ACTA MATHEMATICA VIETNAMICA
Volume 19, Number 1, 1994, pp. 111–137

STRONG CONSISTENCY OF LEAST SQUARES ESTIMATES IN POLYGONAL REGRESSION WITH RANDOM EXPLANATORY VARIABLES*

NGUYEN BAC VAN

Abstract. The strong consistency of generalized least squares estimates in polygonal regression with random explanatory variables is established regardless of the projection support under minimal moment restrictions.

1. Introduction

Regression models with random explanatory variables arise in some important practical situations involving bipartite observations $(X(t), Y(t))$ on n items, and also from the sampling scheme as exemplified in [5] by cross-sectional data. Of particular interest are polygonal regression models, considered in [1]

$$Y'(t) = \sum_{i=1}^{k} b_i'(X(t))q_i I_{S(i)}(X(t)) + e'(t) \tag{1.1}$$

where a prime denotes transpose and where $Y(t)$ are $r \times 1$ response observation vectors, $e(t)$ residuals, $b_i(.)$'s known $l(i) \times 1$ vector-valued functions on the common range space H of $X(t)$, $q_i's$ unknown $l(i) \times r$ matrix parameters. $S(i)'s$ specified disjoint sets in $H, I_{S(i)}$ the indicator of $S(i)$. Constraints may be imposed on the parameters q_i in order to ensure the smoothness to some order of the regression. Models like these arise when the unknown regression function is approximated by linear parametric functions in every domain $S(i)$ and also when the response variable follows a mixture of conditional distributions with linear

Received October 5, 1993
*This paper was supported in part by the National Basic Research Program in Natural Science, Vietnam.
1991 Mathematics Subject Classification:
Primary 62J05; Secondary 60F15.
Keywords and phrases: Polygonal regression models, strong consistency, orthogonal projector, generalized least squares estimates.

conditional mean structure given the explanatory variable in several domains $S(i)$, see [1], Section 5. In [1], pp.30-31, we have defined conditional generalized least squares estimators (CGLSE), we shall now call them GLSE for short and restate the definition in a more general form in Section 3. The purpose of this paper is to establish the strong consistency of GLSE, finite-sample properties of which have been investigated in [1]. The moment restriction needed is necessary and sufficient for the GLSE consistency uniform with respect to the projection support and especially for the GLSE consistency in the extreme case where the projection support has the maximal dimension. For the same extreme case, which necessarily arises when in the definition of least squares estimates the linear hull of the parameter range coincides with the whole corresponding Euclidean space, and for linear fixed-design regression models, Lai, Robbins and Wei have established the strong consistency of least squares estimates under minimal assumptions on the design [4].

2. Notation and conventions

Together with notations and conventions as in (1.1) the following are also to be used throughout the paper.

(i) For matrices

Finite matrix with numerical elements = matrix all elements of which are finite.

$M_{k\times l}$ = linear space of all $k \times l$ real matrices; $f'(.)$ = transpose of matrix function $f(.)$; I = unit matrix; $I_r = r \times r$ unit matrix.

n.n.d. = non - negative definite, p.d. = positive definite.

For any square matrix A:

Det A = determinant of A; Tr A = Trace of A.

For any real matrix $C = (c_{ij})$:

$$||C||^2 = \sum_{i,j} c_{ij}^2. \tag{2.1}$$

For $y = (y_1...y_k)' \in M_{k\times l}$ and $G \subset M_{k\times l}$:

$$[y] = (y_1'...y_k')' \in M_{kl\times 1}; \quad [G] = \{[y] : y \in G\}. \tag{2.2}$$

For $z = (z_1...z_l) \in M_{k\times l}$:

$$vecz = (z_1'...z_l')' \in M_{kl\times 1}.$$

$B \otimes C$ = Kronecker product of matrices, see [6], 1b.8.

Diag $(B_i, i = 1, ..., k)$ = block diagonal matrix.

(ii) For linear spaces

ImB = image of a linear mapping B;

BG = image by B of a set G.

For a $k \times l$ real matrix C and a set $F \subset R^l$:

$$CF = \{Cx : x \quad \text{is} \quad l\times 1 \quad \text{and} \quad x \in F\}. \tag{2.3}$$

Dim$F$ = dimension of the linear space F

$M\{C\}$ = linear space generated by the columns of the real matrix C, also called range space of C (see [6], 1b.6). (2.4)

$\perp$ = orthogonality symbol in an Euclidean space.

For $M_{k\times l}$ endowed with the inner product $(y, z) = [y]'A[z]$, A p.d.,

Pr_L^A = orthogonal projector from $M_{k\times l}$ onto some subspace L. (2.5)

$$Pr_L = Pr_L^I. \tag{2.6}$$

(iii) For random variables

$I_s(.)$ = indicator of the set S

Card$A$ = carinality of the set A

$(\Omega, \mathfrak{J}, P)$ = basic probability space.

P_A and E_A = conditional probability and expectation given the event A.

r.v. = random variable taking values in an arbitrary measurable space.

a.s. = almost surely (sure).

i.i.d. = independent and identically distributed.

$E(.|x)$ = conditional expectation given the r.v.X at some value x.

P^X = probability distribution of the r.v.X.

$\Delta|S = \{A \cap S : A \in \Delta\}$.

(iv) Topical notations

In this paper we will consider a pair of functions $X(t), Y(t)$ defined for $t \in N, N = \{1, 2, ...\}$. The range space of $X(t)$ is an arbitrary space H, whereas $Y(t)$ is $r \times 1$ vector-valued. The index i always ranges $1, ..., k$; $k \geq 1$ being fixed. Disjoint sets $S(i)$ are specified in H. Let us set

$$\{T_{i1}, T_{i2}\} = \{t : t \in N, X(t) \in S(i)\}, T_{i1} < T_{i2} < ....$$

When T_{ij} exists, $j \geq 1$, set

$$(X_{ij}, Y_{ij}) = (X(T_{ij}), Y(T_{ij})). \tag{2.7}$$

Let $b_i(.)$ be an $l(i) \times 1$ vector-valued function on $S(i)$ and $z(.)$ an $r \times r$ p.d. matrix-valued function on H. Let $a(i) \in N$ and set

$$s = a(1) + ... + a(k), l = l(1) + ... + l(k). \tag{2.8}$$

In the following block matrices short symbols $b_{ij} = b_i(X_{ij}), z_{ij} = z(X_{ij})$ are used and the second index j in a pair (ij) always ranges $1, ..., a(i)$.

$U_i' = (...Y_{ij}...) = (Y_{i1}...Y_{ia(i)})$.

$B_i' = (...b_{ij}...), Z_i = diag(z_{ij})$,

$C_i = Z_i^{1/2}(B_i \otimes I_r), T_i = a^{-1}(i)C_i'C_i$;

$U_a' = (U_1'...U_k')$,

$B_a = diag(B_1...B_k)$,

$Z_a = diag(Z_1, ..., Z_k)$,

$C_a = diag(C_1, ..., C_k) = Z_a^{1/2}(B_a \otimes I_r)$,

$A_a = diag(a(i)I_{a(i)r}, i = 1, ..., k)$,

$T_a = diag(T_i) = C_a'A_a^{-1}C_a$.

These block matrices will be of constant use later, the affix a will be dropped when no confusion is possible.

PROPOSITION 2.1. *The following facts are equivalent:*

(a) *det* $T_i > 0$ *or, equivalently,* T_i *is p.d. for all* i,

(b) T *is p.d.,*

(c) $C'C$ *is p.d.,*

(d) *Rank* $C=lr$,

(e) C *determines a one-to-one map from* R^{lr} *into* R^{sr}.

