Kybernetika

Michal Pešta

Asymptotics for weakly dependent errors-in-variables

Kybernetika, Vol. 49 (2013), No. 5, 692--704

Persistent URL: `http://dml.cz/dmlcz/143519`

KYBERNETIKA — VOLUME *49* (2013), NUMBER 5, PAGES 692–704

ASYMPTOTICS FOR WEAKLY DEPENDENT ERRORS–IN–VARIABLES

Michal Pešta

Linear relations, containing measurement errors in input and output data, are taken into account in this paper. Parameters of these so-called *errors-in-variables* (EIV) models can be estimated by minimizing the *total least squares* (TLS) of the input-output disturbances. Such an estimate is highly non-linear. Moreover in some realistic situations, the errors cannot be considered as independent by nature. *Weakly dependent* (α- and φ-mixing) disturbances, which are not necessarily stationary nor identically distributed, are considered in the EIV model. Asymptotic normality of the TLS estimate is proved under some reasonable stochastic assumptions on the errors. Derived asymptotic properties provide necessary basis for the validity of block-bootstrap procedures.

Keywords: errors-in-variables (EIV), dependent errors, total least squares (TLS), asymptotic normality

Classification: 15A51, 15A52, 62E20, 65F15, 62J99

1. INTRODUCTION

The main goal is to establish results concerning asymptotic normality in linear relations, where weakly dependent measurement errors or disturbances in input and output data (errors-in-variables model) occur simultaneously.

1.1. Errors-in-variables model

Errors-in-variables (EIV) model

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times p}{\mathbf{Z}}\,\underset{p\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\varepsilon}} \quad \text{and} \quad \underset{n\times p}{\mathbf{X}} = \underset{n\times p}{\mathbf{Z}} + \underset{n\times p}{\boldsymbol{\Theta}} \tag{E}$$

is considered, where $\boldsymbol{\beta}$ is a vector of *regression parameters* to be estimated, $\mathbf{X}$ and $\mathbf{Y}$ consist of *observable random* variables ($\mathbf{X}$ are covariates and $\mathbf{Y}$ is a response), $\mathbf{Z}$ consists of *unknown constants* and has full rank, and $\boldsymbol{\varepsilon}$ and $\boldsymbol{\Theta}$ are *random errors*.

1.2. Weak dependence

In this paper, the EIV model is not restricted to independent observations due to the fact that in some situations the disturbances cannot be considered as independent by nature.

It is assumed that $\{\xi_n\}_{n=1}^{\infty}$ is a sequence of random elements on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For sub-$\sigma$-fields $\mathcal{A}, \mathcal{B} \subseteq \mathcal{F}$, we define

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|,$$
$$\varphi(\mathcal{A}, \mathcal{B}) := \sup_{A \in \mathcal{A}, B \in \mathcal{B}, \mathbb{P}(A)>0} |\mathbb{P}(B|A) - \mathbb{P}(B)|.$$

Intuitively, α and φ measure the dependence of the events in $\mathcal{B}$ on those in $\mathcal{A}$. There are many ways how to describe weak dependence or, in other words, *asymptotic independence* of random variables [3]. We concentrate on two approaches. Considering a filtration $\mathcal{F}_m^n := \sigma(\xi_i \in \mathcal{F}, m \leq i \leq n)$, sequence $\{\xi_n\}_{n=1}^{\infty}$ of random elements (e. g., variables) is said to be *strong mixing* (α-mixing) if

$$\alpha(n) := \sup_{k \in \mathbb{N}} \alpha(\mathcal{F}_1^k, \mathcal{F}_{k+n}^{\infty}) \to 0, \quad n \to \infty; \tag{1}$$

moreover, it is said to be *uniformly strong mixing* (φ-mixing) if

$$\varphi(n) := \sup_{k \in \mathbb{N}} \varphi(\mathcal{F}_1^k, \mathcal{F}_{k+n}^{\infty}) \to 0, \quad n \to \infty. \tag{2}$$

Uniformly strong mixing—introduced by Rosenblatt—implies strong mixing [11], which was proposed by Ibragimov. Coefficients of dependence $\alpha(n)$ and $\varphi(n)$ measure how much dependence exists between events separated by at least n observations or time periods.

Anderson [1] comprehensively analyzed a class of m-dependent processes. They are φ-mixing, if they are finite order ARMA processes with innovations satisfying *Doeblin's condition* [2, p. 168]. Finite order processes, which do not satisfy Doeblin's condition, can be shown to be α-mixing [10, pp. 312–313]. [14] provides general conditions under which stationary Markov processes are α-mixing. Since functions of mixing processes are themselves mixing [3], time-varying functions of any of the processes just mentioned are mixing as well.

It is obvious that $\alpha(\mathcal{A}, \mathcal{B}) = \alpha(\mathcal{B}, \mathcal{A})$ for arbitrary sub-σ-fields $\mathcal{A}, \mathcal{B} \subseteq \mathcal{F}$. This symmetry does not hold for the φ-dependence. Indeed, [14, pp. 213–214] constructed some strictly stationary Markov chains that are φ-mixing but not "time-reversed" φ-mixing. Therefore, it is not possible to "interchange" the past with the future regarding the definition of the φ-mixing coefficient.

1.3. Error structure of the EIV model

We think of all the vectors as columns. A "row-column" notation for a matrix $\mathbf{A}$ is used in this manner: $\mathbf{A}_{i,\bullet}$ denotes the ith row of matrix $\mathbf{A}$ and $\mathbf{A}_{\bullet,j}$ corresponds to the jth column of matrix $\mathbf{A}$.

Let us consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where all the further mentioned random elements exist in. Proper distributional assumptions of random errors in the EIV model need to be proposed. Two levels of the error structure have to be distinguished. The first level of error structure—*within-individual level*—is that each row $[\mathbf{\Theta}_{i,\bullet}, \varepsilon_i]$ is absolutely continuous with respect to the Lebesgue measure having *zero mean* and non-singular covariance matrix $\sigma^2\mathbf{I}$, where $\sigma^2 > 0$ is unknown (for simplicity). This assumption can be straightforwardly generalized as discussed in, e. g., [12] or [13]. Relationships between individual observations are represented by the second level of error structure—*between-individual level.* Here, the rows $[\boldsymbol{\Theta}_{i,\bullet}, \varepsilon_i]$ form *weakly dependent* sequences with *zero mean*, i. e., zero mean α- or φ-mixing. The reason for this can come from the fact that the measurements, which are "close to each other", influence themselves somehow. Moreover, the influence decreases as the distance between observations increases.

Concerning the *within-individual level*, the mixing sequences of errors are assumed to be pairwise independent. The necessity and possible weakening of this assumption was discussed by [12].

