}^{-1}) = \frac{1}{\lambda_i(\boldsymbol{A})}.$$

In particular, $\lambda_{\max}(\boldsymbol{A}^{-1}) = 1/\lambda_{\min}(\boldsymbol{A})$.

Theorem A.1.3 (Theorem 3.15 of [33]) *If $\boldsymbol{A} = (a_{ij})_{1\leq i,j\leq m}$ is a symmetric $m \times m$ matrix, then:*

$$\lambda_{\min}(\boldsymbol{A})\boldsymbol{x}^T\boldsymbol{x} \leq \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \leq \lambda_{\max}(\boldsymbol{A})\boldsymbol{x}^T\boldsymbol{x},$$

for any $m \times 1$ vector $\boldsymbol{x}$.

Lemma A.1.4 *Let $\boldsymbol{A}$ be a symmetric $m \times m$ matrix. We have:*

$$\lambda_{\min}(\boldsymbol{A}) \leq a_{jj} \leq \lambda_{\max}(\boldsymbol{A}). \tag{A.1.1}$$

Proof. Use Theorem A.1.3 with $\mathbf{x} = \mathbf{e}_j$, where $\mathbf{e}_j$ is the $m \times 1$ vector which has 1 on the j-th entry and 0 elsewhere. □

Definition A.1.5 *Let $\boldsymbol{A}$ be a $n \times m$ matrix. . The spectral norm of $\boldsymbol{A}$ is defined by:*

$$\|\boldsymbol{A}\| = \sup_{\|\boldsymbol{x}\|=1} \|\boldsymbol{A}\boldsymbol{x}\|.$$

Definition A.1.6 *Let $\boldsymbol{A}$ be a $n \times m$ matrix. The Euclidean norm of $\boldsymbol{A}$ is defined by:*

$$\|\boldsymbol{A}\|_E = \left(\sum_{i=1}^{n}\sum_{j=1}^{m} |a_{ij}|^2\right)^{1/2}.$$

Definition A.1.7 *Let $\boldsymbol{A}$ be an $m \times m$ matrix. The numerical radius of $\boldsymbol{A}$ is defined by:*

$$|||\boldsymbol{A}||| = \sup_{\|\boldsymbol{x}\|=1} |\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}|.$$

Definition A.1.8 *Let $\boldsymbol{A}$ be an $m \times m$ matrix.*

We say that $\boldsymbol{A}$ is a nonnegative definite and we write $\boldsymbol{A} \geq 0$ if $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \geq 0$ for any $m \times 1$ vector $\boldsymbol{x}$.

We say that $\boldsymbol{A}$ is a positive definite and we write $\boldsymbol{A} > 0$ if $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} > 0$ for any $m \times 1$ vector $\boldsymbol{x}$.

For the $m \times m$ matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, we use the notation $\boldsymbol{A} \geq \boldsymbol{B}$ if $\boldsymbol{A} - \boldsymbol{B} \geq 0$.

Remark A.1.9 We have the following properties for the norms introduced above: (i) For any $m \times m$ symmetric matrix $\mathbf{A}$, $\|\mathbf{A}\| = [\lambda_{\max}(\mathbf{A}^T\mathbf{A})]^{1/2} = \max_{1\leq i\leq m} \lambda_i(\mathbf{A}^T\mathbf{A})$, where $(\lambda_i)_{1\leq i\leq m}$ are the eigenvalues of $\mathbf{A}^T\mathbf{A}$. If $\mathbf{A} \geq 0$, then $\|\mathbf{A}\| = \lambda_{\max}(\mathbf{A})$.
(ii) $\|\mathbf{A}\|_E = [\text{trace}(\mathbf{A}^T\mathbf{A})]^{1/2}$.
(iii) If $\mathbf{A}$ is an $m \times m$ symmetric then $|||\mathbf{A}||| = \max_{1\leq i\leq m} |\lambda_i(\mathbf{A})|$.

Remark A.1.10 (see Theorem 3.16 in [38])

For any an $n \times m$ matrix $\mathbf{A}$ the Euclidean norm and the spectral norm are equivalent. In addition, if $\mathbf{B}$ is a matrix of dimension $m \times m$, the spectral norm and the numerical radius are also equivalent. More precisely,

$$\begin{aligned} \|\mathbf{A}\| &\leq \|\mathbf{A}\|_E \leq \sqrt{nm}\|\mathbf{A}\|. \\ |||\mathbf{B}||| &\leq \|\mathbf{B}\| \leq 2|||\mathbf{B}|||. \end{aligned}$$

Remark A.1.11 (Theorem 3.21 of [33]) *If* $\mathbf{A}$ *and* $\mathbf{B}$ *are* $m \times m$ *symmetric matrices, then, for any* $1 \leq i \leq m$, *we have*

$$\lambda_i(\mathbf{A}) + \lambda_{\min}(\mathbf{B}) \leq \lambda_i(\mathbf{A}+\mathbf{B}) \leq \lambda_i(\mathbf{A}) + \lambda_{\max}(\mathbf{B}).$$

Lemma A.1.12 *For any* $n \times m$ *matrix* $\boldsymbol{A}$, *we have:*

$$\max_{i\leq n, j\leq m} |a_{ij}| \leq \|\boldsymbol{A}\| \leq \sqrt{nm} \max_{i\leq n, j\leq m} |a_{ij}|.$$

Proof.

$$\|\mathbf{A}\|_E^2 = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}^2 \leq nm \max_{i,j} a_{ij}^2 = \{\sqrt{nm} \max_{i\leq n, j\leq m} |a_{ij}|\}^2.$$

