Charles University in Prague

Faculty of Mathematics and Physics

DOCTORAL THESIS

Michal Pešta

Modern Asymptotic Perspectives on Errors-in-Variables Modeling

Department of Probability and Mathematical Statistics

Supervised by
Prof. RNDr. Jaromír Antoch, CSc.

Study program
Mathematics

Study branch
Probability and Mathematical Statistics

Dedicated to my mother Vierka

and to my father Zdeněk.

Acknowledgements

First of all, I would like to thank my supervisor Prof. Jaromír Antoch for his patient guidance throughout this work. Without him this thesis could not have been completed. His practical advice and suggestions have directed me in my work whenever I have needed it.

I would also like to thank to all my teachers from Charles University in Prague, Czech Republic for introducing me into the world of higher mathematics and mathematical statistics, and providing me important theoretical knowledge. They have inspired me with new ideas and opened my mind for statistical thinking.

Last, but not least, I would like to express my thanks to all my friends, whose encouragement have meant to me so much during the pursuit of my postgraduate degree and the composition of the thesis. My deepest and warmest thanks belong to Barbora for love, inspiration, and patience during our statistical dialogs.

This work was supported by the grants GAUK B-MAT 49809, GAČR 201/09/0755, and SVV 261315/2010.

Statement of Honesty

I hereby declare that I have written this doctoral thesis separately, independently and entirely with using quoted resources. I agree that the University Library shall make it available to borrowers under rules of the Library.

Prague, September 27, 2010

Michal Pešta

Signature

Annotations

Title

Modern Asymptotic Perspectives on Errors-in-Variables Modeling

Author

Mgr. Michal Pešta, M.Sc.
pesta@karlin.mff.cuni.cz

Department

Department of Probability and Mathematical Statistics
Faculty of Mathematics and Physics
Charles University in Prague
Czech Republic

Supervisor

Prof. RNDr. Jaromír Antoch, CSc.
antoch@karlin.mff.cuni.cz

Mailing Address

KPMS MFF UK
Sokolovská 83
186 75 Prague 8
Czech Republic

Abstract

A linear regression model, where covariates and a response are subject to errors, is considered in this thesis. For so-called errors-in-variables (EIV) model, suitable error structures are proposed, various unknown parameter estimation techniques are performed, and recent algebraic and statistical results are summarized.

An extension of the total least squares (TLS) estimate in the EIV model—the EIV estimate—is invented. Its invariant (with respect to scale) and equivariant (with respect to the covariates' rotation, to the change of covariates direction, and to the interchange of covariates) properties are derived. Moreover, it is shown that the EIV estimate coincides with any unitarily invariant penalizing solution to the EIV problem.

It is demonstrated that the asymptotic normality of the EIV estimate is computationally useless for a construction of confidence intervals or hypothesis testing. A proper bootstrap procedure is constructed to overcome such an issue. The validity of the bootstrap technique is proved. A simulation study and a real data example assure of its appropriateness.

Strong and uniformly strong mixing errors are taken into account instead of the independent ones. For such a case, the strong consistency and the asymptotic normality of the EIV estimate are shown. Despite of that, their practical applicability remains problematic.

A suitable block bootstrap method is proposed for the EIV estimate with weakly dependent errors and, consequently, its justification is proved. Again, a computational efficiency is demonstrated by simulations and a real data analysis.

In the end, a nonparametric extension of the EIV model is suggested and a way of a smooth estimation is proposed.

Keywords

errors-in-variables, total least squares, consistency, asymptotic normality, equivariant estimation, bootstrap, weak dependence, block bootstrap

Preface

An intensive research has sprung up for methods to handle measurement errors or disturbances in input and output data simultaneously. *Errors-in-variables* serve as a *regression modeling technique*, where both dependent and independent variables are considered to be measured with errors.

The first chapter introduces our main concern—the errors-in-variables (EIV) model. We will illustratively discuss its *geometrical interpretation*. The EIV model postulates an *optimizing problem*, which solution will be provided. Such an EIV solution is called the *total least squares (TLS) solution* and its algebraic and spectral properties will be summarized. An *error structure* from a stochastic point of view will be incorporated into the EIV model and recent statistical properties of the TLS solution—thought as a suitable estimate—will be provided.

The second chapter deals with a generalization of the TLS estimate—the EIV estimate. The way, how the EIV estimate is constructed, provides us several *invariant* and *equivariant* properties.

Unfortunately, the statistical asymptotic approximations concerning the EIV estimate become problematic and *computational useless*. A feasible solution to this dilemma can be found in the bootstrap approach. A proper *bootstrap procedure* will be proposed and its justification provided in the third chapter.

In some realistic situations, the errors in the EIV model cannot be considered as independent. The error structure—introduced in the first chapter and widely utilized in the third chapter—will be generalized for the case of *weakly dependent errors* in the fourth chapter. Consequently, a strong consistency and an asymptotic normality will be proved for the EIV estimate, when the errors are (uniformly) strong mixing.

In the fifth chapter, a similar task concerning practical applicability of the results from the fourth chapter arises again. Since we are not in the case of independent observations any more, an advanced approach needs to be proposed to make the asymptotic inference computationally efficient. Therefore, a suitable *block bootstrap technique* will be presented. We will also justify its usage by showing its asymptotic equivalence to the approximate normality approach.

On the top of that, *simulation studies* and *real data examples* demonstrate results from the third and the fifth chapter.

Finally, a linearity in the EIV model may sometimes seem restrictive. Hence, a nonlinear attitude to the EIV model is discussed in the sixth chapter. A way of *nonparametric modeling* in the EIV will be suggested as well with a possible way of estimation.

At the end, all the well-known definitions and theorems, which were frequently used through the thesis, are recapitulated in the appendix.

Notation

$a.s.$	...	almost surely
$[\mathbb{P}]$-$a.s.$	...	$[\mathbb{P}]$-almost surely
$\xleftrightarrow{\mathscr{D}([\mathbb{P}]\text{-}a.s.)}$	...	approaching in distribution $[\mathbb{P}]$-almost surely
$\xleftrightarrow{\mathscr{D}(\mathbb{P})}$	...	approaching in distribution in probability $\mathbb{P}$
$\mathbf{e}_i$	...	canonical vector, i-th element equal one and the rest are zeros
$\xrightarrow{[\mathbb{P}]\text{-}a.s.}$	...	convergence $[\mathbb{P}]$-almost surely
$\xrightarrow{\mathscr{D}}$	...	convergence in distribution
$\xrightarrow{\mathbb{P}}$	...	convergence in probability $\mathbb{P}$
$\xrightarrow{\mathbb{P}^*([\mathbb{P}]\text{-}a.s.)}$	...	convergence in probability $\mathbb{P}^*$ $[\mathbb{P}]$-almost surely
$\xrightarrow{\mathbb{P}^*(\mathbb{P})}$	...	convergence in probability $\mathbb{P}^*$ in probability $\mathbb{P}$
$\xrightarrow{D[a,b]}$	...	convergence in Skorokhod space $D[a,b]$
$\mathcal{O}, o$	...	deterministic Landau symbols, confer Appendix A.1
$diag(\mathbf{x})$	...	diagonal square matrix with diagonal elements from vector $\mathbf{x}$
dim	...	dimension of a vector (sub)space
$\mathbb{E}$	...	expectation
iid	...	independent and identically distributed
$\mathbf{I}$	...	identity matrix

$\Im$	...	imaginary part of a complex number
$\mathcal{I}$	...	indicator function
$\mathbb{Z}$	...	integers, i.e., $\{\ldots,-2-1,0,1,2,\ldots\}$
Ker	...	kernel of a linear mapping
δ_{ij}	...	Kronecker delta
$\mathbb{N}$	...	natural numbers, i.e., $\{1,2,\ldots\}$
$\mathbb{N}_0$	...	natural numbers with zero, i.e., $\{0,1,2,\ldots\}$
$[\cdot]$	...	nearest integer function
$\mathbf{t}_{-i}$	...	omitting the ith element from the original vector $\mathbf{t}$
$\mathbb{P}$	...	probability
$Range$	...	range of a linear mapping (column space)
$rank$	...	rank of a matrix
$\mathbb{R}$	...	real numbers
$\Re$	...	real part of a complex number
$D[a,b]$	...	Skorokhod space on interval $[a,b]$
$\mathcal{W}$	...	standard Wiener process
$\mathcal{O}_{\mathbb{P}}, o_{\mathbb{P}}$	...	stochastic Landau symbols, confer Appendix A.1
$\mathbf{t}_{i:j}$	...	subvector $[t_i,t_{i+1},\ldots,t_{j-1},t_j]^\top$ of the original vector $\mathbf{t}$
tr	...	trace of a matrix
$^\top$	...	transpose of a vector or matrix
$\mathbb{V}\mathrm{ar}$	...	variance
$\mathbf{0}$	...	zero vector or matrix

Contents

Chapter 1

Introduction to Errors-in-Variables

Lasciate ogni speranza, voi ch'entrate.
[All hope abandon, ye who enter in.]

DANTE ALIGHIERI

Observing several characteristics—may be thought as variables—straightforwardly postulates a natural question: "What is the relationship between these measured characteristics?" One of many possible attitudes can arise, that some of the characteristics might be explained by a (functional) dependence on the other characteristics. Therefore, we consider the first mentioned variables as *dependent* or *response* and the second ones as *independent*, *explanatory*, or *covariates*.

1.1 Our Goals and Aims

Linear relationships are usually described by a linear regression where the response (regressand) is the only observed variable encumbered by an error. On the other hand, all the covariates (regressors) are considered to be measured *precisely*. Despite of this common approach, some situations can appear where it is more convenient and suitable, sometimes even necessary, to assume that both regressand and regressors are *subject to error* (Madansky, 1959). Hence, a proper statistical model should be chosen to handle errors in all variables, unsurprisingly called errors-in-variables model by Gleser and Watson (1973).

A prime motivation for this thesis is to introduce a *regression model* capable of handling *errors in the observed characteristics*. Our goal is to clearly summarize already developed properties of this model. Consequently, we want to derive additional important properties of the model with errors in measurements, to extend its applicability, and to overcome some problems that exist up to now. We mainly take into account modern asymptotic approaches in order to finalize unsolved questions regarding the usage of regression model with errors in variables. Therefore, we will concentrate on and incorporate:

- computational feasibility,
- invariant, equivariant and consistent estimation,
- finite sample and limiting behavior,
- robust approach,
- weak dependence of errors,
- simulation studies,
- real data analyses.

1.1.1 Main Ideas

As already mentioned in Preface of this thesis, a brief summary of known results is necessary. Our regression model with errors in variables contains some unknown quantities, which needs to be estimated in a reasonable manner. A way of *equivariant estimation* is presented in order to preserve some natural and desired properties. Furthermore, a situation modeled by the independent errors could not be suitable in each real-time scenario and *asymptotics* for the estimates is required in that case. Large sample theory for *dependent variables* is applied. In spite of this, theoretical knowledge does not have to provide solutions for the real problems. *Computational intensive methods*, e.g., various versions of bootstrap, are developed, their correctness proved and, moreover, their applicability demonstrated in the simulations and on the real data as well.

1.2 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 assumed, 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 composed of *random errors* such that the joint distribution of the elements of $[\boldsymbol{\Theta}, \boldsymbol{\varepsilon}]$ is absolutely continuous with respect to the Lebesgue measure.

EIV model (E) for the case when $p = 1$ can be graphically illustrated in the two-dimensional setting as shown in Figure 1.1.

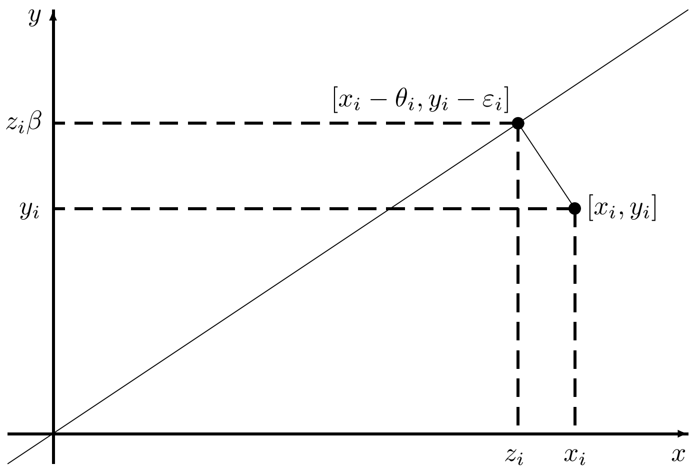

Figure 1.1: Illustration of the EIV model.

EIV model (E) with non-random unknown constants $\mathbf{Z}$ is sometimes called *functional EIV* model. On the other hand, a different approach may handle $\mathbf{Z}$ as random covariates, which is called *structural EIV* model. We purely concentrate on the first mentioned case.

1.3 Algebraic Overview

Our proposed model of dependence contains errors in both the response variable (we think only of one dependent variable) and the explanatory variables as well. But firstly, we just try to find an appropriate fit for some points in the Euclidean space using a *hyperplane*, i.e., approximating several incompatible linear relations. Afterwards, some assumptions on the measurement errors are added and, hence, several statistical asymptotical properties are developed.

Let us consider the *overdetermined system* of linear relations

$$\mathbf{Y} \approx \mathbf{X}\boldsymbol{\beta}, \qquad \mathbf{Y} \in \mathbb{R}^n,\ \mathbf{X} \in \mathbb{R}^{n \times p},\ n > p. \tag{1.1}$$

Relations in (1.1) are deliberately not denoted as equations, because in many cases the exact solution need not exist. Thereby, only an approximation can be found. Hence, one can speak about the "best" solution of the overdetermined system (1.1). Nevertheless, the "best" in which way?

1.3.1 Singular Value Decomposition

Before inquiring into an appropriate solution of (1.1), we should introduce some very important tools for further exploration.

Theorem 1.1 (Singular value decomposition – SVD). *If* $\mathbf{A} \in \mathbb{R}^{n\times p}$*, then there exist orthonormal matrices* $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_n] \in \mathbb{R}^{n\times n}$ *and* $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_p] \in \mathbb{R}^{p\times p}$ *such that*

$$\mathbf{U}^\top \mathbf{A}\mathbf{V} = \mathbf{\Sigma} = diag\{\sigma_1, \ldots, \sigma_q\} \in \mathbb{R}^{n\times p}, \quad \sigma_1 \geq \ldots \geq \sigma_q \geq 0, \quad and \quad q = \min\{n, p\}. \tag{1.2}$$

Proof. See Golub and Van Loan (1996). □

In SVD, the diagonal matrix $\mathbf{\Sigma}$ is uniquely determined by $\mathbf{A}$ (though the matrices $\mathbf{U}$ and $\mathbf{V}$ are not). Previous powerful matrix decomposition allows us to define a *cutting point* $r \in \mathbb{N}_0$ for a given matrix $\mathbf{A} \in \mathbb{R}^{n\times p}$ using its singular values σ_i

$$\sigma_1 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_q = 0, \quad q = \min\{n, p\}.$$

Since the matrices $\mathbf{U}$ and $\mathbf{V}$ in (1.2) are orthonormal, it holds that $rank(\mathbf{A}) = r$ and we may obtain a *dyadic decomposition* (expansion) of the matrix $\mathbf{A}$:

$$\mathbf{A} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top. \tag{1.3}$$

A suitable matrix norm is also required. For our purposes, an entrywise type of matrix norm is preferable and suitable as well. Accordingly, the *Frobenius norm*—a common and widely used representative from the family of entrywise matrix norms—will serve us to investigate the properties of the EIV model. The Frobenius matrix norm for matrix $\mathbf{A} \equiv (a_{ij})_{i,j=1}^{n,p}$ is defined as follows

$$\|\mathbf{A}\|_F := \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{p} a_{ij}^2} = \sqrt{tr(\mathbf{A}^\top \mathbf{A})} = \sqrt{\sum_{i=1}^{q} \sigma_i^2} = \sqrt{\sum_{i=1}^{r} \sigma_i^2}, \quad q = \min\{n, p\}. \tag{1.4}$$

The Frobenius norm can be viewed as a *multivariate* version of the Euclidean vector norm for matrices.

Furthermore, the following approximation theorem plays the main role in the forthcoming derivation, where a matrix is approximated with another one having lower rank.

Theorem 1.2 (Eckart-Young-Mirsky matrix approximation). *Let the SVD of* $\mathbf{A} \in \mathbb{R}^{n\times p}$

be given by $\mathbf{A} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$ *with* $rank(\mathbf{A}) = r$. *If* $k < r$ *and* $\mathbf{A}_k = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$,

$$\min_{rank(\mathbf{B})=k} \|\mathbf{A} - \mathbf{B}\|_F = \|\mathbf{A} - \mathbf{A}_k\|_F = \sqrt{\sum_{i=k+1}^{r} \sigma_i^2}.$$

Proof. See Eckart and Young (1936) and Mirsky (1960). □

Above all, one more technical property needs to be recalled.

Theorem 1.3 (Sturm interlacing property)**.** *Let* $n \geq p$ *and the singular values of* $\mathbf{A} \in \mathbb{R}^{n \times p}$ *are* $\sigma_1 \geq \ldots \geq \sigma_p$. *If* $\mathbf{B}$ *results from* $\mathbf{A}$ *by deleting one column of* $\mathbf{A}$ *and* $\mathbf{B}$ *has singular values* $\sigma_1' \geq \ldots \geq \sigma_{p-1}'$, *then*

$$\sigma_1 \geq \sigma_1' \geq \sigma_2 \geq \sigma_2' \geq \ldots \geq \sigma_{p-1}' \geq \sigma_p \geq 0. \tag{1.5}$$

Proof. See Thompson (1972). □

1.3.2 Total Least Squares Solution

Three basic approximation ways of the linear overdetermined system (1.1) are described in Pešta (2008). The traditional approach penalizes only the misfit in the dependent variable part

$$\min_{\boldsymbol{\epsilon} \in \mathbb{R}^n, \boldsymbol{\beta} \in \mathbb{R}^p} \|\boldsymbol{\epsilon}\|_2 \qquad \text{s.t.} \quad \mathbf{Y} - \boldsymbol{\epsilon} = \mathbf{X}\boldsymbol{\beta} \tag{1.6}$$

and is called the *ordinary least squares* (OLS). Here, the data matrix $\mathbf{X}$ is thought as exactly known and errors occur only in the vector $\mathbf{Y}$. An opposite case to the OLS is represented by the *data least squares* (DLS), which allow corrections only in the explanatory variables (independent input data)

$$\min_{\boldsymbol{\Xi} \in \mathbb{R}^{n \times p}, \boldsymbol{\beta} \in \mathbb{R}^p} \|\boldsymbol{\Xi}\|_F \qquad \text{s.t.} \quad \mathbf{Y} = (\mathbf{X} - \boldsymbol{\Xi})\boldsymbol{\beta}. \tag{1.7}$$

Finally, we concentrate ourselves on the *total least squares* approach minimizing the squares of errors in the values of both dependent and independent variables

$$\min_{[\boldsymbol{\Theta}, \boldsymbol{\varepsilon}] \in \mathbb{R}^{n \times (p+1)}, \boldsymbol{\beta} \in \mathbb{R}^p} \|[\boldsymbol{\Theta}, \boldsymbol{\varepsilon}]\|_F \qquad \text{s.t.} \quad \mathbf{Y} - \boldsymbol{\varepsilon} = (\mathbf{X} - \boldsymbol{\Theta})\boldsymbol{\beta}. \tag{1.8}$$

A graphical illustration of three previous cases can be found in Figure 1.2. One may notice that the TLS "search" for the *orthogonal projection* of the observed data onto the unknown approximation corresponding to a TLS solution. The Frobenius norm is chosen as a suitable (and also a standard) norm to penalize for the errors, because, geometrically speaking, it tries to minimize the orthogonal distance between the observations and fitted hyperplane.

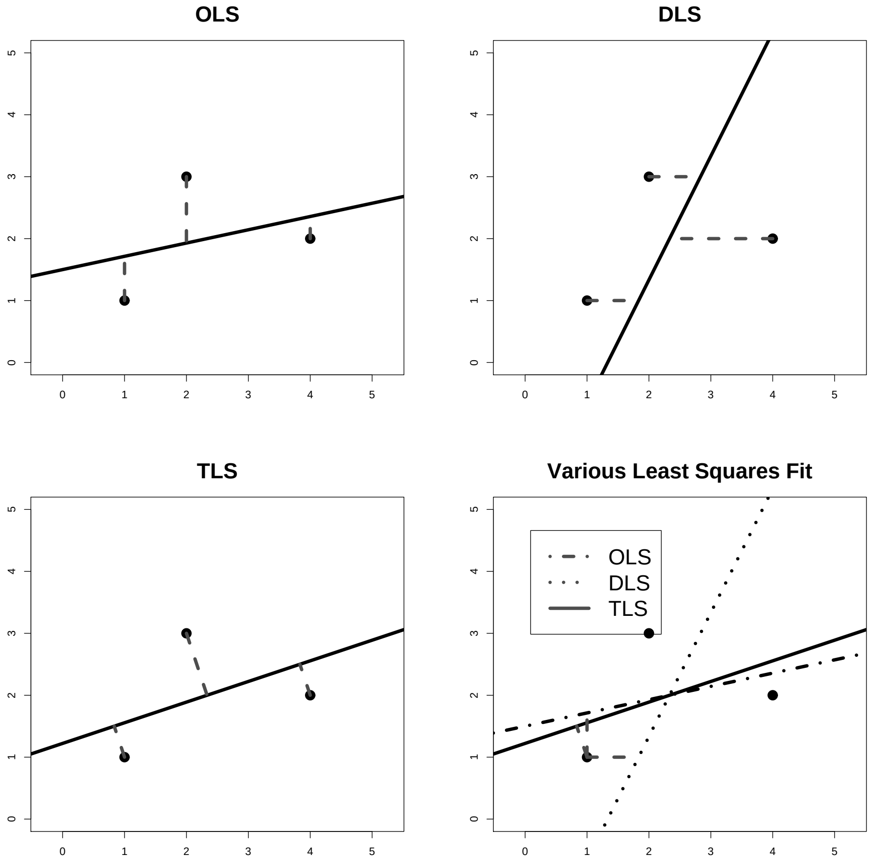

Figure 1.2: Various least squares fits (ordinary, data, and total LS) for the same three data points in the two-dimensional plane that coincides with the regression setup of one response and one explanatory variable.

Once a minimizing $[\widehat{\mathbf{\Theta}}, \widehat{\boldsymbol{\varepsilon}}]$ of the TLS problem (1.8) is found, then any $\boldsymbol{\beta}$ satisfying $\mathbf{Y} - \widehat{\boldsymbol{\varepsilon}} = (\mathbf{X} - \widehat{\mathbf{\Theta}})\boldsymbol{\beta}$ is called a *TLS solution*. The "basic" form of the TLS solution was investigated for the first time by Golub and Van Loan (1980).

Theorem 1.4 (TLS solution of $\mathbf{Y} \approx \mathbf{X}\boldsymbol{\beta}$)**.** *Let the SVD of* $\mathbf{X} \in \mathbb{R}^{n\times p}$ *be given by* $\mathbf{X} = \sum_{i=1}^{p} \sigma_i' \mathbf{u}_i' \mathbf{v}_i'^{\top}$ *and the SVD of* $[\mathbf{X}, \mathbf{Y}] = \sum_{i=1}^{p+1} \sigma_i \mathbf{u}_i \mathbf{v}_i^{\top}$. *If* $\sigma_p' > \sigma_{p+1}$, *then*

$$[\widehat{\mathbf{X}}, \widehat{\mathbf{Y}}] := [\mathbf{X} - \widehat{\mathbf{\Theta}}, \mathbf{Y} - \widehat{\boldsymbol{\varepsilon}}] = \mathbf{U}\widehat{\mathbf{\Sigma}}\mathbf{V}^{\top} \quad \textit{and} \quad \widehat{\mathbf{\Sigma}} = \textit{diag}\,\{\sigma_1, \ldots, \sigma_p, 0\} \tag{1.9}$$

with the corresponding TLS correction matrix

$$[\widehat{\mathbf{\Theta}}, \widehat{\boldsymbol{\varepsilon}}] = \sigma_{p+1}\mathbf{u}_{p+1}\mathbf{v}_{p+1}^{\top} \tag{1.10}$$

solves the TLS problem and

$$\widehat{\boldsymbol{\beta}} = -\frac{1}{\mathbf{e}_{p+1}^\top \mathbf{v}_{p+1}} \left[v_{1,p+1}, \ldots, v_{p,p+1}\right]^\top \tag{1.11}$$

exists and is the unique solution to $\widehat{\mathbf{Y}} = \widehat{\mathbf{X}}\boldsymbol{\beta}$.

Proof. Proof by contradiction, we firstly show that $\mathbf{e}_{p+1}^\top \mathbf{v}_{p+1} \neq 0$. Suppose $v_{p+1,p+1} = 0$, then there exist $\mathbf{0} \neq \mathbf{w} \in \mathbb{R}^p$ such that

$$\left[\mathbf{w}^\top, 0\right] \left[\mathbf{X}, \mathbf{Y}\right]^\top \left[\mathbf{X}, \mathbf{Y}\right] \begin{bmatrix} \mathbf{w} \\ 0 \end{bmatrix} = \sigma_{p+1}^2$$

which yields into $\mathbf{w}^\top \mathbf{X}^\top \mathbf{X} \mathbf{w} = \sigma_{p+1}^2$. But this is a contradiction with the assumption $\sigma_p' > \sigma_{p+1}$, since $\sigma_p'^2$ is the smallest eigenvalue of $\mathbf{X}^\top \mathbf{X}$.

Sturm interlacing property (1.5) and the assumption $\sigma_p' > \sigma_{p+1}$ yield $\sigma_p > \sigma_{p+1}$. Therefore, σ_{p+1} is not a repeated singular value of $[\mathbf{X}, \mathbf{Y}]$ and $\sigma_p > 0$.

If $\sigma_{p+1} \neq 0$, then $rank([\mathbf{X}, \mathbf{Y}]) = p + 1$. We want to find $[\widehat{\mathbf{X}}, \widehat{\mathbf{Y}}]$ such that $\|[\mathbf{X}, \mathbf{Y}] - [\widehat{\mathbf{X}}, \widehat{\mathbf{Y}}]\|_F$ is minimal and $[\widehat{\mathbf{X}}, \widehat{\mathbf{Y}}][\boldsymbol{\beta}^\top, -1]^\top = \mathbf{0}$ for some $\boldsymbol{\beta}$. Therefore, $rank([\widehat{\mathbf{X}}, \widehat{\mathbf{Y}}]) = p$ and applying Eckart-Young-Mirsky Theorem 1.2, one may easily obtain the SVD of $[\widehat{\mathbf{X}}, \widehat{\mathbf{Y}}]$ in (1.9) and the TLS correction matrix (1.10), which must have rank one. Now, it is clear that the TLS solution is given by the last column of $\mathbf{V}$. Finally, since $dim(Ker([\widehat{\mathbf{X}}, \widehat{\mathbf{Y}}])) = 1$, then the TLS solution (1.11) must be unique.

If $\sigma_{p+1} = 0$, then $\mathbf{v}_{p+1} \in Ker([\mathbf{X}, \mathbf{Y}])$ and $[\mathbf{X}, \mathbf{Y}][\boldsymbol{\beta}^\top, -1]^\top = \mathbf{0}$. Hence, no approximation is needed, overdetermined system (1.1) is compatible, and the exact TLS solution is given by (1.11). Uniqueness of this TLS solution follows from the fact that $[\boldsymbol{\beta}^\top, -1]^\top \perp Range([\mathbf{X}, \mathbf{Y}]^\top)$. □

A *closed-form* expression of the TLS solution (1.11) can be derived. If $\sigma_p' > \sigma_{p+1}$, the *existence* and *uniqueness* of the TLS solution has already been shown. Since singular vectors $\mathbf{v}_i$ from (1.9) are eigenvectors of $[\mathbf{X}, \mathbf{Y}]^\top [\mathbf{X}, \mathbf{Y}]$, then $\widehat{\boldsymbol{\beta}}$ also satisfies

$$[\mathbf{X}, \mathbf{Y}]^\top [\mathbf{X}, \mathbf{Y}] \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ -1 \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top \mathbf{Y} & \mathbf{X}^\top \mathbf{Y} \\ \mathbf{Y}^\top \mathbf{X} & \mathbf{Y}^\top \mathbf{Y} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ -1 \end{bmatrix} = \sigma_{p+1}^2 \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ -1 \end{bmatrix} \tag{1.12}$$

and, hence,

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X} - \sigma_{p+1}^2 \mathbf{I})^{-1} \mathbf{X}^\top \mathbf{Y}. \tag{1.13}$$

Previous equation reminds us a form of an estimate in the *ridge regression* setup. Therefore, one may expect that the TLS estimate avoid multicollinearity problems, which are sometimes present in the classical OLS regression (1.6). Expression (1.13) looks almost similar to the

OLS estimator $\widetilde{\boldsymbol{\beta}}$ of (1.6), except the term containing σ_{p+1}^2. This term is missing in the well-known OLS estimator with full rank regression matrix provided by the *Gauss-Markov theorem* as a solution of so-called *normal equations* $\mathbf{X}^\top\mathbf{X}\widetilde{\boldsymbol{\beta}} = \mathbf{X}^\top\mathbf{Y}$.

From a statistical point of view, a situation when $\sigma_p' = \sigma_{p+1}$ occurs for real data is unlikely and also quite irrelevant. In spite of this, Van Huffel and Vandewalle (1991, Chapter 5) investigated this case and concluded the following summary. Suppose $\sigma_q > \sigma_{q+1} = \ldots = \sigma_{p+1}$, $q \leq p$ and denote $\mathbf{Q} := [\mathbf{v}_{q+1}, \ldots, \mathbf{v}_{p+1}]$. Then:

- $\sigma_p' > \sigma_{p+1} \quad \Rightarrow \quad$ the unique TLS solution (1.11) exists;
- $\sigma_p' = \sigma_{p+1} \quad \& \quad \mathbf{e}_{p+1}^\top\mathbf{Q} \neq \mathbf{0} \quad \Rightarrow \quad$ infinitely many TLS solutions of (1.8) exist and one can pick up one of them with the smallest norm;
- $\sigma_p' = \sigma_{p+1} \quad \& \quad \mathbf{e}_{p+1}^\top\mathbf{Q} = \mathbf{0} \quad \Rightarrow \quad$ no solution of (1.8) exists and one needs to define another ("more restrictive") TLS problem.