It has to be emphasized that any form of errors' *stationarity* is not needed to assume. Omitting this, sometimes restrictive, assumption strengthen our results.

Additional *design assumption* is necessary for asymptotics even in the case of independent errors:

$$\boldsymbol{\Delta} := \lim_{n\to\infty} n^{-1}\mathbf{Z}^\top\mathbf{Z} \quad \text{exists and is positive definite.} \tag{D}$$

Importance of the previous design assumption has already been thoroughly discussed in [13].

1.4. Total least squares estimate

Sometimes, full-information approaches like *maximum likelihood* (ML) can provide parameter estimates for the previously mentioned model [8]. Nevertheless, it is requisite to elaborate *distributional-free estimation* method as well, e. g., total least squares introduced by [7].

TLS estimate $\widehat{\boldsymbol{\beta}}$ of an unknown parameter $\boldsymbol{\beta}$ is defined as any vector $\boldsymbol{\beta}$ satisfying (E) such that the *Frobenius norm* of error matrix $\|[\boldsymbol{\Theta}, \boldsymbol{\varepsilon}]\|$ is minimal. Geometrically speaking, the Frobenius norm tries to minimize the *orthogonal* distance between the observations and fitted hyperplane. For a matrix $\mathbf{A} \equiv (A_{ij})_{i,j=1}^{n,m}$, it is defined as

$$\|\mathbf{A}\| \equiv \|\mathbf{A}\|_F := \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{m} A_{ij}^2} = \sqrt{tr(\mathbf{A}^\top\mathbf{A})} = \sqrt{\sum_{i=1}^{\min\{n,m\}} \sigma_i^2} = \sqrt{\sum_{i=1}^{r} \sigma_i^2},$$

where r is the rank of matrix $\mathbf{A}$ and σ_is are its singular values. The Frobenius norm can be viewed as a *multivariate* version of the Euclidean vector norm for matrices.

Let λ be the $(p+1)$-st largest eigenvalue of a matrix $[\mathbf{X}, \mathbf{Y}]^\top[\mathbf{X}, \mathbf{Y}]$ and let $\mathbf{v}_{p+1}$ be the associated eigenvector. [7] showed that if $v_{p+1,p+1} \neq 0$, then the TLS estimate has form

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X} - \lambda\mathbf{I})^{-1}\mathbf{X}^\top\mathbf{Y}.$$

Gleser [6] proved that with *probability tending to one*, as n increases, $v_{p+1,p+1} \neq 0$ and, hence, $\widehat{\boldsymbol{\beta}}$ exists. In the sequel, it has to be kept in mind that λ depends on n.

Finally, remind that the ML estimate of $\boldsymbol{\beta}$ [8] *coincides* with the TLS estimate if the rows of the error matrix are *i.i.d. multivariate normal* with zero mean and non-singular covariance matrix.

1.5. Consistency of the TLS estimate

Strong consistency of the TLS estimate for independent errors is proved by [6] and, moreover, *weak consistency*—again for independent errors, but with less restrictive assumptions—is widely discussed in [4]. When a premise of independence cannot be assumed, strong consistency of the TLS estimate under weak dependence of errors was explored by [12].

Similar situation occurs to the TLS estimate's *asymptotic normality*, which was proved by [5] for the case of independent errors. This result is going to be extended for the weakly dependent errors now.

2. ASYMPTOTIC NORMALITY

Firstly, central limit theorems for weakly dependent variables, that are going to be used in derivation of the TLS estimate's asymptotic normality, are needed to be stated. These *CLTs cannot assume stationarity*, because they will not be applied directly on the (possibly stationary) errors of the EIV model, but on their transformations, which cannot be generally considered as stationary ones.

Let us define $S_n := \sum_{i=1}^{n} \xi_i$ and $\varsigma_n^2 := \mathbb{V}\mathrm{ar}\, S_n$.

All proofs of the following results are in Appendix.

Lemma 2.1 (Central limit theorem for α-mixing)**.** Let $\{\xi_n\}_{n=1}^{\infty}$ be a sequence of zero mean α-mixing random variables with

$$\sup_{n\in\mathbb{N}} \mathbb{E}|\xi_n|^{2+\omega} < \infty \tag{3}$$

and

$$\sum_{n=1}^{\infty} \alpha(n)^{\omega/(2+\omega)} < \infty \tag{4}$$

for some $\omega > 0$. Suppose that

$$\frac{\mathbb{E}S_n^2}{n} \to \varsigma^2 > 0, \quad n \to \infty \tag{5}$$

is satisfied. Then

$$\frac{S_n}{\varsigma_n} \xrightarrow{\mathscr{D}} \mathcal{N}(0,1), \quad n \to \infty.$$

Lemma 2.2 (Central limit theorem for φ-mixing)**.** Let $\{\xi_n\}_{n=1}^{\infty}$ be a sequence of zero mean φ-mixing random variables such that

$$\sup_{n\in\mathbb{N}} \mathbb{E}|\xi_n|^{2+\omega} < \infty \tag{6}$$

for some $\omega > 0$ and

$$\frac{\mathbb{E}S_n^2}{n} \to \varsigma^2 > 0, \quad n \to \infty. \tag{7}$$

Then

$$\frac{S_n}{\varsigma_n} \xrightarrow{\mathscr{D}} \mathcal{N}(0,1), \quad n \to \infty.$$

For a given sequence $\xi_\circ \equiv \{\xi_n\}_{n=1}^{\infty}$ of random elements, the dependence coefficients $\alpha(n)$ is denoted $\alpha(\xi_\circ, n)$. Analogous notation is used for φ-mixing sequences.

The first main result of this section is the asymptotic normality for the TLS estimate, where the errors are α-mixing.

Theorem 2.3 (Asymptotic normality in α-mixing EIV)**.** Let the EIV model hold and assumption (D) be satisfied. Suppose $\{\Theta_{n,1}\}_{n=1}^{\infty}, \ldots, \{\Theta_{n,p}\}_{n=1}^{\infty}$, and $\{\varepsilon_n\}_{n=1}^{\infty}$ are pairwise independent sequences of α-mixing random variables having

$$\alpha(\Theta_{\circ,j}, n) = \mathcal{O}(n^{-1-\delta_j}),\ j = 1, \ldots, p \quad \text{and} \quad \alpha(\varepsilon_\circ, n) = \mathcal{O}(n^{-1-\delta_{p+1}}), \tag{8}$$

as $n \to \infty$ for some $\delta_j > 0$, $j \in \{1, \ldots, p+1\}$. Moreover, assume that

$$\sup_{n\in\mathbb{N}} Z_{n,j}^2 < \infty,\ j \in \{1, \ldots, p\}, \tag{9}$$

$$\sup_{n\in\mathbb{N}} \mathbb{E}|\Theta_{n,j}|^{4+\omega_j} < \infty,\ j \in \{1, \ldots, p\}, \quad \text{and} \quad \sup_{n\in\mathbb{N}} \mathbb{E}|\varepsilon_n|^{4+\omega_{p+1}} < \infty \tag{10}$$

for some $\omega_j > 0$, $j \in \{1, \ldots, p+1\}$ such that

$$\frac{2}{\min_{j=1,\ldots,p+1} \omega_j} < \min_{j=1,\ldots,p+1} \delta_j. \tag{11}$$

