Asymptotic theory with generalized estimating equations for longitudinal data

R.M. Balan*   I. Schiopu-Kratina†
Université de Sherbrooke   Statistics Canada

February 20, 2003

Abstract

We consider the marginal models of Liang and Zeger [7] for the analysis of longitudinal data and we develop a nonparametric theory of statistical inference for such models. We prove the existence and consistency (weak and strong) of the maximum quasi-likelihood estimator using some general results for estimating equations. We also establish the asymptotic normality of this estimator.

Keywords: longitudinal data, marginal model, generalized linear model, maximum quasi-likelihood estimator, consistency, asymptotic normality.

1 Introduction

Longitudinal data sets arise in biostatistics and life-time testing problems when the responses of the individuals are recorded repeatedly over a period of time. By controlling for individual differences, longitudinal studies are well-suited to measure change over time. On the other hand, they require the use of special statistical techniques because the responses on the same individual tend to be strongly correlated. In the seminal paper [7], Liang and Zeger proposed the use of generalized linear models (GLM) for the analysis of longitudinal data and introduced the *marginal models* for which the regression of each marginal response on the explanatory variables is modelled separately from the within-individual correlation.

In a cross-sectional study, a GLM is used when there are reasons to believe that each response y_i depends on an observable vector x_i of covariates. Typically

*Research supported by a post-doctoral fellowship from the Natural Sciences and Engineering Research Council of Canada. *E-mail address*: ralucabalan@hotmail.com

†Corresponding author. *Postal address*: Statistics Canada, HSMD, 16RHC, Ottawa, ON, K1A 0T6, Canada. *Tel*: (613) 951-0331. *E-mail address*: ioana.schiopu-kratina@statcan.ca

this dependence is specified by an unknown parameter β and a link function μ via the relationship $\mu_i(\beta) = \mu(x_i'\beta)$, where $\mu_i(\beta_0)$ is the mean of y_i. For one-dimensional observations, the maximum quasi-likelihood estimator $\hat{\beta}_n$ is defined as the solution of the equation

$$\sum_{i=1}^{n} \frac{\dot{\mu}_i(\beta)}{v_i(\beta)}(y_i - \mu_i(\beta)) = 0. \tag{1}$$

where $\dot{\mu}_i$ is the derivative of μ and $v_i(\beta)$ is the variance of y_i. Note that this equation simplifies considerably if we assume that $v_i(\beta) = \phi_i \dot{\mu}(x_i'\beta)$, with a nuisance scale parameter ϕ_i. In fact (1) is a genuine likelihood equation if the y_i's are independent with densities $c(y_i, \phi_i)\exp\{\phi_i^{-1}[(x_i'\beta)y_i - b(x_i'\beta)]\}$; the asymptotic properties of the maximum likelihood estimator (MLE) have been thouroghly investigated in [2] and [12].

In a longitudinal study, each observation y_i is actually d-dimensional and its components $(y_{i1}, \ldots, y_{id})$ represent repeated measurements at different times for subject i. The approach proposed by Liang and Zeger is to impose the usual assumptions of a GLM for each marginal scalar observation y_{it} (considering the regression on a p-dimensional design vector x_{it}) and to model separately the correlation within-individual. If these correlation matrices are known (but the entire likelihood is not specified), then the d-dimensional version of (1) becomes a *generalized estimating equation* (GEE).

In this article we prove the existence, consistency and asymptotic normality of a sequence of estimators, defined as solutions (roots) of GEEs. We work within the nonparametric set-up of Liang and Zeger, which makes our results stronger than those of [2], [12] even for GLM ($d = 1$). Throughout this article, we consider that the residuals form a martingale difference sequence, which is a generalization of the independence assumption used in [2], [7], [10], [15].

Since the GEE is not the derivative of an equation, most of the technical difficulties surface when proving the asymptotic existence of roots (REEs) of such general estimating equations. General results available in the literature for the existence of REEs involve conditions which are difficult to verify (e.g. Theorem 12.1 of [4]). In this article we use a refinement of Theorem 1 of [15], which also appears in [10] in a slightly different formulation. To apply it, we introduced conditions (E-p) (respectively (E-a.s)), which require the weak (respectively strong) equicontinuity of the derivatives of the GEE functions with respect to the multidimensional parameter. For GLM, these conditions are satisfied if the link functions are equicontinuous and a boundedness condition (B) (on the extreme eigenvalues of the design matrix) holds. We note that our condition (B) is weaker than the corresponding condition (S_δ) considered in [2].

In order to verify (E-p) for GEE, we employ a technique borrowed from the proof of tightness of multiparameter processes with continuous sample paths, and we impose some conditions on the rate of growth of some scalar functions associated with the link functions. These conditions are satisfied in the unidimensional case (i.e. $p = 1$) and in the case of the longitudinal linear model. In order to verify (E-a.s), we impose the rather strong assumption that the recorded observations are bounded, which is satisfied for categorical observations with a finite number of values. This assumption is not needed for the longitudinal linear model.

In order to obtain the asymptotic normality in our more general context, we assume that the residuals are bounded in $L^{2+\delta}$. This condition does not appear in [2], [12] for GLM. Finally, our Lemma 2 leads to a direct proof of the strong consistency of the least square estimator (LSE); see [6].

The paper is organized as follows: in Section 2 we introduce the framework and the assumptions and we state the main results. In Section 3 we give the formal proofs of these results, while in Section 4 we examine the conditions which will lead to the verification of the assumptions. Appendix A includes some general results for estimating equations and in Appendix B we give some auxiliary matrix analysis results which we found useful.

2 Statements of the results

If A is a $p \times p$ matrix, we will denote with $\| A \|$ its spectral norm, with $\| A \|_E$ its Euclidean norm and with $\mathrm{tr}(A)$ its trace. If A is a symmetric matrix, we denote with $\lambda_{\min}(A), \lambda_{\max}(A)$ its minimum, maximum eigenvalues. For a p-dimensional vector x, we will use the Euclidean norm $\| x \| := (x'x)^{1/2} = \mathrm{tr}(xx')^{1/2}$.

For any matrix A, $\| A \| = \{\lambda_{\max}(A'A)\}^{1/2}$ and $\| A \|_E = \{\mathrm{tr}(A'A)\}^{1/2}$. In particular, if A is symmetric and nonnegative definite, then $\| A \| = \lambda_{\max}(A)$.

Throughout the sequel, we will use the notation $A \leq B$ if $B - A$ is nonnegative definite; in this case, $\mathrm{tr}(A) \leq \mathrm{tr}(B)$. Moreover, if A is symmetric and B is symmetric and nonnegative definite such that $-B \leq A \leq B$, then $\| A \| \leq \| B \|$.

We let $A^{1/2L}$ ($A^{1/2R}$) be the left (respectively right) square root of the positive definite matrix A, i.e. $A^{1/2L} A^{1/2R} = A$ and $A^{1/2L} = (A^{1/2R})'$. We set $A^{-1/2L} = (A^{1/2L})^{-1}$ and $A^{-1/2R} = (A^{1/2R})^{-1}$.

Let $y_i := (y_{i1}, \ldots, y_{id})'; i = 1, \ldots, n$ be a longitudinal data set consisting of n respondents, where the components of y_i represent measurements at different times from subject i. In the marginal model that we consider the correlation matrix of y_i is denoted by R_i and the marginal expectations and variances are specified in terms of the regresion parameter β through

$$\mu_{it}(\beta) := E_\beta(y_{it}) = \mu(x'_{it}\beta), \quad \mathrm{Var}_\beta(y_{it}) = \phi_i \dot{\mu}(x'_{it}\beta)$$

where x_{it} are $p \times 1$ vectors of covariates and $\phi_i > 0$ are dispersion parameters. The link function μ is assumed to be continuously differentiable with $\dot{\mu} > 0$.

Examples:
1. in the logistic regression for binary data, $\mu(y) = \exp(y)/[1 + \exp(y)]$;
2. in the log regression for count data, $\mu(y) = \exp(y)$;
3. in the linear regression for continuous data, $\mu(y) = y$.