A more restrictive TLS problem, than the original one mentioned previously, is called a *non-generic* TLS problem. Simply, additional restriction

$$[\boldsymbol{\varepsilon}, \boldsymbol{\Xi}]\,\mathbf{Q} = \mathbf{0}. \tag{1.14}$$

is added to the former constraints (1.8). Restriction (1.14) tries to "project" out "unimportant" or "redundant" data from the original TLS problem (1.8). Moreover, restriction (1.14) assures a uniqueness of the solution for such an updated TLS problem (TLS problem with additional restriction). A detailed discussion can be found in Van Huffel and Vandewalle (1991, Chapter 3).

1.4 Error Structure

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 $[\boldsymbol{\Theta}_{i,\bullet}, \varepsilon_i]$ has *zero mean* and nonsingular covariance matrix $\sigma^2\mathbf{I}$, where $\sigma^2 > 0$ is unknown (for simplicity). This assumption can be straightforwardly generalized as discussed below. Relationships between individual observations are represented by the second level of error structure—*between-individual level*. Here, the rows $[\boldsymbol{\Theta}_{i,\bullet}, \varepsilon_i]$ are *iid*. This assumption may seem quite restrictive and unrealizable in some situations and, therefore, it will be generalized in Chapter 4.

A homoscedastic covariance structure of the within-individual errors $[\boldsymbol{\Theta}_{i,\bullet}, \varepsilon_i]$ can be *generalized by knowing the heteroscedastic covariance matrix* $\boldsymbol{\Gamma} > \mathbf{0}$ in advance. Mathematically speaking, the homoscedastic covariance matrix $\sigma^2\mathbf{I}$ can be replaced by more general one $\boldsymbol{\Gamma} \in \mathbb{R}^{(p+1)\times(p+1)}$. Then, the observation data are just multiplied by its square root as already discussed in Van Huffel and Vandewalle (1991, Section 8.4) or Gleser (1981, Section 5), i.e., new transformed data are

$$[\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}] = [\mathbf{X}, \mathbf{Y}]\boldsymbol{\Gamma}^{-1/2}. \tag{1.15}$$

This transformation of the original data is purely linear, which is not restrictive at all in our case. Since no specific distributional assumptions on the errors are assumed in the whole thesis except the existence of some moments of errors, linear transformation $[\boldsymbol{\Theta}_{i,\bullet}, \varepsilon_i]\boldsymbol{\Gamma}^{-1/2}$ does not require any additional distributional assumption. Therefore, the whole asymptotic inference remain also valid even for the heteroscedastic case. The only property that needs to be satisfied is independence of the transformed errors.

Consequently, TLS estimate (1.13) can be rewritten and a *generalized total least squares* (GTLS) estimate is obtained

$$\widehat{\boldsymbol{\beta}}_{GTLS} = (\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}} - \sigma^2_{p+1,[\widetilde{\mathbf{X}},\widetilde{\mathbf{Y}}]}\mathbf{I})^{-1}\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{Y}}, \tag{1.16}$$

where $\sigma_{p+1,[\widetilde{\mathbf{X}},\widetilde{\mathbf{Y}}]}$ is the $(p+1)$-st singular value of transformed data matrix $[\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}]$. If the covariance matrix $\boldsymbol{\Gamma}$ is unknown, it can be estimated using *repeated observations*, but afterwards, more complicated design of the experiment is necessary:

$$\underset{n\times 1}{\mathbf{Y}_\iota} = \underset{n\times p}{\mathbf{Z}}\,\underset{p\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\varepsilon}_\iota} \quad \text{and} \quad \underset{n\times p}{\mathbf{X}_\iota} = \underset{n\times p}{\mathbf{Z}} + \underset{n\times p}{\boldsymbol{\Theta}_\iota}, \quad \iota = 1, \ldots, r; \tag{1.17}$$

where $r \in \mathbb{N}$ stands for the number of *replications*. Extra information—needed for an estimation of the covariance matrix—comes exactly from the replications. Then, a general covariance matrix $\boldsymbol{\Sigma}$ for the within-individual errors can be estimated as in Healy (1975), e.g., by

$$\widehat{\boldsymbol{\Gamma}} := \frac{1}{n(r-1)} \sum_{i=1}^{r} \sum_{j=1}^{r} [\mathbf{X}_i, \mathbf{Y}_i]^\top \left[\left(\delta_{ij} - \frac{1}{r}\right)\mathbf{I}\right] [\mathbf{X}_j, \mathbf{Y}_j], \tag{1.18}$$

where δ_{ij} denotes Kronecker delta. Previous equation (1.18) can be rewritten using notation purely from replication model (1.17) as

$$\widehat{\boldsymbol{\Gamma}} = \frac{1}{nr} \sum_{i=1}^{r} [\boldsymbol{\Theta}_i, \boldsymbol{\varepsilon}_i]^\top [\boldsymbol{\Theta}_i, \boldsymbol{\varepsilon}_i] - \frac{1}{nr(r-1)} \sum_{i=1}^{r} \sum_{\substack{j=1\\ j\neq i}}^{r} [\boldsymbol{\Theta}_i, \boldsymbol{\varepsilon}_i]^\top [\boldsymbol{\Theta}_j, \boldsymbol{\varepsilon}_j],$$

which illustrates the meaning of the estimate.

A similar situation arises on the between-individual level of errors. If there exist *equilibration matrices* $\boldsymbol{\Upsilon} \in \mathbb{R}^{n\times n}$ and $\boldsymbol{\Gamma} \in \mathbb{R}^{(p+1)\times(p+1)}$, which provide *iid* errors $\boldsymbol{\Upsilon}^{-1/2}[\boldsymbol{\Theta}, \boldsymbol{\varepsilon}]\boldsymbol{\Gamma}^{-1/2}$ (on both levels of errors), the inference should be performed for the transformed data

$\boldsymbol{\Upsilon}^{-1/2}[\mathbf{X},\mathbf{Y}]\boldsymbol{\Gamma}^{-1/2}$. Hence, TLS estimate (1.13) become an *equilibrated total least squares* (ETLS) estimate

$$\widehat{\boldsymbol{\beta}}_{ETLS} = (\widetilde{\mathbf{X}}^\top\boldsymbol{\Upsilon}^{-1}\widetilde{\mathbf{X}} - \sigma^2_{p+1,\boldsymbol{\Upsilon}^{-1/2}[\mathbf{X},\mathbf{Y}]\boldsymbol{\Gamma}^{-1/2}}\mathbf{I})^{-1}\widetilde{\mathbf{X}}^\top\boldsymbol{\Upsilon}^{-1}\widetilde{\mathbf{Y}}, \tag{1.19}$$

where $\sigma_{p+1,\boldsymbol{\Upsilon}^{-1/2}[\mathbf{X},\mathbf{Y}]\boldsymbol{\Gamma}^{-1/2}}$ is the $(p+1)$-st singular value of transformed data matrix $\boldsymbol{\Upsilon}^{-1/2}[\mathbf{X},\mathbf{Y}]\boldsymbol{\Gamma}^{-1/2}$.

1.5 Partial Errors-in-Variables Model

Partial errors-in-variables (PEIV) model is a regression model where some explanatory variables are *subject to error* and some are measured *exactly*. It is an extension of EIV model (E):

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times s}{\mathbf{W}}\,\underset{s\times 1}{\boldsymbol{\alpha}} + \underset{n\times p}{\mathbf{Z}}\,\underset{p\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\boldsymbol{\varepsilon}} \quad \text{and} \quad \underset{n\times p}{\mathbf{X}} = \underset{n\times p}{\mathbf{Z}} + \underset{n\times p}{\boldsymbol{\Theta}}, \tag{1.20}$$

where $\mathbf{W}$ are *observable true* and $\mathbf{Z}$ are *unobservable true* constants, both having full rank. Regression parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ needs to be estimated. The rest of the notation is the same as for the classical EIV model.

PEIV model (1.20) is quite important, because it can incorporate non-random (fixed) *intercept* into the regression model. This can be directly proceeded by setting one column of matrix $\mathbf{W}$ equal to $[1,\ldots,1]^\top$. On the other hand, if $p=0$, then we just have a *classical linear regression model*.

In order to obtain parameter estimates, we *separate* exact and approximate observations, or better to say, we decompose "space of observations"—fixed and random predictors, and response—into mutually orthogonal subspaces. Then, we *project out* exact observations using projection matrix $\mathbf{R} := \mathbf{I} - \mathbf{W}(\mathbf{W}^\top\mathbf{W})^{-1}\mathbf{W}^\top$. Notice that $\mathbf{R} = \mathbf{R}^\top$ and $\mathbf{R} = \mathbf{R}\mathbf{R}$. We end up with

$$\mathbf{R}\mathbf{Y} = \mathbf{R}\mathbf{Z}\boldsymbol{\beta} + \mathbf{R}\boldsymbol{\varepsilon} \tag{1.21}$$

and, consequently, minimizing the orthogonal distance

$$\min_{\boldsymbol{\beta},[\widetilde{\mathbf{Y}},\widetilde{\mathbf{X}}]} \left\|\mathbf{R}[\mathbf{Y},\mathbf{X}] - [\widetilde{\mathbf{Y}},\widetilde{\mathbf{X}}]\right\|_F \quad \text{s.t.} \quad \widetilde{\mathbf{Y}} = \widetilde{\mathbf{X}}\boldsymbol{\beta}$$

yields the estimate for parameter $\boldsymbol{\beta}$, i.e.,

$$\widehat{\boldsymbol{\beta}}_{LS-TLS} = (\mathbf{X}^\top\mathbf{R}\mathbf{X} - \sigma^2_{p+1,\mathbf{R}[\mathbf{X},\mathbf{Y}]}\mathbf{I})^{-1}\mathbf{X}^\top\mathbf{R}\mathbf{Y}, \tag{1.22}$$

where $\sigma_{p+1,\mathbf{R}[\mathbf{X},\mathbf{Y}]}$ is the $(p+1)$-st singular value of the projected data matrix $\mathbf{R}[\mathbf{X},\mathbf{Y}]$.

Now, an ordinary least squares estimate is calculated for parameter $\boldsymbol{\alpha}$ in the "transformed" model

$$\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{LS-TLS} = \mathbf{W}\boldsymbol{\alpha} + \widetilde{\boldsymbol{\varepsilon}}$$

with some unspecified errors $\widetilde{\boldsymbol{\varepsilon}}$. Finally,

$$\widehat{\boldsymbol{\alpha}} = (\mathbf{W}^\top\mathbf{W})^{-1}\mathbf{W}^\top(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_{LS-TLS}).$$

Henceforth, estimate $\widehat{\boldsymbol{\beta}}_{LS-TLS}$ from (1.22) is called a *least squares-total least squares* (LS-TLS) estimate.

Golub et al. (1987) showed that the above procedure for PEIV model (1.20) finds the minimum of the following rank-deficiency optimization problem

$$\min_{[\boldsymbol{\Theta},\boldsymbol{\varepsilon}]\in\mathbb{R}^{n\times(p+1)}} \|[\boldsymbol{\Theta},\boldsymbol{\varepsilon}]\|_F \qquad \text{s.t.} \quad \mathit{Range}(\mathbf{Y}-\boldsymbol{\varepsilon}) \subseteq \mathit{Range}(\mathbf{X}-\boldsymbol{\Theta}) \subseteq \mathit{Range}([\mathbf{W},\mathbf{X}-\boldsymbol{\Theta}]). \tag{1.23}$$

The obtained solution of minimizing (1.23) is called a *mixed least squares – total least squares* (mixed LS-TLS) estimate and is given by

$$\begin{bmatrix} \widehat{\boldsymbol{\alpha}} \\ \widehat{\boldsymbol{\beta}} \end{bmatrix}_{LS-TLS} = \left([\mathbf{W},\mathbf{X}]^\top[\mathbf{W},\mathbf{X}] - \sigma_{p+1}^2 \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \right)^{-1} [\mathbf{W},\mathbf{X}]^\top\mathbf{Y},$$

see, e.g., Gleser (1981).

The previous results can be thought as a *justification* that it is sufficient to take into account only EIV model (E). Hence, all properties for the classical EIV model, which are going to be derived, can be straightforwardly generalized for PEIV model (1.20) as well, since the inference for the classical linear regression is well known.

1.6 Summary of Large Sample Properties

One should not only pay attention to the existence or form of the TLS solution, but also to its properties, e.g., statistical ones. An asymptotical behavior of an estimator is one of its basic characteristics. The asymptotical properties can provide some information about the quality (i.e., efficiency) of the estimator.

Sometimes, full-information approaches like *maximum likelihood* (ML) can provide parameter estimates for the previously mentioned model. Nevertheless, it is requisite to go for *distributional-free estimation* method, e.g., total least squares, due to the impossibility to satisfy the distributional assumptions in the model. Healy (1975) developed a *maximum* *likelihood* theory for the EIV model under some *normality assumptions* on the errors $[\boldsymbol{\Theta}, \boldsymbol{\varepsilon}]$. Finally, let us remind that the ML estimate of $\boldsymbol{\beta}$ (Healy, 1975) *coincides* with the TLS estimate if the rows of the error matrix are *iid multivariate normal* with zero mean and non-singular covariance matrix.

First of all, Okamoto (1973) and Gleser (1981, Section 2) proved that with *probability tending to one*, as n increases, $v_{p+1,p+1} \neq 0$ (and, hence, $\widehat{\boldsymbol{\beta}}$ exists). Furthermore, Gallo (1982b) remarked that the ordinary least squares estimate in the EIV model is *inconsistent* and TLS estimate (1.13) should be taken into account instead.

Additional *design assumption* is necessary for asymptotics:

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

Importance of the previous design assumption has already been thoroughly discussed in, e.g., Pešta (2009b). If the limit in (D) is infinite, then the variance of TLS estimate (1.13) tends to zero, which is inadmissible.

1.6.1 Consistency

Firstly, we provide a theorem showing the strong consistency of the TLS estimator.

Theorem 1.5 (Strong consistency in EIV with independent errors)**.** *If* $\lim_{n\to\infty} \frac{1}{n}\mathbf{Z}^\top\mathbf{Z}$ *exists, then*

$$\lim_{n\to\infty} \frac{\sigma_{p+1}^2}{n} = \sigma^2 \quad a.s. \tag{1.24}$$

Moreover, if assumption (D) *is satisfied, then*

$$\lim_{n\to\infty} \widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} \quad a.s. \tag{1.25}$$

Proof. See Gleser (1981, Corollary 3.1 and Lemma 3.3). □

Previous Theorem 1.5 also provides a strongly consistent estimate of the *homoscedastic variance* parameter σ^2, i.e.:

$$\widehat{\sigma}^2 := \frac{\sigma_{p+1}^2}{n}. \tag{1.26}$$

Assuming (D), Gleser (1981, Lemma 3.1) showed the following very important relation

$$\boldsymbol{\Delta}_n := n^{-1}(\mathbf{X}^\top\mathbf{X} - \sigma_{p+1}^2\mathbf{I}) \xrightarrow{a.s.} \boldsymbol{\Delta}, \quad n \to \infty. \tag{1.27}$$

The assumptions in the previous theorem are somewhat restrictive and need not be satisfied, e.g., univariate errors-in-variables model with the values of the independent variable vary linearly with the sample size (Gallo, 1982b). Therefore, these assumptions need to be weakened yielding the following theorem.

Theorem 1.6 (Weak consistency in EIV with independent errors)**.** *Suppose that the distribution of the rows of* $[\boldsymbol{\Theta}, \boldsymbol{\varepsilon}]$ *possesses finite fourth moment. If*

$$\begin{aligned} \frac{1}{\sqrt{n}}\lambda_{min}\left(\mathbf{Z}^\top\mathbf{Z}\right) &\to \infty, \quad n \to \infty, \\ \frac{\lambda_{min}^2\left(\mathbf{Z}^\top\mathbf{Z}\right)}{\lambda_{max}\left(\mathbf{Z}^\top\mathbf{Z}\right)} &\to \infty, \quad n \to \infty; \end{aligned}$$

then

$$\widehat{\boldsymbol{\beta}} \xrightarrow{\mathbb{P}} \boldsymbol{\beta}, \quad n \to \infty. \tag{1.28}$$

Proof. Can be easily derived using Theorem 2 by Gallo (1982a). □

Notation λ_{min} (respectively, λ_{max}) denotes the minimal (respectively, maximal) eigenvalue. It has to be remarked on the *fourth moment finiteness* of the rows of $[\boldsymbol{\Theta}, \boldsymbol{\varepsilon}]$, that this mathematically means for all $i \in \{1, \ldots, n\}$

$$\mathbb{E} \prod_{\sum_j r_j = 4} \omega_{ij}^{r_j} < \infty, \quad \omega_{ij} \in \left\{\boldsymbol{\Theta}_{i,1}, \ldots \boldsymbol{\Theta}_{i,p}, \varepsilon_i\right\}, \; r_j \in \mathbb{N}.$$

The assumptions in the previous theorems ensure that the values of the independent variables "spread out" fast enough. Gallo (1982a) proved that the previous "intermediate" assumptions are implied by the assumptions in the theorem for strong consistency.

Results for strong and weak consistency were strengthen by Kukush et al. (2002, Theorem 2) considering milder assumptions, but rather complicated to verify.

1.6.2 Asymptotic Normality

Finally, an asymptotic distribution for further statistical inference has to be shown.

Theorem 1.7 (Asymptotic normality in EIV with independent errors)**.** *Suppose that the distribution of the rows of* $[\boldsymbol{\Theta}, \boldsymbol{\varepsilon}]$ *possesses finite fourth moment. If assumption* (D) *is satisfied, then* $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ *has an asymptotic zero-mean multivariate normal distribution as* $n \to \infty$.

Proof. See Gallo (1982b, Chapter 3). □

The *covariance* matrix of the multivariate normal distribution from the previous theorem is not shown here due to its very *complicated form*, which can be calculated from Theorem 4.1 by Gleser (1981) or derived from the proof of Theorem 3.3 by Gallo (1982b).

Moreover, if the *third and fourth moments* of the distribution of the rows of $[\mathbf{\Theta}, \boldsymbol{\varepsilon}]$ are the same as those of the normal distribution and assumption (D) holds, then

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathscr{D}} \mathscr{N}_p\left(\mathbf{0}, \sigma^2(1 + \boldsymbol{\beta}^\top \boldsymbol{\beta})\left\{\mathbf{\Delta}^{-1} + \sigma^2 \mathbf{\Delta}^{-1}\left([\mathbf{I}, \boldsymbol{\beta}][\mathbf{I}, \boldsymbol{\beta}]^\top\right)^{-1}\mathbf{\Delta}^{-1}\right\}\right), \quad n \to \infty. \quad (1.29)$$

1.7 Discussion and Conclusions

In this chapter, the TLS problem from algebraical point of view is summarized and a connection with the errors-in-variables—a statistical model—was shown. Unification of algebraical and numerical results with statistical ones were demonstrated.

The TLS optimizing problem was defined here with the OLS and DLS alternatives. Its solution was found using spectral information of the system; and the existence and uniqueness of this solution were discussed. The errors-in-variables model as a correspondence to the orthogonal regression was introduced. Moreover, a comparison of the classical regression approach with the errors-in-variables setup was shown. Extensions of the EIV model were proposed yielding GTLS, ETLS and mixed LS-TLS. Finally, large sample properties such as the strong and weak consistency, and the asymptotical normality of the TLS estimate—an estimate in the errors-in-variables model—were recapitulated.

Chapter 2

Estimation in Errors-in-Variables

In ancient times they had no statistics
so they had to fall back on lies.

STEPHEN B. LEACOCK

The estimation in errors-in-variables model (E) was performed via penalizing the orthogonal squared misfit. This attitude leads into the minimizing the Frobenius norm of the error matrix. Immediate doubts arise whether this criterion is suitable and in which sense. What happens if we do not consider squared distances? Moreover, a change of measurement units should have no impact on the estimate and an interchange of variables should provide estimate with its exchanged components.

2.1 Unitarily Invariant Matrix Norms

A broad class of matrix norms is introduced in order to penalize errors from the EIV model in a general manner.

Definition 2.1 (Unitarily invariant matrix norm)**.** A matrix norm $\|\cdot\|$ is *unitarily invariant* if

$$\|\mathbf{UAV}\| = \|\mathbf{A}\|$$

for all $\mathbf{A} \in \mathbb{R}^{n\times p}$ and all unitary matrices (see Definition A.1) $\mathbf{U} \in \mathbb{R}^{n\times n}$ and $\mathbf{V} \in \mathbb{R}^{p\times p}$.

A *complete characterization* of the unitarily invariant matrix norms was given by von Neumann (1937) using the following definition and consequent theorem.

Definition 2.2 (Symmetric gauge function)**.** A real-valued function $\varsigma : \mathbb{R}^p \to \mathbb{R}$ is *symmetric gauge function* if $\varsigma(\cdot)$ is a vector norm satisfying

$$\varsigma(\mathbf{PDx}) = \varsigma(\mathbf{x})$$

for all $\mathbf{x} \in \mathbb{R}^p$, all diagonal matrices $\mathbf{D} \in \mathbb{R}^{p\times p}$ having ± 1 on the diagonal, and all permutation matrices $\mathbf{P} \in \mathbb{R}^{p\times p}$ (see Definition A.2).

Theorem 2.1 (Characterization of unitarily invariant matrix norms)**.** *Any unitarily invariant matrix norm* $\|\cdot\|$ *is of the form* $\varsigma([\sigma_1(\cdot), \ldots, \sigma_p(\cdot)]^\top)$, *where* $\sigma_1(\cdot) \geq \ldots \geq \sigma_p(\cdot) \geq 0$ *are singular values of the corresponding matrix.*

Proof. See von Neumann (1937). □

Theorem 2.1 states that a unitarily invariant matrix norm is a symmetric gauge function of the singular values of its argument.

Eckart-Young-Mirsky matrix approximation Theorem 1.2 can be extended, which provides a broader class of estimates. This generalization uses a similar lower-rank matrix approximation.

Theorem 2.2 (Generalized Schmidt matrix approximation)**.** *Let the SVD of* $\mathbf{A} \in \mathbb{R}^{n\times p}$ *be given by* $\mathbf{A} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$ *with* $rank(\mathbf{A}) = r$. *If* $k < r$ *and* $\mathbf{A}_k = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$, *then for any unitarily invariant matrix norm* $\|\cdot\|$ *holds*

$$\min_{rank(\mathbf{B})=k} \|\mathbf{A} - \mathbf{B}\| = \|\mathbf{A} - \mathbf{A}_k\|.$$

Proof. See Mirsky (1960). □

2.2 EIV Estimate

Having general penalization criterion provided by unitarily invariant matrix norms, a solution of overdetermined system (1.1) is found. Henceforth, an estimate to EIV model (E) can be constructed. It can be expected that this new estimate should be more general in some sense than the TLS estimate, but we end up with a surprise.

Theorem 2.3 (Errors-in-variables estimate)**.** *Suppose that* $\|\cdot\|$ *is an arbitrary unitarily invariant matrix norm. Let the SVD of* $\mathbf{X} \in \mathbb{R}^{n\times p}$ *be given by* $\mathbf{X} = \sum_{i=1}^{p} \sigma_i' \mathbf{u}_i' \mathbf{v}_i'^\top$ *and the SVD of* $[\mathbf{X}, \mathbf{Y}] = \sum_{i=1}^{p+1} \sigma_i \mathbf{u}_i \mathbf{v}_i^\top$. *If* $\sigma_p' > \sigma_{p+1}$, *then the optimizing problem*

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p, [\boldsymbol{\Theta},\varepsilon]\in\mathbb{R}^{n\times(p+1)}} \|[\boldsymbol{\Theta}, \varepsilon]\| \qquad s.t. \quad \mathbf{Y} - \boldsymbol{\varepsilon} = (\mathbf{X} - \boldsymbol{\Theta})\boldsymbol{\beta} \tag{2.1}$$

has always a unique solution $\{\widehat{\boldsymbol{\beta}}, [\widehat{\boldsymbol{\Theta}}, \widehat{\varepsilon}]\}$ *for* $\{\boldsymbol{\beta}, [\boldsymbol{\Theta}, \varepsilon]\}$ *consisting of an errors-in-variables (EIV) estimate*

$$\widehat{\boldsymbol{\beta}} = -\frac{1}{\mathbf{e}_{p+1}^\top \mathbf{v}_{p+1}} [v_{1,p+1}, \ldots, v_{p,p+1}]^\top \tag{2.2}$$

and a correction (residual) matrix

$$[\widehat{\boldsymbol{\Theta}}, \widehat{\varepsilon}] = \sigma_{p+1} \mathbf{u}_{p+1} \mathbf{v}_{p+1}^\top. \tag{2.3}$$

The norm of the correction matrix is $\|\widehat{\boldsymbol{\Theta}}, \widehat{\varepsilon}\| = \sigma_{p+1}$ *and the corresponding fitted matrix is*

$$[\widehat{\mathbf{X}}, \widehat{\mathbf{Y}}] := [\mathbf{X} - \widehat{\boldsymbol{\Theta}}, \mathbf{Y} - \widehat{\varepsilon}] = \mathbf{U}\widehat{\boldsymbol{\Sigma}}\mathbf{V}^\top, \tag{2.4}$$

where $\widehat{\boldsymbol{\Sigma}} = diag\{\sigma_1, \ldots, \sigma_p, 0\}$. *Moreover, EIV estimate* (2.2) *has an alternative (closed) form*

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top \mathbf{X} - \lambda \mathbf{I})^{-1} \mathbf{X}^\top \mathbf{Y}, \tag{2.5}$$

where λ *is the smallest eigenvalue of matrix* $[\mathbf{X}, \mathbf{Y}]^\top [\mathbf{X}, \mathbf{Y}]$.

Proof. The proof of (2.2)–(2.4) is just a copy of Theorem 1.4 with only one modification—generalized Schmidt matrix approximation, Theorem 2.2, is used instead of Eckart-Young-Mirsky Theorem 1.2.

Furthermore, relation (1.9) is proved at this moment and, hence, $\|\sigma_{p+1}\mathbf{u}_{p+1}\mathbf{v}_{p+1}^\top\| = \sigma_{p+1}$ for arbitrary unitarily invariant matrix norm $\|\cdot\|$. The previous equation is true due to Definition 2.1, and the orthonormality of vectors $\mathbf{u}_{p+1}$ and $\mathbf{v}_{p+1}$.

Finally, the smallest eigenvalue of positive semidefinite matrix $[\mathbf{X}, \mathbf{Y}]^\top [\mathbf{X}, \mathbf{Y}]$ is the squared smallest singular value of matrix $[\mathbf{X}, \mathbf{Y}]$ due to the SVD (Theorem 1.1) of $[\mathbf{X}, \mathbf{Y}]$ and the eigen decomposition property (Theorem A.1) of $[\mathbf{X}, \mathbf{Y}]^\top [\mathbf{X}, \mathbf{Y}]$. Combining this fact with (1.12) leads to (2.5). □

Repeatedly, Okamoto (1973) and Gleser (1981) proved that $\sigma'_p > \sigma_{p+1}$ with probability one if the elements of errors are absolutely continuous with respect to the Lebesgue measure as stated in the description of the EIV model in Section 1.2. Hence, the estimate $\widehat{\boldsymbol{\beta}}$ *exists with probability* one and *coincides* with the TLS estimate. This property is very plausible and, moreover, some even more plausible properties of our EIV estimate will be derived.

2.2.1 Unification and Coincidence in the Estimation with Examples

Theorem 2.3 has much larger impact and consequences than it might be thought. Let us consider *q-Schatten matrix norms* (also called Schatten q-norms)

$$\|\mathbf{A}\|_q = \left\{ tr \left[\left(\mathbf{A}^\top \mathbf{A}\right)^{q/2} \right] \right\}^{1/q}, \quad q \geq 1. \tag{2.6}$$

The Schatten matrix norms from (2.6) can be alternatively and equivalently defined (see, e.g., Bhatia (1996, Chapter 4)) by the spectral properties of a matrix:

$$\|\mathbf{A}\|_q = \left(\sum_{i=1}^{\min\{n,p\}} \sigma_i^q \right)^{1/q}, \quad q \geq 1, \tag{2.7}$$

where an arbitrary matrix $\mathbf{A} = (a_{ij})_{i,j=1}^{n,p}$ with its singular values $\sigma_1 \geq \ldots \geq \sigma_{\min\{n,p\}} \geq 0$ is taken into account. Neumann's characterization of the unitarily invariant matrix norms (Theorem 2.1) implies that the q-Schatten matrix norms (2.7) are a huge subclass of the unitarily invariant matrix norms when considering symmetric gauge functions $\varsigma(\boldsymbol{\sigma}) = \|\boldsymbol{\sigma}\|_{\mathcal{L}_q}$, where $\|\cdot\|_{\mathcal{L}_q}$ is the $\mathcal{L}_q$-vector norm and $\boldsymbol{\sigma} = [\sigma_1, \ldots, \sigma_{\min\{n,p\}}]^\top$. They can be seen as a way of defining $\mathcal{L}_p$-norms for matrices.

According to Theorem 2.3, all the estimates in the EIV model based upon arbitrary q-Schatten matrix norm are simply *identical* (coincide). Special types of the q-Schatten matrix norms are the *nuclear matrix norm* ($q = 1$)

$$\|\mathbf{A}\|_1 = tr\sqrt{\mathbf{A}^\top \mathbf{A}} = \sum_{i=1}^{\min\{n,p\}} \sigma_i,$$

and the Frobenius norm ($q = 2$). In the context of this thesis, the nuclear matrix norm cannot be confused with an operator norm defined latter in (2.8).

The class of the q-Schatten matrix norms can even be enlarged. Rao (1980) describes a broader subclass of the unitarily invariant matrix norms—*q-Ky Fan k-norms*

$$\|\mathbf{A}\|_q^{(k)} = \left(\sum_{i=1}^{k} \sigma_i^q \right)^{1/q}, \quad q \geq 1,\ 1 \leq k \leq \min\{n,p\}.$$

They are generated by symmetric gauge functions

$$\varsigma([\sigma_1, \ldots, \sigma_{\min\{n,p\}}]^\top) = \left(\sum_{i=1}^{k} \sigma_i^q \right)^{1/q},$$

which can be thought of a *trimmed* version of the q-Schatten matrix norm. Therefore, even a "*robust*" version of an estimate for the EIV model in the sense of various "trimmed $\mathcal{L}_p$-

norms" still coincides with the only EIV estimate. After that, one would expect that the EIV estimate should behave robustly against *leverage observations*, which is analyzed in Van Huffel and Vandewalle (1991, Chapter 9).

The q-Schatten matrix norms are really special types of the q-Ky Fan k-norms when $k = \min\{n, p\}$. Moreover, *operator (spectral) matrix norm*

$$\|\mathbf{A}\|_\infty = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_E}{\|\mathbf{x}\|_E} = \sigma_1, \tag{2.8}$$

where $\|\cdot\|_E$ denotes the Euclidean vector norm, is a special type of the q-Ky Fan k-norm ($k = 1$) as well. The operator norm sometimes stands as a definition of the ∞-Schatten matrix norm.