Using Remark A.1.10, it follows that $\|\mathbf{A}\| \leq \sqrt{nm}\max_{ij} |a_{ij}|$.
For the other inequality, using (A.1.1), we have:

$$\|\mathbf{A}\|^2 = \lambda_{\max}(\mathbf{A}^T\mathbf{A}) \geq \max_{j\leq m}(\mathbf{A}^T\mathbf{A})_{jj} = \max_{j\leq m}\sum_{i=1}^{n} a_{ij}^2 \geq \max_{j\leq m}\max_{i\leq n} a_{ij}^2$$

Hence, $\max_{i,j} |a_{ij}| \leq \|\mathbf{A}\|$. $\square$

Lemma A.1.13 *For any $m \times m$ matrix $\boldsymbol{A}$ and for any m-dimensional vector $\boldsymbol{v}$, we have:*

$$\|\boldsymbol{A}\boldsymbol{v}\|^2 \geq (\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x})^2\|\boldsymbol{v}\|^2,$$

where $\boldsymbol{x}$ is the eigenvector vector of norm 1 corresponding to $\lambda_{\min}(\boldsymbol{A}^T\boldsymbol{A})$

Proof. Using Theorem A.1.3, we have:

$$\mathbf{v}^T\mathbf{A}^T\mathbf{A}\mathbf{v} \geq \lambda_{\min}(\mathbf{A}^T\mathbf{A})\|\mathbf{v}\|^2 = \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{x}\|\mathbf{v}\|^2 \geq (\mathbf{x}^T\mathbf{A}\mathbf{x})^2\|\mathbf{v}\|^2,$$

where for the last inequality we used Lemma 1 of [37]

Lemma A.1.14 *Let $\boldsymbol{A}$ be an $m \times m$ symmetric matrix, $\boldsymbol{I}$ the identity matrix and C a constant. Then:*

(*i*) $\lambda_{\min}(\boldsymbol{A}) \geq C$ *is equivalent to* $\boldsymbol{A} \geq C\boldsymbol{I}$,

(*ii*) $\lambda_{\max}(\boldsymbol{A}) \leq C$ *is equivalent to* $\boldsymbol{A} \leq C\boldsymbol{I}$

Proof. (i). Assume that $\lambda_{\min}(A) \geq C$.

Let $\mathbf{x}$ be any $m \times 1$ vector Then $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq \lambda_{\min}(\mathbf{A})\mathbf{x}^T\mathbf{x} \geq C\mathbf{x}^T\mathbf{x}$

Conversely, since $\mathbf{A} \geq C\mathbf{I}$ then $\mathbf{A} - C\mathbf{I} \geq 0$ and hence, the eigenvalues of $\mathbf{A} - C\mathbf{I}$ are non-negative. It follows that $\lambda_i(\mathbf{A}) - C \geq 0$, for all $i \leq m$ and hence $\lambda_{\min}(\mathbf{A}) \geq C$.

Part (ii) is very similar and is omitted here. □

Lemma A.1.15 *For any symmetric $m \times m$ matrices $\boldsymbol{A}_i$ and for any $m \times 1$ vectors $\boldsymbol{x}_i$, we have:*

$$\left|\sum_{i=1}^{n} \boldsymbol{x}_i^T\boldsymbol{A}_i\boldsymbol{x}_i\right| \leq \max\{|\min_{i\leq n}\lambda_{\min}(\boldsymbol{A}_i)|, |\max_{i\leq n}\lambda_{\max}(\boldsymbol{A}_i)|\} \cdot \sum_{i=1}^{n}\boldsymbol{x}_i^T\boldsymbol{x}_i.$$

In particular, if $\boldsymbol{A}_i$ are diagonal matrices, then:

$$\left|\sum_{i=1}^{n} \boldsymbol{x}_i^T\boldsymbol{A}_i\boldsymbol{x}_i\right| \leq \{\max_{i\leq n}\max_{j\leq m}|a_{jj}^{(i)}|\} \cdot \sum_{i=1}^{n}\boldsymbol{x}_i^T\boldsymbol{x}_i.$$

where $\boldsymbol{A}_i = (a_{jk}^{(i)})_{j,k=1,\ m}$.

Proof. By Theorem A.1.3, for all $i \leq n$. we have:

$$\lambda_{\min}(\mathbf{A}_i)\mathbf{x}_i^T\mathbf{x}_i \leq \mathbf{x}_i^T\mathbf{A}_i\mathbf{x}_i \leq \lambda_{\max}(\mathbf{A}_i)\mathbf{x}_i^T\mathbf{x}_i.$$

Hence:

$$\min_{i\leq n}\lambda_{\min}(\mathbf{A}_i)\sum_{i=1}^{n}\mathbf{x}_i^T\mathbf{x}_i \leq \sum_{i=1}^{n}\mathbf{x}_i^T\mathbf{A}_i\mathbf{x}_i \leq \max_{i\leq n}\lambda_{\max}(\mathbf{A}_i)\sum_{i=1}^{n}\mathbf{x}_i^T\mathbf{x}_i.$$

Using the fact that $a \leq x \leq b$ implies $|x| \leq \max\{|a|, |b|\}$ we obtain our conclusion. For the second statement, note that if $\mathbf{A}$ is diagonal then the eigenvalues of $\mathbf{A}_i$ are in fact the elements on the diagonal (see Theorem 3.2 (c) in [33]). Hence:

$$\lambda_{\min}(\mathbf{A}_i) = a^{(i)}_{j_i j_i}, \quad \lambda_{\max}(\mathbf{A}_i) = a^{(i)}_{k_i k_i},$$

for some indices j_i, k_i. Hence, there exist some indices $i_0, i_1 \leq n$ such that:

$$\begin{aligned} |\min_{i\leq n}\lambda_{\min}(\mathbf{A}_i)| &= |\min_{i\leq n} a^{(i)}_{j_i j_i}| = |a^{(i_0)}_{j_{i_0} j_{i_0}}| \leq \max_{i\leq n}\max_{j\leq m}|a^{(i)}_{jj}| \\ |\max_{i\leq n}\lambda_{\max}(\mathbf{A}_i)| &= |\max_{i\leq n} a^{(i)}_{j_i j_i}| = |a^{(i_1)}_{k_{i_1} k_{i_1}}| \leq \max_{i\leq n}\max_{j\leq m}|a^{(i)}_{jj}|. \ \square \end{aligned}$$