Let $\mu_i(\beta) := E_\beta(y_i), V_i(\beta) := \mathrm{Var}_\beta(y_i)$ and $\epsilon_i(\beta) := y_i - \mu_i(\beta))$. If the matrices R_i are known, then the maximum quasi-likelihood estimator $\hat{\beta}_n$ is the solution of the equation (see [13], p.315)

$$\sum_{i=1}^{n} \dot{\mu}_i(\beta)' V_i(\beta)^{-1} \epsilon_i(\beta) = 0. \tag{2}$$

Note that $\dot{\mu}_i(\beta) = D_i(\beta) X_i$ and $V_i(\beta) = \phi_i D_i(\beta)^{1/2} R_i D_i(\beta)^{1/2}$, where $D_i(\beta)$ is a $d \times d$ diagonal matrix whose (t,t) element is $\dot{\mu}(x_{it}'\beta)$ and X_i is a $d \times p$ matrix whose t-th row is x_{it}'.

In the sequel the unknown parameter β lies in an open set $B \subseteq \mathbf{R}^p$ and β_0 is the true value of this parameter. Our work is under the following assumption:

Assumption (A)

(i) $L := \max_{t=1,\ldots,d} \sup_{i \geq 1} \parallel x_{it} \parallel < \infty$

(ii) $\phi_i = 1, \forall i$

(iii) $R_i = R, \forall i$, where $R = (r_{tl})_{t,l=1,\ldots,d}$ is a (known) symmetric positive definite matrix

(iv) μ is twice continuously differentiable

(v) $\inf_i \dot{\mu}(x_{it}'\beta_0) > 0, \forall t = 1, \ldots, d$

(vi) $\epsilon_i := \epsilon_i(\beta_0), i \geq 1$ is a (d-dimensional) martingale difference sequence, i.e. $E(\epsilon_i | \mathcal{F}_{i-1}) = 0, \forall i \geq 1$, where $\mathcal{F}_i$ is the σ-field generated by $\epsilon_1, \ldots, \epsilon_i$.

The quasi-likelihood equation (2) can be written as

$$s_n(\beta) := \sum_{i=1}^{n} u_i(\beta) = 0 \tag{3}$$

where $u_i(\beta) = X_i' D_i(\beta)^{1/2} G D_i(\beta)^{-1/2} \epsilon_i(\beta) = \sum_{t,l=1}^{d} g_{tl} x_{it} h_{itl}(\beta) \epsilon_{il}(\beta)$, with $G := R^{-1} = (g_{tl})_{t,l=1,\ldots,d}$ and $h_{itl}(\beta) := [\dot{\mu}(x_{it}'\beta)/\dot{\mu}(x_{il}'\beta)]^{1/2}$, which is well-defined by (A). Note that $u_i := u_i(\beta_0), i \geq 1$ is a (p-dimensional) martingale difference sequence. The function $s_n(\beta)$ is called the quasi-score function.

We denote with $Z_i(\beta)$ the $d \times p$ matrix whose t-th row is $z_{it}(\beta)' := \sqrt{\dot{\mu}(x_{it}'\beta)}x_{it}'$, i.e. $Z_i(\beta) = D_i(\beta)X_i$. Then the quasi-information matrix is

$$\begin{aligned} F_n(\beta) := \mathrm{Var}[s_n(\beta)] &= \sum_{i=1}^{n} E[u_i(\beta)u_i(\beta)'] = \sum_{i=1}^{n} X_i' D_i(\beta)^{1/2} G D_i(\beta)^{1/2} X_i \\ &= \sum_{i=1}^{n} Z_i(\beta)' G Z_i(\beta) = \sum_{i=1}^{n} \sum_{t,l=1}^{d} g_{tl} z_{it}(\beta) z_{il}(\beta)' \end{aligned}$$

Note that $E[\dot{s}_n(\beta)] = -F_n(\beta)$, since $\dot{s}_n(\beta) = H_n(\beta) - F_n(\beta)$ with

$$H_n(\beta) := \sum_{i=1}^{n} \sum_{t,l=1}^{d} g_{tl} x_{it} \dot{h}_{itl}(\beta) \epsilon_{il}(\beta).$$

As it is the normal practice in this set-up, we will drop the argument β_0 in $D_i(\beta_0), Z_i(\beta_0), F_n(\beta_0)$, etc.
The asymptotic behaviour of the extreme eigenvalues of the matrix F_n is closely related to those of the matrices $B_n := \sum_{i=1}^{n} Z_i' Z_i = \sum_{i=1}^{d} \sum_{t,l=1}^{d} z_{it} z_{it}'$ and $A_n := \sum_{i=1}^{n} X_i' X_i = \sum_{i=1}^{n} \sum_{t=1}^{d} x_{it} x_{it}'$, since

$$\lambda_{\min}(G) \cdot B_n \leq F_n \leq \lambda_{\max}(G) \cdot B_n \tag{4}$$

(using Lemma 5, Appendix B) and

$$m \cdot A_n \leq B_n \leq M \cdot A_n \tag{5}$$

where $m := \min_{t=1,\ldots,d} \inf_i \dot{\mu}(x_{it}'\beta_0), M := \max_{t=1,\ldots,d} \sup_i \dot{\mu}(x_{it}'\beta_0)$. In particular, $\lim_n \lambda_{\min}(F_n) = \infty$ if and only if $\lim_n \lambda_{\min}(A_n) = \infty$; in this case $\lambda_{\min}(F_n) > 0$ (i.e. F_n is positive definite) for $n \geq N$.

We let $B_\delta(\beta_0) := \{\beta; \| \beta - \beta_0 \| < \delta\}$ be neighbourhoods of β_0. We will use the following conditions:

(D) *Divergence*: $\lim_{n\to\infty} \lambda_{\min}(A_n) = \infty$.

(E-p) *Equicontinuity in probability*: for every $\epsilon > 0$

$$\lim_{\delta \to 0} \limsup_{n\to\infty} P(\sup_{\beta \in B_\delta(\beta_0)} \| F_n^{-1}(\dot{s}_n(\beta) - \dot{s}_n(\beta_0)) \| \geq \epsilon) = 0.$$

(E-a.s) *Equicontinuity almost surely*: there exists an open neighbourhood U of β_0 such that $(F_n^{-1}\dot{s}_n(\beta))_{n \geq N}$ is equicontinuous on U at β_0 a.s.

(V) *Convergence of the normed conditional variance*: $Q_i \xrightarrow{P} 0$, where $Q_i := V_i^{-1/2L} E[\epsilon_i \epsilon_i' | \mathcal{F}_{i-1}] V_i^{-1/2R} - I$.

(N) *Continuity*: for every $\epsilon > 0$

$$\lim_{\delta \to 0} \limsup_{n \to \infty} P(\sup_{\beta \in B_\delta(\beta_0)} \| W_n(\beta) - I \| \geq \epsilon) = 0$$

where $W_n(\beta) := -F_n^{-1/2L} \dot{s}_n(\beta) F_n^{-1/2R}$.

Remarks: (D) is a sufficient condition for the strong consistency of the least squares estimator in the linear model (Theorem 1 of [6]). This condition has also been considered for the strong consistency of the MLE in a GLM with regressors with a compact range (Corollary 1 of [2]; Theorem 3 of [12]).

Conditions (E-p) (respectively (E-a.s)) are new in this framework. These conditions allow us to use an inverse function argument (see the proof of Theorem 4, Appendix A) to obtain the asymptotic (respectively almost sure) existence and the weak (respectively strong) consistency of a solution of equation (3). Unlike [2] and [12], we can not use convexity arguments to obtain this solution since we do not assume that $s_n(\beta)$ is a gradient.

Condition (V) allows us to prove the asymptotic normality of the quasi-score function using a martingale central limit theorem. This condition is automatically satisfied in the case of independent observations.