Why are the unitarily invariant matrix norms so important and preferable? The answer is given by the singular value decomposition (Theorem 1.1)—especially in the *geometric interpretation* of the SVD. When considering the SVD $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$, then $\mathbf{U} \in \mathbb{R}^{n\times n}$ and $\mathbf{V} \in \mathbb{R}^{p\times p}$ are unitary matrices whose columns consist of the orthonormal vectors $\mathbf{u}_i \in \mathbb{R}^n$, $i = 1, \ldots, n$ and $\mathbf{v}_j \in \mathbb{R}^p$, $j = 1, \ldots, p$ with respect to the standard scalar product, which yield orthogonal bases of the Euclidean spaces on $\mathbb{R}^n$ and $\mathbb{R}^p$. Hence, for every matrix $\mathbf{A} \in \mathbb{R}^{n\times p}$ with its SVD as above, there exists a linear mapping $\mathscr{L} : \mathbb{R}^p \to \mathbb{R}^n$ such that

$$\mathscr{L}(\mathbf{v}_i) = \sigma_i \mathbf{u}_i, \quad i = 1, \ldots, \min\{n, p\}$$

and

$$\mathscr{L}(\mathbf{v}_i) = \mathbf{0}, \quad i = \min\{n, p\} + 1, \ldots, \max\{n, p\},$$

where $\sigma_1, \ldots, \sigma_{\min\{n,p\}}$ are the diagonal elements of $\mathbf{\Sigma} \in \mathbb{R}^{n\times p}$. Loosely speaking, the geometric content of the SVD lies in a fact that information from the data (input matrix $\mathbf{A}$) is decomposed by the SVD into *rotation of the data* in the Euclidean spaces (represented by the unitary matrices $\mathbf{U}$ and $\mathbf{V}$) and *magnitude of the data* (represented by the singular values). If the unitarily invariant matrix norms are chosen for the penalizing of the errors in the EIV model, all the information are just hidden in the singular values, because the rotation of the data has no effect on the value of the unitarily invariant matrix norm. This "information extraction" will provide pleasant invariant and equivariant properties of the EIV estimate.

2.3 Invariancy and Equivariancy

Up until now, our EIV estimate is a *reasonable* estimate of the unknown parameter $\boldsymbol{\beta}$, because it minimizes the errors, or in other words, it finds the best fit. Its asymptotic properties, mentioned in Section 1.6, provide the second argument for being a good estimate. In spite of this, we need to ensure ourselves that the EIV estimate incorporates some *natural* and *expectable properties*, i.e.:

- multiplying the input data by a positive constant does not affect the estimate (e.g., changing the measurement units),
- interchange of the explanatory variables has the effect of an exchange of the estimate's components in the corresponding (permuted) order (each estimate's component should correspond to one covariate, which impact is estimated by that component),
- change of the regressor's sign implies multiplication of the corresponding estimate's component by minus one (e.g., changing directions of the measurements yields "opposite" estimate),
- rotation of the explanatory variables provides the correspondingly rotated estimate (vector basis should not determine the estimate).

Previously described geometrical properties are going to be mathematically formulated and, consequently, proved to hold for our EIV estimate.

Let us define $\mathbf{T}([\mathbf{X},\mathbf{Y}])$ as an estimate for an unknown parameter $\boldsymbol{\beta}$ constructed from input data $[\mathbf{X},\mathbf{Y}]$.

Definition 2.3 (Scale invariant estimate)**.** An estimate $\mathbf{T}$ is called *scale invariant* if it satisfies

$$\mathbf{T}(a[\mathbf{X},\mathbf{Y}]) = \mathbf{T}([\mathbf{X},\mathbf{Y}])$$

for any constant $a > 0$.

Corollary 2.4. *EIV estimate* (2.5) *is a scale invariant estimate.*

Proof. Finding a new EIV estimate for the multiplied data $a[\mathbf{X},\mathbf{Y}]$ leads to solving the following optimizing problem

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p,[\boldsymbol{\Theta},\varepsilon]\in\mathbb{R}^{n\times(p+1)}} \|[a\boldsymbol{\Theta},a\varepsilon]\| \qquad \text{s.t.} \quad a\mathbf{Y}-a\boldsymbol{\varepsilon}=(a\mathbf{X}-a\boldsymbol{\Theta})\boldsymbol{\beta}. \tag{2.9}$$

However, minimizing problem (2.9) is clearly equivalent (provides the same solution) to the original optimizing problem (2.1). □

A handy lemma for the further simplification of proofs is going to be derived.

Lemma 2.5 (Equivariantness)**.** *Suppose* $\mathbf{J}\in\mathbb{R}^{p\times p}$ *is a unitary matrix and* $\|\cdot\|$ *is unitarily invariant matrix norm. If* $\{\widehat{\boldsymbol{\beta}},[\widehat{\boldsymbol{\Theta}},\widehat{\varepsilon}]\}$ *is a solution to optimizing problem* (2.1)*, then*

$\{\mathbf{J}^\top\widehat{\boldsymbol{\beta}}, [\widehat{\boldsymbol{\Theta}}, \widehat{\varepsilon}]\}$ *is a solution to the optimizing problem*

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p, [\boldsymbol{\Theta},\varepsilon]\in\mathbb{R}^{n\times(p+1)}} \|[\boldsymbol{\Theta}, \varepsilon]\widetilde{\mathbf{J}}\| \qquad s.t. \quad \mathbf{Y} - \boldsymbol{\varepsilon} = (\mathbf{X} - \boldsymbol{\Theta})\mathbf{J}\boldsymbol{\beta}, \tag{2.10}$$

where

$$\widetilde{\mathbf{J}} = \begin{bmatrix} \mathbf{J} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix}.$$

Proof. Since $\mathbf{J}$ is unitary, then $\mathbf{J}^{-1} = \mathbf{J}^\top$ and, moreover, $\widetilde{\mathbf{J}}$ is unitary as well. Due to the basic property of the unitarily invariant matrix norms, optimizing problem (2.1) is equivalent to

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^p, [\boldsymbol{\Theta},\varepsilon]\in\mathbb{R}^{n\times(p+1)}} \|[\boldsymbol{\Theta}, \varepsilon]\widetilde{\mathbf{J}}\| \qquad \text{s.t.} \quad \mathbf{Y} - \boldsymbol{\varepsilon} = (\mathbf{X} - \boldsymbol{\Theta})\mathbf{J}\mathbf{J}^\top\boldsymbol{\beta}. \tag{2.11}$$

Once a solution $\{\widehat{\boldsymbol{\beta}}, [\widehat{\boldsymbol{\Theta}}, \widehat{\varepsilon}]\}$ to (2.11) is found, then $\{\mathbf{J}^\top\widehat{\boldsymbol{\beta}}, [\widehat{\boldsymbol{\Theta}}, \widehat{\varepsilon}]\}$ has to be a solution to (2.10). □

Definition 2.4 (Interchange equivariant estimate)**.** An estimate $\mathbf{T}$ is called *interchange equivariant* if it satisfies

$$\mathbf{T}([\mathbf{X}, \mathbf{Y}]\widetilde{\mathbf{P}}_\pi) = \mathbf{P}_{\pi^{-1}}\mathbf{T}([\mathbf{X}, \mathbf{Y}]) \tag{2.12}$$

for any permutation $\pi : \{1, \ldots, p\} \to \{1, \ldots, p\}$ with its inverse π^{-1} and the corresponding permutation matrices $\mathbf{P}_\pi$ and $\mathbf{P}_{\pi^{-1}}$, where

$$\widetilde{\mathbf{P}}_\pi = \begin{bmatrix} \mathbf{P}_\pi & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix}.$$

Note that if $\mathbf{P}_\pi$ is a permutation matrix, then $\widetilde{\mathbf{P}}_\pi$ is a permutation one as well. Moreover, its inverse is exactly

$$\widetilde{\mathbf{P}}_{\pi^{-1}} = \begin{bmatrix} \mathbf{P}_{\pi^{-1}} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix}.$$

Number one in the last $\widetilde{\mathbf{P}}_\pi$ entry—position $(p+1, p+1)$—stands for the fixed “location” of the response, i.e., keeps the regressand untouched.

The definition from (2.12) can be alternatively rewritten

$$\mathbf{P}_\pi\mathbf{T}([\mathbf{X}\mathbf{P}_\pi, \mathbf{Y}]) = \mathbf{T}([\mathbf{X}, \mathbf{Y}])$$

in order to get insight into the meaning of equivariancy for this case.

All permutation matrices of the same dimension together with a matrix multiplication form a *symmetric group* under matrix multiplication with the identity matrix as the identity element. Multiplying a matrix by permutation matrix $\mathbf{P}_\pi$ will permute the columns of the matrix by the *inverse* of π, i.e., π^{-1}. Therefore, the components of $\mathbf{T}([\mathbf{X},\mathbf{Y}])$ are reordered in the inverse manner, i.e., $\mathbf{T}$ is multiplied by $\mathbf{P}_{\pi^{-1}}$ from left hand side.

Corollary 2.6. *EIV estimate* (2.5) *is an interchange equivariant estimate.*

Proof. Since each permutation matrix is a unitary one, Lemma 2.5 straightforwardly completes the proof. □

Definition 2.5 (Direction equivariant estimate)**.** An estimate $\mathbf{T}$ is called *direction equivariant* if it satisfies

$$\mathbf{T}([\mathbf{X},\mathbf{Y}]\widetilde{\mathbf{D}}) = -\mathbf{D}\mathbf{T}([\mathbf{X},\mathbf{Y}]) \tag{2.13}$$

for any diagonal matrix $\mathbf{D} \in \mathbb{R}^{p\times p}$ having ± 1 on the diagonal, where

$$\widetilde{\mathbf{D}} = \begin{bmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix}.$$

Formula (2.13) can also be alternatively rewritten

$$\mathbf{D}\mathbf{T}([\mathbf{X}\mathbf{D},\mathbf{Y}]) = \mathbf{T}([\mathbf{X},\mathbf{Y}]).$$

Again, all diagonal matrices with ± 1 on the diagonal of the same dimension together with a matrix multiplication form again a *symmetric group* under matrix multiplication with the identity matrix as the identity element. Multiplying a matrix times $\mathbf{D}$ will change the direction (orientation) of the columns of matrix according to the sign of the corresponding diagonal element. Therefore, the signs of the components of $\mathbf{T}([\mathbf{X},\mathbf{Y}])$ are changed in the inverse manner, i.e., $\mathbf{T}$ is multiplied by $-\mathbf{D}$ from left hand side.

Corollary 2.7. *EIV estimate* (2.5) *is a direction equivariant estimate.*

Proof. Since each diagonal matrix having ± 1 on the diagonal is a unitary one, Lemma 2.5 straightforwardly completes the proof. □

Definition 2.6 (Rotation equivariant estimate)**.** An estimate $\mathbf{T}$ is called *rotation equivariant* if it satisfies

$$\mathbf{T}([\mathbf{X},\mathbf{Y}]\widetilde{\mathbf{R}}) = \mathbf{R}^\top\mathbf{T}([\mathbf{X},\mathbf{Y}]) \tag{2.14}$$

for any rotation matrix $\mathbf{R} \in \mathbb{R}^{p\times p}$ (see Definition A.3), where

$$\widetilde{\mathbf{R}} = \begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix}.$$

Similar situation occurs to the rotation matrices. Here, if $\mathbf{R}$ is a rotation matrix, then $\widetilde{\mathbf{R}}$ is a rotation one as well. Moreover, its inverse is exactly

$$\widetilde{\mathbf{R}}^\top = \begin{bmatrix} \mathbf{R}^\top & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix}.$$

Formula (2.14) can be alternatively rewritten in order to demonstrate the meaning of equivariancy

$$\mathbf{R}\mathbf{T}([\mathbf{X}\mathbf{R}, \mathbf{Y}]) = \mathbf{T}([\mathbf{X}, \mathbf{Y}])$$

For the third time, all rotation matrices having the same dimension together with a matrix multiplication form a *symmetric group* with the identity matrix as the identity element. Multiplying a matrix times $\mathbf{R}$ will rotate the columns of the matrix in the *"inverse"* manner—the same rotation, but with the opposite orientation. Therefore, an estimate $\mathbf{T}([\mathbf{X}, \mathbf{Y}])$ needs to be rotated *reversely*, i.e., $\mathbf{T}$ is multiplied by $\mathbf{R}$ from left hand side.

Corollary 2.8. *EIV estimate* (2.5) *is a rotation equivariant estimate.*

Proof. Since each rotation matrix is a unitary one, Lemma 2.5 straightforwardly completes the proof. □

Finally, Corollaries 2.4, 2.6, 2.7, and 2.8 can be generalized in one common way, which captures, among other things, all the above mentioned transformations. This alternative generalization will be stated in the forthcoming theorem and, moreover, we prove this theorem by a different approach than by solving the equivalent optimizing problems, i.e., using the spectral properties of unitary matrices and the closed form of the EIV estimate from (2.5).

Theorem 2.9. *Suppose $a > 0$ is a positive constant and $\mathbf{J} \in \mathbb{R}^{p\times p}$ is a unitary matrix. If the EIV estimate $\widehat{\boldsymbol{\beta}}$ from data $[\mathbf{X}, \mathbf{Y}]$ exists in the closed form* (2.5)*, then the EIV estimate from transformed data*

$$[\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}] \equiv a[\mathbf{X}, \mathbf{Y}] \begin{bmatrix} \mathbf{J} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix}$$

equals $\mathbf{J}^\top \widehat{\boldsymbol{\beta}}$.

Proof. Let us define

$$\widetilde{\mathbf{J}} := \begin{bmatrix} \mathbf{J} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix}.$$

Since $\mathbf{J}$ is a unitary matrix, then $\widetilde{\mathbf{J}}$ is a unitary one as well. Suppose that the SVD (Theorem 1.1) of data matrix $[\mathbf{X}, \mathbf{Y}]$ is

$$[\mathbf{X}, \mathbf{Y}] = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top,$$

where $\mathbf{U}$, $\mathbf{V}$ are unitary matrices and $\boldsymbol{\Sigma}$ is a diagonal one with singular values on its diagonal ordered in the non-increasing order.

Moreover, let us denote $\widetilde{\mathbf{V}} := \widetilde{\mathbf{J}}^\top \mathbf{V}$. This matrix is also unitary, because

$$\widetilde{\mathbf{V}}^\top \widetilde{\mathbf{V}} = \mathbf{V}^\top \widetilde{\mathbf{J}}\widetilde{\mathbf{J}}^\top \mathbf{V} = \mathbf{I} = \widetilde{\mathbf{J}}^\top \mathbf{V}\mathbf{V}^\top \widetilde{\mathbf{J}} = \widetilde{\mathbf{V}}\widetilde{\mathbf{V}}^\top.$$

The unitarity of $\widetilde{\mathbf{V}}$ and the SVD of $[\mathbf{X}, \mathbf{Y}]$ implies that $a\boldsymbol{\Sigma}$ is a diagonal matrix with the singular values of the transformed data matrix $[\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}]$ on its diagonal ordered in the non-increasing order due to the uniqueness of the SVD

$$[\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}] = a[\mathbf{X}, \mathbf{Y}]\widetilde{\mathbf{J}} = a\mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top \widetilde{\mathbf{J}} = \mathbf{U}(a\boldsymbol{\Sigma})\widetilde{\mathbf{V}}^\top.$$

Since the smallest eigenvalue of square matrix $[\mathbf{X}, \mathbf{Y}]^\top [\mathbf{X}, \mathbf{Y}]$ is σ_{p+1}^2, or in a shortened notation $\lambda \equiv \lambda_{min}([\mathbf{X}, \mathbf{Y}]^\top [\mathbf{X}, \mathbf{Y}]) = \sigma_{p+1}^2$, then the smallest eigenvalue of $[\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}]^\top [\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}]$ is $\lambda_{min}([\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}]^\top [\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}]) = a^2\sigma_{p+1}^2 \equiv a^2\lambda$, because the SVD and the eigen decomposition (Theorem A.1) provides

$$[\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}]^\top [\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}] = \widetilde{\mathbf{V}}(a\boldsymbol{\Sigma})^\top \mathbf{U}^\top \mathbf{U}(a\boldsymbol{\Sigma})\widetilde{\mathbf{V}}^\top = \widetilde{\mathbf{V}}(a^2\boldsymbol{\Sigma}^\top \boldsymbol{\Sigma})\widetilde{\mathbf{V}}^{-1}.$$

Henceforth, the EIV estimate from transformed data $[\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}]$ is calculated according to (2.5). Indeed, the EIV estimate from $[\mathbf{X}, \mathbf{Y}]$ exists in the closed form, so that

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}([\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}]) &= \left(\widetilde{\mathbf{X}}^\top \widetilde{\mathbf{X}} - \lambda_{min}([\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}]^\top [\widetilde{\mathbf{X}}, \widetilde{\mathbf{Y}}])\mathbf{I}\right)^{-1} \widetilde{\mathbf{X}}^\top \widetilde{\mathbf{Y}} \\
&= \left((a\mathbf{X}\mathbf{J})^\top (a\mathbf{X}\mathbf{J}) - a^2\lambda\mathbf{I}\right)^{-1} (a\mathbf{X}\mathbf{J})^\top (a\mathbf{Y}) \\
&= a^{-2}\left(\mathbf{J}^\top (\mathbf{X}^\top \mathbf{X} - \lambda\mathbf{I})\mathbf{J}\right)^{-1} a^2\mathbf{J}^\top \mathbf{X}^\top \mathbf{Y} \\
&= \mathbf{J}^{-1}(\mathbf{X}^\top \mathbf{X} - \lambda\mathbf{I})^{-1}(\mathbf{J}^\top)^{-1}\mathbf{J}^\top \mathbf{X}^\top \mathbf{Y} = \mathbf{J}^\top \widehat{\boldsymbol{\beta}}([\mathbf{X}, \mathbf{Y}]) \equiv \mathbf{J}^\top \widehat{\boldsymbol{\beta}}.
\end{aligned}$$

□

2.4 Summary

A penalization method of the errors in our EIV model via *unitarily invariant matrix norm* was postulated. It creates more *general optimizing* problem than the total least squares minimizing. A solution was found, which provides a well-known estimate for the EIV model. This new estimate—the *EIV estimate*—coincides with the "old-fashioned" TLS estimate. In spite of this, the EIV estimate can be seen as a *generalization* of the TLS estimate, because it is derived from much more general set-up.

Special cases of the EIV estimate were demonstrated as a correspondence to widely used matrix norms (a distributional-free approach in estimation), which serve for the error penalization. The EIV estimate *unifies* a broad class of regression estimates, e.g., $\mathcal{L}_p$-estimation or trimming. Therefore, it may be considered as a specific *robust* alternative in the EIV modeling, despite the fact that it is the same estimate.

The EIV estimate has three groups of nice and reasonable properties: keeping the errors as small as possible, desired asymptotic properties, and transformation invariancy-equivariancy. This third argument for being a natural and reasonable estimate implies that a numerical representation of data does not affect the properties neither the quality of the estimate. Only the information hidden in the data has the impact on it. Formally speaking, scale invariant, interchange equivariant, direction equivariant, and rotation equivariant estimates in the EIV set-up are defined. Consequently, it was shown that our EIV estimate disposes of the previous demanded properties.

Chapter 3

Bootstrap Versus Asymptotics

Je n'ai pas besoin de cette hypothèse.
[I have no need of that hypothesis.]

Pierre-Simon Laplace

The solution to the errors-in-variables problem computed through penalizing an arbitrary unitarily invariant matrix norm of the errors is *highly nonlinear*. Because of this, many statistical procedures for constructing confidence intervals and testing hypotheses cannot be applied. One possible solution to this dilemma is *bootstrapping*. A nonparametric bootstrap technique could fail. The proper nonparametric bootstrap procedure is provided and its correctness proved. On the other hand, a residual bootstrap is not valid and suitable in this case. The results are illustrated through a simulation study. An application of this approach to calibration data is presented.

3.1 Introduction

Classical normal asymptotics could bring some serious pitfalls in parameter's inference in the errors-in-variables model as it will be pointed further. Therefore, a competitive alternative method needs to be invented and implemented for the characterization of large sample (limiting) behavior of the estimates. Bootstrapping seems to be a plausible choice. According to our knowledge, a bootstrap inference for the EIV was not explored, in spite of the fact, that it remains the only possibility for data analysis in some realistic situations.

3.1.1 Motivation for Bootstrapping EIV Estimate

The previous asymptotical results summarized in Section 1.6 can be considered as "ancient". In spite of that, there remain three crucial issues concerning asymptotic normality in (1.29):

the variance 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. Therefore, a bootstrap procedure may be helpful, e.g., for a construction of *confidence regions* for the unknown parameter $\boldsymbol{\beta}$.

The bootstrap approach was introduced by Efron (1979) and extensively investigated by Bickel and Freedman (1981) for the case of linear regression models.

For the rest of this thesis, we link $\|\cdot\|$ together with the Frobenius matrix norm, because it provides the same EIV estimate as any other unitarily invariant matrix norm.

3.2 Nonparametric Bootstrap

A *nonparametric bootstrap* inherits its name from the fact that neither distributional assumptions nor a regression model are assumed while resampling is being performed. The nonparametric resampling refers to the simplest scheme of resampling rows of the data and, therefore, is also often called the *case sampling*.

The idea of the nonparametric bootstrap lies in the resampling of the row data $[\mathbf{X}_{i,\bullet}, Y_i]$ with replacement in order to obtain new *bootstrapped data* $[\mathbf{X}^*, \mathbf{Y}^*]$. A detailed explanation and procedure with the particular steps will be presented later on. From the "starred" data $[\mathbf{X}^*, \mathbf{Y}^*]$, a quantity of interest is computed, e.g., an estimate of the unknown parameter. It is hoped and wished that the distribution of the new bootstrapped quantity mimics the distribution of the original statistics, which we are concerned with.

Since we are interested in the EIV estimate $\widehat{\boldsymbol{\beta}}$ and $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ has multivariate normal distribution, it will be necessary to construct the bootstrapped version of $\widehat{\boldsymbol{\beta}}$, e.g.,

$$\widehat{\boldsymbol{\beta}}^* = \left(\mathbf{X}^{*\top}\mathbf{X}^* - \lambda^*\mathbf{I}\right)^{-1}\mathbf{X}^{*\top}\mathbf{Y}^*,$$

where λ^* is the smallest eigenvalue of $[\mathbf{X}^*, \mathbf{Y}^*]^\top[\mathbf{X}^*, \mathbf{Y}^*]$. Afterwards, it is mandatory to asymptotically compare the distribution of $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ and $\sqrt{n}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})$ in a proper mathematical way to be sure that the empirical distribution of the bootstrap estimate $\widehat{\boldsymbol{\beta}}^*$ can be used instead of the unknown or computationally unreachable distribution of $\widehat{\boldsymbol{\beta}}$.

Unfortunately, it will be demonstrated that there may exist a *nondegenerative* distribution between the asymptotic distribution of $\sqrt{n}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})\big|[\mathbf{X}, \mathbf{Y}]$ and $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. Therefore, considering $\widehat{\boldsymbol{\beta}}^*$ would lead into a different limit approximation than the desired one and we need to introduce the *corrected version of bootstrap estimate*:

$$\widetilde{\boldsymbol{\beta}}^* := \widehat{\boldsymbol{\beta}} - \left(\mathbf{X}^{*\top}\mathbf{X}^* - \lambda^*\mathbf{I}\right)^{-1}\left([\mathbf{I},\widehat{\boldsymbol{\beta}}]\left[\mathbf{I},\widehat{\boldsymbol{\beta}}^*\right]^\top\right)^{-1} [\mathbf{I},\widehat{\boldsymbol{\beta}}]\left([\mathbf{X}^*,\mathbf{Y}^*]^\top[\mathbf{X}^*,\mathbf{Y}^*] - [\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]\right)\begin{bmatrix}\widehat{\boldsymbol{\beta}}\\ -1\end{bmatrix}. \quad (3.1)$$

The reason for such a complicated correction is that the linear EIV problem has a highly nonlinear solution.

Later on, the asymptotical closeness of $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ and $\sqrt{n}(\widetilde{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})$ will be clarified and proved. The bootstrap estimate $\widetilde{\boldsymbol{\beta}}^*$ can be viewed as a *bias corrected version* of the original but improper bootstrap estimate $\widehat{\boldsymbol{\beta}}^*$.

An algorithm for the nonparametric bootstrap is shown in Procedure 3.1 and its validity will be proved in Theorem 3.12.

Procedure 3.1 Nonparametric bootstrap for the EIV estimate.

Input: Data consisting of n row vectors of observations $[\mathbf{X}_{i,\bullet}, Y_i]$.

Output: Empirical bootstrap distribution of $\widehat{\boldsymbol{\beta}}$, i.e., the empirical distribution where probability mass $1/B$ concentrates at each of ${}_{(1)}\widetilde{\boldsymbol{\beta}}^*, \ldots, {}_{(B)}\widetilde{\boldsymbol{\beta}}^*$.

```
calculate TLS estimate β̂ ← (XᵀX − λI)⁻¹XᵀY
for b = 1 to B do // repeat in order to obtain empirical distribution of β̂
    [(b)X*, (b)Y*]_{n×(p+1)} resampled with replacement from rows [X, Y]
    (b)λ* is the (p+1)-st eigenvalue of [(b)X*, (b)Y*]ᵀ[(b)X*, (b)Y*]
    re-estimate (b)β̂* ← ((b)X*ᵀ (b)X* − (b)λ* I)⁻¹ (b)X*ᵀ (b)Y*
    put
```

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

end for

Remark 3.1. It is silently presupposed that the bootstrapped sample is of the same size as the original one, i.e., $[\mathbf{X},\mathbf{Y}] \equiv [\mathbf{X}(n),\mathbf{Y}(n)]$ and $[\mathbf{X}^*,\mathbf{Y}^*] \equiv [\mathbf{X}^*(n),\mathbf{Y}^*(n)]$. In general, it may be considered resampled bootstrapped data $[\mathbf{X}^*,\mathbf{Y}^*] \equiv [\mathbf{X}^*(m),\mathbf{Y}^*(m)]$ of sample size m with replacement from original data $[\mathbf{X},\mathbf{Y}] \equiv [\mathbf{X}(n),\mathbf{Y}(n)]$ of sample size n. Thereby, an additional condition needs to be postulated on the rate of the sample sizes:

$$m = \mathcal{O}(n), \quad n \to \infty \qquad \& \qquad n = \mathcal{O}(m), \quad m \to \infty.$$

For theoretical asymptotical results in this thesis, there is no need to distinguish between the same sample size (of the original and the bootstrapped data) and two different, but asymptotically equivalent (the same asymptotical order) sample sizes. On the other hand, there could be a computational improvement considering different samples size of the bootstrapped data, but we will not focus on this topic here.

3.2.1 Justification of the Nonparametric Bootstrap Asymptotics

We would like to show that $\sqrt{n}(\widetilde{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})$ and $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ *asymptotically coincide*. This means that using the bootstrap distribution is no worse than using the asymptotic normal approximation. However, it does not mean that the bootstrap distribution better approximates the finite sample distribution of$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. To state this mathematically, Belyaev (1995) introduced *conditional weak convergence almost surely* and *in probability.*

Suppose that $\{\boldsymbol{\xi}_n, \boldsymbol{\xi}_n^*, \boldsymbol{\zeta}_n, \boldsymbol{\zeta}_n^*, \boldsymbol{\chi}_n\}_{n=1}^{\infty}$ are sequences of random vectors/matrices, which elements exist on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The components of these sequences do not necessarily have to have the same dimension, e.g., $\boldsymbol{\zeta}_n$ and $\boldsymbol{\zeta}_{n+1}$ can have different dimensions for some $n \in \mathbb{N}$. Let us define a *conditional probability* given $\boldsymbol{\zeta}_n$

$$\mathbb{P}^*_{\boldsymbol{\zeta}_n}[\cdot] := \mathbb{E}_{\mathbb{P}}[\mathcal{I}(\cdot)|\boldsymbol{\zeta}_n].$$

Definition 3.1 (Conditional weak convergence almost surely and in probability)**.** Let $\{\boldsymbol{\xi}_n, \boldsymbol{\xi}_n^*, \boldsymbol{\zeta}_n\}_{n=1}^{\infty}$ be sequences of random vectors/matrices. If for every real-valued bounded continuous function f holds

$$\mathbb{E}\left[f(\boldsymbol{\xi}_n^*)\right] - \mathbb{E}\left[f(\boldsymbol{\xi}_n)\right] \xrightarrow[n\to\infty]{} 0,$$

then $\boldsymbol{\xi}_n^*$ and $\boldsymbol{\xi}_n$ are said to be approaching each other in distribution. In short we write

$$\boldsymbol{\xi}_n^* \xleftrightarrow[n\to\infty]{\mathscr{D}} \boldsymbol{\xi}_n.$$

If for every real-valued bounded continuous function f holds

$$\mathbb{E}\left[f(\boldsymbol{\xi}_n^*)|\boldsymbol{\zeta}_n\right] - \mathbb{E}\left[f(\boldsymbol{\xi}_n)\right] \xrightarrow[n\to\infty]{[\mathbb{P}]\text{-}a.s.} 0,$$

then $\boldsymbol{\xi}_n^*$ conditioned on $\boldsymbol{\zeta}_n$ and $\boldsymbol{\xi}_n$ are said to be approaching each other in distribution $[\mathbb{P}]$-almost surely along $\boldsymbol{\zeta}_n$. In short we write

$$\boldsymbol{\xi}_n^*|\boldsymbol{\zeta}_n \xleftrightarrow[n\to\infty]{\mathscr{D}([\mathbb{P}]\text{-}a.s.)} \boldsymbol{\xi}_n.$$

If for every real-valued bounded continuous function f holds

$$\mathbb{E}\left[f(\boldsymbol{\xi}_n^*)|\boldsymbol{\zeta}_n\right] - \mathbb{E}\left[f(\boldsymbol{\xi}_n)\right] \xrightarrow[n\to\infty]{\mathbb{P}} 0,$$

then $\boldsymbol{\xi}_n^*$ conditioned on $\boldsymbol{\zeta}_n$ and $\boldsymbol{\xi}_n$ are said to be approaching each other in distribution in probability $\mathbb{P}$ along $\boldsymbol{\zeta}_n$. In short we write

$$\boldsymbol{\xi}_n^*\big|\boldsymbol{\zeta}_n \xleftrightarrow[n\to\infty]{\mathscr{D}(\mathbb{P})} \boldsymbol{\xi}_n.$$

In the same manner as above, we may define the distributional convergence on the "conditional" (resampled) level to a random variable $\boldsymbol{\xi}_0$ ("constant" law).