Lemma A.1.16 *For any $m \times m$ matrix $\mathbf{A}$, whose columns are $\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(m)}$ and for any $m \times 1$-dimensional vector $\mathbf{x}$, we have:*

$$|\mathbf{x}^T\mathbf{A}\mathbf{x}| \leq \left\{\sum_{j=1}^{m}|\mathbf{x}^T\mathbf{a}^{(j)}|^2\right\}^{1/2} \cdot \|\mathbf{x}\|.$$

Proof. Using Cauchy-Schwarz inequality, we have:

$$\begin{aligned} |\mathbf{x}^T\mathbf{A}\mathbf{x}| &= \left|\sum_{j=1}^{m}\left(\sum_{i=1}^{m}x_i a_{ij}\right)x_j\right| = \left|\sum_{j=1}^{m}(\mathbf{x}^T\mathbf{a}^{(j)})x_j\right| \\ &\leq \left\{\sum_{j=1}^{m}(\mathbf{x}^T\mathbf{a}^{(j)})^2\right\}^{1/2}\left\{\sum_{j=1}^{m}x_j^2\right\}^{1/2}. \ \square \end{aligned}$$

The next lemma, was essential for the development of the asymptotic results in Chapter 5 and basically enables us to relate the trace of $\Sigma\mathbf{B}_n$ to the trace of $\mathbf{B}_n$.

Lemma A.1.17 *Assume that* $\mathbf{A}$ *and* $\mathbf{B}$ *are two* $m \times m$ *symmetric matrices such that* $\mathbf{A} > 0$ *and* $\mathbf{B} \geq 0$. *Then:*

$$\frac{\text{trace}(\mathbf{B})}{\text{trace}(\mathbf{A}^{-1})} \leq \text{trace}(\mathbf{AB}) = \text{trace}(\mathbf{BA}) \leq \text{trace}(\mathbf{A})\text{trace}(\mathbf{B})$$

Proof. The second inequality follows from Theorem 6.5, of [38]. Since the eigenvalues of $\mathbf{B}$ are the same as the eigenvalues of $\mathbf{A}^{1/2}\mathbf{B}\mathbf{A}^{-1/2}$, it follows that $\text{trace}(\mathbf{B}) = \text{trace}(\mathbf{A}^{1/2}\mathbf{B}\mathbf{A}^{-1/2})$. We applying again Theorem 6.5, of [38], and since $\mathbf{A}^{1/2}\mathbf{B}\mathbf{A}^{1/2} \geq 0$, we obtain:

$$\begin{aligned}\text{trace}(\mathbf{B}) &= \text{trace}(\mathbf{A}^{1/2}\mathbf{B}\mathbf{A}^{1/2}\mathbf{A}^{-1}) \leq \text{trace}(\mathbf{A}^{1/2}\mathbf{B}\mathbf{A}^{1/2})\text{trace}(\mathbf{A}^{-1}) \\ &= \text{trace}(\mathbf{AB})\text{trace}(\mathbf{A}^{-1}).\square\end{aligned}$$

A.2 Random Matrices

An $m \times m$ random matrix $\mathbf{A}$ is a collection $(a_{ij})_{1 \leq i,j \leq m}$ of random variables, defined on the same probability space $(\Omega, \mathcal{F}, P)$. We define $\mathrm{E}(\mathbf{A})$ to be the $m \times m$ matrix with elements $\mathrm{E}(a_{ij}), 1 \leq i, j \leq m$.

Lemma A.2.1 *Let* $\boldsymbol{A}$, $\{\boldsymbol{A}_n\}_{n \geq 1}$ *be* $m \times m$ *random matrices defined on the same probability space* $(\Omega, \mathcal{F}, P)$ *such that their inverses exist a.s. Suppose that:*

$$\boldsymbol{A}_n \longrightarrow \boldsymbol{A} \text{ a.s., element-wise.}$$

(i) Then $\boldsymbol{A}_n^{-1} \longrightarrow \boldsymbol{A}^{-1}$ *a.s. element-wise.*

(ii) If the matrices are symmetric, then $\lambda_i(\boldsymbol{A}_n) \longrightarrow \lambda_i(\boldsymbol{A})$ *a.s., for any* $i \leq m$, *where* $\lambda_1(\boldsymbol{A}) \leq \ldots \leq \lambda_m(\boldsymbol{A})$ *denote the eigenvalues of the matrix* $\boldsymbol{A}$.

Proof. Because the inverses exist, $\det(\mathbf{A}_n) \neq 0$, for any $n \geq 1$. Part (i) follows because the computation of elements of $\mathbf{A}_n^{-1}$ involves only addition and multiplication of elements. Theorem 3.14 in [33] proves (ii).