Condition (N) is an extension to longitudinal data of a similar condition that has been introduced for the asymptotic normality of the MLE in a GLM (Theorem 3 of [2]; Theorem 1 of [12]).

We state now our asymptotic results.

Theorem 1 *Under (D) and (E-p), there exists a sequence $(\hat{\beta}_n)_n$ of random variables such that*
(i) $P(s_n(\hat{\beta}_n) = 0) \to 1$ *and*
(ii) $\hat{\beta}_n \xrightarrow{P} \beta_0$.

Theorem 2 *Under (D) and (E-a.s), there exists a sequence $(\hat{\beta}_n)_n$ of random variables and a random number n_0 such that*
(i) $P(s_n(\hat{\beta}_n) = 0 \text{ for all } n \geq n_0) = 1$ *and*
(ii) $\hat{\beta}_n \to \beta_0$ *a.s.*

Lemma 1 *Suppose that* $\sup_{i \geq 1} E[\, \| \epsilon_i \|^{2+\delta} \, |\mathcal{F}_{i-1}] < \infty$ *for some $\delta > 0$. Under (D) and (V), we have:*

$$F_n^{-1/2L} s_n \to_d N(0, I). \tag{6}$$

Theorem 3 *Suppose that* $\sup_{i \geq 1} E[\, \| \epsilon_i \|^{2+\delta} \, |\mathcal{F}_{i-1}] < \infty$ *for some $\delta > 0$. Under (D), (V) and (N), we have*

$$F_n^{1/2R}(\hat{\beta}_n - \beta_0) \to_d N(0, I)$$

for any sequence $(\hat{\beta}_n)_n$ with $P(s_n(\hat{\beta}_n) = 0) \to 1$ and $\hat{\beta}_n \to_P \beta_0$.

3 Proofs

We will need the following version of the (multivariate) martingale strong law of large numbers.

Lemma 2 *Let $(s_n)_{n\geq 1}$ be a (p-dimensional) zero-mean, square-intergrable martingale and $(A_n)_{n\geq 1}$ be a sequence of nonnegative definite $p \times p$ matrices with $A_n \leq A_{n+1}$ and $\lim_n \lambda_{\min}(A_n) = \infty$. If there exists a constant $c > 0$ such that*

$$\mathrm{Var}(s_n) \leq cA_n, \text{ for all } n \geq N$$

then $A_n^{-1}s_n \to 0$ a.s.

Proof: We will consider only the case $c = 1$. The general case can be reduced to the case $c = 1$ for the matrices $B_n := cA_n$.

Let $u_n := s_n - s_{n-1}$ ($s_0 = 0$). By Theorem 12.4 of [4], it is enough to prove that $\sum_{n\geq N} E[\| A_n^{-1}u_n \|^2] < \infty$. This follows from Lemma 4 (Appendix B) since $E[\| A_n^{-1}u_n \|^2] = \mathrm{tr}\{A_n^{-1}E[u_n u_n']A_n^{-1}\}$ and $\mathrm{Var}(s_n) = \sum_{i=1}^n E[u_i u_i']$. $\square$

Remark: The previous lemma leads to a direct proof for the strong consistency of the LSE $\hat{\beta}_n$ in the classical linear model $y_i = x_i'\beta + \epsilon_i, i \geq 1$, when the residuals $(\epsilon_i)_{i\geq 1}$ form an L^2-bounded martingale difference sequence: if $\lambda_{\min}(\sum_{i=1}^n x_i x_i') \to \infty$, then $\hat{\beta}_n - \beta_0 = (\sum_{i=1}^n x_i x_i')^{-1}(\sum_{i=1}^n x_i\epsilon_i) \to 0$ a.s. since $\mathrm{Var}(\sum_{i=1}^n x_i\epsilon_i) \leq C\sum_{i=1}^n x_i x_i'$, where $C := \sup_i E[\epsilon_i^2]$.

Proofs of Theorems 1, 2: These results will follow by Theorem 4, respectively Theorem 5 (Appendix A) once we verify Assumptions 1, 2, respectively Assumptions 1', 2', for the functions

$$G_n(\beta) := F_n^{-1}s_n(\beta), \quad n \geq N.$$

Under (D), $G_n(\beta_0) = F_n^{-1}s_n \to 0$ a.s. (by Lemma 2), i.e. Assumption 1' is verified; consequently Assumption 1 is verified. Note that G_n are continuously differentiable on B. Since $\dot{G}_n(\beta) = F_n^{-1}\dot{s}_n(\beta)$, Assumption 2.(a) is exactly (E-p) and Assumption 2'.(a') is exactly (E-a.s).

For Assumption 2.(b) we note that $\dot{G}_n(\beta_0) = F_n^{-1}H_n - I$, where I is the identity matrix. We will prove that

$$F_n^{-1}H_n \to 0 \text{ a.s} \tag{7}$$

It will follow that with probability 1, there exists a random number $N_0 (> N)$ such that $\| F_n^{-1}H_n \| < 1/2$ for all $n \geq N_0$. Hence, with probability 1, for every $n \geq N_0$, $\dot{G}_n(\beta_0)$ is invertible with

$$\| \dot{G}_n(\beta_0)^{-1} \| \leq \frac{1}{1 - \| F_n^{-1}H_n \|} \leq 2.$$

Assumption 2.(b) is verified with $\lambda = 1/4$.

In order to prove (7), we let $H_{n,j}$ be the j-th column of H_n. Note that $H_{n,j} = \sum_{i=1}^n v_{i,j}$, where $v_{i,j} := \sum_{t,l=1}^d g_{tl} x_{it} \dot{h}_{itl,j}(\beta_0)\epsilon_{il}, i \geq 1$ is a (p-dimensional) martingale difference sequence; here $\dot{h}_{itl,j}(\beta_0)$ denotes the derivative of h_{itl} with respect to β_j at β_0, which is well-defined by (A). Note that $\dot{h}_{itl}(\beta) = h_{itl}^{(1)}(\beta)x'_{it} - h_{itl}^{(2)}(\beta)x'_{il}$ with

$$h_{itl}^{(1)}(\beta) = \frac{\ddot{\mu}(x'_{it}\beta)}{2\dot{\mu}(x'_{it}\beta)^{1/2}\dot{\mu}(x'_{il}\beta)^{1/2}}, \quad h_{itl}^{(2)}(\beta) = \frac{\ddot{\mu}(x'_{il}\beta)\dot{\mu}(x'_{it}\beta)^{1/2}}{2\dot{\mu}(x'_{il}\beta)^{3/2}}. \qquad (8)$$

Hence, there exists a constant C such that $|\dot{h}_{itl,j}(\beta_0)| \leq C, \forall i$. Using Lemma 5 (Appendix B), we have

$$E[v_{i,j}v'_{i,j}] = \sum_{t_1,t_2=1}^d q_{it_1t_2} x_{it_1} x'_{it_2} \leq \lambda_{\max}(Q_i) \cdot \sum_{t=1}^d x_{it}x'_{it}$$

where Q_i is the matrix with entries $q_{it_1t_2} := \sum_{l_1,l_2} g_{t_1l_1} g_{t_2l_2} \dot{h}_{it_1l_1,j}(\beta_0)\dot{h}_{it_2l_2,j}(\beta_0)$ $\mathrm{cov}(y_{il_1}, y_{il_2})$. Note that $|q_{it_1t_2}| \leq C_1$ for a constant C_1; hence $\lambda_{\max}(Q_i) \leq dC_1$, by Lemma 6 (Appendix B).

Using (4) and (5), $\mathrm{Var}[H_{n,j}] = \sum_{i=1}^n E[v_{i,j}v'_{i,j}] \leq dC_1 \cdot A_n \leq dC_1 \cdot [m\lambda_{\min}(G)]^{-1} \cdot F_n$. Hence $F_n^{-1}H_{n,j} \to 0$ a.s., using again Lemma 2. □

Proof of Lemma 1: Note that (6) is equivalent to: $\forall y \in \mathbf{R}^p$

$$\frac{y's_n}{\sqrt{y'F_ny}} \to_d N(0,1) \qquad (9)$$

This follows by the Cramèr-Wold theorem and the invariance property under orthogonal transformations of a sequence of asymptotically normal random vectors (see (3.4) of [2]): we have $(y's_n)/\sqrt{y'F_ny} = y'P'_nF_n^{-1/2L}s_n$, where $P_ny := (F_n^{1/2R}y)/\sqrt{y'F_ny}$ is an orthogonal transformation.