Definition 3.2 (Conditional weak convergence almost surely and in probability to a constant law)**.** Let $\{\boldsymbol{\xi}_n^*, \boldsymbol{\zeta}_n\}_{n=1}^{\infty}$ be sequences of random vectors/matrices and $\boldsymbol{\xi}_0$ be a random vector/matrix. If for every real-valued bounded continuous function f holds

$$\mathbb{E}\left[f(\boldsymbol{\xi}_n^*)|\boldsymbol{\zeta}_n\right] \xrightarrow[n\to\infty]{[\mathbb{P}]\text{-}a.s.} \mathbb{E}\left[f(\boldsymbol{\xi}_0)\right],$$

then $\boldsymbol{\xi}_n^*$ conditioned on $\boldsymbol{\zeta}_n$ is said to converge to $\boldsymbol{\xi}_0$ in distribution $[\mathbb{P}]$-almost surely along $\boldsymbol{\zeta}_n$. In short we write

$$\boldsymbol{\xi}_n^*\big|\boldsymbol{\zeta}_n \xrightarrow[n\to\infty]{\mathscr{D}([\mathbb{P}]\text{-}a.s.)} \boldsymbol{\xi}_0.$$

If for every real-valued bounded continuous function f holds

$$\mathbb{E}\left[f(\boldsymbol{\xi}_n^*)|\boldsymbol{\zeta}_n\right] \xrightarrow[n\to\infty]{\mathbb{P}} \mathbb{E}\left[f(\boldsymbol{\xi}_0)\right],$$

then $\boldsymbol{\xi}_n^*$ conditioned on $\boldsymbol{\zeta}_n$ is said to converge to $\boldsymbol{\xi}_0$ in distribution in probability $\mathbb{P}$ along $\boldsymbol{\zeta}_n$. In short we write

$$\boldsymbol{\xi}_n^*\big|\boldsymbol{\zeta}_n \xrightarrow[n\to\infty]{\mathscr{D}(\mathbb{P})} \boldsymbol{\xi}_0.$$

Approaching in distribution to each other is often called *weakly approaching* to each other (almost surely or in probability along some sequence).

The appropriateness of Definition 3.1 is ensured by the portmanteau lemma (see, e.g., Billingsley (1999)), which provides equivalent characterizations of the convergence in distribution.

In order to properly and clearly define and, consequently, describe convergence in probability $\mathbb{P}^*$, i.e., convergences on the conditional ("starred") level, we define two types of convergence in probability $\mathbb{P}^*$: $[\mathbb{P}]$-almost surely and in probability $\mathbb{P}$. For the details see, e.g., Belyaev and Sjöstedt-de Luna (2000).

Definition 3.3 (Convergence in conditional probability)**.** Let $\{\boldsymbol{\xi}_n, \boldsymbol{\xi}_n^*, \boldsymbol{\zeta}_n\}_{n=1}^{\infty}$ be sequences of random vectors/matrices. To say that $\boldsymbol{\xi}_n^* - \boldsymbol{\xi}_n$ converges in probability $\mathbb{P}_{\boldsymbol{\zeta}_n}^*$ to zero $[\mathbb{P}]$-almost surely as n tends to infinity, i.e.,

$$\boldsymbol{\xi}_n^* - \boldsymbol{\xi}_n \xrightarrow[n\to\infty]{\mathbb{P}_{\boldsymbol{\zeta}_n}^*([\mathbb{P}]\text{-}a.s.)} \mathbf{0},$$

means

$$\forall \epsilon > 0 : \mathbb{P}\left[\lim_{n\to\infty} \mathbb{P}_{\boldsymbol{\zeta}_n}^*[\|\boldsymbol{\xi}_n^* - \boldsymbol{\xi}_n\| \geq \epsilon] = 0\right] = 1. \tag{3.2}$$

To say that $\boldsymbol{\xi}_n^* - \boldsymbol{\xi}_n$ converges in probability $\mathbb{P}_{\boldsymbol{\zeta}_n}^*$ to zero in probability $\mathbb{P}$ as n tends to infinity, i.e.,

$$\boldsymbol{\xi}_n^* - \boldsymbol{\xi}_n \xrightarrow[n\to\infty]{\mathbb{P}_{\boldsymbol{\zeta}_n}^*(\mathbb{P})} \mathbf{0},$$

means

$$\forall \epsilon > 0, \forall \tau > 0 : \lim_{n\to\infty} \mathbb{P}\left[\mathbb{P}_{\boldsymbol{\zeta}_n}^*[\|\boldsymbol{\xi}_n^* - \boldsymbol{\xi}_n\| \geq \epsilon] \geq \tau\right] = 0. \tag{3.3}$$

Alternatively, (3.2) and (3.3) from Definition 3.3 can be read as

$$\forall \epsilon > 0 : \left\{\mathbb{P}_{\boldsymbol{\zeta}_n}^*[\|\boldsymbol{\xi}_n^* - \boldsymbol{\xi}_n\| \geq \epsilon] \xrightarrow[n\to\infty]{[\mathbb{P}]\text{-}a.s.} 0\right\}$$

and

$$\forall \epsilon > 0 : \left\{\mathbb{P}_{\boldsymbol{\zeta}_n}^*[\|\boldsymbol{\xi}_n^* - \boldsymbol{\xi}_n\| \geq \epsilon] \xrightarrow[n\to\infty]{\mathbb{P}} 0\right\},$$

respectively.

In the same manner as above, we may define the convergence in probability $\mathbb{P}^*$ on the resampled level to a random variable $\boldsymbol{\xi}_0$ (does not depend on n).

Definition 3.4 (Convergence in conditional probability to a "constant" variable)**.** Let $\{\boldsymbol{\xi}_n^*, \boldsymbol{\zeta}_n\}_{n=1}^{\infty}$ be sequences of random vectors/matrices and $\boldsymbol{\xi}_0$ is a random vector/matrix. To say that $\boldsymbol{\xi}_n^*$ converges to $\boldsymbol{\xi}_0$ in probability $\mathbb{P}_{\boldsymbol{\zeta}_n}^*$ $[\mathbb{P}]$-almost surely as n tends to infinity, i.e.,

$$\boldsymbol{\xi}_n^* \xrightarrow[n\to\infty]{\mathbb{P}_{\boldsymbol{\zeta}_n}^*([\mathbb{P}]\text{-}a.s.)} \boldsymbol{\xi}_0,$$

means that $\boldsymbol{\xi}_n^* - \boldsymbol{\xi}_0$ converges in probability $\mathbb{P}_{\boldsymbol{\zeta}_n}^*$ to zero $[\mathbb{P}]$-almost surely as n tends to infinity. To say that $\boldsymbol{\xi}_n^*$ converges to $\boldsymbol{\xi}_0$ in probability $\mathbb{P}_{\boldsymbol{\zeta}_n}^*$ in probability $\mathbb{P}$ as n tends to infinity, i.e.,

$$\boldsymbol{\xi}_n^* \xrightarrow[n\to\infty]{\mathbb{P}^*_{\boldsymbol{\zeta}_n}(\mathbb{P})} \boldsymbol{\xi}_0,$$

means that $\boldsymbol{\xi}_n^* - \boldsymbol{\xi}_0$ converges in probability $\mathbb{P}^*_{\boldsymbol{\zeta}_n}$ to zero in probability $\mathbb{P}$ as n tends to infinity.

Our error assumption (see Section 1.4) was that the rows of errors are assumed to be *iid* (the simplest case). Sometimes, it is assumed that the rows of data $[\mathbf{X}, \mathbf{Y}]$ are *iid*. Considering independent and identically distributed observations seems to be more restrictive than proposing this assumption only on errors. The reason is that we are in the parametric regression setup and we do not want to restrict the covariates too much. Moreover, our main interest lies in the *mean structure* of the parametric regression model and *iid* data somehow contradict the existence of some mean structure. Indeed, identically distributed responses imply constant moments of the responses for each observation, which totally suppresses the existence of a mean structure in the regression model. Therefore, *iid* observations can be considered as a hypothetical assumption without realistic applicability in the situation, where the mean structure consisting of covariates $\mathbf{Z}_{i,\bullet}$ is assumed.

Since the *iid* assumption for input data is not admissible, then the classical machinery of *Mallows metric* (see Bickel and Freedman (1981)) or *Wasserstein metric* (see Dobrushin (1970)) for proving bootstrap's validity cannot be straightforwardly applied. A different statistical apparatus has to be derived and applied in order to justify the appropriateness of the proper nonparametric bootstrap procedure.

Relations between weakly approaching to each other and weakly convergent sequences of probability laws will become handy later on. The Prokhorov's theorem can be extended into our setup as shown below.

Lemma 3.1. *Assume that $\{\boldsymbol{\xi}_n\}_{n=1}^{\infty}$ is tight. Then the following statements are equivalent:*

(i)

$$\boldsymbol{\xi}_n^* | \boldsymbol{\zeta}_n \xleftrightarrow[n\to\infty]{\mathscr{D}(\mathbb{P})} \boldsymbol{\xi}_n.$$

(ii) For each subsequence $\{n_i\}_{i=1}^{\infty}$ such that

$$\boldsymbol{\xi}_{n_i} \xrightarrow[i\to\infty]{\mathscr{D}} \boldsymbol{\xi}_0$$

for some random vector/matrix $\boldsymbol{\xi}_0$,

$$\boldsymbol{\xi}_{n_i}^* | \boldsymbol{\zeta}_{n_i} \xrightarrow[i\to\infty]{\mathscr{D}(\mathbb{P})} \boldsymbol{\xi}_0$$

too.

(iii) *For each subsequence* $\{n_i\}_{i=1}^{\infty}$ *there exists a subsequence* $\{n_{i_k}\}_{k=1}^{\infty}$ *such that* $\boldsymbol{\xi}^*_{n_{i_k}}$ *conditional on* $\boldsymbol{\zeta}_{n_{i_k}}$ *converges in distribution in probability* $\mathbb{P}$ *to the distributional limit of* $\boldsymbol{\xi}_{n_{i_k}}$ *as* $k \to \infty$.

Proof. A simple generalization of Belyaev and Sjöstedt-de Luna (2000, Lemma 1), where conditional law of $\boldsymbol{\xi}^*_n|\boldsymbol{\zeta}_n$ instead of $\boldsymbol{\xi}^*_n$ is considered in the proof as proposed by Zagdański (2005, Proof of Theorem 4.1). □

Important results concerning previously defined types of convergences summarized in the forthcoming Theorem 3.2 will play a crucial role in the following proofs. We need to extend the Slutsky's theorem (see Appendix A.2, Theorem A.2) for our "bootstrap world", i.e., to have a stability property for conditional distributions.

Theorem 3.2 (Slutsky's extended theorem)**.** *Suppose that* $\{\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}^*_n, \boldsymbol{\chi}_n\}_{n=1}^{\infty}$ *are sequences of random vectors/matrices. Then,*

$$\boldsymbol{\xi}^*_n|\boldsymbol{\chi}_n \xrightarrow[n\to\infty]{\mathscr{D}([\mathbb{P}]\text{-}a.s.)} \boldsymbol{\xi}_0 \tag{3.4}$$

and

$$\boldsymbol{\zeta}^*_n \xrightarrow[n\to\infty]{\mathbb{P}^*_{\boldsymbol{\chi}_n}([\mathbb{P}]\text{-}a.s.)} \boldsymbol{\zeta}_0, \tag{3.5}$$

where $\boldsymbol{\xi}_0$ *is a random matrix/vector and* $\boldsymbol{\zeta}_0$ *is a non-random element, implies (for suitable vector/matrix dimensions):*

(i) $[\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}^*_n]|\boldsymbol{\chi}_n \xrightarrow[n\to\infty]{\mathscr{D}([\mathbb{P}]\text{-}a.s.)} [\boldsymbol{\xi}_0, \boldsymbol{\zeta}_0]$;

(ii) $[\boldsymbol{\zeta}^*_n, \boldsymbol{\xi}^*_n]|\boldsymbol{\chi}_n \xrightarrow[n\to\infty]{\mathscr{D}([\mathbb{P}]\text{-}a.s.)} [\boldsymbol{\zeta}_0, \boldsymbol{\xi}_0]$;

(iii) $\boldsymbol{\xi}^*_n + \boldsymbol{\zeta}^*_n|\boldsymbol{\chi}_n \xrightarrow[n\to\infty]{\mathscr{D}([\mathbb{P}]\text{-}a.s.)} \boldsymbol{\xi}_0 + \boldsymbol{\zeta}_0$;

(iv) $\boldsymbol{\xi}^*_n\boldsymbol{\zeta}^*_n|\boldsymbol{\chi}_n \xrightarrow[n\to\infty]{\mathscr{D}([\mathbb{P}]\text{-}a.s.)} \boldsymbol{\xi}_0\boldsymbol{\zeta}_0$;

(v) $\boldsymbol{\zeta}^*_n\boldsymbol{\xi}^*_n|\boldsymbol{\chi}_n \xrightarrow[n\to\infty]{\mathscr{D}([\mathbb{P}]\text{-}a.s.)} \boldsymbol{\zeta}_0\boldsymbol{\xi}_0$;

(vi) $(\boldsymbol{\zeta}^*_n)^{-1}\boldsymbol{\xi}^*_n|\boldsymbol{\chi}_n \xrightarrow[n\to\infty]{\mathscr{D}([\mathbb{P}]\text{-}a.s.)} \boldsymbol{\zeta}_0^{-1}\boldsymbol{\xi}_0$, *provided that* $\boldsymbol{\zeta}^*_n$ *and* $\boldsymbol{\zeta}_0$ *are invertible;*

(vii) $\boldsymbol{\xi}^*_n(\boldsymbol{\zeta}^*_n)^{-1}|\boldsymbol{\chi}_n \xrightarrow[n\to\infty]{\mathscr{D}([\mathbb{P}]\text{-}a.s.)} \boldsymbol{\xi}_0\boldsymbol{\zeta}_0^{-1}$, *provided that* $\boldsymbol{\zeta}^*_n$ *and* $\boldsymbol{\zeta}_0$ *are invertible.*

Moreover,

$$\boldsymbol{\xi}^*_n|\boldsymbol{\chi}_n \xrightarrow[n\to\infty]{\mathscr{D}(\mathbb{P})} \boldsymbol{\xi}_0 \tag{3.6}$$

and

$$\boldsymbol{\zeta}_n^* \xrightarrow[n\to\infty]{\mathbb{P}^*_{\boldsymbol{\chi}_n}(\mathbb{P})} \boldsymbol{\zeta}_0, \tag{3.7}$$

where $\boldsymbol{\xi}_0$ *is a random matrix/vector and* $\boldsymbol{\zeta}_0$ *is a non-random element, implies (for suitable vector/matrix dimensions):*

(viii) $[\boldsymbol{\xi}_n^*, \boldsymbol{\zeta}_n^*]\big|\boldsymbol{\chi}_n \xrightarrow[n\to\infty]{\mathscr{D}(\mathbb{P})} [\boldsymbol{\xi}_0, \boldsymbol{\zeta}_0]$;

(ix) $[\boldsymbol{\zeta}_n^*, \boldsymbol{\xi}_n^*]\big|\boldsymbol{\chi}_n \xrightarrow[n\to\infty]{\mathscr{D}(\mathbb{P})} [\boldsymbol{\zeta}_0, \boldsymbol{\xi}_0]$;

(x) $\boldsymbol{\xi}_n^* + \boldsymbol{\zeta}_n^*\big|\boldsymbol{\chi}_n \xrightarrow[n\to\infty]{\mathscr{D}(\mathbb{P})} \boldsymbol{\xi}_0 + \boldsymbol{\zeta}_0$;

(xi) $\boldsymbol{\xi}_n^*\boldsymbol{\zeta}_n^*\big|\boldsymbol{\chi}_n \xrightarrow[n\to\infty]{\mathscr{D}(\mathbb{P})} \boldsymbol{\xi}_0\boldsymbol{\zeta}_0$;

(xii) $\boldsymbol{\zeta}_n^*\boldsymbol{\xi}_n^*\big|\boldsymbol{\chi}_n \xrightarrow[n\to\infty]{\mathscr{D}(\mathbb{P})} \boldsymbol{\zeta}_0\boldsymbol{\xi}_0$;

(xiii) $(\boldsymbol{\zeta}_n^*)^{-1}\boldsymbol{\xi}_n^*\big|\boldsymbol{\chi}_n \xrightarrow[n\to\infty]{\mathscr{D}(\mathbb{P})} \boldsymbol{\zeta}_0^{-1}\boldsymbol{\xi}_0$, *provided that* $\boldsymbol{\zeta}_n^*$ *and* $\boldsymbol{\zeta}_0$ *are invertible;*

(xiv) $\boldsymbol{\xi}_n^*(\boldsymbol{\zeta}_n^*)^{-1}\big|\boldsymbol{\chi}_n \xrightarrow[n\to\infty]{\mathscr{D}(\mathbb{P})} \boldsymbol{\xi}_0\boldsymbol{\zeta}_0^{-1}$, *provided that* $\boldsymbol{\zeta}_n^*$ *and* $\boldsymbol{\zeta}_0$ *are invertible.*

Proof. First, we show that

$$[\boldsymbol{\xi}_n^*, \boldsymbol{\zeta}_0]\big|\boldsymbol{\chi}_n \xleftrightarrow[n\to\infty]{\mathscr{D}([\mathbb{P}]\text{-}a.s.)} [\boldsymbol{\xi}_0, \boldsymbol{\zeta}_0], \tag{3.8}$$

i.e., for arbitrary bounded continuous function f holds

$$\mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\chi}_n}} f([\boldsymbol{\xi}_n^*, \boldsymbol{\zeta}_0]) - \mathbb{E}_{\mathbb{P}} f([\boldsymbol{\xi}_0, \boldsymbol{\zeta}_0]) \xrightarrow[n\to\infty]{[\mathbb{P}]\text{-}a.s.} 0. \tag{3.9}$$

Let $f([\cdot,\cdot])$ be such arbitrary bounded continuous function. Now consider the function of a single argument $g(\cdot) := f([\cdot, \boldsymbol{\zeta}_0])$. This will obviously be a bounded and continuous non-random function as well. By assumption (3.4), we will have that

$$\mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\chi}_n}} g(\boldsymbol{\xi}_n^*) - \mathbb{E}_{\mathbb{P}} g(\boldsymbol{\xi}_n) \xrightarrow[n\to\infty]{[\mathbb{P}]\text{-}a.s.} 0.$$

However, the latter expression is equivalent to (3.9). Therefore, we now know that $[\boldsymbol{\xi}_n^*, \boldsymbol{\zeta}_0]$ conditioned on $\boldsymbol{\chi}_n$ and $[\boldsymbol{\xi}_0, \boldsymbol{\zeta}_0]$ approach each other in distribution $[\mathbb{P}]$-almost surely along $\boldsymbol{\chi}_n$.

Secondly, consider $\|[\boldsymbol{\xi}_n^*, \boldsymbol{\zeta}_n^*] - [\boldsymbol{\xi}_n^*, \boldsymbol{\zeta}_0]\| = \|\boldsymbol{\zeta}_n^* - \boldsymbol{\zeta}_0\|$. This expression converges in probability $\mathbb{P}^*_{\boldsymbol{\chi}_n}$ to zero $[\mathbb{P}]$-almost surely due to assumption (3.5). Thus we have demonstrated two facts: (3.8) and

$$[\boldsymbol{\xi}_n^*, \boldsymbol{\zeta}_n^*] - [\boldsymbol{\xi}_n^*, \boldsymbol{\zeta}_0] \xrightarrow[n\to\infty]{\mathbb{P}^*_{\boldsymbol{\chi}_n}([\mathbb{P}]\text{-}a.s.)} \mathbf{0}. \tag{3.10}$$

Since each bounded continuous function is bounded Lipschitz and vice versa, consider any bounded Lipschitz function $h(\cdot,\cdot)$:

$$\begin{aligned}&\exists K, M > 0, \quad \forall \mathbf{x}_1, \mathbf{x}_2, \mathbf{y}_1, \mathbf{y}_2 :\\ &\qquad h([\mathbf{x}_1, \mathbf{x}_2]) \leq M \quad \& \quad |h([\mathbf{x}_1, \mathbf{x}_2]) - h([\mathbf{y}_1, \mathbf{y}_2])| \leq K\|[\mathbf{x}_1, \mathbf{x}_2] - [\mathbf{y}_1, \mathbf{y}_2]\|.\end{aligned}$$

Take some arbitrary $\epsilon > 0$ and majorize

$$\begin{aligned}&|\mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\chi}_n}} h([\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}^*_n]) - \mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\chi}_n}} h([\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}_0])| \leq \mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\chi}_n}} |h([\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}^*_n]) - h([\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}_0])|\\ &= \mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\chi}_n}} [|h([\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}^*_n]) - h([\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}_0])|\mathcal{I}\{\|[\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}^*_n] - [\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}_0]\| < \epsilon\}]\\ &\quad + \mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\chi}_n}} [|h([\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}^*_n]) - h([\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}_0])|\mathcal{I}\{\|[\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}^*_n] - [\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}_0]\| \geq \epsilon\}]\\ &\leq \mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\chi}_n}} [K\|[\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}^*_n] - [\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}_0]\|\mathcal{I}\{\|[\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}^*_n] - [\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}_0]\| < \epsilon\}]\\ &\quad + \mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\chi}_n}} [2M\mathcal{I}\{\|[\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}^*_n] - [\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}_0]\| \geq \epsilon\}]\\ &\leq K\epsilon\mathbb{P}^*_{\boldsymbol{\chi}_n} [\|[\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}^*_n] - [\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}_0]\| < \epsilon] + 2M\mathbb{P}^*_{\boldsymbol{\chi}_n} [\|[\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}^*_n] - [\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}_0]\| \geq \epsilon]\\ &\leq K\epsilon + 2M\mathbb{P}^*_{\boldsymbol{\chi}_n} [\|[\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}^*_n] - [\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}_0]\| \geq \epsilon] \quad [\mathbb{P}]\text{-}a.s.\end{aligned}$$

Hence,

$$\begin{aligned}&|\mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\chi}_n}} h([\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}^*_n]) - \mathbb{E}_{\mathbb{P}} h([\boldsymbol{\xi}_0, \boldsymbol{\zeta}_0])|\\ &\leq |\mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\chi}_n}} h([\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}^*_n]) - \mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\chi}_n}} h([\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}_0])| + |\mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\chi}_n}} h([\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}_0]) - \mathbb{E}_{\mathbb{P}} h([\boldsymbol{\xi}_0, \boldsymbol{\zeta}_0])|\\ &\leq K\epsilon + 2M\mathbb{P}^*_{\boldsymbol{\chi}_n} [\|[\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}^*_n] - [\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}_0]\| \geq \epsilon]\\ &\quad + |\mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\chi}_n}} h([\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}_0]) - \mathbb{E}_{\mathbb{P}} h([\boldsymbol{\xi}_0, \boldsymbol{\zeta}_0])| \quad [\mathbb{P}]\text{-}a.s.\end{aligned} \tag{3.11}$$

We take the limit in this expression as $n \to \infty$. Since (3.11) holds for arbitrary $\epsilon > 0$, the second term will go to zero $[\mathbb{P}]$-almost surely due to (3.10) and Definition 3.3. The third term (does not depend on ϵ) will also converge to zero $[\mathbb{P}]$-almost surely by (3.9). Thus,

$$\lim_{n\to\infty} |\mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\chi}_n}} h([\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}^*_n]) - \mathbb{E}_{\mathbb{P}} h([\boldsymbol{\xi}_0, \boldsymbol{\zeta}_0])| \leq K\epsilon \quad [\mathbb{P}]\text{-}a.s.$$

Since ϵ was arbitrary, we conclude that the limit must in fact be equal to zero $[\mathbb{P}]$-almost surely. Therefore, *(i)* is proved.

In order to prove result *(viii)*, consider again $\|[\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}^*_n] - [\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}_0]\| = \|\boldsymbol{\zeta}^*_n - \boldsymbol{\zeta}_0\|$. This expression converges in probability $\mathbb{P}^*_{\boldsymbol{\chi}_n}$ in probability $\mathbb{P}$ to zero due to assumption (3.7). Thus

$$[\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}^*_n] - [\boldsymbol{\xi}^*_n, \boldsymbol{\zeta}_0] \xrightarrow[n\to\infty]{\mathbb{P}^*_{\boldsymbol{\chi}_n}(\mathbb{P})} \mathbf{0}, \tag{3.12}$$

which can be rewritten according to Definition 3.3 as

$$\forall\epsilon > 0 : \left\{\mathbb{P}^*_{\boldsymbol{\mathcal{X}}_n}\left[\|[\boldsymbol{\xi}_n^*, \boldsymbol{\zeta}_n^*] - [\boldsymbol{\xi}_n^*, \boldsymbol{\zeta}_0]\| \geq \epsilon\right] \xrightarrow[n\to\infty]{\mathbb{P}} 0\right\}$$

Let us fix $\epsilon > 0$. Similarly as in the proof of part *(i)*, it can be demonstrated that

$$\mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\mathcal{X}}_n}} f([\boldsymbol{\xi}_n^*, \boldsymbol{\zeta}_0]) - \mathbb{E}_{\mathbb{P}} f([\boldsymbol{\xi}_0, \boldsymbol{\zeta}_0]) \xrightarrow[n\to\infty]{\mathbb{P}} 0, \tag{3.13}$$

when convergence $[\mathbb{P}]$-almost surely is just replaced by convergence in probability $\mathbb{P}$. Indeed, using inequality (3.11), we get for arbitrary $\tau > 0$ (and above chosen fixed and sufficiently small $\epsilon > 0$)

$$\begin{aligned}
&\mathbb{P}\left[|\mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\mathcal{X}}_n}} h([\boldsymbol{\xi}_n^*, \boldsymbol{\zeta}_n^*]) - \mathbb{E}_{\mathbb{P}} h([\boldsymbol{\xi}_0, \boldsymbol{\zeta}_0])| \geq \tau\right] \\
&\leq \mathbb{P}\left[K\epsilon \geq \tau\right] + \mathbb{P}\left[2M\mathbb{P}^*_{\boldsymbol{\mathcal{X}}_n}\left[\|[\boldsymbol{\xi}_n^*, \boldsymbol{\zeta}_n^*] - [\boldsymbol{\xi}_n^*, \boldsymbol{\zeta}_0]\| \geq \epsilon\right] \geq \tau\right] \\
&\quad + \mathbb{P}\left[|\mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\mathcal{X}}_n}} h([\boldsymbol{\xi}_n^*, \boldsymbol{\zeta}_0]) - \mathbb{E}_{\mathbb{P}} h([\boldsymbol{\xi}_0, \boldsymbol{\zeta}_0])| \geq \tau\right].
\end{aligned}$$

If we take the limit in previous inequality as n goes to infinity, the first term is zero and the second term will go to zero due to (3.12) and Definition 3.3. The third term will also converge to zero by (3.13). Thus

$$\lim_{n\to\infty} \mathbb{P}\left[|\mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\mathcal{X}}_n}} h([\boldsymbol{\xi}_n^*, \boldsymbol{\zeta}_n^*]) - \mathbb{E}_{\mathbb{P}} h([\boldsymbol{\xi}_0, \boldsymbol{\zeta}_0])| \geq \tau\right] = 0.$$

Since τ was arbitrary, *(viii)* is proved.

Assertions *(ii)–(vii)* are just corollaries of *(i)*, when continuous mapping theorem (see, e.g., van der Vaart (1998, Theorem 2.3)) is applied. Similarly, assertions *(ix)–(xiv)* are consequences of *(viii)*. □

From now on, we mean by a *bootstrap version* of $\boldsymbol{\xi} \equiv [\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n]^\top$ its (randomly) resampled sequence with replacement—denoted by $\boldsymbol{\xi}^* \equiv [\boldsymbol{\xi}_1^*, \ldots, \boldsymbol{\xi}_n^*]^\top$—with the same length, where for each $i \in \{1, \ldots, n\}$ holds $\mathbb{P}^*_{\boldsymbol{\xi}}[\boldsymbol{\xi}_i^* = \boldsymbol{\xi}_j] = 1/n$, $j = 1, \ldots, n$. So, $\boldsymbol{\xi}_i^*$ has a discrete uniform distribution on $\{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_n\}$ for every $i = 1, \ldots, n$.

A base stone for the consistency of the nonparametric bootstrap in the EIV model lies in the *bootstrap weak law of large numbers* (BWLLN). It will be postulated for independent variables for this time.

Theorem 3.3 (Bootstrap weak law of large numbers)**.** *Let $\{\xi_n\}_{n=1}^\infty$ be a sequence of independent random variables. If*

$$\sup_{n\in\mathbb{N}} \mathbb{E}_{\mathbb{P}} \xi_n^2 < \infty, \tag{3.14}$$

then

$$n^{-1}\sum_{i=1}^{n}\xi_i^* - n^{-1}\sum_{i=1}^{n}\xi_i \xrightarrow[n\to\infty]{\mathbb{P}^*_{\boldsymbol{\xi}}(\mathbb{P})} 0, \tag{3.15}$$

where $\boldsymbol{\xi}^* \equiv [\xi_1^*, \ldots, \xi_n^*]^\top$ *is the bootstrapped version of* $\boldsymbol{\xi} \equiv [\xi_1, \ldots, \xi_n]^\top$.

Proof. By the strong law of large numbers for independent random variables with assumption (3.14), we know that

$$n^{-1}\sum_{i=1}^{n}(\xi_i - \mathbb{E}_{\mathbb{P}}\xi_i) \xrightarrow[n\to\infty]{[\mathbb{P}]\text{-}a.s.} 0.$$

The Markov's inequality with (3.14) implies uniform equiboundedness in probability $\mathbb{P}$ of ξ_n^2. The conditional variance of the bootstrapped sample mean goes to zero as n increases to infinity, because

$$\begin{aligned}\mathbb{V}\mathrm{ar}_{\mathbb{P}^*_{\boldsymbol{\xi}}}\left(n^{-1}\sum_{i=1}^{n}\xi_i^*\right) &= n^{-1}\mathbb{V}\mathrm{ar}_{\mathbb{P}^*_{\boldsymbol{\xi}}}\xi_1^* = n^{-1}\left[\mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\xi}}}\xi_1^{*2} - (\mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\xi}}}\xi_1^*)^2\right]\\ &= n^{-1}\left[\sum_{k=1}^{n}n^{-1}\xi_k^2 - \left(\sum_{k=1}^{n}n^{-1}\xi_k\right)^2\right] = \mathcal{O}_{\mathbb{P}}(n^{-1}), \quad n\to\infty.\end{aligned}$$

Hence, the weak law of large numbers in the "starred" world (for resampled variables) provides

$$n^{-1}\sum_{i=1}^{n}\left(\xi_i^* - \mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\xi}}}\xi_i\right) = n^{-1}\sum_{i=1}^{n}\xi_i^* - n^{-1}\sum_{i=1}^{n}\xi_i \xrightarrow[n\to\infty]{\mathbb{P}^*_{\boldsymbol{\xi}}(\mathbb{P})} 0,$$

because ξ_i^* are conditionally *iid*. □

Convergence of matrices in Frobenius norm implies a type of spectral convergence, which can be mathematically formalized in the consequent lemma.