PROOF. Since $a(i)T_i = C_i'C_i$ is n.n.d., det $T_i > 0$ means T_i is p.d. The positive definiteness of $T = diag(T_i)$ is equivalent to that of all T_i, or also to that of all $C_i'C_i$, or again to that of $C'C$. Since $C'C$ is of order $lr \times lr$ and $RankC = RankC'C$, (c) is equivalent to (d). Finally $RankC = lr$ means that, for any $lr \times 1$ vector $x, Cx = 0$ is equivalent to $x = 0$, hence (d) is equivalent to (e). Q.E.D.

3. Generalized least squares estimates (GLSE)

We shall now give the definition of GLSE, starting from the data-bases $\{(X(t), Y(t)), t = 1, ..., n\}$ according to model (1.1). We first reduce the model to a compact form.

PROPOSITION 3.1. *Set*

$$d(i) = CardS(i) \cap \{X(1), ..., X(n)\}. \tag{3.1}$$

Replace a(i) by d(i) in the topical notations 2 (iv) and then $U_a, B_a, ...$ *will respectively be replaced by* $U_d, B_d, ...$ *Then, when all d(i) are positive, the model (1.1) for t=1,...,n is equivalent to the model*

$$U_d = B_d q + e \tag{3.2}$$

where $q = (q_1' ... q_k')'$ *is an* $l \times r$ *matrix parameter and* e *is some residual.*

PROOF. Keep topical notations. Then the equalities (1.1) for $t = 1, ..., n$, as a whole, are equivalent to $Y_{ij}' = b_i'(X_{ij})q_i + e_{ij}', j = 1, ..., d(i), i = 1, ..., k, e_{ij}$

being residuals. The latter are rewritten as

$$U_i' = (...q_i' b_{ij}...) + e_i' = q_i' B_i' + e_i', i = 1, ..., k,$$

with residuals e_i'. These, as a whole, in turn are rewritten as $U_d' = q' B_d' + e'$. Q.E.D.

PROPOSITION 3.2. *Let A, B, C, y, z be matrices and b some vector such that the following operations are meaningful. There holds*

(i) $(b \otimes A)B = b \otimes AB$,

(ii) $[ABC'] = (A \otimes C)[B]$,

in particular $[AB] = (A \otimes I)[B]$;

(iii) $vec(ABC') = (C \otimes A)vecB$,

in particular $vec(AB) = (I \otimes A)vec\,B$;

(iv) $[y]'[z] = (vec\,y)'vec\,z$.

PROOF. Check (i) and (iv) directly. (ii) figures in [2], p.84, and (iii) in [3], A1.50. Q.E.D.

PROPOSITION 3.3. *With notations 2 (iv), Rank* $B = l$ *is equivalent to Rank* $C = lr$.

PROOF. From 2 (iv) it is seen that B and C are respectively of order $s \times l$ and $sr \times lr$, and that $z(.)$ being p.d., so is Z. Then for any $l \times r$ real matrix x, from Proposition 3.2 (ii), $Bx = 0$ means $(B \otimes I_r)[x] = 0$ or, equivalently, $z^{1/2}(B \otimes I_r)[x] = 0$ which means $C[x] = 0$. Now Rank $B = l$ means $Bx = 0$ if and only if $x = 0_{l \times r}$ or, equivalently, $C[x] = 0$ if and only if $[x] = 0_{lr \times l}$ or again, equivalently, Rank $C = lr$. Q.E.D.

PROPOSITION 3.4. *Let F be a finite-dimensional real vector space endowed with an inner product v and u(.) be the induced norm. Let L be a subspace and* m_0 *some fixed vector of F.*

(i) Then, for given $U \in F$, *there exists a unique element* $p = p(U)$ *of the affine manifold* $L + m_0$ *such that for every* $y \in L + m_0$ *there holds*

$$u(U - p) \leq u(U - y).$$

We shall denote $p(U) = Pr^v_{L+m_0}U$ *and call it the orthogonal projection of* U *on* $L + m_0$.

(ii) If Pr^v_L *is the orthogonal projector of* F *onto* L *and* Id *denotes the identity mapping of* F, *then*

$$Pr^v_{L+m_0}U = Pr^v_L U + (Id - Pr^v_L)m_0. \tag{3.3}$$

(iii) Let D *be any set in* F *and* $D_0 = \{y - z : y, z \in D\}$. *If* L *contains* D_0 *then* $L + m_0$ *and* $Pr^v_{L+m_0}U$ *remain the same for all* $m_0 \in D$.

This Proposition generalizes Lemma 5(i) and (ii) in [1], the proof remains the same.

PROOF. (i) For $U \in F$ we have the decomposition

$$U = U_0 + V_0, U_0 \in L, V_0 \perp L$$

and, similarly,

$$m_0 = e_0 + f_0, e_0 \in L, f_0 \perp L.$$

Hence $U_0 + f_0 = (U_0 - e_0) + m_0 \in L + m_0$. For any $y \in L + m_0$ we have

$$y = y_0 + f_0 \quad \text{with} \quad y_0 = e_0 + (y - m_0) \in L.$$

Hence $U - y = (U_0 - y_0) + U - (U_0 + f_0)$
with $U_0 - y_0 \in L$ and $U - (U_0 + f_0) = V_0 - f_0 \perp L$.
Therefore, by setting $p = U_0 + f_0$, we have $u(U - y) \geq u(U - p)$ with equality if and only if $U_0 - y_0 = 0$ or, equivalently, $y = U_0 + f_0 = p$.

(ii) Since $U_0 = Pr^v_L U$ and $f_0 = m_0 - e_0 = (Id - Pr^v_L)m_0$, (ii) follows.

(iii) For m_0 and m in D, we have $m - m_0 \in D_0 \subset L$, hence $L + m = L + (m - m_0) + m_0 = L + m_0$ and from (ii) we have

$$Pr^v_{L+m}U - Pr^v_{L+m_0}U = m - m_0 - Pr^v_L(m - m_0) = 0. \quad \text{Q.E.D.}$$

We are now in a position to define GLSE in the model (1.1). First, using Proposition 3.1 we reduce the model to the form $U_d = B_d q + e$, then we apply Proposition 3.4.

DEFINITION 3.1. Let Q be the range of the parameter q in $M_{l\times r}$. Define $D = \{B_d q : q \in Q\}$ and $D_0 = \{y - z : y, z \in D\}$. Let $\mathrm{Im}B_d$ be the image of $M_{l\times r}$ by the map $u \to B_d u, u \in M_{l\times r}$. Let L be any linear subspace of $\mathrm{Im}B_d$ containing D_0. Denote $m = B_d q$ and consider the equation in $\hat{q}$:

$$Pr_{L+m}^{Z_d} U_d = B_d \hat{q}, \tag{3.4}$$

where the left member was given by (3.3) with $Pr_L^{Z_d}$ explained in (2.5).

Then any solution $\hat{q}$ will be called a generalized least squares (GLS) value of q and, if unique, a GLSE for q.

PROPOSITION 3.5. *GLS values for q always exist and are independent of m provided all d(i) given by (3.1), are positive. A GLSE for q exists if and only if, all (d(i) being positive, Rank $B_d = l$.*

PROOF. When all $d(i)$ are positive, from topical notations 2 (iv), U_d, B_d and Z_d are defined. From Proposition 3.4 (i) the left-hand side of (3.4) is an element of $L + m$. But $L + m \subset \mathrm{Im}B_d$ since $L \subset \mathrm{Im}B_d$ and $m \in D \subset \mathrm{Im}B_d$. The left-hand side of (3.4) belonging to $\mathrm{Im}B_d$, there always exists some value $\hat{q}$ in $M_{l\times r}$ satisfying the equation (3.4). If $\hat{q}_0$ is a solution of (3.4), then $\hat{q}$ is a solution if and only if $B_d\hat{q} - B_d\hat{q}_0 = B_d(\hat{q} - \hat{q}_0) = 0$ or, equivalently, $\hat{q} = \hat{q}_0 + h$, where h is an arbitrary $l \times r$ real matrix such that $B_d h = 0$.