In order to prove (9) we will use the martingale central limit theorem with conditional Liapunov condition (Corollary 3.1 of [3]). Therefore, we have to verify the following two conditions:

$$\frac{1}{(y'F_ny)^{1+\delta/2}} \sum_{i=1}^n E[\ |y'u_i|^{2+\delta}|\mathcal{F}_{i-1}] \to 0. \qquad (10)$$

$$\frac{1}{y'F_ny} \sum_{i=1}^n E[(y'u_i)^2|\mathcal{F}_{i-1}] \to_P 1 \qquad (11)$$

We have $u_i = Z'_iGD_i^{-1/2}\epsilon_i$ and

$$\sum_{i=1}^n E[\ |y'u_i|^{2+\delta}|\mathcal{F}_{i-1}] \quad \leq \quad \sum_{i=1}^n \parallel Z_iy \parallel^{2+\delta} \cdot \parallel GD_i^{-1/2} \parallel^{2+\delta} \cdot E[\ \parallel \epsilon_i \parallel^{2+\delta} |\mathcal{F}_{i-1}]$$

$$\leq \quad C_1 \cdot \sum_{i=1}^{n} \| Z_i y \|^{2+\delta} \quad \text{(for a constant C}_1\text{)}$$

since GD_i is a matrix whose entries are bounded in modulus. On the other hand, $y'F_n y = \sum_{i=1}^n (Z_i y)' G (Z_i y) \geq \lambda_{\min}(G) \sum_{i=1}^n \| Z_i y \|^2$, using Lemma 5 (Appendix B) with $p = 1$. Hence, condition (10) is verified:

$$\frac{1}{(y'F_n y)^{1+\delta/2}} \sum_{i=1}^{n} E[\,|y'u_i|^{2+\delta}|\mathcal{F}_{i-1}] \leq C_1 \cdot \frac{1}{[\lambda_{\min}(G)]^{1+\delta/2}} \cdot \frac{\sum_{i=1}^n \| Z_i y \|^{2+\delta}}{(\sum_{i=1}^n \| Z_i y \|^2)^{1+\delta/2}}$$

$$\leq C_1 C_2^{\delta} \cdot \frac{1}{[\lambda_{\min}(G)]^{1+\delta/2}} \cdot (\sum_{i=1}^{n} \| Z_i y \|^2)^{-\delta/2} \to 0$$

where C_2 is a constant with $\| Z_i y \| \leq C_2, \forall i$ and we have used the fact that $\sum_{i=1}^n \| Z_i y \|^2 = y'B_n y \geq \lambda_{\min}(B_n) \| y \|^2 \to \infty$, by (D).

To verify (11), we note that $y'F_n y = E[(y's_n)^2] = \sum_{i=1}^n E[(y'u_i)^2]$ and

$$\sum_{i=1}^{n} E[(y'u_i)^2|\mathcal{F}_{i-1}] - y'F_n y = \sum_{i=1}^{n} y'\{E[u_i u_i'|\mathcal{F}_{i-1}] - E[u_i u_i']\} y = \sum_{i=1}^{n} w_i' Q_i w_i$$

with $w_i := V_i^{1/2R} D_i^{-1/2} G Z_i y$ and Q_i as defined in (V). We have

$$a_n \sum_{i=1}^{n} w_i' w_i \leq \sum_{i=1}^{n} w_i' Q_i w_i \leq b_n \sum_{i=1}^{n} w_i' w_i$$

where $a_n := \min_{i \leq n} \lambda_{\min}(Q_i)$ and $b_n := \max_{i \leq n} \lambda_{\max}(Q_i)$. Hence $|\sum_{i=1}^n w_i' Q_i w_i| \leq c_n \sum_{i=1}^n w_i' w_i = c_n y' F_n y$, where $c_n := \max\{|a_n|, |b_n|\}$. The proof is complete since $c_n \to_P 0$, by (V). $\Box$

We need the following result for the proof of Theorem 3.

Lemma 3 *Let $(A_n)_n$ be a sequence of $p \times p$ random matrices and Y_n p-dimensional random vectors such that $X_n := A_n Y_n, n \geq 1$ are square integrable in norm. If $A_n \to_P I$ and $C := \sup_n E[\, \| X_n \|^2] < \infty$, then $Y_n = X_n + o_P(1)$.*

Proof: Let $B_n := I - A_n$. Let $d_0 \in (0,1)$ be arbitrary (to be chosen later). For any $\eta \in (0,1)$ there exists n_0 such that $P(\| B_n \| < d_0) \geq 1 - \eta, \forall n \geq n_0$. On the event $\{\| B_n \| \leq d_0\}$, A_n is nonsingular, $A_n^{-1} = \sum_{k \geq 0} B_n^k$,

$$\| A_n^{-1} - I \| \leq \sum_{k \geq 1} \| B_n \|^k = \frac{\| B_n \|}{1 - \| B_n \|} \leq \frac{d_0}{1 - d_0} := d$$

and $\| Y_n - X_n \| \leq \| A_n^{-1} - I \| \cdot \| X_n \| \leq d \| X_n \|$.

By Chebyshev's inequality, $P(\| X_n \|< M) \geq 1-(C/M^2), \forall M>0$. Hence

$$
\begin{aligned}
P(\| Y_n - X_n \|< dM) &\geq P(\| B_n \|< d_0 \ , \ \| X_n \|< M) \\
&\geq P(\| B_n \|< d_0) + P(\| X_n \|< M) - 1 \\
&\geq 1 - \eta - \frac{C}{M^2}, \quad \forall n \geq n_0
\end{aligned}
$$

Finally, let $M_1 > 0$ and $\epsilon \in (0,1)$ be arbitrary. Pick $M>0$ with $M^2 > C/\epsilon$. The conclusion follows from the above inequality with $\eta := \epsilon - (C/M^2), d := M_1/M$ and $d_0 := d/(1+d)$. □

Proof of Theorem 3: We focus on the event $\{s_n(\hat{\beta}_n)=0\}$ whose probability goes to 1. By the mean-value theorem for vector-valued functions

$$
-s_n = \left[\int_0^1 \dot{s}_n(\beta_0 + t(\hat{\beta}_n - \beta_0)) \ dt\right](\hat{\beta}_n - \beta_0)
$$

and

$$
F_n^{-1/2L} s_n = \left[\int_0^1 W_n(\beta_0 + t(\hat{\beta}_n - \beta_0)) dt\right] F_n^{1/2R}(\hat{\beta}_n - \beta_0)
$$

with W_n as defined in (N). We claim that

$$
\int_0^1 W_n(\beta_0 + t(\hat{\beta}_n - \beta_0)) \ dt \to_P I \tag{12}
$$

To see this, let $\epsilon > 0, \eta > 0$ be arbitrary. By (N), there exist δ and n_1 such that $P(\| W_n(\beta) - I \|< \epsilon, \ \forall \beta \in B_\delta(\beta_0)) \geq 1 - \eta/2, \forall n \geq n_1$. Since $\hat{\beta}_n \to_P \beta_0$, there exists an integer n_2 such that $P(\hat{\beta}_n \in B_\delta(\beta_0)) \geq 1-\eta/2, \forall n \geq n_2$. We have

$$
\begin{aligned}
&P(\| \int_0^1 W_n(\beta_0 + t(\hat{\beta}_n - \beta_0)) \ dt - I \|< \epsilon) \geq \\
&\quad P(\| W_n(\beta) - I \|< \epsilon, \forall \beta \in B_\delta(\beta_0) \text{ and } \hat{\beta}_n \in B_\delta(\beta_0)) \geq \\
&\quad P(\| W_n(\beta) - I \|< \epsilon, \forall \beta \in B_\delta(\beta_0)) + P(\hat{\beta}_n(\beta_0) \in B_\delta(\beta_0)) - 1 \geq 1 - \eta
\end{aligned}
$$