Lemma 3.4. *If* $\mathbf{A}$ *and* $\mathbf{B}$ *are square symmetric matrices, then*

$$|\lambda_{min}(\mathbf{A}) - \lambda_{min}(\mathbf{B})| \leq \|\mathbf{A}-\mathbf{B}\|.$$

Proof. See Gallo (1982b, Lemma 2.3). □

Remember that still $\|\cdot\| \equiv \|\cdot\|_F$.

Another technical, but very useful lemma is going to be derived. This lemma is a corollary of the Jensen's inequality and will help us to set the upper bounds for the higher moments of random variables' sums.

Lemma 3.5. *For $s \geq 1$, $m \in \mathbb{N}$, and $\lambda_i \geq 0$, $i \in \{1, \ldots, m\}$ holds*

$$\left(\sum_{i=1}^{m} \lambda_i\right)^s \leq m^{s-1} \sum_{i=1}^{m} \lambda_i^s.$$

Proof. Since $u \mapsto u^s$ is a convex function in $u \geq 0$ for every $s \geq 1$, then Jensen's inequality implies

$$\left(\sum_{i=1}^{m} \frac{1}{m} \lambda_i\right)^s \leq \frac{1}{m} \sum_{i=1}^{m} \lambda_i^s.$$

□

Suppose that all the errors $\{\Theta_{n,1}\}_{n=1}^{\infty}, \ldots, \{\Theta_{n,p}\}_{n=1}^{\infty}$, and $\{\varepsilon_n\}_{n=1}^{\infty}$ exist on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let the original rows of the error matrix $[\boldsymbol{\Theta}, \boldsymbol{\varepsilon}]$ be independent random vectors with a common probability law in $\mathbb{R}^{p+1}$ having the fourth power of Euclidean norm integrable. Moreover, the *bootstrap (empirical) probability* is the conditional probability

$$\mathbb{P}^*[\cdot] \equiv \mathbb{P}^*_{[\mathbf{X},\mathbf{Y}]}[\cdot] = \mathbb{E}_{\mathbb{P}}\left[\mathcal{I}\{\cdot\}|[\mathbf{X}, \mathbf{Y}]\right].$$

To avoid further ambiguity we make a convention: when a probability for convergence almost surely is not specified, it is meant that convergence $[\mathbb{P}]$-almost surely is considered. Similarly, if convergence in probability is not specified, we assume convergence in probability $\mathbb{P}$.

The resampled vectors $[\mathbf{X}^*_{i,\bullet}, Y_i^*]$, $i = 1, \ldots, n$ are *conditionally independent*, given the original data $[\mathbf{X}, \mathbf{Y}]$, because of the random resampling. Moreover, bootstrapped errors $[\boldsymbol{\Theta}^*_{i,\bullet}, \epsilon_i^*]$, $i = 1, \ldots, n$ are also conditionally independent, given the original data $[\mathbf{X}, \mathbf{Y}]$, with common probability distribution. Recall that $\widehat{\boldsymbol{\beta}}$ minimizes $\sum_{i=1}^{n} \|[\boldsymbol{\Theta}_{\bullet,i}, \varepsilon_i]\|_2^2$. Loosely speaking, $\widehat{\boldsymbol{\beta}}$ is to $\mathbb{P}^*$ as $\boldsymbol{\beta}$ to the "true" unconditional probability $\mathbb{P}$.

Justification of the nonparametric bootstrap procedure for the *nuisance homoscedasticity parameter* σ^2 will be provided at first. This theorem is pivotal, but not for the fact that it characterizes distributional closeness of the bootstrapped variance estimate $\widehat{\sigma}^{2*}$ and the original variance estimate $\widehat{\sigma}^2$ asymptotically. The essential importance of the following theorem lies in providing a distributional property of a spectral element of the resampled data such as the squared smallest singular value of matrix $[\mathbf{X}^*, \mathbf{Y}^*]$, which is also the smallest eigenvalue of $[\mathbf{X}^*, \mathbf{Y}^*]^\top [\mathbf{X}^*, \mathbf{Y}^*]$ due to the SVD of $[\mathbf{X}^*, \mathbf{Y}^*]$.

Theorem 3.6. *Let assume the EIV model and the assumption* (D) *be satisfied. Let*

$$\widehat{\sigma}^{2*} := \frac{\lambda^*}{n}.$$

If

$$\sup_{n\in\mathbb{N}} Z_{n,j}^2 < \infty, \tag{3.16}$$

$$\sup_{n\in\mathbb{N}} \mathbb{E}|\Theta_{n,j}|^4 < \infty, \quad \textit{and} \quad \sup_{n\in\mathbb{N}} \mathbb{E}|\varepsilon_n|^4 < \infty \tag{3.17}$$

for each $j \in \{1,\ldots,p\}$*, then*

$$\widehat{\sigma}^{2*} \xrightarrow[n\to\infty]{\mathbb{P}^*(\mathbb{P})} \sigma^2.$$

Proof. Let us consider

$$\frac{1}{n}[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}] = \frac{1}{n}\sum_{i=1}^{n}[\mathbf{X}_{i,\bullet},Y_i]^\top[\mathbf{X}_{i,\bullet},Y_i] = \frac{1}{n}\sum_{i=1}^{n}\begin{bmatrix} \boldsymbol{\Xi}_i & \boldsymbol{\zeta}_i^\top \\ \boldsymbol{\zeta}_i & \eta_i \end{bmatrix},$$

where

$$\begin{aligned} \boldsymbol{\Xi}_i &:= \mathbf{Z}_{i,\bullet}^\top\mathbf{Z}_{i,\bullet} + \mathbf{Z}_{i,\bullet}^\top\boldsymbol{\Theta}_{i,\bullet} + \boldsymbol{\Theta}_{i,\bullet}^\top\mathbf{Z}_{i,\bullet} + \boldsymbol{\Theta}_{i,\bullet}^\top\boldsymbol{\Theta}_{i,\bullet}, \\ \boldsymbol{\zeta}_i &:= \mathbf{Z}_{i,\bullet}^\top\mathbf{Z}_{i,\bullet}\boldsymbol{\beta} + \mathbf{Z}_{i,\bullet}^\top\varepsilon_i + \boldsymbol{\Theta}_{i,\bullet}^\top\mathbf{Z}_{i,\bullet}\boldsymbol{\beta} + \boldsymbol{\Theta}_{i,\bullet}^\top\varepsilon_i, \end{aligned}$$

and

$$\eta_i := (\mathbf{Z}_{i,\bullet}\boldsymbol{\beta})^2 + 2\varepsilon_i\mathbf{Z}_{i,\bullet}\boldsymbol{\beta} + \varepsilon_i^2.$$

Assumptions (3.16) and (3.17) ensure that all the elements of the matrix $[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]$ satisfy requirements of Theorem 3.3. That is indeed true, because Lemma 3.5 provides

$$\begin{aligned} \sup_{n\in\mathbb{N}} \mathbb{E}_{\mathbb{P}}|X_{n,j}|^4 = \sup_{n\in\mathbb{N}} \mathbb{E}_{\mathbb{P}}|Z_{n,j}+\Theta_{n,j}|^4 &\leq 2^3 \sup_{n\in\mathbb{N}} \mathbb{E}_{\mathbb{P}}\left(|Z_{n,j}|^4 + |\Theta_{n,j}|^4\right) \\ &\leq 2^3\left(\sup_{n\in\mathbb{N}}|Z_{n,j}|^2\right)^2 + 2^3\sup_{n\in\mathbb{N}}\mathbb{E}_{\mathbb{P}}|\Theta_{n,j}|^4 < \infty, \quad j\in\{1,\ldots,p\} \end{aligned}$$

and

$$\begin{aligned} \sup_{n\in\mathbb{N}} \mathbb{E}_{\mathbb{P}}|Y_n|^2 = \sup_{n\in\mathbb{N}} \mathbb{E}_{\mathbb{P}}|Z_{n,\bullet}\boldsymbol{\beta}+\varepsilon_n|^2 &\leq 2^3\sup_{n\in\mathbb{N}}\mathbb{E}_{\mathbb{P}}\left(|Z_{n,\bullet}\boldsymbol{\beta}|^4 + |\varepsilon_n|^4\right) \\ &\leq 2^3p^4 \max_{j\in\{1,\ldots,p\}}\left\{\beta_j^4\left(\sup_{n\in\mathbb{N}}|Z_{n,j}|^2\right)^2\right\} + 2^3\sup_{n\in\mathbb{N}}\mathbb{E}_{\mathbb{P}}|\varepsilon_n|^4 < \infty. \end{aligned}$$

Thereby, the elementwise convergence in probability $\mathbb{P}^*$ along $[\mathbf{X},\mathbf{Y}]$ in probability $\mathbb{P}$ holds due to the bootstrap WLLN (Theorem 3.3):

$$\frac{1}{n}[\mathbf{X}^*, \mathbf{Y}^*]^\top [\mathbf{X}^*, \mathbf{Y}^*] - \frac{1}{n}[\mathbf{X}, \mathbf{Y}]^\top [\mathbf{X}, \mathbf{Y}] \xrightarrow[n\to\infty]{\mathbb{P}^*(\mathbb{P})} \mathbf{0}.$$

Convergence of the matrix elements implies convergence of the matrices in Frobenius norm, i.e.,

$$\frac{1}{n}\|[\mathbf{X}^*, \mathbf{Y}^*]^\top [\mathbf{X}^*, \mathbf{Y}^*] - [\mathbf{X}, \mathbf{Y}]^\top [\mathbf{X}, \mathbf{Y}]\| \xrightarrow[n\to\infty]{\mathbb{P}^*(\mathbb{P})} 0.$$

By Lemma 3.4,

$$\frac{1}{n}|\lambda_{min}([\mathbf{X}^*, \mathbf{Y}^*]^\top [\mathbf{X}^*, \mathbf{Y}^*]) - \lambda_{min}([\mathbf{X}, \mathbf{Y}]^\top [\mathbf{X}, \mathbf{Y}])| \xrightarrow[n\to\infty]{\mathbb{P}^*(\mathbb{P})} 0,$$

where $\lambda_{min}([\mathbf{X}, \mathbf{Y}]^\top [\mathbf{X}, \mathbf{Y}]) \equiv \lambda$ and $\lambda_{min}([\mathbf{X}^*, \mathbf{Y}^*]^\top [\mathbf{X}^*, \mathbf{Y}^*]) \equiv \lambda^*$. With respect to (1.24), it holds that

$$\frac{\lambda}{n} \xrightarrow[n\to\infty]{\mathbb{P}} \sigma^2,$$

which completes the proof. □

Remark 3.2. Since convergence in probability implies convergence in distribution (even in the conditional world), it is possible to say that

$$\widehat{\sigma}^{*2}\big|[\mathbf{X}, \mathbf{Y}] \xleftrightarrow[n\to\infty]{\mathscr{D}(\mathbb{P})} \widehat{\sigma}^2.$$

Theorem 3.6 also says that along all sample sequences in probability $\mathbb{P}$, given $[\mathbf{X}, \mathbf{Y}]$, the conditional law of $\widehat{\sigma}^{*2}$ converges weakly to the point mass at σ^2 as n tends to infinity.

Assumption (D) can be seen as a convergence of a specific sum in the Cauchy sense, i.e., a limit of the averaged partial sums (Cesaro limit). A forthcoming technical lemma allows us to derive various implications of design assumption (D).

Lemma 3.7. *If*

$$\lim_{n\to\infty} n^{-2+\delta} \sum_{i=1}^{n} a_i$$

exists and is finite for some $\delta > 0$*, then*

$$\sum_{n=1}^{\infty} \frac{a_n}{n^2} < \infty.$$

Proof. Due to the Abel's partial summation (Itō, 1993, p. 1412), we have

$$\sum_{i=1}^{n}\frac{a_i}{i^2}=\sum_{i=1}^{n-1}\left(\sum_{j=1}^{i-1}a_j\right)\left(\frac{1}{i^2}-\frac{1}{(i+1)^2}\right)+\frac{1}{n^2}\sum_{i=1}^{n}a_i,\quad \forall n>1. \tag{3.18}$$

If n tends to infinity, the last term of (3.18) tends to zero due to the lemma's assumption. Moreover, the infinite sum formed from the first summand on the right hand side of (3.18) is convergent if and only if

$$\sum_{i=1}^{\infty}i^{-3}\sum_{j=1}^{i-1}a_j=\sum_{i=1}^{\infty}i^{-1-\delta}\left(i^{-2+\delta}\sum_{j=1}^{i-1}a_j\right)$$

is convergent, but the right hand side of previous equation is convergent according to the Abel's convergence criterion ($i^{-2+\delta}\sum_{j=1}^{i-1}a_j \xrightarrow[j\to\infty]{} 0$). Hence, $\sum_{n=1}^{\infty}\frac{a_n}{n^2}$ converges as well. □

Various averaged linear combinations of the errors from the EIV model converge [$\mathbb{P}$]-almost surely, which will become important later on.

Lemma 3.8. *Let assume the EIV model and the assumption* (D) *be satisfied. Then*

(i) $n^{-1}\boldsymbol{\Theta}^\top\boldsymbol{\Theta}\to\sigma^2\mathbf{I}$ [$\mathbb{P}$]-*a.s.*, $n\to\infty$;

(ii) $n^{-1}\boldsymbol{\varepsilon}^\top\boldsymbol{\varepsilon}\to\sigma^2$ [$\mathbb{P}$]-*a.s.*, $n\to\infty$;

(iii) $n^{-1}\boldsymbol{\Theta}^\top\boldsymbol{\varepsilon}\to\mathbf{0}$ [$\mathbb{P}$]-*a.s.*, $n\to\infty$;

(iv) $n^{-1}\mathbf{Z}^\top\boldsymbol{\varepsilon}\to\mathbf{0}$ [$\mathbb{P}$]-*a.s.*, $n\to\infty$;

(v) $n^{-1}\mathbf{Z}^\top\boldsymbol{\Theta}\to\mathbf{0}$ [$\mathbb{P}$]-*a.s.*, $n\to\infty$.

Proof. According to the SLLN for identically distributed variables with keeping in mind that the covariance matrix of the rows of $[\boldsymbol{\Theta},\boldsymbol{\varepsilon}]$ is $\sigma^2\mathbf{I}$, we have

$$n^{-1}\boldsymbol{\Theta}^\top\boldsymbol{\Theta}=n^{-1}\sum_{i=1}^{n}\boldsymbol{\Theta}_{i,\bullet}^\top\boldsymbol{\Theta}_{i,\bullet}\xrightarrow[n\to\infty]{[\mathbb{P}]\text{-}a.s.}\mathbb{E}\boldsymbol{\Theta}_{1,\bullet}^\top\boldsymbol{\Theta}_{1,\bullet}=\sigma^2\mathbf{I} \tag{3.19}$$

and

$$n^{-1}\boldsymbol{\varepsilon}^\top\boldsymbol{\varepsilon}=n^{-1}\sum_{i=1}^{n}\varepsilon_i^2\xrightarrow[n\to\infty]{[\mathbb{P}]\text{-}a.s.}\sigma^2. \tag{3.20}$$

Similarly,

$$n^{-1}\boldsymbol{\Theta}^\top\boldsymbol{\varepsilon}=n^{-1}\sum_{i=1}^{n}\boldsymbol{\Theta}_{i,\bullet}^\top\varepsilon_i\xrightarrow[n\to\infty]{[\mathbb{P}]\text{-}a.s.}\mathbb{E}\boldsymbol{\Theta}_{1,\bullet}^\top\varepsilon_1=\mathbf{0}. \tag{3.21}$$

Secondly, applying the SLLN for non-identically distributed variables (but with the same zero mean), one gets

$$n^{-1}\mathbf{Z}^\top\boldsymbol{\varepsilon} = n^{-1}\sum_{i=1}^{n}\mathbf{Z}_{i,\bullet}^\top\varepsilon_i \xrightarrow[n\to\infty]{[\mathbb{P}]\text{-}a.s.} \mathbb{E}\mathbf{Z}_{1,\bullet}^\top\varepsilon_1 = \mathbf{0}, \tag{3.22}$$

because the following sum is convergent by applying (D) and Lemma 3.7

$$\sum_{n=1}^{\infty}\frac{\mathbb{V}\text{ar}\left\{Z_{nj}\varepsilon_n\right\}}{n^2} = \sigma^2\sum_{n=1}^{\infty}\frac{Z_{nj}^2}{n^2} < \infty, \quad j = 1,\ldots,p.$$

Analogously, using the SLLN for non-identically distributed variables,

$$n^{-1}\mathbf{Z}^\top\boldsymbol{\Theta} = n^{-1}\sum_{i=1}^{n}\mathbf{Z}_{i,\bullet}^\top\boldsymbol{\Theta}_{i,\bullet} \xrightarrow[n\to\infty]{[\mathbb{P}]\text{-}a.s.} \mathbb{E}\mathbf{Z}_{1,\bullet}^\top\boldsymbol{\Theta}_{1,\bullet} = \mathbf{0}. \tag{3.23}$$

□

A pylon for the proof of central limit theorem for the bootstrapped sample is created by an extension of the well-known Berry-Esseen theorem.

Theorem 3.9 (Berry-Esseen-Katz theorem)**.** *Let g be a non-negative, even, non-decreasing function on $[0,\infty)$ satisfying:*

(i) $\lim_{x\to\infty} g(x) = \infty$*,*

(ii) $x/g(x)$ is defined for all $x \in \mathbb{R}$ and non-decreasing on $[0,\infty)$.

Assume that $\{\xi_n\}_{n=1}^{\infty}$ are iid random variables such that $\mathbb{E}\xi_1 = 0$ and $\mathbb{V}ar\xi_1 = \varsigma^2 > 0$. If

$$\mathbb{E}\xi_1^2 g(\xi_1) < \infty,$$

then there exists a constant $C > 0$, such that for all $n \in \mathbb{N}$ holds

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left[\frac{1}{\sqrt{n\varsigma^2}}\sum_{i=1}^{n}\xi_i \le x\right] - \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{t^2}{2}\right\}dt\right| \le \frac{C\mathbb{E}\xi_1^2 g\left(\xi_1/\varsigma\right)}{\varsigma^2 g(\sqrt{n})}.$$

Proof. See Katz (1963). □

If it is known that a statistic has approximately normal distribution and we are able to construct a bootstrap version of this statistic, one may be interested in the asymptotical comparison of the bootstrap distribution with the original one. A handy tool for performing such a comparison—showing approximate closeness—can be a *bootstrap central limit theorem.* Besides that, it can assure of the bootstrap appropriateness.

Theorem 3.10 (Bootstrap central limit theorem for independent variables)**.** *Let* $\{\xi_n\}_{n=1}^{\infty}$ *be a sequence of zero mean independent random variables satisfying*

$$\sup_{n\in\mathbb{N}} \mathbb{E}_{\mathbb{P}}\xi_n^4 < \infty. \tag{3.24}$$

Suppose that $\boldsymbol{\xi}^* \equiv [\xi_1^*, \ldots, \xi_n^*]^\top$ *is the bootstrapped version of* $\boldsymbol{\xi} \equiv [\xi_1, \ldots, \xi_n]^\top$ *and denote*

$$\bar{\xi}_n := n^{-1}\sum_{i=1}^{n}\xi_i, \quad \bar{\xi}_n^* := n^{-1}\sum_{i=1}^{n}\xi_i^*, \quad \textit{and} \quad \varsigma_n^2 := \sum_{i=1}^{n}\mathbb{V}ar_{\mathbb{P}}\xi_i.$$

If

$$\liminf_{n\to\infty}\frac{\varsigma_n^2}{n} = \varsigma^2 > 0, \tag{3.25}$$

then

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}_{\boldsymbol{\xi}}^*\left[\frac{n}{\sqrt{\varsigma_n^2}}\left(\bar{\xi}_n^* - \bar{\xi}_n\right) \le x\right] - \mathbb{P}\left[\frac{n}{\sqrt{\varsigma_n^2}}\bar{\xi}_n \le x\right]\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0. \tag{3.26}$$

Proof. The Lyapunov condition for sequence of random variables $\{\xi_n\}_{n=1}^{\infty}$ is satisfied due to (3.24) and (3.25), i.e., for fixed $\omega > 0$:

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

Thereupon, the CLT for $\{\xi_n\}_{n=1}^{\infty}$ holds and

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left[\frac{n}{\sqrt{\varsigma_n^2}}\bar{\xi}_n \le x\right] - \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{t^2}{2}\right\}\mathrm{d}t\right| \xrightarrow[n\to\infty]{} 0.$$

Henceforth, to prove this theorem, it suffices to show the following three statements:

(i)

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}_{\boldsymbol{\xi}}^*\left[\frac{\sqrt{n}}{\sqrt{\mathbb{V}\mathrm{ar}_{\mathbb{P}_{\boldsymbol{\xi}}^*}\xi_1^*}}\left(\bar{\xi}_n^* - \mathbb{E}_{\mathbb{P}_{\boldsymbol{\xi}}^*}\bar{\xi}_n^*\right) \le x\right] - \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{t^2}{2}\right\}\mathrm{d}t\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0;$$

(ii)

$$\mathbb{V}\mathrm{ar}_{\mathbb{P}_{\boldsymbol{\xi}}^*}\xi_1^* - n^{-1}\varsigma_n^2 \xrightarrow[n\to\infty]{[\mathbb{P}]\text{-}a.s.} 0;$$

(iii)

$$\mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\xi}}}\bar{\xi}^*_n = \bar{\xi}_n \quad [\mathbb{P}]\text{-}a.s.$$

Proving *(iii)* is trivial, because the bootstrapped variables $\{\xi^*_n\}_{n=1}^{\infty}$ are conditionally *iid* and, therefore,

$$\mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\xi}}}\bar{\xi}^*_n = \mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\xi}}}\xi^*_1 = n^{-1}\sum_{i=1}^{n}\xi_i = \bar{\xi}_n \quad [\mathbb{P}]\text{-}a.s.$$

Let us calculate the conditional variance of the bootstrapped ξ^*_1:

$$\mathbb{V}\text{ar}_{\,\mathbb{P}^*_{\boldsymbol{\xi}}}\xi^*_1 = \mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\xi}}}\xi^{*2}_1 - (\mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\xi}}}\xi^*_1)^2 = n^{-1}\sum_{i=1}^{n}\xi_i^2 - \left(n^{-1}\sum_{i=1}^{n}\xi_i\right)^2 \quad [\mathbb{P}]\text{-}a.s.$$

The strong law of large numbers for independent non-identically distributed random variables with (3.24) provide

$$\bar{\xi}_n - n^{-1}\sum_{i=i}^{n}\mathbb{E}_{\mathbb{P}}\xi_i = \bar{\xi}_n \xrightarrow[n\to\infty]{[\mathbb{P}]\text{-}a.s.} 0$$

and

$$0 \xleftarrow[n\to\infty]{[\mathbb{P}]\text{-}a.s.} n^{-1}\sum_{i=1}^{n}\xi_i^2 - \left(n^{-1}\sum_{i=1}^{n}\xi_i\right)^2 - n^{-1}\sum_{i=1}^{n}\mathbb{E}_{\mathbb{P}}\xi_i^2 = \mathbb{V}\text{ar}_{\,\mathbb{P}^*_{\boldsymbol{\xi}}}\xi^*_1 - n^{-1}\varsigma_n^2.$$

The last result of the SLLN is true, because (3.24) implies

$$\sum_{n=1}^{\infty}\frac{\mathbb{V}\text{ar}_{\,\mathbb{P}}\xi_n^2}{n^2} \leq \sum_{n=1}^{\infty}\frac{\mathbb{E}_{\mathbb{P}}\xi_n^4}{n^2} \leq \left[\sup_{\iota\in\mathbb{N}}\mathbb{E}_{\mathbb{P}}\xi_\iota^4\right]\sum_{n=1}^{\infty}n^{-2} < \infty.$$

Thus *(ii)* is proved.

Berry-Esseen-Katz Theorem 3.9 with $g(x) = |x|^{\epsilon}$, $\epsilon > 0$ for the bootstrapped sequence of *iid* (with respect to $\mathbb{P}^*$) random variables $\{\xi^*_n\}_{n=1}^{\infty}$ results into

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}^*_{\boldsymbol{\xi}}\left[\frac{\sqrt{n}}{\sqrt{\mathbb{V}\text{ar}_{\,\mathbb{P}^*_{\boldsymbol{\xi}}}\xi^*_1}}\left(\bar{\xi}^*_n - \mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\xi}}}\bar{\xi}^*_n\right) \leq x\right] - \int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\exp\left\{-\frac{t^2}{2}\right\}\mathrm{d}t\right|$$
$$\leq Cn^{-\epsilon/2}\mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\xi}}}\left|\frac{\xi^*_1 - \mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\xi}}}\xi^*_1}{\sqrt{\mathbb{V}\text{ar}_{\,\mathbb{P}^*_{\boldsymbol{\xi}}}\xi^*_1}}\right|^{2+\epsilon} \quad [\mathbb{P}]\text{-}a.s., \quad (3.27)$$

where $C > 0$ is an absolute constant.

The Minkowski inequality and Lemma 3.5 provide an upper bound for the nominator from the right-hand side of (3.27):

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\xi}}}|\xi_1^* - \mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\xi}}}\xi_1^*|^{2+\epsilon} = n^{-1}\sum_{i=1}^n \left|\xi_i - n^{-1}\sum_{j=1}^n \xi_j\right|^{2+\epsilon} \\
&\leq n^{-1}\left\{\left(\sum_{i=1}^n |\xi_i|^{2+\epsilon}\right)^{1/(2+\epsilon)} + n^{-(1+\epsilon)/(2+\epsilon)}\left|\sum_{j=1}^n \xi_j\right|\right\}^{2+\epsilon} \\
&\leq 2^{1+\epsilon} n^{-1}\sum_{i=1}^n |\xi_i|^{2+\epsilon} + 2^{1+\epsilon}\left|n^{-1}\sum_{i=1}^n \xi_i\right|^{2+\epsilon} \qquad [\mathbb{P}]\text{-}a.s.
\end{aligned}$$

The right-hand side of the previously derived upper bound is uniformly bounded in probability $\mathbb{P}$, because of the Markov's inequality and (3.24). In very deed, for fixed $\tau > 0$

$$\mathbb{P}\left[n^{-1}\sum_{i=1}^n |\xi_i|^{2+\epsilon} \geq \tau\right] \leq \tau^{-1} n^{-1}\sum_{i=1}^n \mathbb{E}_{\mathbb{P}}|\xi_i|^{2+\epsilon} \leq \tau^{-1}\sup_{\iota\in\mathbb{N}} \mathbb{E}_{\mathbb{P}}|\xi_\iota|^{2+\epsilon} < \infty, \quad \forall n \in \mathbb{N}$$

and

$$\mathbb{P}\left[\left|n^{-1}\sum_{i=1}^n \xi_i\right| \geq \tau\right] \leq \tau^{-1} n^{-1}\mathbb{E}_{\mathbb{P}}\left|\sum_{i=1}^n \xi_i\right| \leq \tau^{-1}\sup_{\iota\in\mathbb{N}} \mathbb{E}_{\mathbb{P}}|\xi_\iota| < \infty, \quad \forall n \in \mathbb{N}.$$

Since $\mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\xi}}}|\xi_1^* - \mathbb{E}_{\mathbb{P}^*}\xi_1^*|^{2+\epsilon}$ is bounded in probability $\mathbb{P}$ uniformly over n and the denominator of the right-hand side of (3.27) is uniformly bounded away from zero due to (3.25), then the left-hand side of (3.27) converges in probability $\mathbb{P}$ to zero as n tends to infinity. So, *(i)* is proved as well. □

Assumption (3.24) may seem a little bit restrictive. It may be weakened to

$$\sum_{n=1}^\infty \frac{\mathbb{E}_{\mathbb{P}}\xi_n^4}{n^2} < \infty \quad \text{and} \quad \sup_{n\in\mathbb{N}} \mathbb{E}_{\mathbb{P}}|\xi_n|^{2+\omega} < \infty, \text{ for some } \omega > 0;$$

as found out while going through the proof. On the other hand, the previous weaker premises are rather complicated to verify for transformed errors in the proof of bootstrap consistency, which will be noticed later on. The EIV model will force us to suppose (3.24) anyway, despite its limitation.

Belyaev (1995) strengthened assumption (3.24) and replaced by

$$\sup_{n\in\mathbb{N}} \mathbb{E}_{\mathbb{P}}|\sqrt{n}\xi_n|^{4+\omega} < \infty, \tag{3.28}$$

for some $\omega > 0$. Afterwards, bootstrap CLT 3.10 provides a stronger result:

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}^*_{\boldsymbol{\xi}}\left[\frac{n}{\sqrt{\varsigma_n^2}}\left(\bar{\xi}_n^* - \bar{\xi}_n\right) \leq x\right] - \mathbb{P}\left[\frac{n}{\sqrt{\varsigma_n^2}}\bar{\xi}_n \leq x\right]\right| \xrightarrow[n\to\infty]{[\mathbb{P}]\text{-}a.s.} 0, \tag{3.29}$$

whereas the convergence in distribution in probability is replaced by the convergence in distribution almost surely. In spite of this, assumption (3.28) can be considered as too much restrictive.

Our situation would become much easier, when *iid* variables are assumed. Let us have a look at the proof of Theorem 3.10. Additionally assume that $\mathbb{E}_{\mathbb{P}}|\xi_1|^{2+\omega} < \infty$ for some $\omega > 0$. Hence, the right-hand side of (3.27) converge [$\mathbb{P}$]-almost surely to zero. Moreover, the finiteness of $(2+\omega)$-th moment is enough to show *(ii)* as well. Therefore, (3.29) holds in this case.

A utilization of the Cramér-Wold device helps us to derive a bootstrap version of the CLT for random vectors.