The solution is unique when and only when $B_d h = 0$ entails $h = 0$ or, equivalently, Rank $B_d = l$.

From Proposition 3.4(iii) it follows that the solutions $\hat{q}$ of (3.4) are independent of m when m varies in D. Q.E.D.

REMARK 3.1. In the topical notations 2 (iv) let us use a particular function $z(.)$ such that $z(x) = d_i^{-1}(x) I_r$ for $x \in S(i), i = 1, \ldots, k$, where $d_i(.)$ is a known positive function on $S(i)$. Then, setting

$$V_i = diag\{d_i(X_{ij}), j = 1, \ldots, d(i)\}, V = diag(V_1, \ldots, V_k),$$

we have $Z_d = V^{-1} \otimes I_r$. From Proposition 3.2 for u and y in the range space of U_d we have

$$[u]'Z_d[y] = [u]'(V^{-1} \otimes I_r)[y] = [u]'[V^{-1}y] =$$

$$= (vec\ u)'vec\ (V^{-1}y) = (vec\ u)'(I_r \otimes V^{-1})vec\ y.$$

Thus, for this choice of function $z(.)$, the inner product used for defining the projection in the range space of U_d coincides with the inner product used in [1] for defining conditional GLSE, see [1], formula (33), p.30.

Hence Definition 3.1 includes conditional GLSE examined in [1].

REMARK 3.2. Consider the extreme case $L = \mathrm{Im} B_d$ in Definition 3.1. Then $L + m = L$ since $m = B_d q \in L$. Hence equation (3.4) becomes $Pr_L^{Z_d} U_d = B_d \hat{q}$.

In particular, if the function $z(.)$ equals I_r for all $x \in H$, then the preceding equation is written as $Pr_L^I U_d = B_d \hat{q}$. In [4], for linear fixed-design regression models, just least squares estimates given by this simpler equation were examined.

4. Expression of GLSE error

In this section, we will solve equation (3.4) in $\hat{q} - q$. In the sequel, the following notations will be used, see (2.2), (2.3).

$$e = U_d - B_d q \qquad \text{(see (3.2))}, \qquad c = d(1) + \ldots + d(k); \qquad (4.1)$$

$[G] = Z_d^{1/2}[L]$ with G in $M_{c\times r}$, the range space of U_d;

$[h] = Z_d^{1/2}[e]$ with $h \in M_{c\times r}$;

$g = C_d' A_d^{-1}[h]$ with $A_d = diag(d(i)I_{d(i)r})$, by replacing $a(i)$ by $d(i)$ in topical notations 2(iv) throughout.

PROPOSITION 4.1. *The equation (3.4) defining GLS values $\hat{q}$ is equivalent to*

$$C_d[\hat{q} - q] = JC_d b, \qquad (4.2)$$

where J is the orthogonal projector from R^{cr} onto $[G]$ according to the inner product $u'v(u, v \in R^{cr})$ and where b is defined by the equality

$$C_d' A_d^{-1} C_d b = g. \tag{4.3}$$

Particularly, in the extreme case $L = \mathrm{Im}B_d$, GLS values $\hat{q}$ are defined by the equation

$$C_d[\hat{q} - q] = C_d b. \tag{4.4}$$

PROOF. Since no confusion is possible, we shall drop the affix d. By rewriting the left-hand side of the equation (3.4) according to Proposition 3.4(ii), we see that this equation becomes

$$Pr_L^Z U + (Bq - Pr_L^Z Bq) = B\hat{q} \qquad \text{(see (2.5))},$$

or, equivalently,

$$B(\hat{q} - q) = Pr_L^Z e. \tag{4.5}$$

By the isomorphism $y \to [y]$, the linear space $M_{c\times r}$ endowed with the inner product $(y, z) = [y]'Z[z]$ passes into the vector space R^{cr} with the same inner product, the orthogonal projector Pr_L^Z is transformed into the projector $Pr_{[L]}^Z$ of R^{cr} and, by using Proposition 3.2 (ii), the equation (4.5) becomes

$$(B \otimes I_r)[\hat{q} - q] = Pr_{[L]}^Z [e]. \tag{4.6}$$

The automorphism $[y] \to Z^{1/2}[y]$ of R^{cr} transform R^{cr} with the inner product $[y]'Z[z]$ into R^{cr} with the inner product $[u]'[v], [u] = Z^{1/2}[y], v = Z^{1/2}[z]$; the orthogonal projector $Pr_{[L]}^Z$ is transformed into the orthogonal projector $Pr_{[G]}^I$, i.e. $Pr_{[G]}$ by (2.6), since $[G]$ is the image of $[L]$.

With the topical notations 2(iv), the equation (4.6) is thus transformed into the equivalent one

$$C[\hat{q} - q] = Pr_{[G]}[h], \tag{4.7}$$

since $[h]$ is the image of $[e]$ by the automorphism. By considering B and C together as matrices and as mappings, with the mapping B from $M_{l\times r}$ into $M_{c\times r}$ and the mapping C from R^{lr} into R^{cr}, the inclusion $L \subset \mathrm{Im}B$ with $\mathrm{Im}B =$

$BM_{l\times r}$ is transformed by the product transformation $y \to [y] \to Z^{1/2}[y]$ into the inclusion

$$[G] \subset CR^{lr} = M\{C\} \quad \text{(see (2.4))},$$

for $C = Z^{1/2}(B \otimes I_r)$. In particular, when $L = \text{Im}B$ then $[G] = M\{C\}$. From (4.8) we have

$$Pr_{[G]}[h] = Pr_{[G]}Pr_{M\{C\}}[h].$$

But there exists some vector $b \in R^{lr}$ such that

$$Pr_{M\{C\}}[h] = Cb, \tag{4.9}$$

hence (4.7) is rewritten as (4.2). Now (4.9) is equivalent to $([h] - Cb) \perp M\{C\}$. For any vector $y \in R^{cr}$ with $c = d(1) + ... + d(k)$, let us write $y' = (y_1' ... y_k')$ with $y_i \in R^{d(i)r}$. Recall that $A_d = diag(d(i)I_{d(i)r})$ and that, from 2 (iv), $C = diag(C_i)$.

Then it is seen that the relation $y \perp M\{C\}$ is successively equivalent to the following ones

$$C'y = 0; C_i'y_i = 0 \quad \text{for all} \quad i;$$
$$d^{-1}(i)C_i'y_i = 0 \quad \text{for all} \quad i; C_d'A_d^{-1}y = 0,$$

the $d(i)$'s being assumed to be positive considering Prop. 3.5. Hence (4.9) is equivalent to

$$C'A_d^{-1}([h] - Cb) = 0$$

or again to (4.3). Finally, when $L = \text{Im}B$ then $Cb \in M\{C\} = [G]$ hence $JCb = Cb$ and (4.2) becomes (4.4). Q.E.D.