Note that $E[\| F_n^{-1/2L} s_n \|^2] = \mathrm{tr}\{F_n^{-1/2L} E[s_n s_n'] F_n^{-1/2R}\} = \mathrm{tr}(I) = p, \forall n$. From Lemma 3, $F_n^{1/2R}(\hat{\beta}_n - \beta_0) = F_n^{-1/2L} s_n + o_P(1)$. The result follows from Lemma 1, using Slutsky's theorem. □

4 Verification of the assumptions

In this section we will examine some conditions which lead to the verification of condition (E-p), (E-a.s.) and (N) introduced in Section 2. We will suppose that there exists $K > L \| \beta_0 \|$ such that $\inf_{y \in [-K,K]} \dot{\mu}(y) > 0$, with L as defined in

Assumption (A). If we choose $r \leq (K/L) - \parallel \beta_0 \parallel$ such that $U := B_r(\beta_0) \subseteq B$, then $|x'_{it}\beta| \leq L(r + \parallel \beta_0 \parallel) \leq K,\ \forall \beta \in U$, i.e. $x'_{it}\beta \in [-K, K],\ \forall \beta \in U$.

We begin by noting that if the eigenvalue ratio $\lambda_{\max}(A_n)/\lambda_{\min}(A_n)$ is bounded, then (N) is equivalent to (E-p): to see this, note that $F_n^{-1}H_n \to 0$ a.s. and

$$F_n^{-1/2R}(W_n(\beta) - I)F_n^{1/2R} = -F_n^{-1}H_n + F_n^{-1}(\dot{s}_n(\beta_0) - \dot{s}_n(\beta)).$$

Now we examine conditions (E-p), (E-a.s.). For this purpose, we write

$$\dot{s}_n(\beta) = H_n^{(1)}(\beta) - H_n^{(2)}(\beta) - F_n(\beta)$$

where $H_n^{(1)}(\beta) = \sum_{i=1}^n \sum_{t,l=1}^d g_{tl} h_{itl}^{(1)}(\beta)\epsilon_{il}(\beta)x_{it}x'_{it}$, $H_n^{(2)}(\beta) = \sum_{i=1}^n \sum_{t,l=1}^d g_{tl}$ $h_{itl}^{(2)}(\beta)\epsilon_{il}(\beta)x_{it}x'_{il}$ and $h_{itl}^{(1)}, h_{itl}^{(2)}$ are given by (8). We will use the following notations: $\Delta_{itl}^{(s)}(\beta_1, \beta_2) := h_{itl}^{(s)}(\beta_2)\epsilon_{il}(\beta_2) - h_{itl}^{(s)}(\beta_1)\epsilon_{il}(\beta_1)$ for $s = 1, 2$ and $\Delta_{itl}^{(3)}(\beta_1, \beta_2) := h_{itl}^{(3)}(\beta_2) - h_{itl}^{(3)}(\beta_1)$, where $h_{itl}^{(3)}(\beta) = \sqrt{\dot{\mu}(x'_{it}\beta)}\sqrt{\dot{\mu}(x'_{il}\beta)}$. We let $\Phi_n := \{1, \ldots, n\} \times \{1, \ldots, d\}$.

We consider the following condition:

(B) $\limsup_{n\to\infty} [\lambda_{\max}(A_n)]^{1/2}/\lambda_{\min}(A_n) < \infty$

Under (D), this condition is weaker than condition (S_δ) of [2] (or condition (D_2) of [12]), in which a power $1/2 + \delta$ of $\lambda_{\max}(A_n)$ is considered.

Proposition 1 *Suppose that $\exists C_s, \alpha_s > 0; s = 0, 1, 2, 3$ such that*

$$|\mu_{il}(\beta_2) - \mu_{il}(\beta_1)| \leq C_0 \parallel \beta_2 - \beta_1 \parallel^{\frac{p}{2} + \alpha_0} \tag{13}$$

$$|h_{itl}^{(s)}(\beta_2) - h_{itl}^{(s)}(\beta_1)| \leq C_s \parallel \beta_2 - \beta_1 \parallel^{\frac{p}{2} + \alpha_s} \tag{14}$$

$\forall i, \forall \beta_1, \beta_2 \in U$. Then (B) implies (E-p).

Proof: The result will follow by Theorem 6 (Appendix A) once we prove that $\exists C, \alpha > 0$ such that $\forall n \geq 1, \forall \beta_1, \beta_2 \in U$

$$E[\parallel F_n^{-1}(\dot{s}_n(\beta_2) - \dot{s}_n(\beta_1)) \parallel_E^2] \leq C \parallel \beta_2 - \beta_1 \parallel^{p+\alpha}.$$

Using (B), it is enough to show that $\exists C'_s, \alpha'_s > 0$ such that $\forall n \geq 1, \forall \beta_1, \beta_2 \in U$

$$E[\parallel H_n^{(s)}(\beta_2) - H_n^{(s)}(\beta_1) \parallel_E^2] \leq C'_s \parallel \beta_2 - \beta_1 \parallel^{p+\alpha'_s} \cdot \parallel A_n \parallel \tag{15}$$

for $s = 1, 2$ (and a similar inequality for F_n, without the $E[\cdot]$).
Using the fact that $E[\parallel A \parallel_E^2] = \mathrm{tr}[E(A'A)]$, we have

$$E[\parallel H_n^{(1)}(\beta_2) - H_n^{(1)}(\beta_1) \parallel_E^2] = \mathrm{tr}\{\sum_{(i_1,t_1),(i_2,t_2)\in\Phi_n} s_{(i_1,t_1),(i_2,t_2)}^{(1)} x_{i_1 t_1} x'_{i_2 t_2}\}$$

where $s^{(1)}_{(i_1,t_1),(i_2,t_2)} := (x'_{i_1t_1}x_{i_2t_2})\sum_{l_1,l_2} g_{t_1l_1}g_{t_2l_2}E[\Delta^{(1)}_{i_1t_1l_1}\Delta^{(1)}_{i_2t_2l_2}]$ and $\Delta^{(1)}_{i_1t_1l_1} := \Delta^{(1)}_{i_1t_1l_1}(\beta_1,\beta_2)$. After tedious computations we reach the conclusion that

$$E[\Delta^{(1)}_{i_1t_1l_1}\Delta^{(1)}_{i_2t_2l_2}] = \begin{cases} \delta_{i_1t_1l_1}\delta_{i_2t_2l_2} & \text{if } i_1 \neq i_2 \\ \delta_{it_1l_1}\delta_{it_2l_2} + w_{it_1l_1t_2l_2} & \text{if } i_1 = i_2 = i \end{cases}$$

where $\delta_{itl} := \phi_{itl}(\beta_2) - \phi_{itl}(\beta_1)$ with $\phi_{itl}(\beta) := h^{(1)}_{itl}(\beta)(\mu_{il}(\beta) - \mu_{il}(\beta_0))$, and $w_{it_1l_1t_2l_2} := [h^{(1)}_{it_1l_1}(\beta_2) - h^{(1)}_{it_1l_1}(\beta_1)][h^{(1)}_{it_2l_2}(\beta_2) - h^{(1)}_{it_2l_2}(\beta_1)]\mathrm{cov}(y_{il_1}, y_{il_2})$. Hence

$$E[\parallel H^{(1)}_n(\beta_2) - H^{(1)}_n(\beta_1)\parallel^2_E] = \mathrm{tr}\ \{\sum_{(i_1,t_1),(i_2,t_2)\in\Phi_n} r_{(i_1,t_1),(i_2,t_2)}x_{i_1t_1}x'_{i_2t_2}\}$$