Theorem 3.11 (Bootstrap multivariate central limit theorem for independent vectors)**.** *Let* $\{\boldsymbol{\xi}_n\}_{n=1}^{\infty}$ *be a sequence of zero mean independent q-dimensional random vectors satisfying*

$$\sup_{n\in\mathbb{N}} \mathbb{E}_{\mathbb{P}}|\xi_{j,n}|^4 < \infty, \quad j \in \{1,\ldots,q\}, \tag{3.30}$$

where $\boldsymbol{\xi}_n \equiv [\xi_{1,n},\ldots,\xi_{q,n}]^\top \in \mathbb{R}^q$, $n \in \mathbb{N}$. *Assume that* $\boldsymbol{\Xi}^* \equiv [\boldsymbol{\xi}_1^*,\ldots,\boldsymbol{\xi}_n^*]^\top$ *is the bootstrapped version of* $\boldsymbol{\Xi} \equiv [\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_n]^\top$. *Denote*

$$\bar{\boldsymbol{\xi}}_n := n^{-1}\sum_{i=1}^n \boldsymbol{\xi}_i, \quad \bar{\boldsymbol{\xi}}_n^* := n^{-1}\sum_{i=1}^n \boldsymbol{\xi}_i^*, \quad \textit{and} \quad \boldsymbol{\Gamma}_n := \sum_{i=1}^n \mathbb{V}ar_{\mathbb{P}}\boldsymbol{\xi}_i.$$

If

$$\lim_{n\to\infty}\frac{1}{n}\boldsymbol{\Gamma}_n = \boldsymbol{\Gamma} > \mathbf{0}, \tag{3.31}$$

then

$$n\boldsymbol{\Gamma}_n^{-1/2}\left(\bar{\boldsymbol{\xi}}_n^* - \bar{\boldsymbol{\xi}}_n\right)\Big|\boldsymbol{\Xi} \xleftrightarrow[n\to\infty]{\mathscr{D}(\mathbb{P})} n\boldsymbol{\Gamma}_n^{-1/2}\bar{\boldsymbol{\xi}}_n \tag{3.32}$$

and, moreover,

$$\sqrt{n}\left(\bar{\boldsymbol{\xi}}_n^* - \mathbb{E}_{\mathbb{P}^*}\boldsymbol{\xi}_1^*\right)\Big|\boldsymbol{\Xi} \xleftrightarrow[n\to\infty]{\mathscr{D}(\mathbb{P})} \sqrt{n}\bar{\boldsymbol{\xi}}_n. \tag{3.33}$$

Proof. According to Cramér-Wold theorem (see Appendix A.2, Theorem A.3) it is sufficient to ensure that all the assumptions of one-dimensional bootstrap CLT 3.10 are valid for any linear combination of the elements of random vector $\boldsymbol{\xi}_n$, $n \in \mathbb{N}$.

For arbitrary fixed $\mathbf{t} \in \mathbb{R}^q$ using Lemma 3.5, we get

$$\sup_{n\in\mathbb{N}} \mathbb{E}_{\mathbb{P}}|\mathbf{t}^\top \boldsymbol{\xi}_n|^4 \leq q^3 \sup_{n\in\mathbb{N}} \sum_{j=1}^{q} t_j^4 \mathbb{E}_{\mathbb{P}}|\xi_{j,n}|^4 \leq q^4 \max_{j=1,\ldots,q} t_j^4 \sup_{n\in\mathbb{N}} \mathbb{E}_{\mathbb{P}}|\xi_{j,n}|^4 < \infty.$$

Hence, assumption (3.30) implies assumption (3.24) for random variables $\mathbf{t}^\top \boldsymbol{\xi}_n$, $n \in \mathbb{N}$.

Similar situation arises, when assumption (3.31) implies assumption (3.25) for such an arbitrary linear combination, i.e., positive definiteness of matrix $\boldsymbol{\Gamma}$ yields

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{V}\text{ar}_{\mathbb{P}} \mathbf{t}^\top \boldsymbol{\xi}_i = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbf{t}^\top \left(\mathbb{V}\text{ar}_{\mathbb{P}} \boldsymbol{\xi}_i\right) \mathbf{t} = \mathbf{t}^\top \left(\lim_{n\to\infty} \frac{1}{n} \boldsymbol{\Gamma}_n\right) \mathbf{t} = \mathbf{t}^\top \boldsymbol{\Gamma} \mathbf{t} > 0.$$

Finally, we need to realize that (3.31) holds, (3.32) has already been proved above, $\{\boldsymbol{\xi}_i^*\}_{i=1}^n$ are conditionally *iid*, and

$$\mathbb{E}_{\mathbb{P}^*_{\Xi}} \boldsymbol{\xi}_1^* = n^{-1} \sum_{i=1}^{n} \boldsymbol{\xi}_i = \bar{\boldsymbol{\xi}}_n.$$

□

The main result for the asymptotical validity of the proper nonparametric bootstrap can be stated.

Theorem 3.12 (Coincidence of the nonparametric bootstrap distribution in EIV)**.** *Let assume the EIV model and assumption* (D) *be satisfied. Suppose that*

$$\sup_{n\in\mathbb{N}} Z_{n,j}^2 < \infty, \quad j \in \{1,\ldots,p\}, \tag{3.34}$$

$$\sup_{n\in\mathbb{N}} \mathbb{E}|\Theta_{n,j}|^8 < \infty, \quad j \in \{1,\ldots,p\}, \tag{3.35}$$

and

$$\sup_{n\in\mathbb{N}} \mathbb{E}|\varepsilon_n|^8 < \infty. \tag{3.36}$$

If there exists a positive definite matrix $\beth$ *such that*

$$\lim_{n\to\infty} \frac{1}{n} \mathbb{V}ar_{\mathbb{P}} [\mathbf{X},\mathbf{Y}]^\top [\mathbf{X},\mathbf{Y}] \begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix} = \beth > \mathbf{0}, \tag{3.37}$$

then

$$\sqrt{n}(\widetilde{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) \Big| [\mathbf{X},\mathbf{Y}] \xleftrightarrow[n\to\infty]{\mathscr{D}(\mathbb{P})} \sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}).$$

Proof. Partitioning (1.12) yields

$$\mathbf{X}^\top\mathbf{X}\widehat{\boldsymbol{\beta}} - \mathbf{X}^\top\mathbf{Y} = \lambda\widehat{\boldsymbol{\beta}} \quad \text{and} \quad \mathbf{Y}^\top\mathbf{X}\widehat{\boldsymbol{\beta}} - \mathbf{Y}^\top\mathbf{Y} = -\lambda.$$

The previous equations can be rewritten in the following manner

$$\mathbf{X}^\top\mathbf{Y} = (\mathbf{X}^\top\mathbf{X} - \lambda\mathbf{I})\widehat{\boldsymbol{\beta}} \quad \text{and} \quad \mathbf{Y}^\top\mathbf{Y} = \mathbf{Y}^\top\mathbf{X}\widehat{\boldsymbol{\beta}} + \lambda.$$

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

$$[\mathbf{I}, \boldsymbol{\beta}]\,[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}] \begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix} = [\mathbf{I}, \boldsymbol{\beta}] \left[\mathbf{I}, \widehat{\boldsymbol{\beta}}\right]^\top (\mathbf{X}^\top\mathbf{X} - \lambda\mathbf{I})(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}) \tag{3.38}$$

and, then,

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = -\boldsymbol{\Delta}_n^{-1} \left([\mathbf{I}, \boldsymbol{\beta}] \left[\mathbf{I}, \widehat{\boldsymbol{\beta}}\right]^\top\right)^{-1} [\mathbf{I}, \boldsymbol{\beta}] \left(n^{-1/2}[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]\right) \begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix}. \tag{3.39}$$

Since

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

then

$$[\mathbf{I}, \boldsymbol{\beta}]\mathbb{E}_{\mathbb{P}}[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}] \begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix} = \mathbf{0}. \tag{3.40}$$

Relation (3.39) can be alternatively rewritten using identity (3.40) in a slightly more sophisticated way

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = -\boldsymbol{\Delta}_n^{-1}\left([\mathbf{I},\boldsymbol{\beta}]\left[\mathbf{I},\widehat{\boldsymbol{\beta}}\right]^\top\right)^{-1}[\mathbf{I},\boldsymbol{\beta}] \left(n^{-1/2}\left\{[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]-\mathbb{E}_{\mathbb{P}}[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]\right\}\right)\begin{bmatrix}\boldsymbol{\beta}\\ -1\end{bmatrix}, \quad (3.41)$$

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

The inverse of $[\mathbf{I},\boldsymbol{\beta}]\left[\mathbf{I},\widehat{\boldsymbol{\beta}}\right]^\top$ from equation (3.41) exists with probability tending to one as n increases, because the probability that matrix $[\mathbf{I},\boldsymbol{\beta}]\left[\mathbf{I},\widehat{\boldsymbol{\beta}}\right]^\top = \mathbf{I}+\boldsymbol{\beta}\widehat{\boldsymbol{\beta}}^\top$ is singular tends to zero as n tends to infinity due to the strong consistency of $\widehat{\boldsymbol{\beta}}$, i.e.,

$$\mathbf{I}+\boldsymbol{\beta}\widehat{\boldsymbol{\beta}}^\top \xrightarrow{[\mathbb{P}]\text{-}a.s.} \mathbf{I}+\boldsymbol{\beta}\boldsymbol{\beta}^\top > \mathbf{0}, \quad n\to\infty. \quad (3.42)$$

Similarly, the inverse of $\boldsymbol{\Delta}_n$ exists with probability tending to one due to assumption (D) and the strong consistency result (1.27).

It is desirable to asymptotically compare the distribution of $\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})$ with the empirical (bootstrap) distribution of $\sqrt{n}(\widetilde{\boldsymbol{\beta}}^*-\widehat{\boldsymbol{\beta}})$ conditional on the original data $[\mathbf{X},\mathbf{Y}]$. With respect to (3.1), we get

$$\sqrt{n}(\widetilde{\boldsymbol{\beta}}^*-\widehat{\boldsymbol{\beta}}) = -(\boldsymbol{\Delta}_n^*)^{-1}\left([\mathbf{I},\widehat{\boldsymbol{\beta}}]\left[\mathbf{I},\widehat{\boldsymbol{\beta}}^*\right]^\top\right)^{-1}[\mathbf{I},\widehat{\boldsymbol{\beta}}] \left(n^{-1/2}\left\{[\mathbf{X}^*,\mathbf{Y}^*]^\top[\mathbf{X}^*,\mathbf{Y}^*]-[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]\right\}\right)\begin{bmatrix}\widehat{\boldsymbol{\beta}}\\ -1\end{bmatrix}, \quad (3.43)$$

where

$$\boldsymbol{\Delta}_n^* := n^{-1}(\mathbf{X}^{*\top}\mathbf{X}-\lambda^*\mathbf{I}).$$

Gallo (1982b, Proof of Theorem 3.3) proved that

$$n^{-1/2}\left\{[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]-\mathbb{E}_{\mathbb{P}}[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]\right\}\begin{bmatrix}\boldsymbol{\beta}\\ -1\end{bmatrix}$$

has asymptotically multivariate normal distribution. Let us check the assumptions of Theorem 3.11 for a sequence of zero mean independent $(p+1)$-dimensional random vectors

$$\begin{aligned}\boldsymbol{\xi}_n &:= \left\{[\mathbf{X}_{n,\bullet}, Y_n]^\top[\mathbf{X}_{n,\bullet}, Y_n] - \mathbb{E}_{\mathbb{P}}[\mathbf{X}_{n,\bullet}, Y_n]^\top[\mathbf{X}_{n,\bullet}, Y_n]\right\}\begin{bmatrix}\boldsymbol{\beta}\\-1\end{bmatrix}\\ &= \left\{[\mathbf{X}_{n,\bullet}, Y_n]^\top[\mathbf{X}_{n,\bullet}, Y_n] - \left([\mathbf{I},\boldsymbol{\beta}]^\top\mathbf{Z}_{n,\bullet}^\top\mathbf{Z}_{n,\bullet}[\mathbf{I},\boldsymbol{\beta}] + \sigma^2\mathbf{I}\right)\right\}\begin{bmatrix}\boldsymbol{\beta}\\-1\end{bmatrix}. \quad (3.44)\end{aligned}$$

Since

$$[\mathbf{X}_{n,\bullet}, Y_n]^\top[\mathbf{X}_{n,\bullet}, Y_n] = \begin{bmatrix}\boldsymbol{\Xi}_n & \boldsymbol{\zeta}_n^\top\\ \boldsymbol{\zeta}_n & \eta_n\end{bmatrix}$$

with

$$\begin{aligned}\boldsymbol{\Xi}_n &:= \mathbf{Z}_{n,\bullet}^\top\mathbf{Z}_{n,\bullet} + \mathbf{Z}_{n,\bullet}^\top\boldsymbol{\Theta}_{n,\bullet} + \boldsymbol{\Theta}_{n,\bullet}^\top\mathbf{Z}_{n,\bullet} + \boldsymbol{\Theta}_{n,\bullet}^\top\boldsymbol{\Theta}_{n,\bullet},\\ \boldsymbol{\zeta}_n &:= \mathbf{Z}_{n,\bullet}^\top\mathbf{Z}_{n,\bullet}\boldsymbol{\beta} + \mathbf{Z}_{n,\bullet}^\top\varepsilon_n + \boldsymbol{\Theta}_{n,\bullet}^\top\mathbf{Z}_{n,\bullet}\boldsymbol{\beta} + \boldsymbol{\Theta}_{n,\bullet}^\top\varepsilon_n,\\ \eta_n &:= (\mathbf{Z}_{n,\bullet}\boldsymbol{\beta})^2 + 2\varepsilon_n\mathbf{Z}_{n,\bullet}\boldsymbol{\beta} + \varepsilon_n^2;\end{aligned}$$

and assumptions (3.34)–(3.36) hold, then each row of $\boldsymbol{\xi}_n$ has uniformly bounded the fourth moments. Indeed by Lemma 3.5,

$$\begin{aligned}\sup_{n\in\mathbb{N}}\mathbb{E}_{\mathbb{P}}|X_{n,j}|^8 &= \sup_{n\in\mathbb{N}}\mathbb{E}_{\mathbb{P}}|Z_{n,j}+\Theta_{n,j}|^8 \leq 2^7\sup_{n\in\mathbb{N}}\mathbb{E}_{\mathbb{P}}\left(|Z_{n,j}|^8 + |\Theta_{n,j}|^8\right)\\ &\leq 2^7\left(\sup_{n\in\mathbb{N}}|Z_{n,j}|^2\right)^4 + 2^7\sup_{n\in\mathbb{N}}\mathbb{E}_{\mathbb{P}}|\Theta_{n,j}|^8 < \infty, \quad j\in\{1,\ldots,p\} \quad (3.45)\end{aligned}$$

and

$$\begin{aligned}\sup_{n\in\mathbb{N}}\mathbb{E}_{\mathbb{P}}|Y_n|^8 &= \sup_{n\in\mathbb{N}}\mathbb{E}_{\mathbb{P}}|Z_{n,\bullet}\boldsymbol{\beta}+\varepsilon_n|^8 \leq 2^7\sup_{n\in\mathbb{N}}\mathbb{E}_{\mathbb{P}}\left(|Z_{n,\bullet}\boldsymbol{\beta}|^8 + |\varepsilon_n|^8\right)\\ &\leq 2^7p^8\max_{j\in\{1,\ldots,p\}}\left\{\beta_j^8\left(\sup_{n\in\mathbb{N}}|Z_{n,j}|^2\right)^4\right\} + 2^7\sup_{n\in\mathbb{N}}\mathbb{E}_{\mathbb{P}}|\varepsilon_n|^8 < \infty. \quad (3.46)\end{aligned}$$

Hence, assumption (3.30) is satisfied for $\boldsymbol{\xi}_n$. Moreover, assumption (3.37) implies assumption (3.31), because

$$\begin{aligned}\mathbf{0} &< \lim_{n\to\infty}\frac{1}{n}\mathbb{V}\mathrm{ar}_{\mathbb{P}}\left\{[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]\begin{bmatrix}\boldsymbol{\beta}\\-1\end{bmatrix}\right\}\\ &= \lim_{n\to\infty}\frac{1}{n}\mathbb{V}\mathrm{ar}_{\mathbb{P}}\sum_{i=1}^{n}\left\{[\mathbf{X}_{i,\bullet}, Y_i]^\top[\mathbf{X}_{i,\bullet}, Y_i]\begin{bmatrix}\boldsymbol{\beta}\\-1\end{bmatrix}\right\}\end{aligned}$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbb{V}\text{ar}_{\mathbb{P}} \left\{ [\mathbf{X}_{i,\bullet}, Y_i]^\top [\mathbf{X}_{i,\bullet}, Y_i] \begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix} \right\}.$$

After a calculation of the conditional expectation

$$\mathbb{E}_{\mathbb{P}^*}[\mathbf{X}^*, \mathbf{Y}^*]^\top [\mathbf{X}^*, \mathbf{Y}^*] = \mathbb{E}_{\mathbb{P}^*} \sum_{i=1}^{n} [\mathbf{X}^*_{i,\bullet}, Y^*_i]^\top [\mathbf{X}^*_{i,\bullet}, Y^*_i]$$
$$= \sum_{i=1}^{n} \sum_{k=1}^{n} \frac{1}{n} [\mathbf{X}_{k,\bullet}, Y_k]^\top [\mathbf{X}_{k,\bullet}, Y_k] = [\mathbf{X}, \mathbf{Y}]^\top [\mathbf{X}, \mathbf{Y}],$$

we need to realize with respect to (3.44) that

$$\begin{aligned}
&\bar{\boldsymbol{\xi}}^*_n - \mathbb{E}_{\mathbb{P}^*} \boldsymbol{\xi}^*_1 \\
&= n^{-1} \sum_{i=1}^{n} [\mathbf{X}^*_{i,\bullet}, Y^*_i]^\top [\mathbf{X}^*_{i,\bullet}, Y^*_i] \begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix} \\
&\quad - n^{-1} \sum_{i=1}^{n} \left\{ [\mathbf{I}, \boldsymbol{\beta}]^\top \mathbf{Z}^{*\top}_{i,\bullet} \mathbf{Z}^*_{i,\bullet} [\mathbf{I}, \boldsymbol{\beta}] + \sigma^2 \mathbf{I} \right\} \begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix} \\
&\quad - \mathbb{E}_{\mathbb{P}^*} [\mathbf{X}^*_{1,\bullet}, Y^*_1]^\top [\mathbf{X}^*_{1,\bullet}, Y^*_1] \begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix} \\
&\quad + \mathbb{E}_{\mathbb{P}^*} \left\{ [\mathbf{I}, \boldsymbol{\beta}]^\top \mathbf{Z}^{*\top}_{1,\bullet} \mathbf{Z}^*_{1,\bullet} [\mathbf{I}, \boldsymbol{\beta}] + \sigma^2 \mathbf{I} \right\} \begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix} \\
&= n^{-1} [\mathbf{X}^*, \mathbf{Y}^*]^\top [\mathbf{X}^*, \mathbf{Y}^*] \begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix} - n^{-1} [\mathbf{X}, \mathbf{Y}]^\top [\mathbf{X}, \mathbf{Y}] \begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix}.
\end{aligned}$$

By bootstrap multivariate CLT 3.11, we get

$$n^{-1/2} \left\{ [\mathbf{X}^*, \mathbf{Y}^*]^\top [\mathbf{X}^*, \mathbf{Y}^*] - \mathbb{E}_{\mathbb{P}^*} [\mathbf{X}^*, \mathbf{Y}^*]^\top [\mathbf{X}^*, \mathbf{Y}^*] \right\} \begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix} \Bigg| [\mathbf{X}, \mathbf{Y}]$$
$$\xleftrightarrow[n\to\infty]{\mathscr{D}(\mathbb{P})} n^{-1/2} \left\{ [\mathbf{X}, \mathbf{Y}]^\top [\mathbf{X}, \mathbf{Y}] - \mathbb{E}_{\mathbb{P}} [\mathbf{X}, \mathbf{Y}]^\top [\mathbf{X}, \mathbf{Y}] \right\} \begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix}. \quad (3.47)$$

We would like to prove that

$$n^{-1/2} \left\{ [\mathbf{X}^*, \mathbf{Y}^*]^\top [\mathbf{X}^*, \mathbf{Y}^*] - [\mathbf{X}, \mathbf{Y}]^\top [\mathbf{X}, \mathbf{Y}] \right\} \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ -1 \end{bmatrix} \Bigg| [\mathbf{X}, \mathbf{Y}]$$
$$\xleftrightarrow[n\to\infty]{\mathscr{D}(\mathbb{P})} n^{-1/2} \left\{ [\mathbf{X}, \mathbf{Y}]^\top [\mathbf{X}, \mathbf{Y}] - \mathbb{E}_{\mathbb{P}} [\mathbf{X}, \mathbf{Y}]^\top [\mathbf{X}, \mathbf{Y}] \right\} \begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix}. \quad (3.48)$$

Equiboundedness (3.45) and (3.46) provide stronger results than the assumptions of Theorem 3.3 for $\{X_{n,j}X_{n,k}\}_{n=1}^{\infty}$ and $\{X_{n,j}Y_n\}_{n=1}^{\infty}$ are, where $j, k \in \{1, \ldots, p\}$. Nevertheless, the bootstrap WLLN for independent variables implies

$$n^{-1}\sum_{i=1}^{n}\mathbf{X}_{i,\bullet}^{*\top}\mathbf{X}_{i,\bullet}^{*} - n^{-1}\sum_{i=1}^{n}\mathbf{X}_{i,\bullet}^{\top}\mathbf{X}_{i,\bullet} \xrightarrow[n\to\infty]{\mathbb{P}^*(\mathbb{P})} \mathbf{0} \tag{3.49}$$

and

$$n^{-1}\sum_{i=1}^{n}\mathbf{X}_{i,\bullet}^{*\top}Y_i^{*} - n^{-1}\sum_{i=1}^{n}\mathbf{X}_{i,\bullet}^{\top}Y_i \xrightarrow[n\to\infty]{\mathbb{P}^*(\mathbb{P})} \mathbf{0}. \tag{3.50}$$

The asymptotic normality of $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ provides

$$\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \mathcal{O}_{\mathbb{P}}(n^{-1/2}), \quad n \to \infty.$$

Due to Definition 3.3, (3.49), and (3.50), we get

$$n^{-1/2}\left\{[\mathbf{X}^*, \mathbf{Y}^*]^{\top}[\mathbf{X}^*, \mathbf{Y}^*] - [\mathbf{X}, \mathbf{Y}]^{\top}[\mathbf{X}, \mathbf{Y}]\right\}\begin{bmatrix}\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\\ 0\end{bmatrix} \xrightarrow[n\to\infty]{\mathbb{P}^*(\mathbb{P})} \mathbf{0}.$$

Afterwards, (3.48) is shown.

According to Definition 3.3, we simply obtain that $\widehat{\boldsymbol{\beta}} \xrightarrow[n\to\infty]{[\mathbb{P}]\text{-}a.s.} \boldsymbol{\beta}$ implies

$$\widehat{\boldsymbol{\beta}} \xrightarrow[n\to\infty]{\mathbb{P}^*(\mathbb{P})} \boldsymbol{\beta}. \tag{3.51}$$

Theorem 3.6 can be rewritten in this manner:

$$\frac{\lambda^*}{n} - \frac{\lambda}{n} \xrightarrow[n\to\infty]{\mathbb{P}^*(\mathbb{P})} 0.$$

Moreover by (1.24), we have

$$\frac{\lambda^*}{n} \xrightarrow[n\to\infty]{\mathbb{P}^*(\mathbb{P})} \sigma^2.$$

With respect to (3.49), it also holds that

$$\boldsymbol{\Delta}_n^* - \boldsymbol{\Delta}_n \xrightarrow[n\to\infty]{\mathbb{P}^*(\mathbb{P})} \mathbf{0}$$

and, consequently by (1.27),

$$\boldsymbol{\Delta}_n^* \xrightarrow[n\to\infty]{\mathbb{P}^*(\mathbb{P})} \boldsymbol{\Delta}. \tag{3.52}$$

In other words, the conditional law of $n^{-1}(\mathbf{X}^{*\top}\mathbf{X}^{*} - \lambda^{*}\mathbf{I})$ is close to the unconditional law of $\boldsymbol{\Delta}_n$. The latter inverted concentrates near non-random element $\boldsymbol{\Delta}^{-1}$ as mentioned in (1.27). Therefore, the conditional law of $n(\mathbf{X}^{*\top}\mathbf{X}^{*} - \lambda^{*}\mathbf{I})^{-1}$ is degenerate and also concentrates near $\boldsymbol{\Delta}^{-1}$.

Since

$$\begin{aligned}\widehat{\boldsymbol{\beta}}^{*} - \widehat{\boldsymbol{\beta}} &= (\boldsymbol{\Delta}_n^{*})^{-1} n^{-1}\mathbf{X}^{*\top}\mathbf{Y}^{*} - \boldsymbol{\Delta}_n^{-1} n^{-1}\mathbf{X}^{\top}\mathbf{Y} \\ &= \left\{(\boldsymbol{\Delta}_n^{*})^{-1} - \boldsymbol{\Delta}_n^{-1}\right\} n^{-1}\mathbf{X}^{\top}\mathbf{Y} + (\boldsymbol{\Delta}_n^{*})^{-1}\left\{n^{-1}\mathbf{X}^{*\top}\mathbf{Y}^{*} - n^{-1}\mathbf{X}^{\top}\mathbf{Y}\right\}\end{aligned}$$

and (3.50) holds, then

$$\widehat{\boldsymbol{\beta}}^{*} - \widehat{\boldsymbol{\beta}} \xrightarrow[n\to\infty]{\mathbb{P}^{*}(\mathbb{P})} \mathbf{0},$$

which implies

$$\widehat{\boldsymbol{\beta}}^{*} \xrightarrow[n\to\infty]{\mathbb{P}^{*}(\mathbb{P})} \boldsymbol{\beta}. \tag{3.53}$$

Finally, let us denote

$$\begin{aligned}\boldsymbol{\Upsilon}_n &:= -\boldsymbol{\Delta}_n^{-1}\left([\mathbf{I},\boldsymbol{\beta}]\left[\mathbf{I},\widehat{\boldsymbol{\beta}}\right]^{\top}\right)^{-1}[\mathbf{I},\boldsymbol{\beta}],\\ \boldsymbol{\Upsilon}_n^{*} &:= -(\boldsymbol{\Delta}_n^{*})^{-1}\left([\mathbf{I},\widehat{\boldsymbol{\beta}}]\left[\mathbf{I},\widehat{\boldsymbol{\beta}}^{*}\right]^{\top}\right)^{-1}[\mathbf{I},\widehat{\boldsymbol{\beta}}],\\ \boldsymbol{\upsilon}_n &:= n^{-1/2}\left\{[\mathbf{X},\mathbf{Y}]^{\top}[\mathbf{X},\mathbf{Y}] - \mathbb{E}_{\mathbb{P}}[\mathbf{X},\mathbf{Y}]^{\top}[\mathbf{X},\mathbf{Y}]\right\}\begin{bmatrix}\boldsymbol{\beta}\\ -1\end{bmatrix},\\ \boldsymbol{\upsilon}_n^{*} &:= n^{-1/2}\left\{[\mathbf{X}^{*},\mathbf{Y}^{*}]^{\top}[\mathbf{X}^{*},\mathbf{Y}^{*}] - [\mathbf{X},\mathbf{Y}]^{\top}[\mathbf{X},\mathbf{Y}]\right\}\begin{bmatrix}\widehat{\boldsymbol{\beta}}\\ -1\end{bmatrix}.\end{aligned}$$

In order to finish this proof, it is sufficient to show that

$$\boldsymbol{\Upsilon}_n^{*}\boldsymbol{\upsilon}_n^{*}\big|[\mathbf{X},\mathbf{Y}] \xleftrightarrow[n\to\infty]{\mathscr{D}(\mathbb{P})} \boldsymbol{\Upsilon}_n\boldsymbol{\upsilon}_n. \tag{3.54}$$

We use the Prokhorov's extended characterization of the convergence in distribution stated in Lemma 3.1. Relation (3.54) is valid if and only if for each subsequence $\{n_i\}_{i=1}^{\infty}$ there is a subsequence $\{n_{i_k}\}_{k=1}^{\infty}$ and a random variable $\boldsymbol{\zeta}_0$ such that

$$\boldsymbol{\Upsilon}_{n_{i_k}}^{*}\boldsymbol{\upsilon}_{n_{i_k}}^{*}\big|[\mathbf{X},\mathbf{Y}] \xrightarrow[k\to\infty]{\mathscr{D}(\mathbb{P})} \boldsymbol{\zeta}_0 \quad \text{and} \quad \boldsymbol{\Upsilon}_{n_{i_k}}\boldsymbol{\upsilon}_{n_{i_k}} \xrightarrow[k\to\infty]{\mathscr{D}} \boldsymbol{\zeta}_0. \tag{3.55}$$

Gallo (1982b, Theorem 3.3) proved, that $\boldsymbol{\Upsilon}_n\boldsymbol{\upsilon}_n \xrightarrow[n\to\infty]{\mathscr{D}} \boldsymbol{\zeta}_0$, where $\boldsymbol{\zeta}_0$ has a multivariate zero mean normal distribution. Therefore, sequence $\{\boldsymbol{\Upsilon}_n\boldsymbol{\upsilon}_n\}_{n=1}^{\infty}$ is tight. Let $\{n_i\}_{i=1}^{\infty}$ be an arbitrary subsequence. According to (3.48) and Lemma 3.1, there exist a subsequence $\{n_{i_k}\}_{k=1}^{\infty}$ of $\{n_i\}_{i=1}^{\infty}$ and random vector $\boldsymbol{\upsilon}_0$ for which

$$\boldsymbol{\upsilon}_{n_{i_k}}^*\big|[\mathbf{X},\mathbf{Y}] \xrightarrow[k\to\infty]{\mathscr{D}(\mathbb{P})} \boldsymbol{\upsilon}_0 \quad \text{and} \quad \boldsymbol{\upsilon}_{n_{i_k}} \xrightarrow[k\to\infty]{\mathscr{D}} \boldsymbol{\upsilon}_0.$$

Taking (3.51), (3.52), and (3.53) into account, we get

$$\boldsymbol{\Upsilon}_{n_{i_k}}^* \xrightarrow[k\to\infty]{\mathbb{P}^*(\mathbb{P})} \boldsymbol{\Upsilon}_0 \quad \text{and} \quad \boldsymbol{\Upsilon}_{n_{i_k}} \xrightarrow[k\to\infty]{\mathbb{P}} \boldsymbol{\Upsilon}_0,$$

for some $\boldsymbol{\Upsilon}_0$.

Further, we conclude from the Slutsky's theorem (see Appendix A.2, Theorem A.2) and its modification (Theorem 3.2) that (3.55) holds, which by virtue of Lemmas 3.1 yields (3.54).