Lemma A.2.2 *Let $\{\boldsymbol{A}_n\}_{n\geq 1}$ and $\{\boldsymbol{B}_n\}_{n\geq 1}$ be sequences of $m \times m$ random matrices defined on a probability space $(\Omega, \mathcal{F}, P)$ such that:*

$$\boldsymbol{A}_n - \boldsymbol{B}_n \xrightarrow{a.s} 0 \quad \textit{element-wise}$$

and $|a_{ij}^{(n)}| \leq a$, $|b_{ij}^{(n)}| \leq b$ for all n, a.s., for some constants $a > 0$ and $b > 0$. Then,

$$E(\boldsymbol{A}_n) - E(\boldsymbol{B}_n) \longrightarrow 0 \quad \textit{element-wise.}$$

Proof. $\mathrm{E}(\mathbf{A}_n) = (\mathrm{E}(a_{ij}^{(n)}))_{i,j=1,\ldots m}$ and $\mathrm{E}(\mathbf{B}_n) = (\mathrm{E}(b_{ij}^{(n)}))_{i,j=1,\ldots m}$. Hence,

$$\mathrm{E}(\mathbf{A}_n) - \mathrm{E}(\mathbf{B}_n) = (\mathrm{E}(a_{ij}^{(n)} - b_{ij}^{(n)}))_{i,j=1,\ldots m}$$

and the conclusion follows by the bounded convergence theorem. □

Lemma A.2.3 *Let $\boldsymbol{A}$ be a random matrix defined on a probability space $(\Omega, \mathcal{F}, P)$.*

If there exists a constant $C > 0$ such that $\boldsymbol{A} \geq C\boldsymbol{I}$ a.s, then

$$E(\boldsymbol{A}) \geq C\boldsymbol{I}.$$

Proof. Hypothesis $\mathbf{A} \geq C\mathbf{I}$ a.s. implies that $\mathbf{x}^T\mathbf{A}\mathbf{x} \geq C\mathbf{x}^T\mathbf{x}$ for any $m \times 1$ vector $\mathbf{x}$ a.s. Taking the expectation we obtain $\mathrm{E}(\mathbf{x}^T\mathbf{A}\mathbf{x}) \geq C\mathbf{x}^T\mathbf{x}$, for any $m \times 1$ vector $\mathbf{x}$. But $\mathrm{E}(\mathbf{x}^T\mathbf{A}\mathbf{x}) = \mathrm{E}(\sum_{i,j=1}^m x_i a_{ij} x_j) = \sum_{i,j=1}^m x_i \mathrm{E}(a_{ij}) x_j = \mathbf{x}^T\mathrm{E}(\mathbf{A})\mathbf{x}$.
Hence $\mathbf{x}^T\mathrm{E}(\mathbf{A})\mathbf{x} \geq C\mathbf{x}^T\mathbf{x}$, for any $m \times 1$ vector $\mathbf{x}$, which implies our conclusion. □

Theorem A.2.4 *Let $\boldsymbol{A}(\alpha)$ be an $m \times m$ random matrix, whose elements $a_{ij}(\alpha)$, $1 \leq i,j \leq m$ are a.s. continuously differentiable with respect to $\alpha \in \mathcal{T}$, $\mathcal{T} \subset \mathbb{R}$. Suppose that $\boldsymbol{A}(\alpha)^{-1}$ exists for all $\alpha \in \mathcal{T}$ a.s.*

Then, the elements of $\boldsymbol{A}(\alpha)^{-1}$ are a.s. differentiable with respect to α and:

$$\frac{d}{d\alpha}[\mathbf{A}(\alpha)^{-1}] = -\mathbf{A}(\alpha)^{-1}\left[\frac{d}{d\alpha}\mathbf{A}(\alpha)\right]\mathbf{A}(\alpha)^{-1}, \quad a.s.$$

Here $\frac{d}{d\alpha}[\mathbf{A}(\alpha)^{-1}]$ and $\frac{d}{d\alpha}\mathbf{A}(\alpha)$ denote the matrices whose elements are the derivatives with respect to α of the elements of $\mathbf{A}(\alpha)^{-1}$ and $\mathbf{A}(\alpha)$, respectively.

Proof. Let us denote by $\widetilde{\Omega}$ the event where the hypothesis of the theorem hold. This result is a consequence of Theorem 9.2 in [33], for every $\omega \in \widetilde{\Omega}$.

Appendix B

Martingale Strong Laws of Large Numbers

In this section we include some results about the convergence of p-dimensional martingales.

Definition B.0.5 *We say that a sequence $\{\boldsymbol{s}_n\}_{n\geq 1}$ of p-dimensional random vectors with components $\boldsymbol{s}_n = (s_n^{(1)}, \ldots, s_n^{(p)})$ is a **p-dimensional martingale** if*

$$E(s_n^{(k)} | \boldsymbol{s}_1, \ldots, \boldsymbol{s}_{n-1}) = s_{n-1}^{(k)}, \ \forall k \in \{1, \ldots, p\}, \ \forall n \geq 1.$$

Theorem B.0.6 (Theorem 3 of [22]) *Let $\{\boldsymbol{s}_n\}_{n\geq 1}$ be a p-dimensional square integrable martingale with $\boldsymbol{s}_n = (s_n^{(1)}, \ldots, s_n^{(p)})$. Suppose that $E(s_n^{(k)}) = 0$ for all $k \in \{1, \ldots, p\}$ and let $\boldsymbol{M}_n = Cov(\boldsymbol{s}_n) = E(\boldsymbol{s}_n \boldsymbol{s}_n^T)$ be positive definite.*

If $\lambda_{\max}(\boldsymbol{M}_n) \to \infty$, then for any $\delta > 1/2$,

$$\frac{\boldsymbol{M}_n^{-1/2} \boldsymbol{s}_n}{[\log \lambda_{\max}(\boldsymbol{M}_n)]^\delta} \to 0, \ \ a.s.$$

Corollary B.0.7 *Let $\{s_n\}_{n\geq 1}$ be a zero-mean square integrable martingale, such that $\sigma_n^2 := E(s_n^2) \to \infty$ as $n \to \infty$. Then*

$$\frac{s_n}{(\sigma_n^2)^{1/2} (\log \sigma_n^2)^\varepsilon} \to 0, \ \ a.s., \ \forall \varepsilon > 0.$$

In particular,

$$\frac{\boldsymbol{s}_n}{(\sigma_n^2)^{1/2+\delta}} \to 0, \ \ a.s., \ \forall\delta > 0.$$

Remark B.0.8 Corollary B.0.7 is a martingale generalization of a classical result in probability theory (valid for independent random variables). See page 132 of [11].