If there exists a positive definite matrix $\daleth$ such that

$$n^{-1}\mathbb{V}\mathrm{ar}\,[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}] \begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix} \to \daleth > \mathbf{0}, \quad n \to \infty; \tag{12}$$

then $\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{\mathscr{D}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, $n \to \infty$, where $\boldsymbol{\Sigma}$ is a positive definite matrix depending on $\boldsymbol{\beta}$, $\boldsymbol{\Delta}$, and $\daleth$.

The covariance matrix $\boldsymbol{\Sigma}$ of the normal distribution is indeed omitted. The reason is that even in the simpler case of independent errors, three main pitfalls exist [13]: the covariance of the limiting multivariate normal distribution depends on the unknown parameter $\boldsymbol{\beta}$ and on the unknown matrix $\boldsymbol{\Delta}$, and without the assumption on the third and fourth moments of the rows of $[\boldsymbol{\Theta}, \boldsymbol{\varepsilon}]$, the covariance matrix has a very complicated form. A partial solution to the first two mentioned issues could be plugging consistent estimates instead of the unknown entities. On the contrary, the third issue seems to be a big problem whatsoever, because the third and the fourth errors' moments cannot be estimated from the data.

Therefore, it is a waste of effort to theoretically derive the covariance, because such a formula—which could be unobtainable—is already computationally useless. It is sufficient to know that the TLS estimate has a non-degenerate asymptotic normal distribution with some (unknown) positive definite covariance matrix. Nevertheless, the practical calculation of the asymptotic distribution will be proceeded by the resampling methods anyway.

The second main result of this section is the asymptotic normality of the TLS estimate, where the errors are φ-mixing.

Theorem 2.4 (Asymptotic normality in φ-mixing EIV)**.** Let the EIV model hold and assumption (D) be satisfied. Suppose $\{\Theta_{n,1}\}_{n=1}^{\infty},\ldots,\{\Theta_{n,p}\}_{n=1}^{\infty}$, and $\{\varepsilon_n\}_{n=1}^{\infty}$ are pairwise independent sequences of φ-mixing random variables such that

$$\sum_{n=1}^{\infty}\sqrt{\varphi(\Theta_{\circ,j},n)}<\infty,\ j\in\{1,\ldots,p\}\quad\text{and}\quad\sum_{n=1}^{\infty}\sqrt{\varphi(\varepsilon_{\circ},n)}<\infty. \tag{13}$$

Moreover, assume that

$$\sup_{n\in\mathbb{N}} Z_{n,j}^2<\infty,\ j\in\{1,\ldots,p\}, \tag{14}$$

$$\sup_{n\in\mathbb{N}}\mathbb{E}|\Theta_{n,j}|^{4+\omega_j}<\infty,\ j\in\{1,\ldots,p\},\quad\text{and}\quad\sup_{n\in\mathbb{N}}\mathbb{E}|\varepsilon_n|^{4+\omega_{p+1}}<\infty \tag{15}$$

for some $\omega_j>0$, $j\in\{1,\ldots,p+1\}$. If there exists a positive definite matrix $\daleth$ such that (12) holds, then $\sqrt{n}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)\xrightarrow{\mathscr{D}}\mathcal{N}(\mathbf{0},\boldsymbol{\Sigma})$, $n\to\infty$, where $\boldsymbol{\Sigma}$ is a positive definite matrix depending on $\boldsymbol{\beta}$, $\boldsymbol{\Delta}$, and $\daleth$.

3. CONCLUSIONS

A structure of the EIV model with *weakly dependent errors* is introduced in this paper. Suitable central limit theorems for strong mixing and uniformly strong mixing sequences are postulated. *Strong consistency* of the TLS estimate under both proposed forms of errors' *asymptotic independence* is used for proving *asymptotic normality* of the TLS estimate under α- and φ-mixing errors in the EIV model. Stochastic assumptions for the asymptotic normality of the TLS estimate are derived. Moreover, all the presented results are valid for non-stationary errors.

3.1. Discussion

A straightforward application of the derived asymptotic properties lies in the *justification of the bootstrap methods*. The variance of the asymptotic multivariate normal distribution of the TLS estimate could sometimes be computationally unfeasible. One possible solution to this dilemma is bootstrapping. [13] proved validity of various bootstrap procedures in the case of independent errors in the EIV model, where asymptotic normality of the TLS estimate served as a crucial cornerstone. Asymptotic results from this paper will provide a workable basis for the block-bootstrap techniques when constructing confidence intervals and testing hypotheses in the case of weakly dependent errors of the EIV model.

It is known that a sequence of random variables is φ-mixing implies that this sequence is α-mixing. On the other hand, the CLT for φ-mixing (Lemma 2.2) has weaker assumptions than the CLT for α-mixing (Lemma 2.1). Indeed, Lemma 2.2 does not require any assumptions on *mixing rate* $\varphi(n)$ such as assumption (4) on α-mixing rates. Therefore in Theorem 2.4, we do not have to deal with mixing rate assumption like (8) nor a restriction on the *moment order* like (11) as in Theorem 2.3. On the contrary, Theorem 2.2 requires mixing rate assumption (13), which are inherited from the assumptions for the strong consistency of the TLS estimate. Both asymptotic normality results of the TLS estimate (Theorem 2.3 and Theorem 2.4) require all the assumptions from the strong consistency results (Theorem A.1 and Theorem A.2), because they were used in the proofs of asymptotic normality.

Assumption (12) concerning the *long-run variance* of the TLS estimate is requisite and cannot be omitted, because it assures that the variance of the TLS estimate is bounded away from zero and, simultaneously, does not explode into infinity. This assumption straightforwardly allows to apply the appropriate CLT in order to prove the asymptotic normality of $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$.

All the assumptions and remarks regarding the error structure or the pairwise independence of errors have been already discussed by [12].

Finally, if identically distributed rows of errors (on the between-individual level) are taken into account together with finiteness of their moments up to a suitable order, then moments' assumptions (10) and (15) are trivially satisfied. Then, a *strict stationarity* of the between-individual errors and existence of appropriate moments have to imply these assumptions as well. In other words, moment assumptions (10) and (15) cannot be considered as unattainable. Moreover for strictly stationary errors, even the supremum in definitions (1) and (2) can be simply avoided.