$$+ \sum_{i=1}^{n} \mathrm{tr}\ \{\sum_{t_1,t_2} v_{it_1t_2}x_{it_1}x'_{it_2}\} \tag{16}$$

where $r_{(i_1,t_1),(i_2,t_2)} := (x'_{i_1t_1}x_{i_2t_2})\sum_{l_1,l_2} g_{t_1l_1}g_{t_2l_2}\delta_{i_1t_1l_1}\delta_{i_2t_2l_2}$ and $v_{it_1t_2} := (x'_{it_1}x_{it_2})\sum_{l_1,l_2} g_{t_1l_1}g_{t_2l_2}w_{it_1l_1t_2l_2}$. By Lemma 5 (Appendix B)

$$\sum_{(i_1,l_1),(i_2,l_2)\in\Phi_n} r_{(i_1,t_1),(i_2,t_2)}x_{i_1t_1}x'_{i_2t_2} \leq \frac{1}{2}\lambda_{\max}(\tilde{R}_n)\cdot A_n$$

$$\sum_{t_1,t_2} v_{it_1t_2}x_{it_1}x'_{it_2} \leq \frac{1}{2}\lambda_{\max}(V_i)\cdot\sum_{t=1}^{d} x_{it}x'_{it}$$

where $\tilde{R}_n$ is the matrix with entries $\tilde{r}_{(i_1,t_1),(i_2,t_2)} = r_{(i_1,t_1),(i_2,t_2)} + r_{(i_2,t_2),(i_1,t_1)}$ with $(i_1,t_1),(i_2,t_2) \in \Phi_n$ and V_i is the matrix with entries $v_{it_1t_2}$.

From conditions (13) and (14) and Assumption (A), it follows that there exists $C_4, \alpha_4 > 0$ such that $|\tilde{r}_{(i_1,t_1),(i_2,t_2)}| \leq 2C_4 \parallel \beta_2 - \beta_1 \parallel^{p+\alpha_4}$; using Lemma 6 (Appendix B), it follows that $\lambda_{\max}(\tilde{R}_n) \leq 2dC_4 \parallel \beta_2 - \beta_1 \parallel^{p+\alpha_4}$. Therefore

$$\mathrm{tr}\ \{\sum_{(i_1,t_1),(i_2,t_2)\in\Phi_n} r_{(i_1,t_1),(i_2,t_2)}x_{i_1t_1}x'_{i_2t_2}\} \leq dC_4 \parallel \beta_2 - \beta_1 \parallel^{p+\alpha_4} \mathrm{tr}(A_n) \tag{17}$$

The similar argument applied to the matrix V_i leads us to

$$\mathrm{tr}\ \{\sum_{t_1,t_2} v_{it_1t_2}x_{it_1}x'_{it_2}\} \leq dC_5 \parallel \beta_2 - \beta_1 \parallel^{p+\alpha_5} \mathrm{tr}\{\sum_t x_{it}x'_{it}\} \tag{18}$$

Inequality (15) follows from (16), (17) and (18), since $\mathrm{tr}(A_n) \leq p \parallel A_n \parallel$.
A similar argument can be used for $H^{(2)}_n$ (respectively for F_n) by writing

$$E[\parallel H^{(2)}_n(\beta) - H^{(2)}_n(\beta_0)\parallel^2_E] \ = \ \mathrm{tr}\{\sum_{(i_1,l_1),(i_2,l_2)\in\Phi_n} s^{(2)}_{(i_1,l_1),(i_2,l_2)}x_{i_1l_1}x'_{i_2l_2}\}$$

where $s^{(2)}_{(i_1,l_1),(i_2,l_2)} := \sum_{t_1,t_2}(x'_{i_1t_1}x_{i_2t_2})g_{t_1l_1}g_{t_2l_2}E[\Delta^{(2)}_{i_1t_1l_1}\Delta^{(2)}_{i_2t_2l_2}]$. $\square$

Proposition 2 *Suppose that*

$$\sup_i \parallel y_i \parallel < \infty \quad \text{a.s.} \tag{19}$$

Then (B) implies (E-a.s).

Proof: From the uniform continuity of the functions $\mu, \dot{\mu}, \ddot{\mu}$ on the interval $[-K, K]$ and the boundedness of the regressors x_{it}, we obtain that $(\mu(x'_{it}\beta))_{i\geq 1}$, $(\dot{\mu}(x'_{it}\beta))_{i\geq 1}, (\ddot{\mu}(x'_{it}\beta))_{i\geq 1}$ and hence $(h^{(3)}_{itl}(\beta))_{i\geq 1}$ are equicontinuous on U at β_0. Moreover, since $(\epsilon_i)_{i\geq 1}$ are bounded in norm a.s. (by (19)), it follows that $(h^{(s)}_{itl}(\beta)\epsilon_{il}(\beta))_i; s = 1, 2$ are equicontinuous on U at β_0 a.s. Hence, with probability 1, for every $\epsilon > 0$ there exists a random number $\delta \in (0, r)$ such that $|\Delta^{(s)}_{itl}(\beta, \beta_0)| \leq \epsilon, \forall \beta \in B_\delta(\beta_0), \forall i, \forall t, \forall l, \forall s = 1, 2, 3$.

Using (B) it is enough to show that $\exists C_s > 0$ such that $\forall \beta \in B_\delta(\beta_0), \forall n$

$$\parallel H^{(s)}_n(\beta) - H^{(s)}_n(\beta_0) \parallel \leq C_s \epsilon \parallel A_n \parallel^{1/2}$$

for $s = 1, 2$ (and a similar inequality for F_n).

Using the fact that $\parallel A \parallel = \parallel A'A \parallel^{1/2}$, we have

$$\parallel H^{(1)}_n(\beta) - H^{(1)}_n(\beta_0) \parallel = \parallel \sum_{(i_1,t_1),(i_2,t_2)\in\Phi_n} s^{(1)}_{(i_1,t_1),(i_2,t_2)} x_{i_1t_1} x'_{i_2t_2} \parallel^{1/2}$$

where $s^{(1)}_{(i_1,t_1),(i_2,t_2)} := (x'_{i_1t_1} x_{i_2t_2}) \sum_{l_1,l_2} g_{t_1l_1} g_{t_2l_2} \Delta^{(1)}_{i_1t_1l_1} \Delta^{(1)}_{i_2t_2l_2}$ and $\Delta^{(1)}_{i_1t_1l_1} := \Delta^{(1)}_{i_1t_1l_1}(\beta, \beta_0)$. By Lemma 5 (Appendix B)

$$\frac{1}{2}\lambda_{\min}(\tilde{S}_n) \cdot A_n \leq \sum_{(i_1,t_1),(i_2,t_2)\in\Phi_n} s^{(1)}_{(i_1,t_1),(i_2,t_2)} x_{i_1t_1} x'_{i_2t_2} \leq \frac{1}{2}\lambda_{\max}(\tilde{S}_n) \cdot A_n$$

where $\tilde{S}_n$ is the matrix with entries $\tilde{s}_{(i_1,t_1),(i_2,t_2)} = s^{(1)}_{(i_1,t_1),(i_2,t_2)} + s^{(1)}_{(i_2,t_2),(i_1,t_1)}$ with $(i_1, t_1), (i_2, t_2) \in \Phi_n$. Note that $|\tilde{s}_{(i_1,t_1),(i_2,t_2)}| \leq 2d^2 L^2 (\max_{t,l} |g_{tl}|)^2 \epsilon^2 := 2C_1^2 \epsilon^2$; using Lemma 6 (Appendix B), it follows that $|\lambda_{\max}(\tilde{S}_n)| \leq 2dC_1^2\epsilon^2$ and $|\lambda_{\min}(\tilde{S}_n)| \leq 2dC_1^2\epsilon^2$. Hence for every $\beta \in B_\delta(\beta_0)$ and for every n

$$\parallel H^{(1)}_n(\beta) - H^{(1)}_n(\beta_0) \parallel \leq \sqrt{d} C_1 \epsilon \cdot \parallel A_n \parallel^{1/2}.$$

A similar argument can be used for $H^{(2)}_n$ (respectively for F_n) by writing

$$\parallel H^{(2)}_n(\beta) - H^{(2)}_n(\beta_0) \parallel = \parallel \sum_{(i_1,l_1),(i_2,l_2)\in\Phi_n} s^{(2)}_{(i_1,l_1),(i_2,l_2)} x_{i_1l_1} x'_{i_2l_2} \parallel^{1/2}$$

where $s^{(2)}_{(i_1,l_1),(i_2,l_2)} := \sum_{t_1,t_2} (x'_{i_1t_1} x_{i_2t_2}) g_{t_1l_1} g_{t_2l_2} \Delta^{(2)}_{i_1t_1l_1} \Delta^{(2)}_{i_2t_2l_2}$. $\square$

We conclude this section by discussing some examples.