Theorem B.0.9 *Let $\{\boldsymbol{s}_n\}_{n\geq 1}$ be a p-dimensional martingale, $\boldsymbol{s}_n = (s_n^{(1)}, \ldots, s_n^{(p)})$ and suppose that $E(s_n^{(k)}) = 0$ and $E[(s_n^{(k)})^2] < \infty$, for all $k \leq p$. Let $\boldsymbol{M}_n := Cov(\boldsymbol{s}_n) = E(\boldsymbol{s}_n \boldsymbol{s}_n^T) = \{E(s_n^{(j)} s_n^{(k)})\}_{j,k\leq p}$.*

If $\lambda_{\min}(\boldsymbol{M}_n) \to \infty$, then

$$\frac{s_n}{[\lambda_{\max}(\boldsymbol{M}_n)]^{1/2+\delta}} \to 0 \quad a.s\ , \ \forall\delta > 0.$$

Proof. Let $k \in \{1, \ldots, p\}$ be fixed.

By (A.1.1), $\lambda_{\min}(\mathbf{M}_n) \leq \mathrm{E}[(s_n^{(k)})^2] \leq \lambda_{\max}(\mathbf{M}_n)$, and so $\sigma_{n,k}^2 = \mathrm{Var}(s_n^{(k)}) = \mathrm{E}[(s_n^{(k)})^2] \to \infty$.

We apply Corollary B.0.7 to the martingale $\{s_n^{(k)}\}_{n\geq 1}$ and obtain:

$$\frac{s_n^{(k)}}{(\sigma_{n,k}^2)^{1/2+\delta}} \to 0 \quad \text{a.s}\ , \ \forall\delta > 0.$$

Since $\dfrac{s_n^{(k)}}{(\sigma_{n,k}^2)^{1/2+\delta}} \geq \dfrac{s_n^{(k)}}{[\lambda_{\max}(\mathbf{M}_n)]^{1/2+\delta}}$, the conclusion follows.

Appendix C

Real Analysis Results

Theorem C.0.10 (Lemma A of [7]) *Let $T : \mathcal{T} \subset \mathbb{R}^p \to \mathbb{R}^p$ be a one-to-one continuously differentiable function. Let $\beta_0 \in \mathcal{T}$, $r > 0$ and let $B_r(\beta_0) = \{\beta \in \mathbb{R}^p,\ \|\beta - \beta_0\| \leq r\}$ and $\partial B_r(\beta_0) = \{\beta \in \mathbb{R}^p\colon\ \|\beta - \beta_0\| = r\}$. Assume $B_r(\beta_0) \subset \mathcal{T}$.*

If there exists $\varepsilon > 0$ such that $\varepsilon \leq \inf_{\beta \in \partial B_r(\beta_0)} \|T(\beta) - T(\beta_0)\|$, then $B_\varepsilon(T(\beta_0)) \subset T(B_r(\beta_0))$, where $B_\varepsilon(T(\beta_0)) = \{\boldsymbol{y} \in \mathbb{R}^p;\ \|\boldsymbol{y} - T(\beta_0)\| \leq \varepsilon\}$.

In particular, if.

$$\varepsilon = \|T(\beta_0)\| \leq \inf_{\beta \in \partial B_r(\beta_0)} \|T(\beta) - T(\beta_0)\|,$$

then $0 \in B_\varepsilon(T(\beta_0)) \subset T(B_r(\beta_0))$ i.e. there exists a point $\widehat{\beta} \in B_r(\beta_0)$ such that $T(\widehat{\beta}) = 0$

Remark C.0.11 Since T is one-to-one, the point $\widehat{\beta}$ is unique.

Lemma C.0.12 *If $(s_i)_{i\geq 1}$ and $(t_i)_{i\geq 1}$ are sequences of random variables such that: $\dfrac{\sum_{i=1}^n (s_i - t_i)^2}{\sum_{i=1}^n t_i^2} \xrightarrow{P} 0$, then $\dfrac{\sum_{i=1}^n s_i^2}{\sum_{i=1}^n t_i^2} \xrightarrow{P} 1$.*

Proof. We have $|\sum_{i=1}^n (s_i^2 - t_i^2)| = |\sum_{i=1}^n [(s_i - t_i)^2 + 2t_i(s_i - t_i)]| \leq \sum_{i=1}^n (s_i - t_i)^2 + 2(\sum_{i=1}^n t_i^2)^{1/2} [\sum_{i=1}^n (s_i - t_i)^2]^{1/2}$ and therefore:

$$\left| \frac{\sum_{i=1}^n (s_i^2 - t_i^2)}{\sum_{i=1}^n t_i^2} \right| \leq \frac{\sum_{i=1}^n (s_i - t_i)^2}{\sum_{i=1}^n t_i^2} + 2 \left(\frac{\sum_{i=1}^n (s_i - t_i)^2}{\sum_{i=1}^n t_i^2} \right)^{1/2}.$$

The result follows since, by hypothesis, the right hand side converges to 0, in probability. □