A. APPENDIX

Theorem A.1 (Strong consistency in α-mixing EIV)**.** Let the EIV model hold and assumption (D) be satisfied. Suppose $\{\Theta_{n,1}\}_{n=1}^{\infty}, \ldots, \{\Theta_{n,p}\}_{n=1}^{\infty}$, and $\{\varepsilon_n\}_{n=1}^{\infty}$ are pairwise independent sequences of α-mixing random variables having

$$\alpha(\Theta_{\circ,j}, n) = \mathcal{O}(n^{-q_j/(2q_j-2)-\delta_j}),\ j = 1, \ldots, p \quad \text{and} \quad \alpha(\varepsilon_{\circ}, n) = \mathcal{O}(n^{-q_{p+1}/(2q_{p+1}-2)-\delta_{p+1}}), \tag{16}$$

as $n \to \infty$ for some $\delta_j > 0$ and $1 < q_j \leq 2$, $j \in \{1, \ldots, p+1\}$. If

$$\sup_{n\in\mathbb{N}} Z_{n,j}^2 < \infty,\ j \in \{1, \ldots, p\}, \tag{17}$$

$$\sup_{n\in\mathbb{N}} \mathbb{E}|\Theta_{n,j}|^{2q_j} < \infty,\ j \in \{1, \ldots, p\}, \quad \text{and} \quad \sup_{n\in\mathbb{N}} \mathbb{E}|\varepsilon_n|^{2q_{p+1}} < \infty, \tag{18}$$

then

$$\lim_{n\to\infty} \widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}, \text{ a.s.} \quad \text{and} \quad \lim_{n\to\infty} \frac{\lambda}{n} = \sigma^2, \text{ a.s.} \tag{19}$$

Proof. See [12, Theorem 3.1]. □

Theorem A.2 (Strong consistency in φ-mixing EIV)**.** Let the EIV model hold and assumption (D) be satisfied. Suppose $\{\Theta_{n,1}\}_{n=1}^{\infty}, \ldots, \{\Theta_{n,p}\}_{n=1}^{\infty}$, and $\{\varepsilon_n\}_{n=1}^{\infty}$ are pairwise independent sequences of φ-mixing random variables such that

$$\sum_{n=1}^{\infty} \sqrt{\varphi(\Theta_{\circ,j}, n)} < \infty,\ j \in \{1, \ldots, p\} \quad \text{and} \quad \sum_{n=1}^{\infty} \sqrt{\varphi(\varepsilon_{\circ}, n)} < \infty. \tag{20}$$

If

$$\sum_{n=1}^{\infty} \frac{\mathbb{E}\Theta_{n,j}^4}{n^2} < \infty,\ j \in \{1, \ldots, p\} \quad \text{and} \quad \sum_{n=1}^{\infty} \frac{\mathbb{E}\varepsilon_n^4}{n^2} < \infty, \tag{21}$$

then

$$\lim_{n\to\infty}\widehat{\boldsymbol{\beta}}=\boldsymbol{\beta},\ \text{a.s.}\quad\text{and}\quad\lim_{n\to\infty}\frac{\lambda}{n}=\sigma^2,\ \text{a.s.}\tag{22}$$

P r o o f. See [12, Theorem 3.2]. □

P r o o f. [*Lemma 2.1*]
Define random elements on Skorokhod space $D[0,1]$ as follows:

$$\mathcal{W}_n(t):=\frac{S_{[nt]}}{\varsigma_n},\quad 0\leq t\leq 1,$$

where $[\cdot]$ denotes the nearest integer function. Then weak invariance principle for α-mixing by [9] or [11, Corollary 3.2.1] completes this proof, because a functional distributional limit (a convergence in Skorokhod space) on a bounded interval implies the pointwise distributional limit at the right end-point of the interval. Indeed, there is a problem with measurability of the functionals of discontinuous processes. Space of càdlàg functions with the uniform metric is not separable, but equipped with the Skorokhod metric is [2, Section 14]. Recall the Skorokhod metric $\varrho_{[0,1]}(\cdot,\cdot)$ for càdlàg functions on $[0,1]$ by introducing $\mathcal{F}_{[0,1]}:=\left\{f:[0,1]\xleftrightarrow{1-1}[0,1],\ \text{strictly increasing}\right\}$ and, hence, $\varrho_{[0,1]}(g,h):=\inf_{f\in\mathcal{F}_{[0,1]}}\max\left\{\sup_{t\in[0,1]}|f(t)-1|,\sup_{t\in[0,1]}|g(t)-h(f(t))|\right\}$. If $f\in\mathcal{F}_{[0,1]}$, then $f(1)=1$. Since

$$\max\left\{\sup_{t\in[0,1]}|f(t)-1|,\ \sup_{t\in[0,1]}|\mathcal{W}_n(t)-\mathcal{W}(f(t))|\right\}\geq|\mathcal{W}_n(1)-\mathcal{W}(1)|,$$

where $\mathcal{W}$ is the standard Wiener process, then

$$0\xleftarrow{n\to\infty}\varrho_{[0,1]}(\mathcal{W}_n,\mathcal{W})=\inf_{f\in\mathcal{F}_{[0,1]}}\max\left\{\sup_{t\in[0,1]}|f(t)-1|,\ \sup_{t\in[0,1]}|\mathcal{W}_n(t)-\mathcal{W}(f(t))|\right\}\\ \geq|\mathcal{W}_n(1)-\mathcal{W}(1)|.$$

□

P r o o f. [*Lemma 2.2*]
It is sufficient to show that assumptions (6) and (7) implies Lindeberg condition

$$\forall\delta>0:\quad\lim_{n\to\infty}\frac{1}{\varsigma_n^2}\sum_{i=1}^n\mathbb{E}\xi_i^2\mathcal{I}\{|\xi_i|>\delta\varsigma_n\}=0$$

from [15, Corollary 4]. The first step is to show that conditions (6) and (7) implies so-called Lyapunov condition, i. e., having fixed $\omega>0$:

$$\frac{1}{\varsigma_n^{2+\omega}}\sum_{i=1}^n\mathbb{E}|\xi_i|^{2+\omega}\leq\frac{1}{\varsigma_n^{2+\omega}}\sum_{i=1}^n\sup_{\iota\in\mathbb{N}}\mathbb{E}|\xi_\iota|^{2+\omega}=\frac{n}{\varsigma_n^{2+\omega}}\sup_{\iota\in\mathbb{N}}\mathbb{E}|\xi_\iota|^{2+\omega}\to 0,\quad n\to\infty.$$