PROPOSITION 4.2. *Set $T_d = C_d'A_d^{-1}C_d$. Then, whenever the GLSE exists, it satisfies the formula*

$$[\hat{q} - q] = (T_d^{-1}C_d'A_d^{-1/2})(A_d^{-1/2}JA_d^{1/2})(A_d^{-1/2}C_dT_d^{-1})g. \tag{4.10}$$

PROOF. On account of Proposition 3.5, the d(i)'s are always assumed to be positive, then a GLSE $\hat{q}$ exists if and only if Rank $B_d = l$, or, equivalently, Rank $C_d = lr$ considering Proposition 3.3. From topical notations 2 (iv), by replacing a(i) by d(i) in T_a we obtain T_d. Then, from Proposition 2.1, the positive definiteness of T_d is equivalent to the existence of the GLSE $\hat{q}$. From now on, drop the affix d. Multiplying both sides of (4.2) by $C'A_d^{-1}$ and replacing b by $T^{-1}g$ from (4.3), we get

$$T[\hat{q}-q] = C'A_d^{-1}JCT^{-1}g = C'A_d^{-1/2}(A_d^{-1/2}JA_d^{1/2})A_d^{-1/2}CT^{-1}g.$$

Hence (4.10) follows. Q.E.D.

Let us quote some well-known facts.

PROPOSITION 4.3. *Let A,B,C be real matrices such that the product AB exists. Then (see (2.1))*

(i) $||A||^2 = TrAA'$,
(ii) $||AB||^2 \leq ||A||^2||B||^2$,
(iii) $Tr(A \otimes C) = TrA.TrC$ *if* A *and* C *are square matrices.*

We will now evaluate the Euclidean distance between the GLSE $\hat{q}$ and the true value q of the parameter in the model (1.1).

PROPOSITION 4.4. *Whenever the GLSE exists, there holds*

$$||\hat{q}-q||^2 \leq lr(TrT_d^{-1})^2(\sum_{f,g=1}^{k} d(g)d^{-1}(f))||g||^{-2}. \tag{4.11}$$

PROOF. Apply Proposition 4.3 (ii) to the right-hand side of (4.10) repeatedly. Note that, by dropping d, we have

$$||T^{-1}C'A_d^{-1/2}||^2 = ||A_d^{-1/2}CT^{-1}||^2 =$$
$$= TrT^{-1}C'A_d^{-1}CT^{-1} = TrT^{-1},$$

by using Proposition 4.3 (i). Write J as a partitioned matrix $J = (J_{fg}), f,g = 1,...,k$, where J_{fg} is an $d(f)r \times d(g)r$ matrix. For $A_d^{\pm 1/2} = diag(d^{\pm 1/2}(i)I_{d(i)r})$, we have

$$A_d^{-1/2}JA_d^{1/2} = (d^{-1/2}(f)d^{1/2}(g)J_{fg}), f,g = 1,...,k.$$

Hence

$$||A_d^{-1/2} J A_d^{1/2}||^2 = \sum_{f,g=1}^{k} d^{-1}(f) d(g) ||J_{fg}||^2. \quad (4.12)$$

Recall that (see [6], 1c.4(v)) the orthogonal projector $J = Pr_{[G]}$ is a symmetric and idempotent matrix with $M\{J\} = [G]$, and that for idempotent matrices (see [6], 1b.7) we have $\mathrm{Tr} J = \mathrm{Rank} J$. But $\mathrm{Rank} J = \dim M\{J\} = \dim[G]$. From the inclusion (4.8) we have

$$\dim[G] \leq \dim M\{C\} = \mathrm{Rank}\, C \leq lr$$

since C is of order (cr, lr), see notations (4.1). Therefore

$$||J_{fg}||^2 \leq ||J||^2 = TrJJ' = Tr(J^2) = TrJ \leq lr.$$

Thus the expresssion (4.12) is bounded by $lr \sum_{f,g=1}^{k} d(g) d^{-1}(f)$. Hence (4.11) follows. Q.E.D.

5. Strong consistency of GLSE

In this section, $(X(t), Y(t)), t = 1, 2, \ldots$, are i.i.d. copies of a pair (X, Y) of r.v'.s Y is $r \times 1$ vector-valued, whereas X takes values in an arbitrary measurable space (H, Δ). Disjoint sets $S(i) \in \Delta$ with $P(X \in S(i)) = p_i > 0, i = 1, \ldots, k$, are specified. From notations 2(iii), for each $\omega \in \Omega$ let us denote

$$d(n, \omega, i) = Card S(i) \cap \{X(1), \ldots, X(n)\}, \quad (5.1)$$
$$S = S(1) + \ldots + S(k).$$

Topical notations 2 (iv) will be used and when a(i) and the affix a are involved, they will be replaced by $d(n, \omega, i)$ and the affix d respectively, so we will have T_d, C_d, etc.; further, write

$$(X_i, Y_i) = (X_{i1}, Y_{i1}) \quad \text{in} \quad (2.7). \quad (5.2)$$

We will first describe the distribution features of the sequence $(X_{i1}, Y_{i1}), (X_{i2}, Y_{i2}), \ldots$

Proposition 5.1. *Let $A \in \Delta$ with $P(X \in A) > 0$. Let V be a vector r.v. such that $EI_A(X))$ exists. Then the equalities*

$$E(f(X)I_B(X)) = E(VI_B(X)) \quad \text{for every} \quad B \in \Delta|A \tag{5.3}$$

define a $\Delta|A$ - measurable function $f(.)$ on A, up to a P^X - equivalence. We willl call $f(x)$ a restricted on A conditional expectation of V given X at x. Further, for P-almost all $\omega \in \{X(\omega) \in A\}$ we have

$$f(X) = E(VI_A(X)|X) = E_{(X \in A)}(V|X), \tag{5.4}$$

where the last member denotes a conditional expectation given X relatively to the probability measure $P_{(X \in A)}$.

Proof. In the probability space $(\Omega, \mathfrak{J}, P)$ set $G = \{\omega : X(\omega) \in A\} \in \mathfrak{J}$. It suffices to consider some coordinate V_0 of the vector V, defined on the new probability space $(G, \mathfrak{J}|G, P_G)$, where $P_G = P/P(G)$ on $\mathfrak{J}|G$. The new expectation, denoted by $E_G V_0$, exists since $E(V_0 I_A(X)) = P(G) E_G V_0$. For $B \subset A$ we have

$$E(V_0 I_B(X)) = E(V_0 I_B(X) I_A(X)) = P(G) E_G(V_0 I_B(X)),$$

and, similarly,

$$E(f(X)I_B(X)) = P(G) E_G(f(X)I_B(X)).$$

Thus, for V_0 the equalities (5.3) become $E_G(f(X)I_B(X)) = E_G(V_0 I_B(X))$ for every $B \in \Delta|A$, and then, as known in conditioning theory, they define a $\Delta|A$-measurable function $f(x)$ on $A, f(x)$ is defined up to a P^X-equivalence and is the well-known conditional expectation $E_G(V_0|x)$ of V_0 given X at $x \in A$ relatively to the new probability space $(G, \mathfrak{J}|G, P_G)$. Therefore, for P-almost all $\omega \in \{X(\omega) \in A\} = G$ we have

$$f(X) = E_G(V_0|X) = E_{(X \in A)}(V_0|X). \tag{5.5}$$

On the other hand, rewrite (5.3) for V_0 as

$$E(f(X)I_A(X)I_B(X)) = E(V_0 I_A(X) I_B(X)) \quad \text{for every} \quad B \in \Delta.$$

Then, for P-almost all $\omega \in \{X(\omega) \in A\}$ we have

$$f(X) = f(X)I_A(X) = E(V_0 I_A(X)|X),$$

from which and (5.5) we get (5.4). Finally, when $E(V_0|X)$ exists, it will satisfy (5.3), hence in this case $f(x)$ coincides P^X-almost everywhere with $E(V_0|x)$ on A, and then $f(x)$ is indeed $E(V_0|x)$ restricted on A. We will call $f(x)$ a restricted on A conditional expectation of V_0 given x, whether $E(V_0|X)$ exists or not. Q.E.D.