Appendix D

Domains of Attraction

D.1 Domain of Attraction of the Normal Law

Definition D.1.1 *Let $\{X_1, \ldots, X_n\}$ be a sample from a distribution X. The distribution of X is said to belong to the domain of attraction of the normal law ($X \in DAN$) if there exists sequences of scalars $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$ such that*

$$\frac{\sum_{i=1}^{n} X_i - a_n}{b_n^2} \xrightarrow{\mathcal{D}} N(0,1).$$

Lemma D.1.2 (see Proposition 1 in [4]) *Let $\{Y_1, \ldots, Y_n\}$ be a sample from a distribution Y, with $Y_i \geq 0$. Then $\frac{\max_{i\leq n} Y_i}{\sum_{i=1}^{n} Y_i} \xrightarrow{P} 0$ if and only if there exists a sequence $(b_n)_{n\geq 1}$ with $b_n \to \infty$ such that $\frac{\sum_{i=1}^{n} Y_i}{b_n} \xrightarrow{P} 1$.*

Lemma D.1.3 *Let $\{X_1, \ldots, X_n\}$ be a sample from a distribution X. Then $X \in DAN$ if and only if $\frac{\max_{i\leq n} X_i^2}{\sum_{i=1}^{n} X_i^2} \xrightarrow{P} 0$.*

Remark D.1.4 From Lemma D.1.3 and Lemma D.1.2, by taking $Y = X^2$, it follows that $X \in DAN$ is equivalent to $\frac{\sum_{i=1}^{n} X_i^2}{b_n} \xrightarrow{P} 1$, for some sequence of scalars $(b_n)_{n\geq 1}$, with $b_n \to \infty$.

Remark D.1.5 By [14], p.313, $X \in DAN$ if and only if the function

$$l_X(x) = \int_{-x}^{x} u^2 dF_X(u)$$

is slowly varying, as $x \to \infty$ (i.e. $\dfrac{l_X(tx)}{l_X(x)} \to 1$, as $x \to \infty$, for all $t > 0$).

Theorem D.1.6 (Theorem 1 in [25]) *If $X \in DAN$, $Y \in DAN$ and X is independent of Y, then $XY \in DAN$. Conversely, if $\mathrm{E}(X^2) < \infty$ and $XY \in DAN$, then $Y \in DAN$.*

Lemma D.1.7 (see e.g. [17]) *Let $\{X_1, \ldots, X_n\}$ be a sample from a distribution X. Then, $X \in DAN$ and $\mathrm{E(X)} = 0$ are equivalent to either one of the following CLT*

$$(a)\ \frac{\sqrt{n}\,\overline{X}}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}} \xrightarrow{\mathcal{D}} N(0,1),$$

$$(b)\ \frac{\sum_{i=1}^{n} X_i}{\sqrt{\sum_{i=1}^{n} X_i^2}} \xrightarrow{\mathcal{D}} N(0,1).$$

Lemma D.1.8 (Lemma 4 in [27]) *Let $\{(U,V),(U_i,V_i), i \geq 1\}$ be i.i.d. random vectors, and U and V be independent. If $U \in DAN$ and $V \in DAN$, then $UV \in DAN$. Conversely, if $\mathrm{E(V^2)} < \infty$ and $UV \in DAN$, then $U \in DAN$.*

Lemma D.1.9 (Lemma 5 in [27]) *Let $\{(U,V),(U_i,V_i), i \geq 1\}$ be i.i.d. random vectors with $\mathrm{E|UV|} < \infty$.*

(a) If $U \in DAN$ and $V \in DAN$, then $U + V \in DAN$, provided that $\mathrm{P(U + V = const)} \neq 1$.

(b) Conversely, if $U + V \in DAN$, $U \in DAN$ and $\mathrm{E}(U^2) < \infty$, then $V \in DAN$ provided that $\mathrm{P}(V = \mathrm{const}) \neq 1$.

D.2 Generalized Domain of Attraction of the Normal Law

Lemma D.2.1 (see Lemma 2.1 in [26]) *Suppose for a sequence of random vectors $\boldsymbol{\xi}_n \in \mathbb{R}^m$, we have $\mathbf{B}_n\boldsymbol{\xi}_n \xrightarrow{\mathcal{D}} G$, as $n \to \infty$, where G is a nondegenerate spherically symmetric random vector in $\mathbb{R}^m$ and $\mathbf{B}_n$ are square nonsingular stochastic matrices. Then $\mathbf{B}_n$ may be taken to be symmetric matrices.*

Theorem D.2.2 (see Theorem 1.1 [34]) *Let $\{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n\}$ be a sample of random vectors in $\mathbb{R}^m$ from a distribution $\boldsymbol{\xi}$. The following statements are equivalent:*

(i) $\boldsymbol{\xi} \in GDAN$,

(ii) $\forall \epsilon > 0, \lim_{n\to\infty} \sup_{\|\theta\|=1} \mathrm{P}\left(\frac{\max_{i\le n} |\langle \boldsymbol{\xi}_i, \theta\rangle|}{(\sum_{i=1}^n \langle \boldsymbol{\xi}_i, \theta\rangle^2)^{1/2}} > \epsilon \right) = 0.$

Remark D.2.3 By taking $\theta = \frac{1}{\sqrt{m}}(1, \ldots, 1)^T$ in Theorem D.2.2. using the characterization in Lemma D.1.3, it follows that $\boldsymbol{\xi} \in GDAN$ implies that the sum of its components is in DAN. Note that the converse of this statement is not true, in general.