Now, Lyapunov condition $\lim_{n\to\infty} \varsigma_n^{-2-\omega} \sum_{i=1}^n \mathbb{E}|\xi_i|^{2+\omega} = 0$ holds and we fix $\delta > 0$. Since $|\xi_i| > \delta\varsigma_n$ implies $|\xi_i/\delta\varsigma_n|^\omega > 1$, we obtain

$$\frac{1}{\varsigma_n^2}\sum_{i=1}^n \mathbb{E}\xi_i^2 \mathcal{I}\{|\xi_i| > \delta\varsigma_n\} \leq \frac{1}{\delta^\omega \varsigma_n^{2+\omega}}\sum_{i=1}^n \mathbb{E}|\xi_i|^{2+\omega}\mathcal{I}\{|\xi_i| > \delta\varsigma_n\}$$
$$\leq \frac{1}{\delta^\omega \varsigma_n^{2+\omega}}\sum_{i=1}^n \mathbb{E}|\xi_i|^{2+\omega} \to 0, \quad n\to\infty.$$

□

P r o o f. [*Theorem 2.3*]

Assumptions of Theorem 2.3 imply the assumptions of Theorem A.1. Indeed, assumptions (17) and (9) coincide. Assumption (8) on α-mixing rates clearly implies assumption (16) for any $\delta_j > 0$ and $1 < q_j \leq 2$, $j \in \{1,\ldots,p+1\}$. Supremum assumption (10) implies (18) for any $\omega_j > 0$ and $1 < q_j \leq 2$, $j \in \{1,\ldots,p+1\}$ as well, because of a corollary of the Jensen's inequality

$$(\mathbb{E}|\xi|^r)^{1/r} \leq (\mathbb{E}|\xi|^s)^{1/s},\ 0 < r < s < \infty. \tag{23}$$

Golub and Van Loan [7, Section 2] proved that $[\widehat{\boldsymbol{\beta}}, -1]^\top$ is the eigenvector of matrix $[\mathbf{X},\mathbf{Y}]^\top [\mathbf{X},\mathbf{Y}]$ with associated eigenvalue λ. Hence,

$$[\mathbf{X},\mathbf{Y}]^\top [\mathbf{X},\mathbf{Y}] \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ -1 \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top\mathbf{X} & \mathbf{X}^\top\mathbf{Y} \\ \mathbf{Y}^\top\mathbf{X} & \mathbf{Y}^\top\mathbf{Y} \end{bmatrix}\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ -1 \end{bmatrix} = \lambda \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ -1 \end{bmatrix}.$$

Previous partitioning yields $\mathbf{X}^\top\mathbf{X}\widehat{\boldsymbol{\beta}} - \mathbf{X}^\top\mathbf{Y} = \lambda\widehat{\boldsymbol{\beta}}$ and $\mathbf{Y}^\top\mathbf{X}\widehat{\boldsymbol{\beta}} - \mathbf{Y}^\top\mathbf{Y} = -\lambda$, which can be rewritten in the following manner $\mathbf{X}^\top\mathbf{Y} = (\mathbf{X}^\top\mathbf{X} - \lambda\mathbf{I})\widehat{\boldsymbol{\beta}}$ and $\mathbf{Y}^\top\mathbf{Y} = \mathbf{Y}^\top\mathbf{X}\widehat{\boldsymbol{\beta}} + \lambda$. Hence,

$$\begin{aligned}[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}] &= \begin{bmatrix} \mathbf{X}^\top\mathbf{X} & (\mathbf{X}^\top\mathbf{X} - \lambda\mathbf{I})\widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\beta}}^\top(\mathbf{X}^\top\mathbf{X} - \lambda\mathbf{I}) & \mathbf{Y}^\top\mathbf{X}\widehat{\boldsymbol{\beta}} + \lambda \end{bmatrix} \\ &= \left[\mathbf{I},\widehat{\boldsymbol{\beta}}\right]^\top (\mathbf{X}^\top\mathbf{X} - \lambda\mathbf{I})\left[\mathbf{I},\widehat{\boldsymbol{\beta}}\right] + \lambda\mathbf{I}.\end{aligned}$$

It can be easily noted that $[\mathbf{I},\boldsymbol{\beta}]\begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix} = \mathbf{0}$. Therefore, we obtain

$$[\mathbf{I},\boldsymbol{\beta}]\,[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]\begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix} = [\mathbf{I},\boldsymbol{\beta}]\left[\mathbf{I},\widehat{\boldsymbol{\beta}}\right]^\top (\mathbf{X}^\top\mathbf{X} - \lambda\mathbf{I})(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}})$$

and, then,

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = -\boldsymbol{\Delta}_n^{-1}\left([\mathbf{I},\boldsymbol{\beta}]\left[\mathbf{I},\widehat{\boldsymbol{\beta}}\right]^\top\right)^{-1}[\mathbf{I},\boldsymbol{\beta}]\left(n^{-1/2}[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]\right)\begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix}. \tag{24}$$

Since $\mathbb{E}[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}] = [\mathbf{I},\boldsymbol{\beta}]^\top\mathbf{Z}^\top\mathbf{Z}[\mathbf{I},\boldsymbol{\beta}] + n\sigma^2\mathbf{I}$, then

$$[\mathbf{I},\boldsymbol{\beta}]\mathbb{E}[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]\begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix} = \mathbf{0}. \tag{25}$$

Relation (24) can be alternatively rewritten using identity (25) in a slightly more sophisticated way

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = -\boldsymbol{\Delta}_n^{-1}\left([\mathbf{I},\boldsymbol{\beta}]\left[\mathbf{I},\widehat{\boldsymbol{\beta}}\right]^\top\right)^{-1}[\mathbf{I},\boldsymbol{\beta}] \left(n^{-1/2}\left\{[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]-\mathbb{E}[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]\right\}\right)\begin{bmatrix}\boldsymbol{\beta}\\-1\end{bmatrix},$$

which will be useful in the forthcoming steps of this proof.

The proof of Theorem 3.1 by [12] also provides the following important relation

$$\boldsymbol{\Delta}_n := n^{-1}(\mathbf{X}^\top\mathbf{X}-\lambda\mathbf{I}) \xrightarrow[n\to\infty]{\text{a.s.}} \boldsymbol{\Delta}. \tag{26}$$

With respect to (26) and Theorem A.1, we have

$$\boldsymbol{\Delta}_n^{-1}\left([\mathbf{I},\boldsymbol{\beta}]\left[\mathbf{I},\widehat{\boldsymbol{\beta}}\right]^\top\right)^{-1}[\mathbf{I},\boldsymbol{\beta}] \xrightarrow{\text{a.s.}} \boldsymbol{\Delta}^{-1}\left([\mathbf{I},\boldsymbol{\beta}]\,[\mathbf{I},\boldsymbol{\beta}]^\top\right)^{-1}[\mathbf{I},\boldsymbol{\beta}], \quad n\to\infty. \tag{27}$$

With probability tending to one, the inverse $\boldsymbol{\Delta}_n$ exists due to (D) and (26), and the inverse of $[\mathbf{I},\boldsymbol{\beta}]\left[\mathbf{I},\widehat{\boldsymbol{\beta}}\right]^\top$ exists because of (19). Moreover, matrix $[\mathbf{I},\boldsymbol{\beta}]\,[\mathbf{I},\boldsymbol{\beta}]^\top = \mathbf{I}+\boldsymbol{\beta}\boldsymbol{\beta}^\top$ is always positive definite and, hence, regular.