By recalling (3.39), (3.40), (3.41), and (3.43), it is concluded that

$$\sqrt{n}(\widetilde{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})\big|[\mathbf{X},\mathbf{Y}] \xleftrightarrow[n\to\infty]{\mathscr{D}(\mathbb{P})} \sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}).$$

□

The main reason for the occurrence of a nondegenerative distribution between the conditional distribution of $\sqrt{n}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})$ and $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is the fact that squared errors present in the expression for $\widehat{\boldsymbol{\beta}}$ do not converge in probability $\mathbb{P}$ to zero as the sample size increases.

Remark 3.3. Partial "consistency" result (3.53) can be wrongly considered as a satisfaction for the validity of the nonparametric bootstrap procedure. It has to be realized that consistency does not immediately imply $\sqrt{n}$-asymptotic normality. Hence, (3.53) cannot assure about $\sqrt{n}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})\big|[\mathbf{X},\mathbf{Y}] \xleftrightarrow[n\to\infty]{\mathscr{D}(\mathbb{P})} \sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. Moreover, the above proof shows us a nice contradiction to this hypothesis.

Analogously to (3.39), it can be derived that

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) = -n\left(\mathbf{X}^{*\top}\mathbf{X}^* - \lambda^*\mathbf{I}\right)^{-1}\left([\mathbf{I},\widehat{\boldsymbol{\beta}}]\left[\mathbf{I},\widehat{\boldsymbol{\beta}}^*\right]^\top\right)^{-1}$$
$$[\mathbf{I},\widehat{\boldsymbol{\beta}}]\left(n^{-1/2}[\mathbf{X}^*,\mathbf{Y}^*]^\top[\mathbf{X}^*,\mathbf{Y}^*]\right)\begin{bmatrix}\widehat{\boldsymbol{\beta}}\\ -1\end{bmatrix}. \quad (3.56)$$

Remark 3.4. The difference between the "incorrect" bootstrap estimate $\widehat{\boldsymbol{\beta}}^*$ and the "proper" one $\widetilde{\boldsymbol{\beta}}^*$ can be derived from (3.1) and (3.56):

$$\widetilde{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}^* = \left(\mathbf{X}^{*\top}\mathbf{X}^* - \lambda^*\mathbf{I}\right)^{-1}\left([\mathbf{I},\widehat{\boldsymbol{\beta}}]\left[\mathbf{I},\widehat{\boldsymbol{\beta}}^*\right]^\top\right)^{-1}[\mathbf{I},\widehat{\boldsymbol{\beta}}][\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]\begin{bmatrix}\widehat{\boldsymbol{\beta}}\\ -1\end{bmatrix}. \quad (3.57)$$

Applying the results and the algebraic machinery from the proof of Theorem 3.12, it can be shown that $\widetilde{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}^* \xrightarrow[n\to\infty]{\mathbb{P}^*(\mathbb{P})} \mathbf{0}$ and, moreover, $\widetilde{\boldsymbol{\beta}}^* \xrightarrow[n\to\infty]{\mathbb{P}^*(\mathbb{P})} \boldsymbol{\beta}$, which demonstrates another (weaker) characterization of the nonparametric bootstrap's consistency. Although, one cannot generally prove that the limiting conditional distribution of $\sqrt{n}(\widetilde{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}^*)$ is just a Dirac distribution. In spite of that, the asymptotic distribution of $\sqrt{n}(\widetilde{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}^*)$ could possibly converge to zero (i.e., vanish) in probability $\mathbb{P}^*$ in probability $\mathbb{P}$ as the sample size increases under some additional assumptions, which might improve the speed of convergences of $\widehat{\boldsymbol{\beta}}$ or $\boldsymbol{\Delta}_n$ as demonstrated by Kukush et al. (2005, Theorem 3).

3.3 Residual Bootstrap

Residuals for the EIV model are defined in this section. A *residual bootstrap* procedure can be proposed, where the residuals are resampled with replacement in order to obtain the empirical (bootstrap) distribution of the original parameters' estimates. The residual bootstrap procedure is more sophisticated than the nonparametric bootstrap one, because a regression model is assumed and, consequently, specific residuals obtained from this model are resampled.

3.3.1 Residuals and Their Disadvantageous Properties

In EIV setup, a natural way how to define a reasonable type of residuals is

$$[\widehat{\boldsymbol{\Theta}}, \widehat{\boldsymbol{\varepsilon}}] := \arg\min \|[\boldsymbol{\Theta}, \boldsymbol{\varepsilon}]\|$$

such that

$$\mathbf{Y} = (\mathbf{X} - \widehat{\boldsymbol{\Theta}})\widehat{\boldsymbol{\beta}} + \widehat{\boldsymbol{\varepsilon}}$$

They are estimates of the disturbances in the EIV model in a particular sense. Finding the smallest errors $[\boldsymbol{\Theta}, \boldsymbol{\varepsilon}]$ in the EIV model is tantamount to looking for the "closest" subspace to data, i.e., to penalize for the errors in the orthogonal direction to the fitted hyperplane.

Recall that Golub and Van Loan (1980, Section 2) proved an interesting property of such residuals: $\|[\widehat{\boldsymbol{\Theta}}, \widehat{\boldsymbol{\varepsilon}}]\|^2 = \lambda$.

The residual bootstrap procedure generally relies on obtaining residuals from the regression model, which are consequently recentered by their average and, then, resampled with replacement. In our case, one should bootstrap rows of centered residuals $\{[\widetilde{\boldsymbol{\Theta}}_{i,\bullet}, \widetilde{\varepsilon}_i]\}_{i=1}^n$, where

$$[\widetilde{\boldsymbol{\Theta}}_{i,\bullet}, \widetilde{\varepsilon}_i] := [\widehat{\boldsymbol{\Theta}}_{i,\bullet}, \widehat{\varepsilon}_i] - n^{-1}\sum_{j=1}^{n}[\widehat{\boldsymbol{\Theta}}_{j,\bullet}, \widehat{\varepsilon}_j], \quad i = 1, \ldots, n.$$

A crucial step for proving the correctness of the bootstrap is a convergence of the distance between estimated residuals and residual errors divided by number of observations to zero.

Detailed discussion can be found in Bickel and Freedman (1981, Section 2). We are going to show that the opposite is true.

Theorem 3.13. *Let the EIV model hold and assumption* (D) *be satisfied. If* $\boldsymbol{\beta} \neq \mathbf{0}$*, then*

$$n^{-1}\|[\widehat{\boldsymbol{\Theta}}, \widehat{\boldsymbol{\varepsilon}}] - [\boldsymbol{\Theta}, \boldsymbol{\varepsilon}]\|^2 \nrightarrow 0 \quad [\mathbb{P}]\text{-}a.s., \quad n \to \infty.$$

Proof. Let us calculate

$$n^{-1}\|[\widehat{\boldsymbol{\Theta}}, \widehat{\boldsymbol{\varepsilon}}] - [\boldsymbol{\Theta}, \boldsymbol{\varepsilon}]\|^2 \geq n^{-1}\|\widehat{\boldsymbol{\varepsilon}} - \boldsymbol{\varepsilon}\|_2^2 = n^{-1}\widehat{\boldsymbol{\varepsilon}}^\top\widehat{\boldsymbol{\varepsilon}} + n^{-1}\boldsymbol{\varepsilon}^\top\boldsymbol{\varepsilon} - 2n^{-1}\boldsymbol{\varepsilon}^\top\widehat{\boldsymbol{\varepsilon}}. \tag{3.58}$$

Denote the SVD of data $[\mathbf{X}, \mathbf{Y}] = \mathbf{U} diag\{\boldsymbol{\rho}\}\mathbf{V}^\top$. Then Theorem 2.6 by Van Huffel and Vandewalle (1991) ensures $\widehat{\boldsymbol{\varepsilon}} = \rho_{p+1}\mathbf{u}_{p+1}v_{p+1,p+1}$, where $\mathbf{u}_{p+1}$ is the last column of $\mathbf{U}$, $v_{p+1,p+1}$ is the $(p+1)$-st element of the last column of $\mathbf{V}$, and ρ_{p+1} is the smallest singular value of $[\mathbf{X}, \mathbf{Y}]$.

Keeping in mind that Gleser (1981, Section 3) proved

$$v_{p+1,p+1}^2 \xrightarrow[n\to\infty]{[\mathbb{P}]\text{-}a.s.} \left(1 + \boldsymbol{\beta}^\top\boldsymbol{\beta}\right)^{-1}, \tag{3.59}$$

realizing that $\rho_{p+1} = \sqrt{\lambda}$ and the columns of the matrix $\mathbf{U}$ are orthonormal, and using (1.26), for the first term in (3.58) we have

$$n^{-1}\widehat{\boldsymbol{\varepsilon}}^\top\widehat{\boldsymbol{\varepsilon}} = n^{-1}v_{p+1,p+1}\mathbf{u}_{p+1}^\top\sqrt{\lambda}\sqrt{\lambda}\mathbf{u}_{p+1}v_{p+1,p+1} \xrightarrow[n\to\infty]{[\mathbb{P}]\text{-}a.s.} \sigma^2\left(1 + \boldsymbol{\beta}^\top\boldsymbol{\beta}\right)^{-1}. \tag{3.60}$$

If the SLLN is applied for the second term in (3.58) (see Lemma 3.8), then

$$n^{-1}\boldsymbol{\varepsilon}^\top\boldsymbol{\varepsilon} \xrightarrow[n\to\infty]{[\mathbb{P}]\text{-}a.s.} \sigma^2. \tag{3.61}$$

Let us concentrate on the third term in (3.58). If the almost sure limit of $n^{-1}\boldsymbol{\varepsilon}^\top\widehat{\boldsymbol{\varepsilon}}$ does not exist, the proof is finished, because then it is impossible for $n^{-1}\|[\widehat{\boldsymbol{\Theta}}, \widehat{\boldsymbol{\varepsilon}}] - [\boldsymbol{\Theta}, \boldsymbol{\varepsilon}]\|^2$ to converge almost surely to zero. On the other hand, let us suppose the existence of $\lim_{n\to\infty} n^{-1}\boldsymbol{\varepsilon}^\top\widehat{\boldsymbol{\varepsilon}} := \varphi \quad [\mathbb{P}]$-a.s. Using the Cauchy-Schwarz inequality and orthogonality of the matrix $\mathbf{U}$, we get

$$\left|n^{-1}\boldsymbol{\varepsilon}^\top\widehat{\boldsymbol{\varepsilon}}\right| = \left|n^{-1}\boldsymbol{\varepsilon}^\top\sqrt{\lambda}\mathbf{u}_{p+1}v_{p+1,p+1}\right| \leq \sqrt{\frac{\lambda}{n}}\left\|n^{-1/2}\boldsymbol{\varepsilon}^\top\right\|_2 |v_{p+1,p+1}|. \tag{3.62}$$

Passing to limit in (3.62) and taking into account (1.26), (3.59), and (3.61), we obtain

$$|\varphi| \leq \sigma^2\left(1 + \boldsymbol{\beta}^\top\boldsymbol{\beta}\right)^{-1/2}. \tag{3.63}$$

Finally, if $\lim_{n\to\infty} n^{-1}\|\widehat{\boldsymbol{\varepsilon}} - \boldsymbol{\varepsilon}\|_2^2$ exists, then

$$\lim_{n\to\infty} n^{-1}\|\widehat{\boldsymbol{\varepsilon}} - \boldsymbol{\varepsilon}\|_2^2 \geq \sigma^2 \left\{ \left(1 + \boldsymbol{\beta}^\top \boldsymbol{\beta}\right)^{-1} + 1 - 2\left(1 + \boldsymbol{\beta}^\top \boldsymbol{\beta}\right)^{-1/2} \right\} > 0,$$

unless $\boldsymbol{\beta} = \mathbf{0}$. Hence, $n^{-1}\|[\widehat{\boldsymbol{\Theta}}, \widehat{\boldsymbol{\varepsilon}}] - [\boldsymbol{\Theta}, \boldsymbol{\varepsilon}]\|^2$ cannot converge [$\mathbb{P}$]-almost surely to zero. □

The EIV residuals' *inconsistency* in the way introduced above does not allow us to consider them distributionally close enough to the original errors. Although, it is not said that a suitably modified algorithm for the residual bootstrap cannot be constructed and its correctness proved.

3.4 Simulation Study

A simulation experiment was performed to study the *finite sample* properties of the bootstrap procedures for EIV. In particular, the interest lies in the *coverage level of confidence intervals* (CIs) based on the EIV nonparametric bootstrap for finite samples. Random samples (5000 each time) were generated from a one-dimensional EIV model with the design points $\left\{\sqrt{1-1/i}\right\}_{i=1}^{n}$ and, hence, $\boldsymbol{\Delta} = \lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}(1-1/i) = 1$. A design with equidistant points cannot be used, because it does not satisfy assumption (D). The previous design can be seen as a compromise between a "logarithmic scale" design and the equidistant one. Therefore, it may be quite useful in engineering science.

First of all, normally distributed errors are taken into account. Bootstrap 95%-CIs based upon *percentile method* (see, e.g., Efron (1982), Hall (1992), or (Davidson and Hinkley, 1997)) were considered for the incorrect nonparametric bootstrap procedure and the proper one as well. $B = 5000$ bootstrap replications were conducted. Percentile confidence intervals are of the form $(\widehat{v}_{\alpha/2}; \widehat{v}_{1-\alpha/2})$ where $\widehat{v}_\gamma$ is the $\gamma 100\%$ quantile of the empirical bootstrap distribution of $\widehat{\beta}$ (scalar). On the other hand, the confidence intervals based upon *standard deviation method* (approximate normal asymptotics) of the form $\widehat{\beta} \pm u_{\alpha/2}\widehat{s}$ were produced as well, where u_γ is the $\gamma 100\%$ quantile of a standard normal distribution and $\widehat{s}$ is the standard deviation of the re-calculated estimate $\widehat{\beta}$.

Finally, the average lengths of CIs were computed and compared with the theoretical ones based on asymptotic normality with known parameters σ^2 and $\boldsymbol{\Delta}$ (without plugging their estimates). Moreover, the *empirical coverages* of CIs were calculated and compared with the theoretical exact value of 95%. The results for percentile-based (bootstrap) and variance-based (approximate asymptotic normality) CIs from our simulation study with normally distributed errors for various "unknown" β are shown in Table 3.1, Table 3.2, and Table 3.3.

From these results it follows that the average length of bootstrap CIs reaches the theoretical value for $n = 20$ satisfactorily and almost coincides with it for $n = 50$. A similar situation occurs, when the approximate asymptotics is compared to the theoretical lengths.

Size	Standard deviation	Average CI length Approximate asymptotic normality Nonparametric bootstrap (proper) Theoretical asymptotic normality	Empirical CI coverage
$n = 20$	$\sigma = 10^{-2}$	1.317×10^{-2}	93.08%
		1.316×10^{-2}	92.78%
		1.240×10^{-2}	–
	$\sigma = 10^{-3}$	1324×10^{-3}	93.70%
		1.322×10^{-3}	93.64%
		1.240×10^{-3}	–
$n = 50$	$\sigma = 10^{-2}$	8.101×10^{-3}	94.86%
		8.104×10^{-3}	94.90%
		7.840×10^{-3}	–
	$\sigma = 10^{-3}$	8.104×10^{-4}	94.14%
		8.108×10^{-4}	94.02%
		7.840×10^{-4}	–

Table 3.1: Simulations of 95% confidence intervals for the nonparametric bootstrap when $\beta = 1$ and the errors are zero mean normally distributed.

Size	Standard deviation	Average CI length Approximate asymptotic normality Nonparametric bootstrap (proper) Theoretical asymptotic normality	Empirical CI coverage
$n = 20$	$\sigma = 10^{-2}$	9.360×10^{-2}	93.10%
		9.347×10^{-2}	92.76%
		8.809×10^{-2}	–
	$\sigma = 10^{-3}$	9.369×10^{-3}	93.08%
		9.355×10^{-3}	93.18%
		8.809×10^{-3}	–
$n = 50$	$\sigma = 10^{-2}$	5.740×10^{-2}	94.46%
		5.740×10^{-2}	94.46%
		5.571×10^{-2}	–
	$\sigma = 10^{-3}$	5.777×10^{-3}	94.88%
		5.778×10^{-3}	94.96%
		5.571×10^{-3}	–

Table 3.2: Simulations of 95% confidence intervals for the nonparametric bootstrap when $\beta = 10$ and the errors are zero mean normally distributed.

		Average CI length	Empirical CI coverage
		Approximate asymptotic normality Nonparametric bootstrap (proper) Theoretical asymptotic normality	
Size	Standard deviation		
$n = 20$	$\sigma = 10^{-2}$	9.348×10^{-3}	92.38%
		9.334×10^{-3}	92.18%
		8.809×10^{-3}	–
	$\sigma = 10^{-3}$	9.370×10^{-4}	92.96%
		9.360×10^{-4}	92.88%
		8.809×10^{-4}	–
$n = 50$	$\sigma = 10^{-2}$	5.758×10^{-3}	94.32%
		5.764×10^{-3}	94.22%
		5.572×10^{-3}	–
	$\sigma = 10^{-3}$	5.759×10^{-4}	93.86%
		5.762×10^{-4}	93.92%
		5.571×10^{-4}	–

Table 3.3: Simulations of 95% confidence intervals for the nonparametric bootstrap when $\beta = .1$ and the errors are zero mean normally distributed.

The estimated variances of the EIV estimate based on proper bootstrap and asymptotic normality are similar. Therefore, this simulation study ensures ourselves that the correct bootstrap method for construction of the CIs in the EIV model is *asymptotically equivalent* to the method based on approximate normal asymptotics.

Furthermore, the empirical coverage of the CIs based on approximate asymptotic normality is nearly 95%, which conforms to the fact that the CI should contain the unknown parameter with a prescribed probability. Similarly, the proper bootstrap CIs maintain the empirical coverage levels close to the nominal level of 95%. Note that in all the simulations, the empirical coverages are slightly below 95% and *become closer* when the *number of observations increases* or the *variability of errors decreases*.

There is *no striking difference* between the incorrect and proper nonparametric bootstrap procedures in the average length of the CIs neither in the empirical coverage. Both approaches seem to behave quite similarly. The lengths of the incorrect nonparametric bootstrap CIs and proper nonparametric bootstrap CIs have to be exactly the same, because the empirical conditional distributions are just shifted. I.e., for each simulated data set and a particular fixed resampling, $\widetilde{\boldsymbol{\beta}}^*$ (generally a vector) is only a shifted version of $\widehat{\boldsymbol{\beta}}^*$.

In all the simulations, the endpoints of confidence intervals from the incorrect nonparametric bootstrap and the proper one are not exactly the same and slightly differ (for a mathematical explanation see Remark 3.4). The difference varies from $\approx 10^{-3}$ to $\approx 10^{-8}$

when passing from one particular simulated data set to another one and evidently depends on the setup of the simulation. On the contrary, when calculating the empirical CI coverages over all the simulated data sets, the difference in the empirical coverage for the incorrect nonparametric bootstrap and the proper one vanishes and the unsuitability of the incorrect nonparametric bootstrap cannot be noticed through these simulations. In principle, $\sqrt{n}(\widetilde{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}^*)$ might asymptotically have a *zero mean nondegenerative distribution*, because the setup of simulations may satisfy some additional assumptions, which guarantee such asymptotic closeness.

Changing the value of the "unknown" (true) parameter β does not effect the conclusions already presented. Thus it may be concluded that the proper nonparametric bootstrap computationally works for any value of $\boldsymbol{\beta}$.

The results also indicate that the *bootstrap provides reasonable CIs* for the unknown parameter and *may be used instead of* the approximate asymptotic normality.

Moreover, if the *cross-moments of the errors* are not identical, up to and including moments of order four, to the corresponding moments of the multivariate normal distribution, the asymptotic normality is *computationally useless* in our situation (Gallo, 1982b, Subsection 3.4). Despite of this disadvantage, the proper nonparametric bootstrap approach provides a satisfiable answer as proved before. To lightly demonstrate that, let us think of two distributions for the errors: the *Student* t-distribution with 8.2 degrees of freedom (in order to be able to consider the existence of the eight moment), which has relatively heavy tails; and the centered *Gumbel* (extreme value) distribution, which is not symmetric (skewed) and have excess kurtosis. Essentially, the empirical CI coverage becomes closer to the theoretical value of 95% for the proper nonparametric bootstrap procedure when the sample size increases and the standard deviation decreases as it can be seen in Table 3.4 and Table 3.5. The CIs based upon the incorrect nonparametric bootstrap behave similarly as described in the case of normally distributed errors.

		Average CI length	Empirical CI coverage
Size	Standard deviation	Nonparametric bootstrap (proper)	
$n = 20$	$\sigma = 10^{-2}$	1.314×10^{-2}	93.08%
	$\sigma = 10^{-3}$	1.306×10^{-3}	92.42%
$n = 50$	$\sigma = 10^{-2}$	8.085×10^{-3}	93.86%
	$\sigma = 10^{-3}$	8.092×10^{-4}	94.26%

Table 3.4: Simulations of 95% confidence intervals for the nonparametric bootstrap when $\beta = 1$ and the errors have Student's t-distribution with 8.2 degrees of freedom multiplied by a constant such that its variance is equal to the squared standard deviation.

Finally, all the previous conclusions concerning percentile nonparametric bootstrap con-

		Average CI length	Empirical CI coverage
Size	Standard deviation	Nonparametric bootstrap (proper)	
$n = 20$	$\sigma = 10^{-2}$	1.313×10^{-2}	92.84%
	$\sigma = 10^{-3}$	1.312×10^{-3}	93.12%
$n = 50$	$\sigma = 10^{-2}$	8.104×10^{-3}	93.76%
	$\sigma = 10^{-3}$	8.047×10^{-4}	93.92%

Table 3.5: Simulations of 95% confidence intervals for the nonparametric bootstrap when $\beta = 1$ and the errors have shifted (centered) Gumbel distribution (with density before centering $\exp -\{\exp(-x-\gamma)\}$, $x \in \mathbb{R}$ and γ is Euler-Mascheroni constant (≈ 0.5772)) with zero mean multiplied by a constant such that its variance is equal to the squared standard deviation.

fidence intervals can be graphically illustrated in the following set of Figures 3.1–3.7, where each one of them consists of four subfigures corresponding to a different simulation setup: sample size $n = 20$ or $n = 50$ and standard deviation 10^{-2} or 10^{-3}, respectively.

The empirical distributions of estimate $\widehat{\beta}$ based upon "asymptotic normality" (AN) and proper nonparametric bootstrap (Boot) are compared in Figures 3.1–3.3 for various values of the "unknown" (true) parameter β, i.e., $\beta = 1$ in Figure 3.1, $\beta = 10$ in Figure 3.2, and $\beta = .1$ in Figure 3.3. The dark gray kernel density estimate (the left part of a subfigure) belongs to the estimated distribution of $\widehat{\beta}$ (which is approximately $\sqrt{n}$-normal) based upon estimates $\{\widehat{\beta}_s\}_{s=1}^S$, where $S = 5000$ is the number of simulations. The light one (the right part of a subfigure) stands for the proper nonparametric bootstrap version, i.e., the smoothed empirical distribution of means of the proper nonparametric bootstrap estimates

$$\left\{\frac{1}{B}\sum_{b=1}^{B}\widetilde{\beta}_{b,s}\right\}_{s=1}^{S},$$

where $\widetilde{\beta}_{b,s}$ is the b-th proper nonparametric bootstrap estimate of β from the s-th simulated data set.

Every subfigure compares not only the whole empirical distributions by their smoothed estimates, but also by the sample mean—dotted horizontal line—and by some specific sample characteristics—solid horizontal lines corresponding to the minimum, first quartile, median, third quartile, and the maximum from the bottom to the top. It can be seen that these sample quantities almost coincide. Hence, the proper nonparametric bootstrap seems to work at least as good as the appropriate asymptotic normality.

A similar situation arise, when comparing the estimated distributions of $\widehat{\beta}$ based upon the theoretical asymptotic normality, i.e., $\widehat{\beta}$ is approximately $\beta + n^{-1/2}\mathcal{N}(0,\cdot)$ (the limit distribution of $\sqrt{n}(\widehat{\beta}-\beta)$ is known, because the errors are generated from a normal distribution);

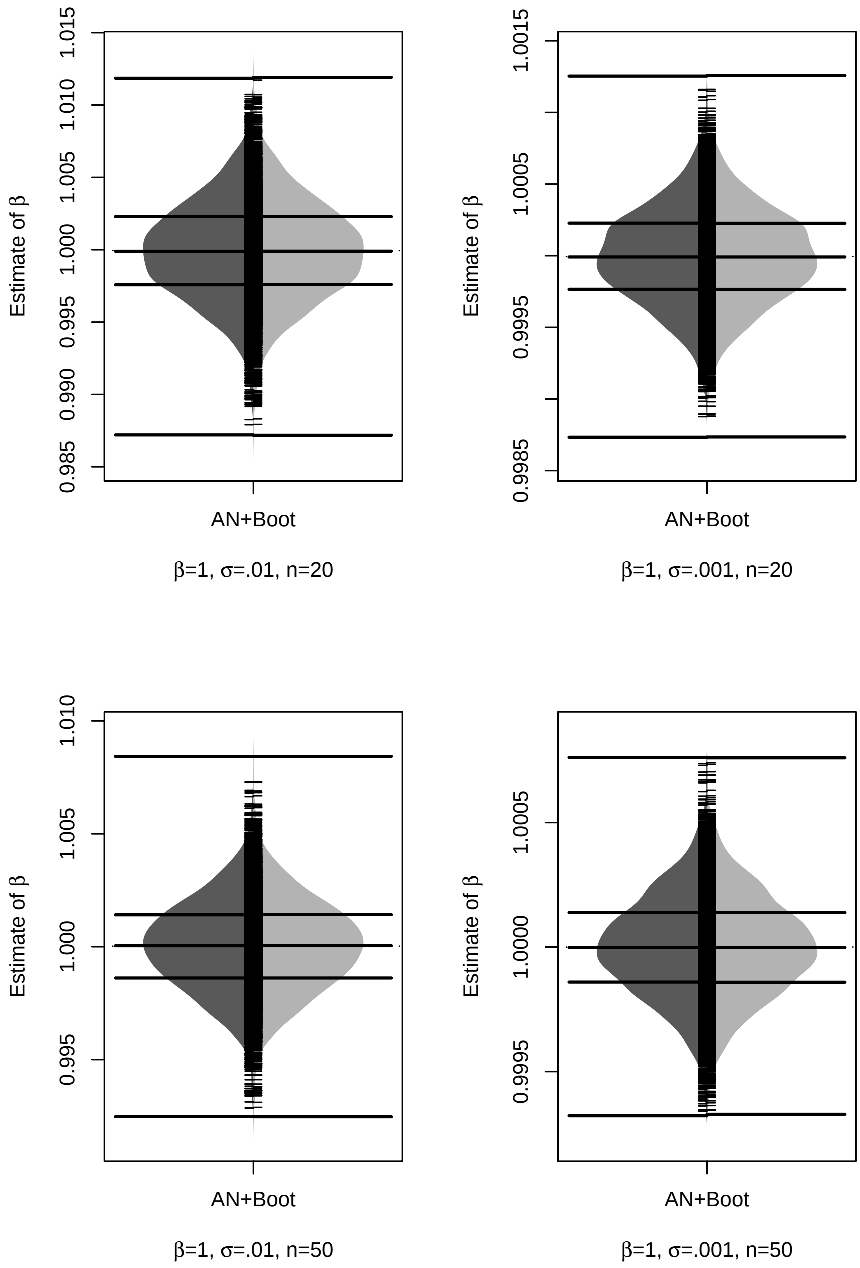

Figure 3.1: Comparisons of the original empirical distribution of $\widehat{\beta}$ with the nonparametric bootstrap version when $\beta = 1$.

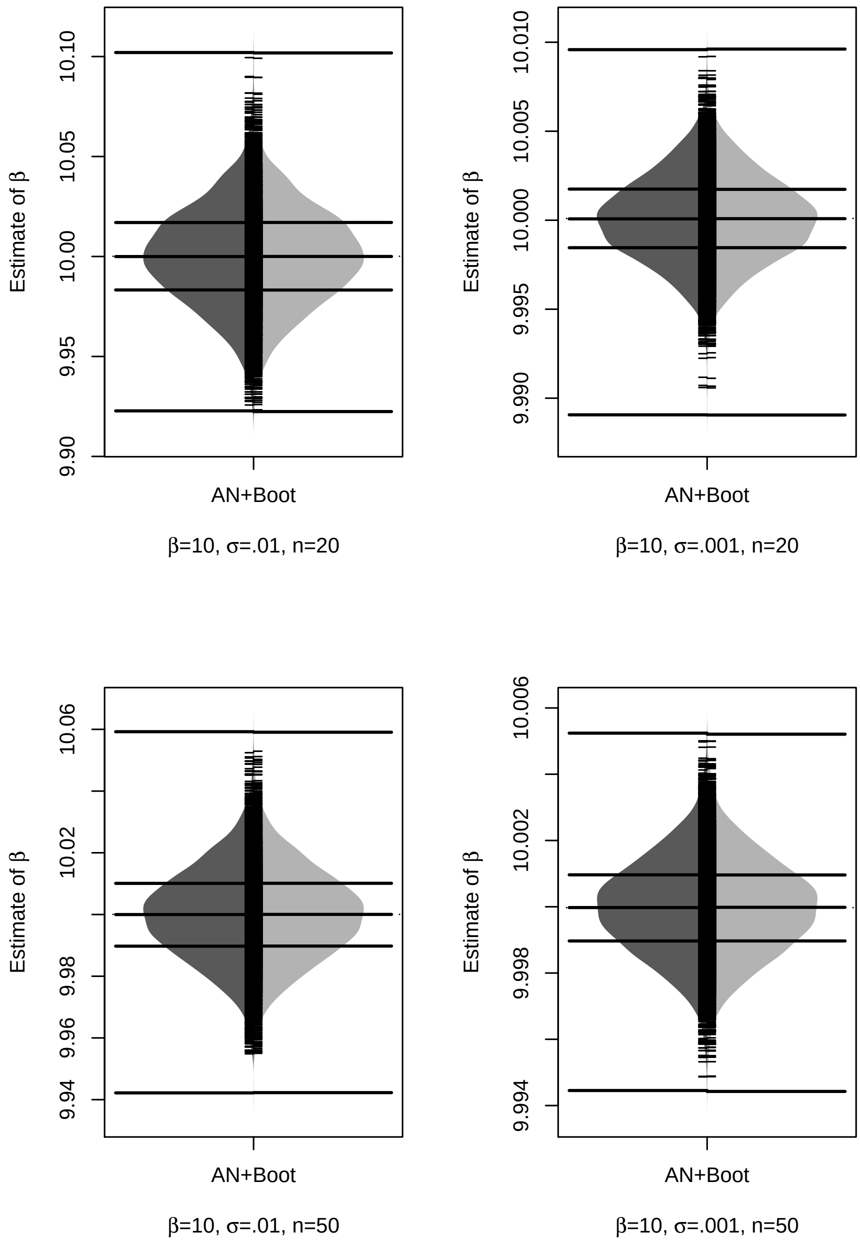

Figure 3.2: Comparisons of the original empirical distribution of $\widehat{\beta}$ with the nonparametric bootstrap version when $\beta = 10$.