Example 1. For $p = 1$, condition (13) of Proposition 1 is automatically satisfied, while condition (14) is satisfied if μ is three times continuously differentiable on $[-K, K]$. This follows by the mean-value theorem, since the functions $h_{itl}^{(s)}$ are continuously differentiable with $|\dot{h}_{itl}^{(s)}(\beta)| \leq C, \forall i, \forall \beta \in U$. Moreover, in this case, (B) is equivalent to $\liminf_n \sum_{i=1}^n \sum_{t=1}^d x_{it}^2 > 0$, which is an immediate consequence of (D).

Example 2. In the case of the linear regression, we have $\mu(y) = y$; hence $h_{itl}^{(s)} \equiv 0$ for $s = 1, 2$ and $\dot{s}_n(\beta) = -F_n(\beta)$. In this case it can be checked directly that (B) implies (E-p) and (E-a.s). Conditions (13) and (14) of Proposition 1, respectively condition (19) of Proposition 2 are no longer needed.

A General result for estimating equations

Let $G_n(\theta) := G_n(\omega, \theta), n \geq 1$ be p-variate random functions of $\theta \in \Theta$, where Θ is an open subset of $\mathbf{R}^p$ which contains the true parameter θ_0.

Assumption 1. $G_n(\theta_0) \to_P 0$.

Assumption 2. There exists an open neighbourhood U of θ_0 such that with probability 1, $G_n(\theta)$ is continuously differentiable on U, $\forall n \geq 1$. Moreover,
(a) $(\dot{G}_n(\theta))_{n \geq 1}$ is "equicontinuous in probability at θ_0", i.e. for every $\epsilon > 0$

$$\lim_{\delta \to 0} \limsup_{n \to \infty} P(\sup_{\theta \in B_\delta(\theta_0)} \| \dot{G}_n(\theta) - \dot{G}_n(\theta_0) \| \geq \epsilon) = 0;$$

(b) with probability 1, there exists a random number N_0 such that $\dot{G}_n(\theta_0)$ is nonsingular $\forall n \geq N_0$ and there exists a *nonrandom* number $\lambda > 0$ with $\lambda < \frac{1}{2} \inf_{n \geq N_0} \| \dot{G}_n(\theta_0)^{-1} \|^{-1}$.

Theorem 4 *Under Assumptions 1 and 2, there exists a sequence $(\hat{\theta}_n)_n$ of random variables such that*
(i) $P(G_n(\hat{\theta}_n) = 0) \to 1$ and
(ii) $\hat{\theta}_n \to_P \theta_0$.

Proof: With probability 1, for every $n \geq N_0$, the functions G_n are one-to-one on U and we define $\hat{\theta}_n$ as the unique zero of the function G_n in U if it exists and as an arbitrary constant otherwise. Let $\eta > 0$ be arbitrary. By Assumption 2.(a) there exist some nonrandom numbers δ, n_0 such that $\forall n \geq n_0$

$$P(\| \dot{G}_n(\theta) - \dot{G}_n(\theta_0) \| < \lambda, \forall \theta \in B_\delta(\theta_0)) \geq 1 - \frac{\eta}{2}$$

By a modified version of the inverse function theorem (p. 221 of [9]; see also Lemma 1 of [15]), the event $\{\| \dot{G}_n(\theta) - \dot{G}_n(\theta_0) \| < \lambda, \forall \theta \in B_\delta(\theta_0)\}$ is contained in the event $\{B_{\lambda\delta}(G_n(\theta_0)) \subseteq G_n(B_\delta(\theta_0))\}$. By Assumption 1, there exists a nonrandom number $n_1(> n_0)$ such that $P(0 \in B_{\lambda\delta}(G_n(\theta_0))) \geq 1-\eta/2, \forall n \geq n_1$. Hence

$$P(0 \in G_n(B_\delta(\theta_0))) \geq P(0 \in B_{\lambda\delta}(G_n(\theta_0)) \subseteq G_n(B_\delta(\theta_0))) \geq$$
$$P(0 \in B_{\lambda\delta}(G_n(\theta_0))) + P(B_{\lambda\delta}(G_n(\theta_0)) \subseteq G_n(B_\delta(\theta_0))) - 1 \geq 1 - \eta$$

for every $n \geq n_1$, i.e. $P(0 \in G_n(B_\delta(\theta_0))) \to 1$.

Finally, we prove that $\hat{\theta}_n \to_P \theta_0$. Suppose that there exist $\eta_0, \delta_0 > 0$ and a subsequence $(n_k)_k$ such that $P(\hat{\theta}_{n_k} \in B_{\delta_0}(\theta_0)) < 1 - \eta_0, \forall k$. Using the same argument as above we get a contradiction. □

In order to obtain the existence of a strongly consistent estimator $\hat{\theta}_n$ such that with probability 1, $G_n(\hat{\theta}_n) = 0$ for all n large, we need to strenghten our assumptions as follows.

Assumption 1'. $G_n(\theta_0) \to 0$ a.s.

Assumption 2'. The same as Assumption 2, except that (a) is replaced by: (a') $(\dot{G}_n(\theta))_{n\geq 1}$ is "equicontinuous on U at θ_0 a.s.", i.e. with probability 1, for every $\epsilon > 0$ there exists a random number $\delta > 0$ such that $B_\delta(\theta_0) \subseteq U$ and

$$\| \dot{G}_n(\theta) - \dot{G}_n(\theta_0) \| < \epsilon, \ \forall \theta \in B_\delta(\theta_0), \ \forall n \geq 1;$$

Theorem 5 *Under Assumptions 1' and 2', there exists a sequence $(\hat{\theta}_n)_n$ of random variables and a random number n_0 such that*
(i) $P(G_n(\hat{\theta}_n) = 0$ for all $n \geq n_0) = 1$ *and*
(ii) $\hat{\theta}_n \to \theta_0$ *a.s.*

Proof: By Assumption 2'.(a'), with probability 1, there exists a random number $\delta > 0$ such that $\| \dot{G}_n(\theta) - \dot{G}_n(\theta_0) \| < \lambda, \forall \theta \in B_\delta(\theta_0), \forall n \geq 1$. By Assumption 1', with probability 1, there exists a random number n_0 such that $0 \in B_{\lambda\delta}(G_n(\theta_0)), \forall n \geq n_0$. Using the same argument as in the proof of Theorem 4 we can conclude that $P(0 \in G_n(B_\delta(\theta_0), \forall n \geq n_0) = 1$. The proof that $\hat{\theta}_n \to \theta_0$ a.s. is again by contradiction. □

The next result will give us a tool for verifying Assumption 2.(a) in practice.