Convergence almost surely from (27) and the Slutsky's theorem reduce the problem of finding a limiting distribution for $\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})$ to finding a limiting distribution for

$$\begin{aligned} n^{-1/2}&\left([\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]-\mathbb{E}[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]\right)\begin{bmatrix}\boldsymbol{\beta}\\-1\end{bmatrix}\\ &= n^{-1/2}\left([\mathbf{I},\boldsymbol{\beta}]^\top\mathbf{Z}^\top[\boldsymbol{\Theta},\boldsymbol{\varepsilon}]+[\boldsymbol{\Theta},\boldsymbol{\varepsilon}]^\top[\boldsymbol{\Theta},\boldsymbol{\varepsilon}]-n\sigma^2\mathbf{I}\right)\begin{bmatrix}\boldsymbol{\beta}\\-1\end{bmatrix}. \end{aligned} \tag{28}$$

Now, it is sufficient to prove the univariate asymptotic normality of

$$n^{-1/2}\sum_{i=1}^n \mathbf{t}^\top\left([\mathbf{Z},\mathbf{Z}\boldsymbol{\beta}]_{i,\bullet}^\top[\boldsymbol{\Theta},\boldsymbol{\varepsilon}]_{i,\bullet}+[\boldsymbol{\Theta},\boldsymbol{\varepsilon}]_{i,\bullet}^\top[\boldsymbol{\Theta},\boldsymbol{\varepsilon}]_{i,\bullet}-\sigma^2\mathbf{I}\right)\begin{bmatrix}\boldsymbol{\beta}\\-1\end{bmatrix}, \quad \forall\mathbf{t}\in\mathbb{R}^{p+1};$$

and apply the Cramér–Wold theorem.

The case of $\mathbf{t}=\mathbf{0}$ is trivial. If $\mathbf{t}\neq\mathbf{0}$, a sum of zero mean α-mixing random variables

$$\begin{aligned} \rho_i :=& \ \mathbf{t}^\top\left([\mathbf{Z},\mathbf{Z}\boldsymbol{\beta}]_{i,\bullet}^\top[\boldsymbol{\Theta},\boldsymbol{\varepsilon}]_{i,\bullet}+[\boldsymbol{\Theta},\boldsymbol{\varepsilon}]_{i,\bullet}^\top[\boldsymbol{\Theta},\boldsymbol{\varepsilon}]_{i,\bullet}-\sigma^2\mathbf{I}\right)\begin{bmatrix}\boldsymbol{\beta}\\-1\end{bmatrix} \\ =& \ \mathbf{Z}_{i,\bullet}\mathbf{t}_{-(p+1)}\boldsymbol{\Theta}_{i,\bullet}\boldsymbol{\beta}+\mathbf{Z}_{i,\bullet}\boldsymbol{\beta}t_{p+1}\boldsymbol{\Theta}_{i,\bullet}\boldsymbol{\beta}-\mathbf{Z}_{i,\bullet}\mathbf{t}_{-(p+1)}\varepsilon_i-\mathbf{Z}_{i,\bullet}\boldsymbol{\beta}t_{p+1}\varepsilon_i \\ &+\boldsymbol{\Theta}_{i,\bullet}\mathbf{t}_{-(p+1)}\boldsymbol{\Theta}_{i,\bullet}\boldsymbol{\beta}+t_{p+1}\varepsilon_i\boldsymbol{\Theta}_{i,\bullet}\boldsymbol{\beta}-\boldsymbol{\Theta}_{i,\bullet}\mathbf{t}_{-(p+1)}\varepsilon_i-t_{p+1}\varepsilon_i^2-\sigma^2\mathbf{t}^\top\begin{bmatrix}\boldsymbol{\beta}\\-1\end{bmatrix} \end{aligned} \tag{29}$$

satisfies all the assumptions of Lemma 2.1. Choosing $\omega = 1/2\min_{j=1,\ldots,p+1}\omega_j$, assumption (3) holds due to (9), (10), and (23) together with the Cauchy–Schwarz inequality and Jensen's inequality:

$$
\begin{aligned}
\sup_{n\in\mathbb{N}}\mathbb{E}|\rho_n|^{2+\omega} &\leq 9^{1+\omega}\sup_{n\in\mathbb{N}}\mathbb{E}\Big\{|\mathbf{Z}_{n,\bullet}\mathbf{t}_{-(p+1)}\boldsymbol{\Theta}_{n,\bullet}\boldsymbol{\beta}|^{2+\omega}+|\mathbf{Z}_{n,\bullet}\boldsymbol{\beta}t_{p+1}\boldsymbol{\Theta}_{n,\bullet}\boldsymbol{\beta}|^{2+\omega}\\
&\quad+|\mathbf{Z}_{n,\bullet}\mathbf{t}_{-(p+1)}\varepsilon_n|^{2+\omega}+|\mathbf{Z}_{n,\bullet}\boldsymbol{\beta}t_{p+1}\varepsilon_n|^{2+\omega}+|\boldsymbol{\Theta}_{n,\bullet}\mathbf{t}_{-(p+1)}\boldsymbol{\Theta}_{n,\bullet}\boldsymbol{\beta}|^{2+\omega}\\
&\quad+|t_{p+1}\varepsilon_n\boldsymbol{\Theta}_{n,\bullet}\boldsymbol{\beta}|^{2+\omega}+|\boldsymbol{\Theta}_{n,\bullet}\mathbf{t}_{-(p+1)}\varepsilon_n|^{2+\omega}+|t_{p+1}\varepsilon_n^2|^{2+\omega}+|\sigma^2(\mathbf{t}_{-(p+1)}^\top\boldsymbol{\beta}-t_{p+1})|^{2+\omega}\Big\}\\
&\leq 9^{1+\omega}\Big\{\sup_{n\in\mathbb{N}}|\mathbf{Z}_{n,\bullet}\mathbf{t}_{-(p+1)}|^{2+\omega}\sup_{n\in\mathbb{N}}\mathbb{E}|\boldsymbol{\Theta}_{n,\bullet}\boldsymbol{\beta}|^{2+\omega}+\sup_{n\in\mathbb{N}}|\mathbf{Z}_{n,\bullet}\boldsymbol{\beta}t_{p+1}|^{2+\omega}\sup_{n\in\mathbb{N}}\mathbb{E}|\boldsymbol{\Theta}_{n,\bullet}\boldsymbol{\beta}|^{2+\omega}\\
&\quad+\sup_{n\in\mathbb{N}}|\mathbf{Z}_{n,\bullet}\mathbf{t}_{-(p+1)}|^{2+\omega}\sup_{n\in\mathbb{N}}\mathbb{E}|\varepsilon_n|^{2+\omega}+\sup_{n\in\mathbb{N}}|\mathbf{Z}_{n,\bullet}\boldsymbol{\beta}t_{p+1}|^{2+\omega}\sup_{n\in\mathbb{N}}\mathbb{E}|\varepsilon_n|^{2+\omega}\\
&\quad+\left[\sup_{n\in\mathbb{N}}\mathbb{E}|\boldsymbol{\Theta}_{n,\bullet}\mathbf{t}_{-(p+1)}|^{4+2\omega}\right]^{1/2}\left[\sup_{n\in\mathbb{N}}\mathbb{E}|\boldsymbol{\Theta}_{n,\bullet}\boldsymbol{\beta}|^{4+2\omega}\right]^{1/2}\\
&\quad+\left[\sup_{n\in\mathbb{N}}\mathbb{E}|t_{p+1}\varepsilon_n|^{4+2\omega}\right]^{1/2}\left[\sup_{n\in\mathbb{N}}\mathbb{E}|\boldsymbol{\Theta}_{n,\bullet}\boldsymbol{\beta}|^{4+2\omega}\right]^{1/2}\\
&\quad+\left[\sup_{n\in\mathbb{N}}\mathbb{E}|\boldsymbol{\Theta}_{n,\bullet}\mathbf{t}_{-(p+1)}|^{4+2\omega}\right]^{1/2}\left[\sup_{n\in\mathbb{N}}\mathbb{E}|\varepsilon_n|^{4+2\omega}\right]^{1/2}\\
&\quad+|t_{p+1}|\sup_{n\in\mathbb{N}}\mathbb{E}|\varepsilon_n|^{4+2\omega}+|\sigma^2(\mathbf{t}_{-(p+1)}^\top\boldsymbol{\beta}-t_{p+1})|^{2+\omega}\Big\}\\
&\leq 9^{1+\omega}\Big\{p^{2+2\omega}\max_{j=1,\ldots,p}|t_j|^{2+\omega}\sum_{j=1}^p\sup_{n\in\mathbb{N}}|Z_{n,j}|^{2+\omega}\max_{j=1,\ldots,p}|\beta_j|^{2+\omega}\sum_{j=1}^p\sup_{n\in\mathbb{N}}\mathbb{E}|\Theta_{n,j}|^{2+\omega}\\
&\quad+p^{2+2\omega}|t_{p+1}|\max_{j=1,\ldots,p}|\beta_j|^{2+\omega}\sum_{j=1}^p\sup_{n\in\mathbb{N}}|Z_{n,j}|^{2+\omega}\max_{j=1,\ldots,p}|\beta_j|^{2+\omega}\sum_{j=1}^p\sup_{n\in\mathbb{N}}\mathbb{E}|\Theta_{n,j}|^{2+\omega}\\
&\quad+p^{1+\omega}\max_{j=1,\ldots,p}|t_j|^{2+\omega}\sum_{j=1}^p\sup_{n\in\mathbb{N}}|Z_{n,j}|^{2+\omega}\sup_{n\in\mathbb{N}}\mathbb{E}|\varepsilon_n|^{2+\omega}\\
&\quad+p^{1+\omega}|t_{p+1}|\max_{j=1,\ldots,p}|\beta_j|^{2+\omega}\sum_{j=1}^p\sup_{n\in\mathbb{N}}|Z_{n,j}|^{2+\omega}\sup_{n\in\mathbb{N}}\mathbb{E}|\varepsilon_n|^{2+\omega}\\
&\quad+p^{3+2\omega}\max_{j=1,\ldots,p}|t_j|^{2+\omega}\max_{j=1,\ldots,p}|\beta_j|^{2+\omega}\sum_{j=1}^p\sup_{n\in\mathbb{N}}\mathbb{E}|\Theta_{n,j}|^{4+2\omega}\\
&\quad+p^{3/2+\omega}|t_{p+1}|^{2+\omega}\max_{j=1,\ldots,p}|\beta_j|^{2+\omega}\left[\sum_{j=1}^p\sup_{n\in\mathbb{N}}\mathbb{E}|\Theta_{n,j}|^{4+2\omega}\right]^{1/2}\left[\sup_{n\in\mathbb{N}}\mathbb{E}|\varepsilon_n|^{4+2\omega}\right]^{1/2}\\
&\quad+p^{3/2+\omega}\max_{j=1,\ldots,p}|t_j|^{2+\omega}\left[\sum_{j=1}^p\sup_{n\in\mathbb{N}}\mathbb{E}|\Theta_{n,j}|^{4+2\omega}\right]^{1/2}\left[\sup_{n\in\mathbb{N}}\mathbb{E}|\varepsilon_n|^{4+2\omega}\right]^{1/2}\\
&\quad+|t_{p+1}|\sup_{n\in\mathbb{N}}\mathbb{E}|\varepsilon_n|^{4+2\omega}+|\sigma^2(\mathbf{t}_{-(p+1)}^\top\boldsymbol{\beta}-t_{p+1})|^{2+\omega}\Big\}<\infty.
\end{aligned}
$$

Bradley [3, Theorem 5.2(a)] provides $\alpha(\rho_\circ,n)\leq\alpha(\varepsilon_\circ,n)+\sum_{j=1}^p\alpha(\mathbf{\Theta}_{\circ,j},n)$. Consequently, assumption (4) holds due to the concavity of function $u\mapsto u^{\omega/(2+\omega)}$, $\omega>0$:

$$\sum_{n=1}^\infty\alpha(\rho_\circ,n)^{\omega/(2+\omega)}\leq\sum_{n=1}^\infty\alpha(\varepsilon_\circ,n)^{\omega/(2+\omega)}+\sum_{j=1}^p\sum_{n=1}^\infty\alpha(\mathbf{\Theta}_{\circ,j},n)^{\omega/(2+\omega)}<\infty,\quad\omega>0;$$

which is true because of (11) and the fact that

$$\alpha(\mathbf{\Theta}_{\circ,j},n)^{\omega/(2+\omega)}=\mathcal{O}\left(n^{-1-\frac{\delta_j\omega-2}{2+\omega}}\right),\quad\delta_j>2/\omega>0,\ j\in\{1,\ldots,p\};$$
$$\alpha(\varepsilon_\circ,n)^{\omega/(2+\omega)}=\mathcal{O}\left(n^{-1-\frac{\delta_j\omega-2}{2+\omega}}\right),\quad\delta_{p+1}>2/\omega>0.$$