PROPOSITION 5.2. *Let (X_{ij}, Y_{ij}) be defined by (2.7) from the sequence of i.i.d.r.v.'s $(X(t), Y(t))$. The following assertions hold*

(i) The family $\{(X_{ij}, Y_{ij}), j = 1, 2, ...; i = 1, ..., k\}$ is a.s. defined and is an independent family.

(ii) For each fixed (i, j) the P-distribution of (X_{ij}, Y_{ij}) coincides with the $P_{(X \in S(i))}$ - distribution of (X, Y).

(iii) If Y is vector-valued and $E(YI_S(X))$ exists, then setting $f(X) = E(YI_S(X)|X)$ a.s. we have

$$f(X_{ij}) = E(Y_{ij}|X_{ij}) \quad \text{a.s. for each fixed } (i, j).$$

Assertions (i) and (ii) are contained in Theorem 2 of [1]. We will give an independent proof.

PROOF. With notations 2(iv) and from (5.1), for each i we have

$$d(n, \omega, i) = Card\{j : j \geq 1, T_{ij} \leq n\}. \quad \text{Hence}$$
$$\{T_{ij} < \infty \quad \text{for all} \quad j = 1, 2, ...\} = \{d(n, \omega, i) \to \infty \quad \text{as} \quad n \to \infty\}.$$

In view of $P(X \in S(i)) > 0$ we have

$$P\{d(n, \omega, i) \to \infty \quad \text{as} \quad n \to \infty\} = 1 \quad \text{for each} \quad i.$$

Hence it follows that $P\{T_{ij} < \infty \text{ for all } (i, j)\} = 1.$

Thus, from (2.7) the family $\{(X_{ij}, Y_{ij})\}$ is a.s. defined.

Consider arbitrary positive integers a(i) and k disjoint sets in N

$$\{n_{i1}, ..., n_{ia(i)}\}, n_{i1} < n_{i2} < ..., i = 1, ..., k,$$

and their union M. Denote

$$A(M) = \{T_{ij} = n_{ij}, i = 1, ..., k; j = 1, ..., a(i)\},$$
$$A' = \{X(n_{ij}) \in S(i), i = 1, ..., k; j = 1, ..., a(i)\},$$
$$A'' = \bigcap_{i=1}^{k} \{X(t) \notin S(i) \quad \text{for all} \quad t \notin M, \quad \text{and together} \quad t < n_{ia(i)}\}.$$

Then $A(M) = A' \cap A'', P(A(M)) = P(A')P(A''))$. Further, denote $K(t) = (X(t), Y(t)), K_{ij} = (X_{ij}, Y_{ij}), K = (X, Y)$.

Then, for arbitrary measurable sets B_{ij} in the range space of K we have

$$\begin{aligned} &P_{A(M)}\{K_{ij} \in B_{ij}, i = 1, ..., k; j = 1, ..., a(i)\} = \\ &= P^{-1}(A')P^{-1}(A'')P\{A' \textstyle\bigcap (K(n_{ij})) \in B_{ij}, i \leq k, j \leq a(i))\}P(A'') \\ &= \prod_{i=1}^{k} \prod_{j=1}^{a(i)} P_{(X \in S(i))}(K \in B_{ij}) \end{aligned}$$

by noting that $K(n_{ij})$ are independent copies of K. Now, we have

$$1 = P\{T_{ij} < \infty, i = 1, ..., k; j = 1, ..., a(i)\} = \sum_{M} P(A(M)),$$

the sum extending only over the class of sets M with $P(A(M)) > 0$. Hence

$$\begin{aligned} &P(K_{ij} \in B_{ij}, i \leq k, j \leq a(i)) = \\ &= \sum_{M} P(A(M))P_{A(M)}(K_{ij} \in B_{ij}, i \leq k, j \leq a(i)) \\ &= \prod_{i=1}^{k} \prod_{j=1}^{a(i)} P_{(X \in S(i))}(K \in B_{ij}). \end{aligned}$$

This equality proves assertion (ii) and completes the the proof of assertion (i). To prove (iii), apply Proposition 5.1. Since $E(YI_S(X))$ exists and $S(i) \subset$

$S, E\{YI_{S(i)}(X)\}$ exists too and from (5.4) we have

$$E\{YI_{S(i)}(X)|X\} = E_{(X\in S(i))}(Y|X) \tag{5.6}$$

for P-almost all $\omega \in \{X(\omega) \in S(i)\}$. Set $f(X) = E(YI_S(X)|X)$. Since $YI_{S(i)}(X) = YI_S(X)I_{S(i)}(X), E\{YI_{S(i)}(X)|x\}$ coincides with $f(x)$ on $S(i)$, therefore from (5.6)

$$f(X) = E_{(X\in S(i))}(Y|X) \tag{5.7}$$

for P-almost all $\omega \in \{X(\omega) \in S(i)\}$. Now, from (ii) the P-distribution of (X_{ij}, Y_{ij}) coincides with the $P_{(X\in S(l))}$-distribution of (X, Y), hence (5.7) is equivalent to $f(X_{ij}) = E(Y_{ij}|X_{ij})$ a.s. Q.E.D.

The behaviour of T_d may be seen from the following four propositions

PROPOSITION 5.3. *Let u and v be random matrices of which v is n.d.d. Then $E(||u||^2 Trv)$ is finite if and only if $E(uu' \otimes v)$ exists and is finite.*

PROOF. For a random n.n.d. matrix F, it follows from Schwarz inequality that EF exists and is finite if and only if $ETrF$ is finite. From [3], A1, 49, $uu' \otimes v$ is n.n.d. and from Proposition 4.3

$$Tr(uu' \otimes v) = Truu'.Trv = ||u||^2 Trv,$$

hence our assertion follows. Q.E.D.

PROPOSITION 5.4. *Let b be a $t \times 1$ random vector and u an $r \times r$ p.d. random matrix. Assume that $E(bb' \otimes u)$ exists and is finite. Then $E(bb' \otimes u)$ is n.n.d, moreover it is p.d. if and only if the probability distribution of b is not concentrated in any proper subspace of R^t.*

PROOF. Let v be any $tr \times 1$ non-random vector. Equivalently, $v = vecV$, where V is an $r \times t$ non-random matrix. Applying Proposition 3.2 (iii) and (i) we have

$$\begin{aligned} & v'(bb' \otimes u)v = v'(bb' \otimes u)vecV = v'vec(uVbb') = \\ & = v'(b \otimes uV)vecb = v'(b \otimes uV)b = v'(b \otimes uVb) = \\ & = (b' \otimes b'V'u)v = (b' \otimes b'V'u)vecV = vec(b'V'uVb) = \\ & = (Vb)'uVb \end{aligned}$$

which is non-negative and which vanishes if and only if $Vb = 0$ since u is p.d. Hence, by taking expectation, $E(bb' \otimes u)$ is n.n.d.

Then, det $(Ebb' \otimes u) = 0$ is successively equivalent to

$$(\exists V \neq 0), E(b'V'uVb) = 0;$$
$$(\exists V \neq 0), Vb = 0 \quad a.s.;$$

(There exists a $1 \times t$ non-null and non-random row V_0), $V_0 b = 0$ a.s. Q.E.D.