Theorem 6 *Let $(X_n(\theta))_{\theta\in\Theta}, n \geq 1$ be multiparameter processes with values in $\mathbf{R}^k$ and continuous sample paths. If there exist $\gamma, C, \alpha > 0$ such that*

$$E[\| X_n(\theta_2) - X_n(\theta_1) \|^\gamma] \leq C \| \theta_2 - \theta_1 \|^{p+\alpha}$$

$\forall n \geq 1, \forall \theta_1, \theta_2 \in \Theta$*, then for every* $\epsilon > 0$

$$\lim_{\delta\to 0} \limsup_{n\to\infty} P(\sup_{\|\theta_2-\theta_1\|<\delta} \| X_n(\theta_2) - X_n(\theta_1) \| \geq \epsilon) = 0.$$

Proof: See problems 2.2.9, 2.4.11 and 2.4.13 of [5]. □

B Some matrix results

The first lemma is a matrix analogue of the following result: if $(b_n)_n$ is a sequence of positive real numbers and $a_n \geq \sum_{i=1}^n b_i$, then $\sum_n (b_n/a_n^2) < \infty$. Our proof is an extension to the multi-dimensional case of the argument kindly provided to us by Peter Daffer for the one-dimensional case.

Lemma 4 *Let $(B_i)_{i\geq 1}$ be a sequence of nonnegative definite matrices such that $A_n \geq \sum_{i=1}^n B_i, \forall n \geq N$, where $(A_n)_n$ are positive definite matrices. Then*

$$\sum_{n\geq N} \mathrm{tr}(A_n^{-1} B_n A_n^{-1}) < \infty.$$

Proof: Let $C_n := A_n - \sum_{i=1}^n B_i, n \geq N$. Then $A_n = \sum_{i=1}^n D_i, \forall n \geq N$, where $D_i := B_i + (1/n)C_n$ is a nonnegative definite matrix. Since $\mathrm{tr}(A_n^{-1} B_n A_n^{-1}) \leq \mathrm{tr}(A_n^{-1} D_n A_n^{-1})$ it is enough to prove that

$$\sum_{n\geq N} \mathrm{tr}(A_n^{-1} D_n A_n^{-1}) < \infty.$$

We have $A_n^{-1} D_n A_n^{-1} = -(A_{n-1}^{-1} - A_n^{-1}) + A_{n-1}^{-1}(I - A_{n-1}A_n^{-1})(I + A_{n-1}A_n^{-1})$ and

$$\begin{aligned} \mathrm{tr}\{A_{n-1}^{-1}(I - A_{n-1}A_n^{-1})(I + A_{n-1}A_n^{-1})\} &\leq \mathrm{tr}\{A_{n-1}^{-1}(I - A_{n-1}A_n^{-1})(I + I)\} \\ &= 2\mathrm{tr}(A_{n-1}^{-1} - A_n^{-1}) \end{aligned}$$

(To see this, write $A_n^{-1} = (A_{n-1} + D_n)^{-1} = A_{n-1}^{-1} - A_{n-1}^{-1}(A_{n-1}^{-1} + D_n^{-1})^{-1} A_{n-1}^{-1}$; hence $I - A_{n-1}A_n^{-1} = (A_{n-1}^{-1} + D_n^{-1})^{-1} A_{n-1}^{-1}$ and $\mathrm{tr}\{A_{n-1}^{-1}(I - A_{n-1}A_n^{-1})^2\} \geq 0$.) Hence, for every $n \geq N + 1$

$$\mathrm{tr}(A_n^{-1} D_n A_n^{-1}) \leq -\mathrm{tr}(A_{n-1}^{-1} - A_n^{-1}) + 2\mathrm{tr}(A_{n-1}^{-1} - A_n^{-1}) = \mathrm{tr}(A_{n-1}^{-1} - A_n^{-1})$$

$$\begin{aligned} \sum_{i=N}^n \mathrm{tr}(A_i^{-1} D_i A_i^{-1}) &\leq \mathrm{tr}(A_N^{-1} D_N A_N^{-1}) + \sum_{i=N+1}^n \mathrm{tr}(A_{i-1}^{-1} - A_i^{-1}) \\ &\leq \mathrm{tr}(A_N^{-1} D_N A_N^{-1}) + \mathrm{tr}(A_N^{-1}) \end{aligned}$$

which concludes the proof. □

The next result gives a matrix analogue for the inequality $\lambda_{\min}(G) \sum_{t=1}^d z_t^2 \leq \sum_{t,l=1}^d g_{tl} z_t z_l \leq \lambda_{\max}(G) \sum_{t=1}^d z_t^2$, which is valid for any symmetric matrix G and for every $z_1, \ldots, z_d \in \mathbf{R}$ (see Theorem 3.15 of [11], or p. 62 of [8]).

Lemma 5 *If $F = (f_{tl})_{t,l=1,\ldots,d}$ is an arbitrary matrix and $x_1, \ldots, x_d \in \mathbf{R}^p$, then $\frac{1}{2}\lambda_{\min}(\tilde{F}) \sum_{t=1}^d x_t x_t' \leq \sum_{t,l=1}^d f_{tl} x_t x_l' \leq \frac{1}{2}\lambda_{\max}(\tilde{F}) \sum_{t=1}^d x_t x_t'$ where $\tilde{F}$ is the matrix with entries $\tilde{f}_{tl} := f_{tl} + f_{lt}$.*

Proof: We have $y'(2\sum_{t,l=1}^{d} f_{tl}x_t x_l')y = y'(\sum_{t,l} f_{tl}x_t x_l' + \sum_{t,l} f_{lt}x_l x_t')y = \sum_{t,l}\tilde{f}_{tl}(x_t'y)(x_l'y) \leq \lambda_{\max}(\tilde{F})\sum_{t=1}^{d}(x_t'y)^2 = \lambda_{\max}(\tilde{F})\cdot y'(\sum_{t=1}^{d} x_t x_t')y$, for every $y \in \mathbf{R}^p$. The other inequality is similar. □

Lemma 6 *If* $A = (a_{tl})_{t,l=1,\ldots,d}$ *is a matrix with* $|a_{tl}| \leq \epsilon, \forall t, \forall l$ *then* $|\lambda| \leq d\epsilon$ *for any eigenvalue* λ *of* A.

Proof: Let λ be an eigenvalue of A and x an eigenvector corresponding to it, with $\| x \| = 1$. Then $|\lambda| = |\lambda| \, \| x \| = \| Ax \| \leq \| A \| \leq \| A \|_E \leq d\epsilon$. □

References

[1] Chen, K., Hu, I. and Ying, Z. (1999). Strong consistency of maximum quasi-likelihood estimators in generalized linear models with fixed and adaptive designs. *Ann. Stat.* **27**, 1155-1163.

[2] Fahrmeir, L. and Kaufmann H. (1985). Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann. Stat.* **13**, 342-368.

[3] Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and Its Applications*, Wiley, New York.

[4] Heyde, C.C. (1997). *Quasi-Likelihood and Its Applications*, Springer, New York.

[5] Karatzas, I. and Shreve, S.E. (1991). *Brownian Motion and Stochastic Calculus*, Springer, New York.

[6] Lai, T.L., Robbins, H., and Wei, C.Z. (1979). Strong consistency of least squares estimates in multiple regression II. *J. Multiv. Anal.* **9**, 343-361.

[7] Liang, K.-Y., and Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13-22.

[8] Rao, C.R. (1973). *Linear Statistical Inference and Its Applications*, Wiley, New York.

[9] Rudin, W. (1976). *Principles of Mathematical Analysis*, 3rd ed., McGraw-Hill, New York.

[10] Schiopu-Kratina, I. (2002). Asymptotic results for generalized estimating equations with data from complex surveys. To appear in *Revue roumaine de math. pures et appl.*

[11] Schott, J.R. (1997). *Matrix Analysis for Statistics*, John Wiley, New York.

[12] Shao, J. (1992). Asymptotic theory in generalized linear models with nuisance scale parametres. *Probab. Theory Relat. Fields* **91**, 25-41.

[13] Shao, J. (1999). *Mathematical Statistics*, Springer, New York.

[14] Wu, C.F.J. (1981). Asymptotic theory of nonlinear least squares estimation. *Ann. Stat.* **9**, 501-513.

[15] Yuan, K.-H., and Jennrich, R.J. (1998). Asymptotics of estimating equations under natural conditions. *J. Multiv. Anal.* **65**, 245-260.