Lemma D.2.4 (see Lemma 2.3 in [26]) *When $\boldsymbol{\xi}$ is full, there is a constant $c > 0$ such that* $\mathrm{P}\left(\lambda_{\min}\left(\sum_{i=1}^n \boldsymbol{\xi}_i\boldsymbol{\xi}_i^T\right) \ge cn\right) \xrightarrow{n\to\infty} 1.$

Theorem D.2.5 (see Theorem 3.4 in [16]) *Let $\boldsymbol{\xi}$ be a symmetric full random variable in $\mathbb{R}^m$ and $\{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n\}$ be a sample from the distribution $\boldsymbol{\xi}$. Then, if $\boldsymbol{\xi} \in GDAN$ we have*

$$n\mathrm{E}\left[\boldsymbol{\xi}_1^{\mathrm{T}}\left(\sum_{i=1}^{n} \boldsymbol{\xi}_i\boldsymbol{\xi}_i^{\mathrm{T}}\right)^{-1}\boldsymbol{\xi}_1\right]^2 \xrightarrow{n\to\infty} 0.$$

Appendix E

List of Assumptions

E.1 Model Assumptions - when the covariates are measured directly

(AH) there exists $C > 0$ such that for $l = 2, 3$, we have:

$$k_{n,r}^{[l]} := \sup_{\beta \in B_r} k_{n,r}^{[l]}(\beta) \leq C, \ \forall r > 0, \quad \forall n \geq 1.$$

(D) $\lambda_{\min}[\mathbf{H}_n^{\text{indep}}(\beta)] \to \infty$, for any $\beta \in \mathcal{T}$.

(D') $\lambda_{\min}(\mathbf{H}_n^{\text{indep}}) \to \infty$.

(G) there exists a constant $C > 0$ such that $\lambda_{\max}(\mathbf{H}_n^{\text{indep}}) \leq Cn, \ \forall n \geq 1$.

(K) $\lim_{r \to 0} \limsup_{n \to \infty} r a_n^{1/2} = 0$.

(K') $\lim_{r \to 0} \limsup_{n \to \infty} \eta_n(r) = 0$.

(H) for any $\beta \in \mathcal{T}$, there exists a constant $C_\beta > 0$ such that $\lambda_{\min}[\overline{\mathbf{R}}_n^{(c)}(\beta)] \geq C_\beta$, P_β a.s. for any $n \geq 1$.

(H') there exists a constant $C > 0$ such that $\lambda_{\min}(\overline{\mathbf{R}}_n^{(c)}) \geq C, \forall n \geq 1$, a.s.

(N_δ) for any $\beta \in \mathcal{T}$ there exists some constants $\delta_\beta > 0$ and $C_\beta > 0$, such that $E_\beta \|\mathbf{A}_i(\beta)^{-1/2}(\mathbf{y}_i - \mu_i(\beta))\|^{2+\delta_\beta} \leq C_\beta$ for all $i \geq 1$.

(N'_δ) there exist some constants $\delta > 0$ and $C > 0$ such that $\mathrm{E}\|\mathbf{A}_i^{-1/2}(\mathbf{y}_i - \mu_i)\|^{2+\delta} < C$, for all $i \geq 1$.

(S'_δ) there exist an integer $N \geq 1$ and some constants $\delta > 0, c_0 > 0$ such that:

$$\lambda_{\min}(\mathbf{H}_n^{\text{indep}}) \geq c_0[\lambda_{\max}(\mathbf{H}_n^{\text{indep}})]^{1/2+\delta}, \quad \forall n \geq N.$$

E.2 Assumptions on the Substitute of the Correlation Matrix in GEE

(A) $\mathcal{R}_n(\beta)$ is positive definite for all $n \geq 1$ and $\beta \in \mathcal{T}$; the entries of $\mathcal{R}_n(\beta)$ are continuously differentiable with respect to β, for all $n \geq 1$ and $\beta \in \mathcal{T}$.

(B) the entries of $\mathcal{R}_n(\beta)$ are $\mathcal{F}_n$ - measurable for all $n \geq 1$ and $\beta \in \mathcal{T}$.

(C) $\mathcal{R}^*_{n-1}(\beta) - \overline{\mathbf{R}}_n^{(c)}(\beta) \xrightarrow{P_\beta} 0$ element-wise, for all $\beta \in \mathcal{T}$.

(E) there exists a constant $C > 0$ such that $\lambda_{\max}(\mathcal{R}_n) \leq C$ fora ll $n \geq 1$ a.s.

(R) there exists $K(\beta) > 0$ such that $\lambda_{min}[\mathcal{R}^*_n(\beta)] \geq K(\beta)$ P_β a.s. for all $\beta \in \mathcal{T}$ and $n \geq 1$.

(R') there exists a constant $C > 0$ such that $\lambda_{\min}(\mathcal{R}_n) \geq C$, for all $n \geq 1$, a.s.

E.3 The Error-in Variables Model Assumption

Assume that $\boldsymbol{\xi}$ is a symmetric, full random variable in $\mathbb{R}^m$

$(A1)$ $\boldsymbol{\xi}_1$ lies in the generalized domain of attraction of the normal law (GDAN). $\boldsymbol{\xi}_1 - \boldsymbol{\mu}$ has a symmetric distribution, whenever the matrix $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}$ has at least one element on the diagonal which is ∞

$(A2)$ $E(\varepsilon_{1j}) = 0,\ E(\delta_{1j}) = 0,\ E(\varepsilon_{1j}^4) < \infty,\ E(\delta_{1j}^4) < \infty$ with $1 \leq j \leq m$, and $\boldsymbol{\Sigma}_{\text{error}} := \begin{pmatrix} \boldsymbol{\Sigma}_{\varepsilon} & \boldsymbol{\Sigma}_{\varepsilon\delta} \\ \boldsymbol{\Sigma}_{\varepsilon\delta}^T & \boldsymbol{\Sigma}_{\delta} \end{pmatrix}$ is positive definite

$(A3)$ $\quad (\boldsymbol{\xi}_i)_{1\leq i\leq n}$ and $\{(\boldsymbol{\varepsilon}_i, \boldsymbol{\delta}_i)\}_{1\leq i\leq n}$ are independent.