PROPOSITION 5.5. *Assume that $X_{i1}, X_{i2}, \ldots$ are i.i.d. copies of the r.v. X_i. Let $b_i(X_i)$ and $z(X_i)$ be r.v.'s of which $b_i(.)$ is $l(i) \times 1$ vector-valued and $z(.)$ is p.d. matrix-valued. Then under the condition*

$$E\{||b_i(X_i)||^2 Trz(X_i)\} < \infty, \tag{5.8}$$

with topical notation 2 (iv), the matrix T_i tends a.s. to $E\{b_i(X_i)b_i'(X_i) \otimes z(X_i)\}$ as the non-random $a(i) \to +\infty$. Moreover, this limit L_i is p.d. if

The probability distribution of $b_i(X_i)$ is not concentrated in any proper subspace of $R^{l(i)}$. (5.9)

PROOF. From notations 2(iv), writing $b_{ij} = b_i(X_{ij}), z_{ij} = z(X_{ij})$ we have

$$(B_i' \otimes I_r)Z_i' = (\ldots(b_{ij} \otimes I_r)z_{ij}\ldots) = (\ldots b_{ij} \otimes z_{ij}\ldots) \tag{5.10}$$

by using Proposition 3.2 (i), the index j ranging $1, \ldots, a(i)$. Hence

$$a(i)T_i = (B_i' \otimes I_r)Z_i(B_i \otimes I_r)$$
$$= (\ldots b_{ij} \otimes I_r \ldots)(\ldots b_{ij} \otimes z_{ij} \ldots)'.$$

Thus

$$T_i = a^{-1}(i) \sum_{j=1}^{a(i)} b_i(X_{ij})b_i'(X_{ij}) \otimes z(X_{ij}).$$

From Proposition 5.3 condition (5.8) means $E\{b_i(X_i)b_i'(X_i) \otimes z(X_i)\}$ exists and is finite.

From Kolmogorov strong law of large numbers, as the non-random $a(i) \to +\infty, T_i$ tends a.s. to the asserted limit. The condition (5.9) ensuring that the limit is p.d. follows from Proposition 5.4. Q.E.D.

PROPOSITION 5.6. *Let $g(a(i),\omega)$ be some matrix function of ω and a non-random integer-valued variable $a(i)$*

(i) If $g(a(i),\omega) \to 0$ a.s. as $a(i) \to +\infty$, then $g(d(n,\omega,i),\omega) \to 0$ a.s. as $n \to +\infty$.

(ii) If $g(d(n,\omega,i),\omega) \to 0$ for some ω as $n \to +\infty$, then $g(a(i),\omega) \to 0$ as $a(i) \to +\infty$, providing $d(n,\omega,i) \to +\infty$ as $n \to +\infty$.

PROOF.

(i) Set $F = \{\omega : g(a(i),\omega)$ does not tend to zero as $a(i) \to +\infty\}$,
$G = \{\omega : g(d(n,\omega,i),\omega)$ does not tend to zero as $n \to +\infty\}$,
$\Omega' = \{\omega : d(n,\omega,i) \to +\infty$ as $n \to +\infty\}$.

Consider some $\omega \in G \cap \Omega'$. Then

$$(\exists \epsilon > 0), (\exists \{n_m(\omega)\} \to +\infty), ||g(d(n_m,\omega,i),\omega)|| \geq \epsilon, m = 1,2,...$$

Since $\omega \in \Omega', d(n_m,\omega,i) \to +\infty$ as $m \to +\infty$. For the considered ω put $a_m = d(n_m,\omega,i)$, then there holds

$$(\exists \epsilon > 0), (\exists \{a_m\} \to +\infty), ||g(a_m,\omega)|| \geq \epsilon, m = 1,2,...$$

This means the considered $\omega \in F$, i.e. $G \cap \Omega' \subset F$. Hence $G \subset F \cup (\Omega - \Omega')$. But $P(\Omega - \Omega') = 0$ since $P(X \in S(i)) > 0$, considering (5.1). Thus (i) follows.

(ii) For the considered ω there holds

$$(\forall \epsilon > 0), (\exists n_0(\omega)), (\forall n \geq n_0), ||g(d(n,\omega,i),\omega|| \leq \epsilon.$$

From (5.1), as n increases $d(n,\omega,i)$ is non-decreasing, hence $d(n,\omega,i) \geq d(n_0,\omega,i)$ for $n \geq n_0(\omega)$; moreover, whenever $d(n,\omega,i)$ increases, it increases by the unit. But, from assumption, $\omega \in \Omega'$ hence as n ranges $n_0(\omega), n_0(\omega) +$

$1, ..., d(n, \omega, i)$ ranges $d(n_0, \omega, i), d(n_0, \omega, i) + 1, ...$ ad infinitum. Therefore by putting $a_0(i) = d(n_0, \omega, i)$ for the considered ω there holds

$$(\forall \epsilon > 0), (\exists a_0(i)), (\forall a(i) \geq a_0(i)), ||g(a(i), \omega)|| \leq \epsilon,$$

which means $g(a(i), \omega) \to 0$ as $a(i) \to +\infty$. Q.E.D.

Next, we will give a result on the existence of GLSE generalizing Theorem 4 in [1], p.27.

THEOREM 5.1. *Consider the condition*

$$\begin{cases} \text{The } P_{(X \in S(i))} - \text{distribution of the assumed } r.v. b_i(X) \text{ is not} \\ \text{concentrated in any proper subspace of } R^{l(i)}, \ i = 1, ..., k. \end{cases} \tag{5.11}$$

and the event { the GLSE $\hat{q}$ exists as soon as the sample size n is sufficiently large}. Formally, this event is

$$\Omega_0 = \{\omega : \exists n_0(\omega), \forall n \geq n_0, \text{ the GLSE } \hat{q} \text{ exists}\}.$$

Then $P\Omega_0 = 1$ or 0 according to (5.11) which is satisfied or not, respectively.

PROOF. By implicitly assuming that $b_i(.)$ is $\Delta|S(i)$-measurable, from Proposition 5.2 (i) and (ii), $b_i(X_{i1}), b_i(X_{i2}), ...$ are i.i.d copies of $b_i(X_i)$, see (5.2), and the P-distribution of $b_i(X_i)$, which coincides with the $P_{(X \in S(i))}$-distribution of $b_i(X)$, is not concentrated in any proper subspace of $R^{l(i)}$ when in the first part we suppose that condition (5.11) is satisfied. According to notations 2 (iv), from Corollary 2 in [1], p.27, there holds

$$P(\text{Rank } B_i' = l(i)) \uparrow 1 \quad \text{as } a(i) \text{ non-random } \uparrow +\infty.$$

Since $B_a = diag(B_i)$ we have Rank $B_a = \sum_{i=1}^{k}$ Rank B_i. Hence, with $l = l(1) + ... + l(k)$, there holds

$$P(Rank\ B_a = l) = P\{\bigcap_1^k (Rank\ B_i = l(i))\} = \prod_1^k P(Rank\ B_i = l(i)) \uparrow 1$$

as the non-random $a = (a(1), ..., a(k) \uparrow \infty$, the $B_i's$ being independent in view of Proposition 5.2 (i).