Using (12) and (28), let us calculate

$$\begin{aligned}\frac{1}{n}\mathbb{E}\left(\sum_{i=1}^n\rho_i\right)^2&=\frac{1}{n}\mathbb{E}\left\{\mathbf{t}^\top\left([\mathbf{Z},\mathbf{Z}\boldsymbol{\beta}]^\top[\mathbf{\Theta},\boldsymbol{\varepsilon}]+[\mathbf{\Theta},\boldsymbol{\varepsilon}]^\top[\mathbf{\Theta},\boldsymbol{\varepsilon}]-n\sigma^2\mathbf{I}\right)\begin{bmatrix}\boldsymbol{\beta}\\-1\end{bmatrix}\right\}^2\\
&=\frac{1}{n}\mathbb{E}\mathbf{t}^\top\left([\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]-\mathbb{E}[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]\right)\begin{bmatrix}\boldsymbol{\beta}\\-1\end{bmatrix}\\
&\quad[\boldsymbol{\beta}^\top,-1]\left([\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]-\mathbb{E}[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]\right)\mathbf{t}\\
&=\mathbf{t}^\top\left\{\frac{1}{n}\mathbb{V}\mathrm{ar}\,[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]\begin{bmatrix}\boldsymbol{\beta}\\-1\end{bmatrix}\right\}\mathbf{t}\to\mathbf{t}^\top\beth\mathbf{t}>0,\quad n\to\infty;\end{aligned}$$

and assumption (5) is satisfied as well. Thus, Lemma 2.1 implies asymptotically zero mean normal distribution $n^{-1/2}\sum_{i=1}^n\rho_i$. □

P r o o f. [*Theorem 2.4*]

This proof contains very similar ideas as the proof of Theorem 2.3.

Assumptions of Theorem 2.4 imply the assumptions of Theorem A.2. Indeed, assumption (20) is implied by assumption (13). Assumption (21) directly follows from (15), because (23) and (15) yield

$$\sup_{n\in\mathbb{N}}\mathbb{E}\xi_n^4\leq\left(\sup_{n\in\mathbb{N}}\mathbb{E}|\xi_n|^{4+\omega}\right)^{4/(4+\omega)}<\infty$$

for $\xi_n\in\{\Theta_{n,1},\ldots,\Theta_{n,p},\varepsilon_n\}$ and $\omega=\min_{j=1,\ldots,p+1}\omega_j$. Hence,

$$\sum_{n=1}^\infty\frac{\mathbb{E}\xi_n^4}{n^2}\leq\sum_{n=1}^\infty\frac{\sup_{\iota\in\mathbb{N}}\mathbb{E}\xi_\iota^4}{n^2}=\sup_{\iota\in\mathbb{N}}\mathbb{E}\xi_\iota^4\sum_{n=1}^\infty\frac{1}{n^2}<\infty.$$

Therefore, the consistency results for φ-mixing errors can be used.

With respect to the Slutsky's theorem, to the Cramér–Wold theorem, and to the proof of Theorem 2.3, it is necessary to find the limiting distribution of $\{\rho_n\}_{n=1}^\infty$ from (29). In light of Lemma 2.2, we only need to check whether sequences $\{\rho_n\}_{n=1}^\infty$ is φ-mixing sequences, i.e., $\varphi(\rho_\circ,n)\to0$ as $n\to\infty$. This follows directly from [3, Theorem 5.2(d)] and assumptions $\varphi(\Theta_{\circ,j},n)\to0$ for $j=1,\ldots,p$ and $\varphi(\varepsilon_\circ,n)\to0$ as $n\to\infty$. The rest of the assumptions of Lemma 2.2 is included in the assumptions of Lemma 2.1 and has been completely checked on sequence $\{\rho_n\}_{n=1}^\infty$ in the proof of Theorem 2.3. □

ACKNOWLEDGEMENT

This paper was written with the support of the Czech Science Foundation project "DYME / Dynamic Models in Economics" No. P402/12/G097.

(Received April 27, 2012)

REFERENCES

[1] T. W. Anderson: An Introduction to Multivariate Statistical Analysis. John Wiley and Sons, New York 1958.

[2] P. Billingsley: Convergence of Probability Measures. First edition. John Wiley and Sons, New York 1968.

[3] R. C. Bradley: Basic properties of strong mixing conditions. A survey and some open questions. Probab. Surveys *2* (2005), 107–144.

[4] P. P. Gallo: Consistency of regression estimates when some variables are subject to error. Comm. Statist Theory Methods *11* (1982), 973–983.

[5] P. P. Gallo: Properties of Estimators in Errors-in-Variables Models. Ph.D. Thesis, University of North Carolina, Chapel Hill 1982.

[6] L. J. Gleser: Estimation in a multivariate "errors in variables" regression model: Large sample results. Ann. Statist. *9* (1981), 24–44.

[7] G. H. Golub and C. F. Van Loan: An analysis of the total least squares problem. SIAM J. Numer. Anal. *17* (1980), 6, 883–893.

[8] J. D. Healy: Estimation and Tests for Unknown Linear Restrictions in Multivariate Linear Models. Ph.D. Thesis, Purdue University 1975.

[9] N. Herrndorf: A functional central limit theorem for strongly mixing sequence of random variables. Probab. Theory Rel. Fields *69* (1985), 4, 541–550.

[10] I. A. Ibragimov and Y. V. Linnik: Independent and Stationary Sequences of Random Variables. Wolters–Noordhoff 1971.

[11] Z. Lin and C. Lu: Limit Theory for Mixing Dependent Random Variables. Springer-Verlag, New York 1997.

[12] M. Pešta: Strongly consistent estimation in dependent errors-in-variables. Acta Univ. Carolin. – Math. Phys. *52* (2011), 1, 69–79.

[13] M. Pešta: Total least squares and bootstrapping with application in calibration. Statistics: J. Theor. and Appl. Statistics *46* (2013), 5, 966–991.

[14] M. Rosenblatt: Markov Processes: Structure and Asymptotic Behavior. Springer-Verlag, Berlin 1971.

[15] S. A. Utev: The central limit theorem for φ-mixing arrays of random variables. Theory Prob. Appl. *35* (1990), 131–139.

Michal Pešta, Charles University in Prague, Faculty of Mathematics and Physics, Department of Probability and Mathematical Statistics, Sokolovská 83, 186 75 Praha 8. Czech Republic.
e-mail: pesta@karlin.mff.cuni.cz