Bibliography

[1] Balan, R. M., Dumitrescu L. and Schiopu-Kratina, I. Asymptotically optimal estimating equation with strongly consistent solutions for longitudinal data. *Mathematical Methods of Statistics, to appear.*

[2] Balan, R. M. and Schiopu-Kratina, I. (2005). Asymptotic Results with Generalized Estimating Equations for Longitudinal Data. *Annals of Statistics.* **33**, 522-541.

[3] Billingsley, P. (1995). *Probability and Measure*, Third Edition, John Wiley, New York.

[4] Breiman, L. (1965). On some limit theorems similar to the arcsine law. *Theory of Probability and its Applications.* **10**, 323-331.

[5] Buonaccorsi, J., Demidenko, E., Tosteson, T. (2000). Estimation in longitudinal random effects models with measurement error. *Statistica Sinica.* **10**, 885-903.

[6] Carroll, R. J., Ruppert, D., Stefanski, L., Crainiceanu, C. M. (2006). Measurement Error in Nonlinear Models: A Modern Perspective. Chapman and Hall.

[7] 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**, 1155-1163.

[8] Cheng, C.-L. and Van Ness, J.W. (1999). *Statistical Regression with Measurement Error.* Arnold, London.

[9] Crowder, M. (1995). On the use of a working correlation matrix in using generalised linear models for repeated measures. *Biometrika.* **82**, 407-410.

[10] Crowder, M. (2001). On repeated measures analysis with misspecified covariance structure. *Journal of the Royal Statistical Society.* **63**, 55-62.

[11] Chung, K. L. (2001). *A Course in Probability Theory*, Third Edition, Academic Press.

[12] Dumitrescu, L. (2010). Estimation for a longitudinal linear model with measurement errors, *Electronic Journal of Statistics.* **4**, 486-524.

[13] Fahrmeir, L. and Kaufmann H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Annals of Statistics.* **13**, 342-368.

[14] Feller, W. (1971). *An Introduction to Probability Theory and its Applications*, Vol. **2**. Second Edition. New York: Wiley.

[15] Fuller, W. A. (1987). *Measurement Error Models.* Wiley.

[16] Gine, E. and Gotze, F. (2004). On standard normal convergence of the multivariate student t-statistic for symmetric random vectors. *Electronic Communications in Probability.* **9**, 162-171.

[17] Gine, E., Gotze, F. and Mason, D.M. (1997). When is the Student t-statistic asymptotically standard normal? *Annals of Probability.* **25**, 1514-1531.

[18] Godambe, V. P., Heyde, C. C. (1987). Quasi-likelihood and optimal estimation *International Statistical Review.* **55**, 231-274.

[19] Gleser, L.J. (1992). The importance of assessing measurement reliability in multivariate regression. *Journal of the American Statistical Association.* **87**, 696-707.

[20] Hahn, M. and Klass, M. (1980). Matrix normalization of sums of random vectors in the domain of attraction of the multivariate normal. *The Annals of Probability.* **8**, 262-280.

[21] Heyde. C. C. (1997). *Quasi-Likelihood and its Application.* Springer, New York.

[22] Kaufmann, H. (1987). On the strong law of large numbers for multivariate martingales. *Stochastic Processes and Their Applications.* **26**, 73-85.

[23] Jian, J., Luan, Y. and Wang, Y-G. (2007). Iterative estimating equations: linear convergence and asymptotic properties. *The Annals of Statistics.* **35**, 2233-2260.

[24] Liang, K.-Y., and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika.* **73**, 13-22.

[25] Maller, R. A. (1981). A theorem on products of random variables, with application to regression. *Australian Journal of Statistics.* **23**, 177-185.

[26] Maller, R. A. (1993). Quadratic negligibility and the asymptotic normality of operator normed sums. *Journal of Multivariate Analysis.* **44**, 191-219.

[27] Martsynyuk, Y. V. (2007). Central limit theorems in linear structural error-in-variables models with explanatory variables in the domain of attraction of the normal law. *Electronic Journal of Statistics.* **1**, 195-222.

[28] McCullagh, P. and Nelder, J. A. (1983). *Generalized Linear Models.* Chapman and Hall, London.

[29] Merkouris, T. (2007). Transform martingale estimating functions. *The Annals of Statistics.* **35**, 1975-2000.

[30] Prentice, R. L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics.* **44**, 1033-1048.

[31] Prentice, R. L. and Zhao, L. P. (1991). Estimating equations for the parameters in means and covariances of multivariate discrete and continuous responses. *Biometrics.* **47**, 825-839.

[32] Qu, A., Linsday, B. G. and Li, B. (2000). Improving generalised estimating equations using quadratic inference functions. *Biometrika.* **87**, 823-836.

[33] Schott, J. R. (2005). *Matrix Analysis for Statistics*, Second edition. Wiley, Hoboken New Jersey.

[34] Sepanski, S. J. (1994). Probabilistic characterizations of the generalized domain of attraction of the multivariate normal law. *Journal of Theoretical Probability.* **7**, 857-866.

[35] Sutradhar, B. C. and Das, K. (1999). On the efficiency of regression estimators in generalised linear models for longitudinal data. *Biometrika.* **86**, 459-465.

[36] Tosteson, T., Buonaccorsi, J., and Demidenko, E. (1998). Covariate measurement error and the estimation of random effect parameters in mixed model for longitudinal data. *Statistics in Medicine.* **17**, 1959-1971.

[37] Xie, M. and Yang, Y. (2003). Asymptotics for Generalized Estimating Equations with Large Cluster Sizes. *Ann. Statist.* **31**, 310-347.

[38] Zhang, F. (1999). *Matrix Theory: Basic Results and Techniques.* Springer.