Therefore

$$(\forall \epsilon > 0), (\exists (a_1, ..., a_k)).(\forall a > (a_1, ..., a_k)), \qquad (5.12)$$
$$P(Rank\ B_a = l) > 1 - \epsilon.$$

Replace $a(i)$ by $d(n, \omega, i)$, then denoting

$$\Omega(a, n) = \{\omega : d(n, \omega, i) = a(i), i = 1, ..., k\},$$

we have

$$P(Rank\ B_d = l) \geq P\{\sum_{a>(a_1,...,a_k)} \Omega(a, n) \cap (Rank\ B_d = l)\} \geq$$

$$\geq \sum_{a>(a_i,...,a_k)} P(\Omega(a, n)).P_{\Omega(a,n)}(Rank\ B_a = l).$$

But, on account of Theorem 1 (i) and (ii) in [1], p.8. the $P_{\Omega(a,n)}$-distribution of the family $\{X_{ij}, i = 1, ..., k, j = 1, ..., a(i)\}$ coincides with the P-distribution of the same family according to Proposition 5.2 (i) and (ii), therefore B_a being function of this family, we have

$$P_{\Omega(a,n)}(Rank\ B_a = l) = P(Rank\ B_a = l)$$

and from (5.12)

$$P(Rank\ B_d = l) \geq (1 - \epsilon) \sum_{a>(a_1,...,a_k)} P(\Omega(a, n) \geq$$

$$\geq (1 - \epsilon)P\{d(n, \omega, i) > a_i, i = 1, ..., k\}$$

By letting $n \to +\infty$ and passing to the limit we have

$$\liminf P(Rank\ B_d = l) \geq 1 - \epsilon$$

for every $\epsilon > 0$, hence $\lim_{n\to\infty} P(Rank\ B_d = l) = 1$. Now consider

$$\Omega_0 = \{\omega : \exists m(\omega), \forall n \geq m, Rank\ B_d = l\} =$$
$$= \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \{Rank\ B_d = l\}. \qquad (5.13)$$

Put $A_m = \bigcap_{n=m}^{\infty}\{Rank\ B_d = l\}$, then the sequence $\{A_m\}$ is non-decreasing. Hence $P(\Omega_0) = \lim_{m\to\infty} P(A_m)$. Further, since $Rank\ B_d = \sum_1^k Rank\ B_i$ is non-decreasing as $(a(1),...,a(k))$ increases and since $d(n,\omega,i)$'s are non-decreasing as n increases, it follows that $Rank\ B_d$ is non-decreasing as n increases. We thus have

$$A_m = \bigcap_{n=m}^{\infty}(Rank\ B_d = l) = \{Rank\ B_d = l \text{ for } n = m\}.$$

Hence $\lim_{m\to\infty} PA_m = 1$ and $P\Omega_0 = 1$. Since $Rank\ B_a = l$ or, equivalently $Rank\ B_i = l(i)$ for each i, entails $a(i) \geq l(i) > 0$, the event $\{Rank\ B_d = 1\}$ entails all $d(n,\omega,i)$ are positive and, from Proposition 3.5, is equivalent to the existence of GLSE. Thus the first part is proved. Now suppose (5.11) is not satisfied. Then from Corollary 2 (ii) in [1], p.27, there is some index i such that the event $\{Rank\ B_i = l(i)\}$ is a null one for every non-random $a(i) > 0$. Hence $(Rank\ B_a = l)$ is a null event for every non-random $a > 0$. Since

$$(Rank\ B_d = l) = \sum_{a\geq(l(1),...,l(k))} \Omega(a,n) \cap (Rank\ B_a = l)$$

$(Rank\ B_d = l)$ is a null event, hence so is

$$\Omega_0 = \bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty}(Rank\ B_d = l). \quad Q.E.D.$$

REMARK 5.1. If the vector $b_i(X_i)$ lies a.s. in some $m(i)$-dimensional proper subspace of $R^{l(i)}$ with a basis $\{u_1,...,u_{m(i)}\}$, then denoting by u the $l(i)\times m(i)$ matrix $(u_1...u_{m(i)})$, we have

$$b_i(X_i) = uc_i(X_i)\ a.s.,$$

where $c_i(.)$ is an $m(i)\times 1$ vector function. Hence we have the better parametrization

$$b_i'(X_i)q_i = c_i'(X_i)u'q_i = c_i'(X_i)f_i,$$

where the new matrix parameter $f_i = u'q_i$ is only of order $m(i)\times r$, with $m(i) < l(i)$. Thus Assumption (5.11) is natural for a rational parametrization.

The following theorem gives an upper evaluation of the GLSE error norm in comparison with residuals. We first note that for $X_i = X_{i1}$, in view of Proposition 5.2 (ii), in Proposition 5.5 the condition

$$E||b_i(X_i)||^2 Trz(X_i) < \infty, i = 1, ..., k,$$

is equivalent to

$$E_{(X \in S(i))}\{||b_i(X)||^2 Trz(X)\} < \infty, i = 1, ..., k. \tag{5.14}$$

THEOREM 5.2. *Together with Assumption (5.11) also assume (5.14). Then there exists a finite positive constant R such that a.s., as soon as the sample size is sufficiently large, for each subspace L involved in Definition 3.1 the corresponding GLSE $\hat{q}$ exists and*

$$\sup_L ||\hat{q} - q|| \leq R||g||,$$

where $g = C'_d A_d^{-1} Z_d^{1/2}[e]$ (see (4.1)) and $\sup_L$ is taken over the set of subspaces L involved.

PROOF. Under the terminological phrase "a.s. as soon as the sample size is sufficiently large" we always mean the following logical proposition

$$(\exists F, PF = 1)(\forall \omega \in F)(\exists n_0(\omega))(\forall n \geq n_0).$$

From Proposition 2.1 and 3.3, the equality $Rank\ B_d = l$ is equivalent to the positive definiteness of T_d. Hence, from (5.13) in the proof of Theorem 5.1, a.s. as soon as n is large T_d is p.d. From Assumptions (5.11) and (5.14) and Propositions 5.5 and 5.6 it follows that T_d tends a.s. to the p.d. limit $L_0 = diag(L_i, i = 1, ..., k)$ as $n \to +\infty$, where L_i is the a.s. limit of T_i in Proposition 5.5. Therefore $Tr(T_d^{-1})$ tends a.s. to $Tr(L_0^{-1})$ as $n \to +\infty$. On the other hand, in Proposition 4.4, with $d(i) = d(n, \omega, i), p_i = P(X \in S(i))$, the ratio $d(g)d^{-1}(f)$ tends a.s. to $p_g p_t^{-1}$ as $n \to +\infty$. Now the right-hand side of (4.11), Proposition 4.4, which will be written briefly as $K^2||g||^2$, is quite independent of the subspace L involved in Definition 3.1 besides Proposition

4.4 is valid as soon as $Rank\ B_d = l$ considering Proposition 3.5, hence by using the a.s. event Ω_0 given by (5.13) we have successively

$$(\forall\omega \in \Omega_0), (\exists m(\omega)), (\forall n \geq m(\omega)), Rank\ B_d = l,$$
$$(\forall\omega \in \Omega_0), (\exists m(\omega)), (\forall n \geq m(\omega)), ||\hat{q} - q|| \leq K||g||,$$

the latter is also valid for $\sup_L ||\hat{q} - q||$ since by the former $m(\omega)$ is independent of L. From above K^2 tends a.s. to a finite positive constant as $n \to +\infty$, besides K^2 being independent of L hence we can manage so that

$$(\exists F \subset \Omega_0, PF = 1), (\forall\omega \in F), (\exists m(\omega)), (\forall n \geq m(\omega)), K \leq R,$$

where R is a suitable finite positive constant. Therefore,

$$(\exists F, PF = 1), (\forall\omega \in F), (\exists m(\omega)),$$
$$(\forall n \geq m(\omega)), \sup_L ||\hat{q} - q|| \leq R||g||. \text{ Q.E.D.}$$

We now state the strong conssistency of GLSE.

THEOREM 5.3. *Let, in the model (1.1), $(X(t), Y(t), t = 1, 2, ...)$ be i.i.d. copies of a pair (X, Y). Let $E\{YI_S(X)\}$ exist, and assume*

$$E\{Y'I_S(X)|X\} = \sum_{i=1}^{k} b_i'(X)q_iI_{S(i)}(X) \text{ a.s.} \tag{5.15}$$

Then under the conditions (5.11) and (5.14) $\sup_L ||\hat{q} - q|| \to 0$ a.s. as $n \to +\infty$ if and only if

$$E_{(X\in S(i))}\{b_i(X) \otimes z(X)Y\} \text{ exists, } i = 1, ..., k, \tag{5.16}$$

$\sup_L$ being taken over the set of all subspaces L involved in Definition 3.1 of GLSE $\hat{q}$.

PROOF. With notations 2 (iv), consider

$$C_a' A_a^{-1} Z_a^{1/2} = diag\{a^{-1}(i)(B_i' \otimes I_r)Z_i\}.$$

By denoting

$$e_i = U_i - B_i q_i, e_a = U_a - B_a q = (e_1' ... e_k')',$$
$$b_{ij} = b_i(X_{ij}), z_{ij} = z(X_{ij}),$$

there hold

$$[e_i]' = (...Y_{ij}' - b_{ij}' q_i ...), [e_a]' = ([e_1]' ... [e_k]'),$$
$$C_a' A_a^{-1} Z_a^{1/2} [e_a] = (... a^{-1}(i) [e_i]' Z_i (B_i \otimes I_r) ...)',$$

and on account of (5.10)

$$a^{-1}(i)[e_i]' Z_i (B_i \otimes I_r) = a^{-1}(i) \sum_{j=1}^{a(i)} (Y_{ij}' - b_{ij}' q_i)(b_{ij}' \otimes z_{ij}).$$

From Proposition 5.2 (i) and (ii), $(X_{i1}, Y_{i1}), ...$ are P-i.i.d. hence, by Kolmogorov strong law of large numbers,

$$\text{the existence and vanishing of } E(b_i(X_i) \otimes z(X_i))(Y_i - q_i' b_i(X_i)), i = 1, ..., k, \tag{5.17}$$

are necessary and sufficient for $C_a' A_a^{-1} Z_a^{1/2} [e_a]$ to tend a.s. to the null vector as $a = (a(1), ..., a(k))$ increases indefinitely. Replacing $a(i)$ by $d(n, \omega, i)$ and applying Proposition 5.6 we see that, with notations (4.1), Claim (5.17) is equivalent to asserting $g \to 0$ a.s. as $n \to +\infty$. On the other hand, using Proposition 3.2 (i) and (ii) we have

$$\begin{aligned} &(b_i(X_i) \otimes z(X_i)) Y_i = b_i(X_i) \otimes z(X_i) Y_i, \\ &(b_i(X_i) \otimes z(X_i)) q_i' b_i(X_i) = b_i \otimes z q_i' b_i = (b_i \otimes z q_i' b_i)[I_1] = \\ &= [b_i b_i' q_i z] = (b_i(X_i) b_i'(X_i) \otimes z(X_i))[q_i]. \end{aligned} \tag{5.18}$$

From Proposition 5.3 Assumption (5.14) means the existence and finiteness of $E\{b_i(X_i) b_i'(X_i) \otimes z(X_i)\}$ for all i, hence it entails

$$E\{(b_i(X_i) \otimes z(X_i)) q_i' b_i(X_i)\} \text{ exists and is finite, } i = 1, ..., k, \tag{5.19}$$

since q_i's are non-random. Further, the conditional structure (5.15) is equivalent to

$$E(Y_i'|X_i) = b_i'(X_i)q_i \text{ a.s., } i = 1, ..., k; \tag{5.20}$$

indeed (5.15) entails (5.20) by Proposition 5.2 (iii), conversely the P-distribution of (X_i, Y_i) coincides with the $P_{(X \in S(i))}$-distribution of (X, Y) and by (5.4), Proposition 5.1, $E_{(X \in S(i))}(Y|X) = E\{YI_{S(i)}(X)|X\}P_{(X \in S(i))}$-a.s., hence (5.20) is equivalent to

$$E\{Y'I_{S(i)}(X)|X\} = b_i'(X)q_iI_{S(i)}(X) \text{ a.s. } i = 1, ..., k,$$

then (5.20) entails (5.15) since $Y'I_S(X) = \sum_1^k Y'I_{S(i)}(X)$.

Let us now prove our assertion. First, condition (5.16) is equivalent to

$$E\{(b_i(X_i) \otimes z(X_i))Y_i\} \text{ exists, } i = 1, ..., k, \tag{5.21}$$

considering (5.18). Then the structure (5.15) and condition (5.16), equivalently, the structure (5.20) and condition (5.21) entail

$$\begin{aligned} E(b_i(X_i) \otimes z(X_i))Y_i =& E\{(b_i(X_i) \otimes z(X_i))E(Y_i|X_i)\} \\ =& E\{(b_i \otimes z(X_i))q_i'b_i\}. \end{aligned}$$

From (5.19) the last member, hence the first are finite. Hence Claim (5.17) is fulfilled. Thus $g \to 0$ a.s. as $n \to +\infty$. By Theorem 5.2, $\sup_L ||\hat{q} - q|| \to 0$ a.s. as $n \to +\infty$. Let us prove the converse. From Proposition 4.1, in the extreme case $L = \mathrm{Im}B_d$, GLS values are defined by the equation $C_d[\hat{q} - q] = C_d b$, therefore $C_d'A_d^{-1}C_d[\hat{q} - q] = C_d'A_d^{-1}C_d b$ i.e. by (4.3) $T_d[\hat{q} - q] = g$. Hence, from Proposition 4.3

$$||g||^2 \leq ||\hat{q} - q||^2 Tr(T_d^2) \leq ||\hat{q} - q||^2 (Tr\ T_d)^2, \tag{5.22}$$

see [3], A1.70. Under Assumption (5.14), from Propositions 5.5 and 5.6, T_d tends a.s. to a finite limit as $n \to +\infty$, hence so is TrT_d. By assumption $\hat{q}$ is strongly consistent, hence, from (5.22), $g \to 0$ as $n \to +\infty$, which from above is equivalent to Claim (5.17). On account of (5.19) it follows that $E(b_i(X_i) \otimes z(X_i))Y_i$ exists or, equivalently, condition (5.16) is satisfied. Q.E.D.

REMARK 5.2. From the proof, the above main theorem can be stated more concisely and formally as follows.

The GLSE, defined according to model (1.1) and Definition 3.1 on the basis of i.i.d. observations on a pair (X, Y) satisfying Assumption (5.11) and (5.14) is strongly consistent uniformly in L on the set of all subspaces L involved if and only if the representation

$$b_i(X) \otimes z(X)Y = (b_i(X)b_i'(X) \otimes z(X))[q_i] + r_i,$$
$$E_{(X \in S(i))} r_i = 0, i = 1, ..., k \qquad (5.23)$$

holds.

Especially, under the polygonal conditional mean structure (5.15), the representation (5.23) is equivalent to the existence of $E_{(X \in S(i))}\{b_i(X) \otimes z(X)Y\}, i = 1, ..., k$.

REFERENCES

[1] N. Bac-Van, *Estimation and prediction in regression models with random explanatory variables*, Dissertationes mathematicae, CCCXVII, Warszawa 1992, 42pp.
[2] J. Coursol, *Technique statistique des modèles linéaires, (1) Aspects théoriques*, Les course du C.I.M.P.A., série statistique, 1980, 277pp.
[3] K.M.S. Humak, *Statistische Methoden der Modellbildung*, Band I, Akademie-Verlag, Berlin 1977.
[4] T.L. Lai, H. Robbins, C.Z. Wei, *Strong consistency of least squares estimates in multiple regression II*, J. Multivariate Analysis **9** (1979), 343–361.
[5] Y.P. Mack and H.-G Müller, *Derivative estimation in nonparametric regression with random predictor variable*, Sankhyà Ser. A **51** (1989), 59-72.
[6] C.R. Rao, *Linear statistical inference and its applications*, 2nd ed., Wiley, New York, 1973.

DEPARTMENT OF MATHEMATICS
HO CHI MINH CITY UNIVERSITY
227 NGUYEN VAN CU Q5
HO CHI MINH CITY, VIETNAM