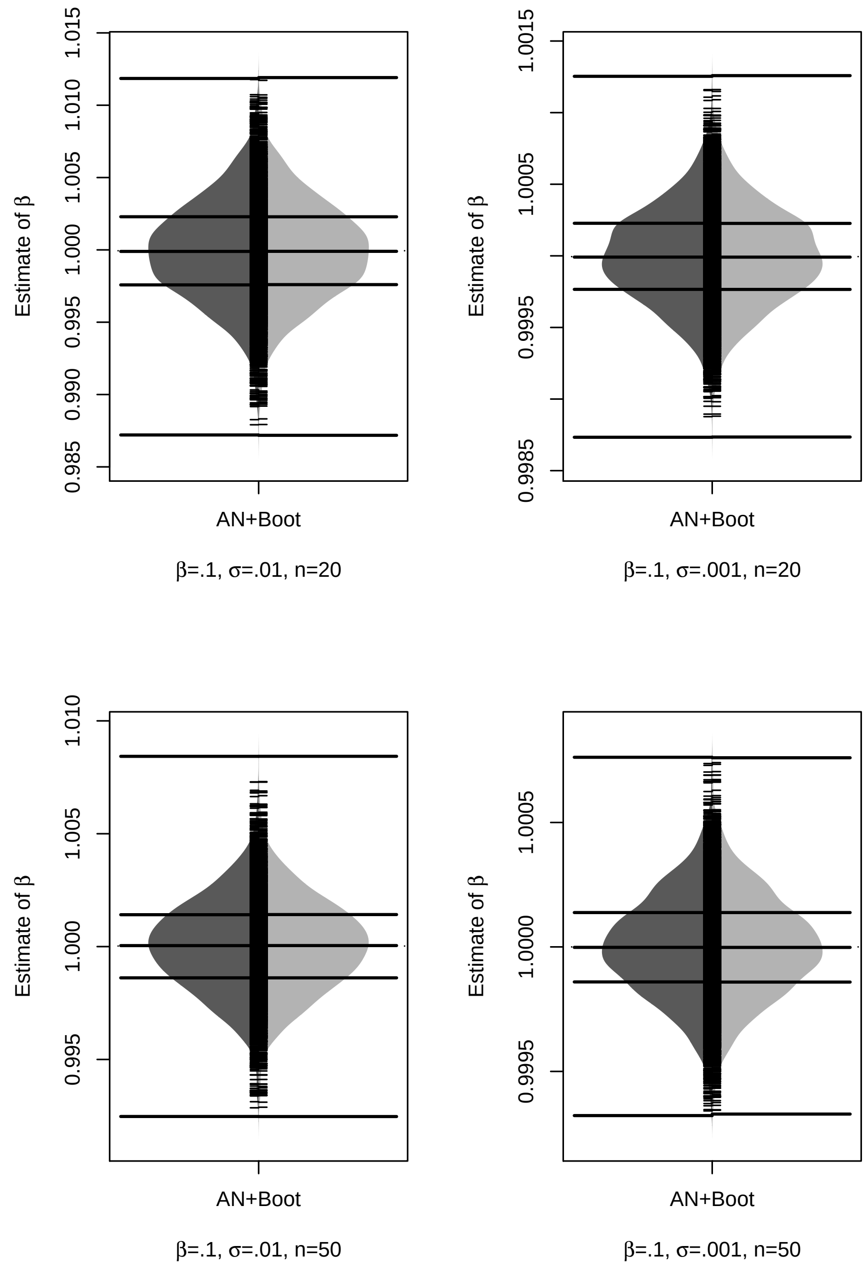

Figure 3.3: Comparisons of the original empirical distribution of $\widehat{\beta}$ with the nonparametric bootstrap version when $\beta = .1$.

and upon the proper nonparametric bootstrap. Figure 3.4 corresponds to a simulation setup when $\beta = 1$, Figure 3.5 to $\beta = 10$, and Figure 3.5 to $\beta = .1$.

It may be noticed, that in some cases, the sample variance of $\widehat{\beta}$ estimated by the proper nonparametric bootstrap procedure is slightly larger than the theoretical one (see also average CI lengths in Tables 3.1–3.3). In spite of that, as the sample size increases and the standard deviation of errors decreases, the empirical coverage becomes closer to the theoretical value of 95%. One of the possible reasons (and the most probable one) for such an overestimation of the variance is hidden in the monotonicity of the design points (covariates $\mathbf{Z}$), that can lead into monotonic convergence of $n^{-1}\mathbf{Z}^\top\mathbf{Z}$ to $\mathbf{\Delta}$ as $n \to \infty$. We will see in the simulation study of Chapter 5, that this negligible overestimation vanishes when the design points are not monotonically ordered.

On the top of that, the proper and the incorrect nonparametric bootstrap procedure are compared in Figure 3.7 (only the case of $\beta = 1$ is presented for simplicity).

There is no visible striking difference between the empirical distribution based on the incorrect nonparametric bootstrap and the corrected version. On the contrary, it has to be kept in mind that a possible asymptotic nondegenerative zero mean distribution can occur between the proper nonparametric bootstrap distribution and the incorrect one.

R software (v2.10.0) is used for all the computations in the thesis with the default random number generator and *set.seed(1982)*, which was chosen according to the birth year of the thesis' author.

3.5 Data Analysis – Calibration

Let us consider the following *calibration problem*: a company has two industrial devices, where the first one is calibrated according to some institute of standards and the second one is just a casual device. We want to calibrate the second device according to the first one. Consequently, other devices of the same type are needed to be calibrated as well. For some reasons, e.g., economic, it is only possible to calibrate one device by the authorities.

Our data set contains 21 couples of speed values of two hammer rams, where the first forging hammer is calibrated. We set the same power level on both hammers and measure the speed of each hammer ram repeatedly changing only the power level. Our measurements of the speed are encumbered with errors of the same variability in both cases, because we use the same device for measuring the speed and both forging hammers are of the same type. Since the power set for the forging hammer is directly proportional to the speed of the hammer ram, our goal is to set a *correction coefficient*, which the second (ordinary) forging hammer's power needs to be corrected (multiplied) by. Therefore, our EIV model is very suitable for this setup—a linear dependence and errors in both measured speeds (with the same variance). Non-equal error variance in other type of experiments is not a big issue as already discussed in Section 1.4.

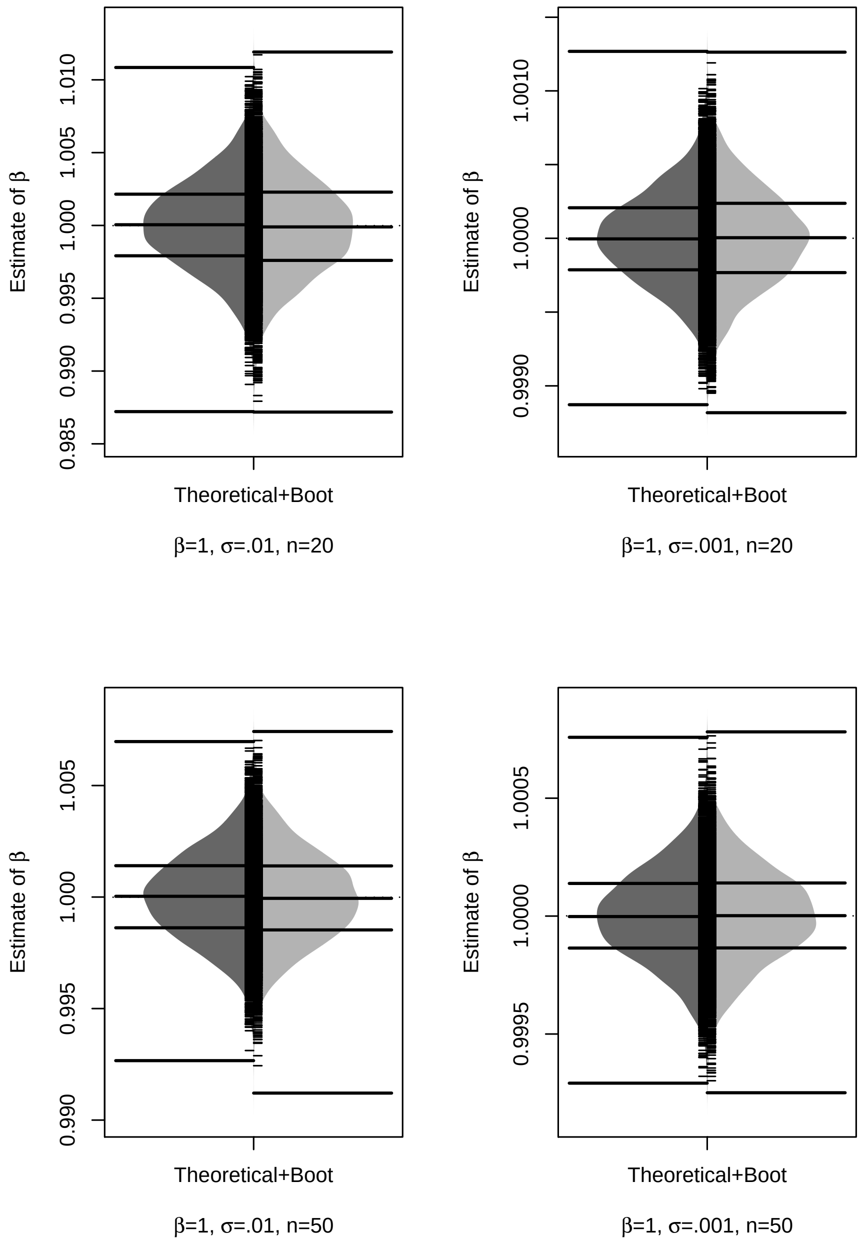

Figure 3.4: Comparisons of the empirical distributions of $\widehat{\beta}$ based on the theoretical asymptotic normality and the nonparametric bootstrap when $\beta = 1$.

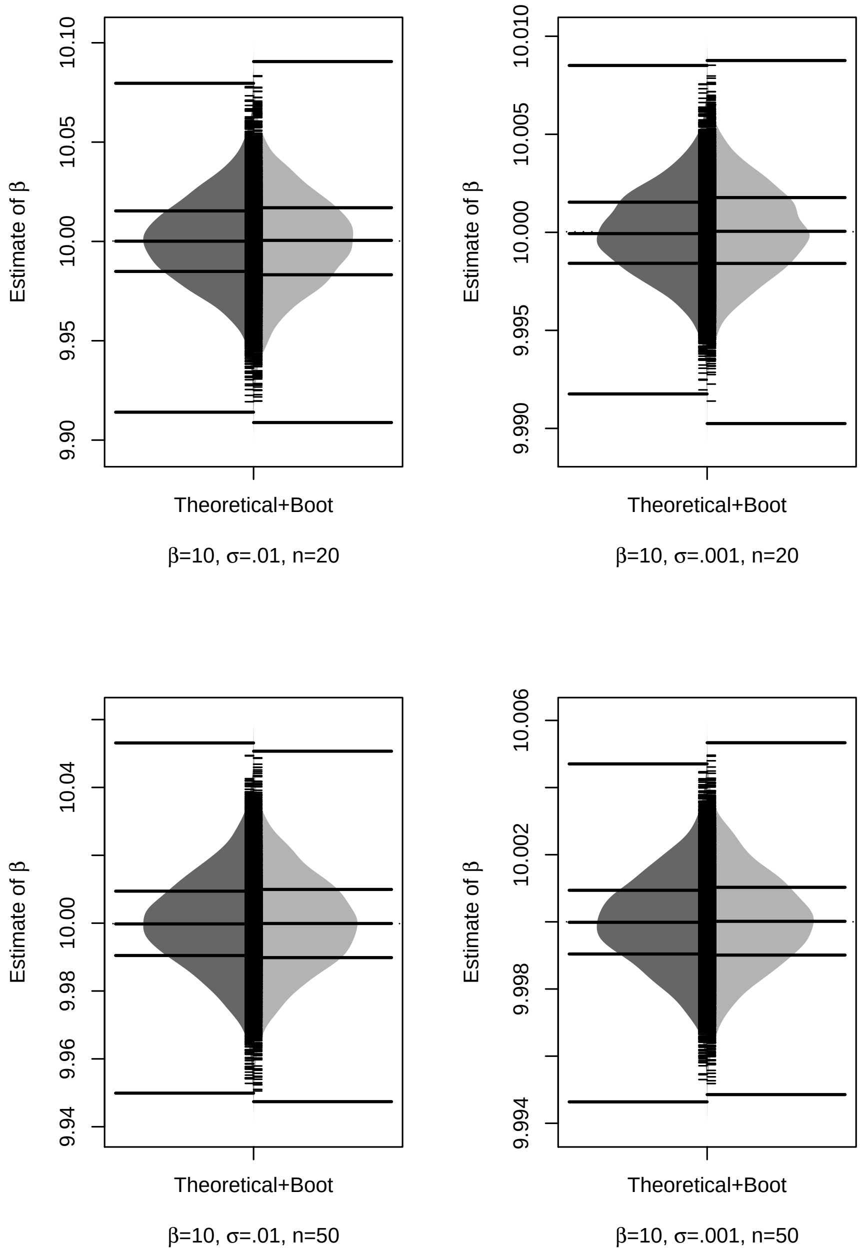

Figure 3.5: Comparisons of the empirical distributions of $\widehat{\beta}$ based on the theoretical asymptotic normality and the nonparametric bootstrap when $\beta = 10$.

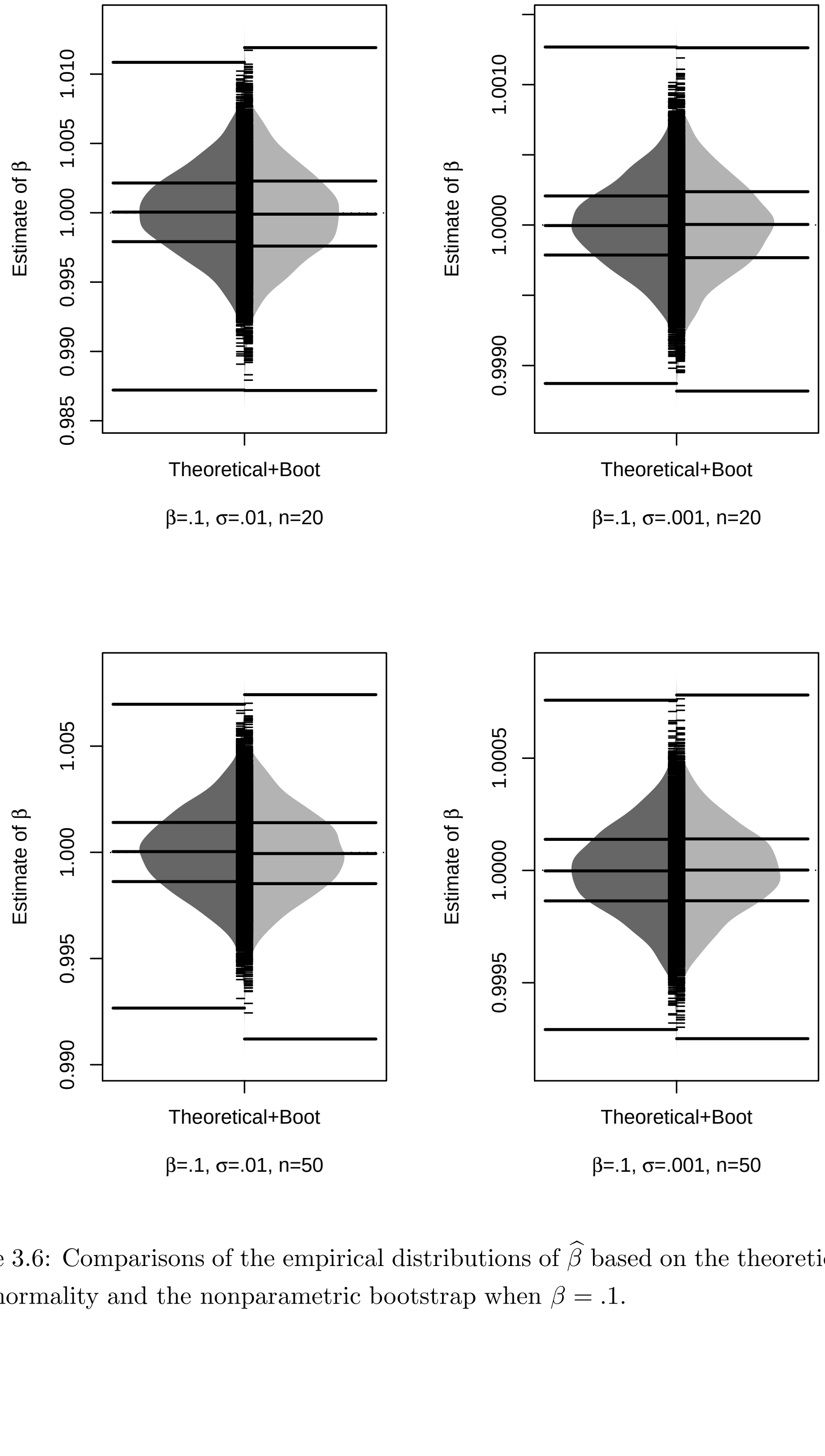

Figure 3.6: Comparisons of the empirical distributions of $\widehat{\beta}$ based on the theoretical asymptotic normality and the nonparametric bootstrap when $\beta = .1$.

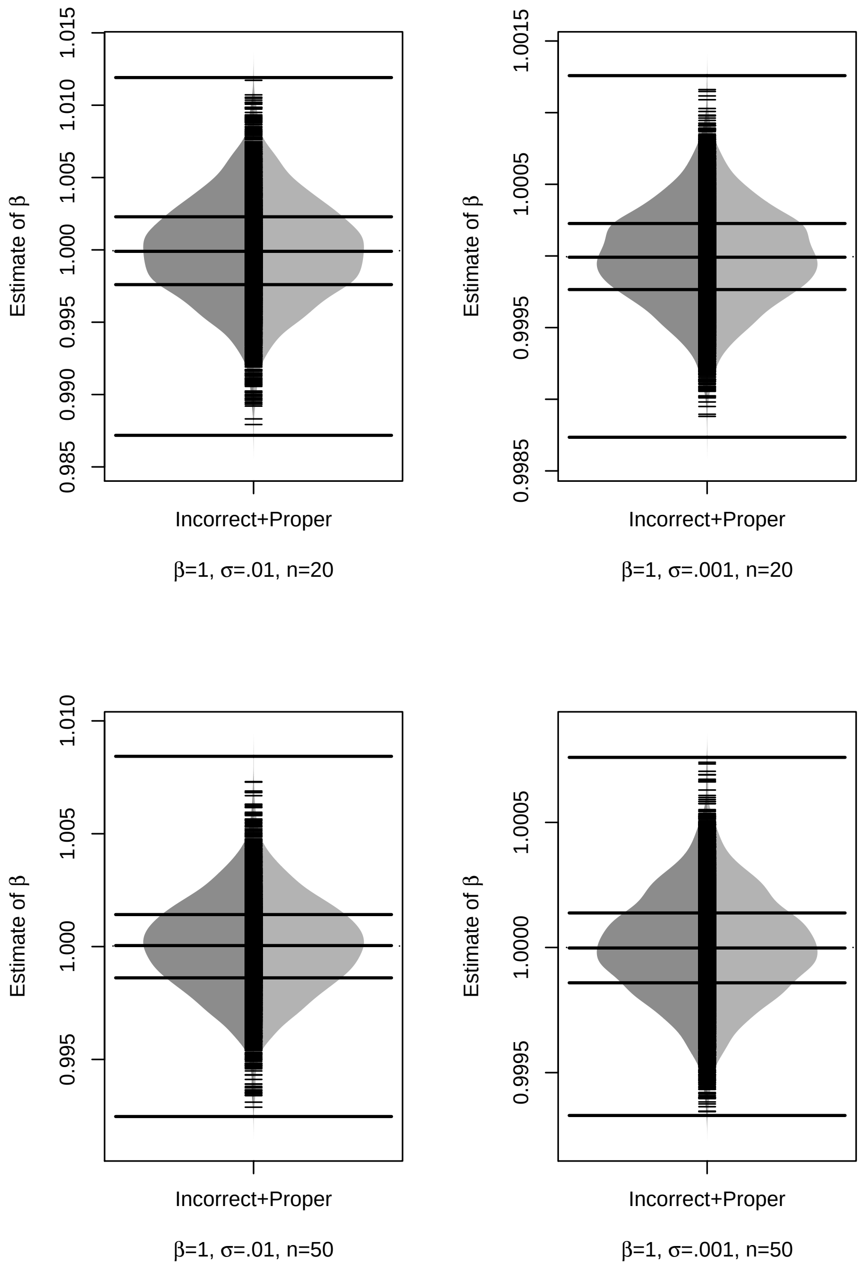

Figure 3.7: Comparisons of the empirical distributions of $\widehat{\beta}$ based on the incorrect nonparametric bootstrap and the proper nonparametric bootstrap when $\beta = 1$.

The hammer rams data with the corresponding EIV estimate (correction) are shown in Figure 3.8.

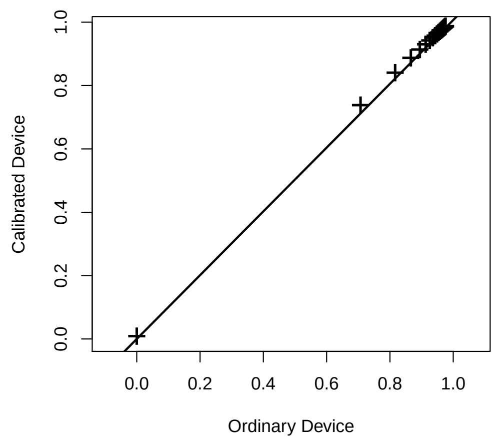

Figure 3.8: Measured speeds of two forging hammers of hammer rams—calibrated and ordinary (uncalibrated)—with an estimate of the corresponding correction coefficient.

One of our main interests lies in a construction of the CI for the calibration correction parameter in order to know how precise we are. Moreover, a test whether a correction of the second devise is necessary would be demanded, i.e., to test whether the correction parameter is equal to one. The estimated correction parameter for our forging hammer is 1.0068 and various types of CIs (approximate normality and proper nonparametric bootstrap) are summarized in Table 3.6.

Method	Confidence interval	CI length
Approximate asymptotic normality	[1.0043;1.0092]	0.0049
Nonparametric bootstrap (proper)	[1.0049;1.0093]	0.0045

Table 3.6: Hammer ram data set, $\widehat{\beta} = 1.0068$.

If we wanted to *test a hypothesis* whether $\beta = 1$ (is correction needed?) against the general alternative that $\beta \neq 1$, we should reject the null hypothesis according to the nonparametric bootstrapping and, similarly, according to the asymptotic normality. In spite of this, it is suggested that we should not rely on the asymptotic normality, because the proper formula for the asymptotic variance requires the third and the fourth moments of errors, which cannot simply be provided neither estimated. Hence, the second forging hammer should be calibrated according to the test based on the proper nonparametric bootstrapping.

It is difficult or sometimes even impossible to estimate the third and the fourth moments of the errors and, consequently, to test whether they are the same as those of the multivariate normal distribution. Therefore, one should *use the bootstrap approach for real data* as the preferable option.

3.6 Conclusions

A linear EIV model with its EIV estimate is considered in this chapter. Its disadvantageous asymptotic properties are picked out. Our problem is nicely linearly defined, but comes with a highly nonlinear estimate and inference. Therefore, methods based on the asymptotic normality might be computationally useless and fail. One of the reasons is that the assumption, where the *third and the fourth moments* of errors coincide with the third and the fourth moments of the normal distribution, cannot be verified neither assured.

Furthermore, one should absolutely *not rely on the normal asymptotics' approximation* in EIV model regarding the EIV estimate due to the unwieldy limiting variance in the case of the unknown third and fourth moments of errors.

Two nonparametric bootstrap approaches were proposed. *Justification* for use of the *proper nonparametric bootstrap* is given. The incorrect version of nonparametric bootstrap is shown to be invalid. We proved that the valid nonparametric bootstrap gives a proper answer with large samples in the EIV setup and, moreover, we showed that in finite samples the bootstrap performs at least as well as the conventional asymptotics.

Finally, a simulation study was conducted in order to demonstrate theoretical conclusions. Moreover, an application of our approach was performed on the calibration data, which the EIV provides an appealing approach to. Indeed, in the calibration problem the attitude of orthogonal regression seems to be appropriate and very flexible due to the fact that the predictive and the predicted variables switch their roles at a particular moment.

3.6.1 Discussion

Firstly, one needs to realize that the EIV estimate's inference *does not require any distributional assumptions* on the errors except the uniform boundedness of the fourth moments. The *crucial assumption* in the whole thesis is the "$\boldsymbol{\Delta}$" assumption (D), which can be restrictive in some situations. On the other hand, one can check whether it might be satisfied by calculating the partial sums of $i^{-1}\mathbf{X}_{1:i,\bullet}^\top\mathbf{X}_{1:i,\bullet}$, $i = 1,\ldots,n$. Moreover, a design of the experiment can be set properly to fulfill this assumption or a transformation of variables can be performed. On the top of that, it would be helpful to invent a test for the verification of design assumption (D).

It was discussed that the identically distributed observations suppress an existence of the mean structure in the regression model. Hence, they are quite hypothetical without realistic applicability. On the other hand, the identically distributed errors provide a reasonable and workable background for the mean structure in the EIV model.

It should be remembered, if the third and the fourth moments of errors are unknown, asymptotic normality does not provide a formula for the estimator's variance.

A heteroscedasticity is not an issue for the nonparametric bootstrap, because it is based on case sampling and the data can be transformed into the homoscedastic case as described in Section 1.4. Probably the biggest problem concerning residual bootstrap would lie in the fact that it *cannot handle* a nondiagonal covariance matrix. The reason for this is that the residuals are estimates of the errors and the nondiagonal covariance matrix brings dependence into the errors. Thus this would be another reason against the usage and applicability of some king of the residual bootstrap.

However, the nonparametric bootstrap is sometimes open to criticism. Its biggest disadvantage against the residual one is the *inefficiency* if the homoscedasticity is satisfied (Davidson and Hinkley, 1997). Indeed, the residual bootstrap design retains the information about regressors and response from the sample.

The *equiboundedness of the fourth moments* (3.24) in the bootstrap CLT is needed, because the second conditional moment is necessary for the existence of $\mathbb{V}\mathrm{ar}_{\mathbb{P}^*_{\boldsymbol{\xi}}}\xi_1^*$ and, consequently, the equiboundedness of the second moment of the second conditional moment is used for the convergence $[\mathbb{P}]$-almost surely of $\mathbb{V}\mathrm{ar}_{\mathbb{P}^*_{\boldsymbol{\xi}}}\xi_1^*$. Since the proper (corrected) bootstrapped estimate $\widetilde{\boldsymbol{\beta}}^*$ depends on the incorrect bootstrapped estimate $\widehat{\boldsymbol{\beta}}^*$, where the squared errors occur, and the bootstrap CLT assumes the equiboundedness of the fourth moments, then we have to impose the equiboundedness of the eighth moments on the errors of the EIV model.

The fourth moments' equiboundedness in the bootstrap CLT can be weakened by assuming identically distributed variables. On the other hand, the assumption of *iid* errors is quite restrictive and, moreover, it is a waste of too strong assumptions. Indeed, the errors of the EIV model are multiplied by the changing (non-constant) unknown covariates $\mathbf{Z}_{i,\bullet}$, $i = 1, \ldots, n$ in the expression for estimate $\widehat{\boldsymbol{\beta}}$. In order to apply the bootstrap CLT on these transformed errors, one cannot assume that those transformed errors remain identically distributed (if they originally were *iid*) nor they become identically distributed.

When is the nonparametric bootstrap procedure consistent (valid) in the way that the estimates are distributionally comparable in distribution $[\mathbb{P}]$-almost surely? As it was pointed out previously in this chapter, assumption (3.24) from the bootstrap CLT can be replaced by a stronger assumption (also a quite restrictive one) formulated in (3.28) and, hence, the bootstrap CLT would change into (3.29). From the point of view of the applicability, there is no *practical* difference between the convergence in distribution in probability $\mathbb{P}$ and the convergence in distribution $[\mathbb{P}]$-almost surely, when the theoretical results are applied to the data. Therefore, there is no practical need to consider another version of Theorem 3.10 for the nonparametric bootstrap validity with more restrictive assumptions implying a stronger type of the distributional coincidence—convergence in distribution [$\mathbb{P}$]-almost surely.

Finally, an interesting fact needs to be remarked concerning the conditional variance of sum of the bootstrapped observations. As it may be noticed in (3.26), the normalizing conditional variance $\mathbb{V}\mathrm{ar}_{\mathbb{P}^*_{\boldsymbol{\xi}}}\xi_1^*$ can be replaced by $n^{-1}\varsigma_n^2$ as shown in the proof of the bootstrap central limit Theorem 3.10. Hence, this makes the results of the bootstrap CLT even stronger and more applicable.

Chapter 4

Asymptotics for Weakly Dependent Errors

It is better to be roughly right than precisely wrong.

JOHN W. TUKEY

Our EIV model concerns linear relations, where measurement errors in input and output data occur simultaneously. Due to the fact that in some situations these disturbances cannot be considered as independent by nature, a proper error structure is required and, consequently, suitable statistical inference needs to be derived.

Errors-in-variables model with *dependent* errors is considered. A strong consistency of the EIV estimate for *weakly dependent* (α- and φ-mixing) measurements—encumbered with not necessarily stationary errors—is proved. Thereafter, an asymptotic normality of the EIV estimate is derived for such cases.

4.1 Weak Dependence

We are not in the case of independent observations any more and, therefore, the dependence between measurement errors needs to be specified. 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 $\mathcal{A}$. Henceforth, let us define a filtration $\mathcal{F}_m^n := \sigma(\xi_i \in \mathcal{F}, m \leq i \leq n)$.

There are many ways how to describe weak dependence or, in other words, asymptotic independence of random variables (Bradley, 2005). In this thesis we concentrate on two approaches. A sequence $\{\xi_n\}_{n=1}^{\infty}$ of random elements (e.g., variables) is said to be *strong mixing* (α-mixing) if

$$\alpha(n) := \sup_{k \in \mathbb{N}} \alpha(\mathcal{F}_1^k, \mathcal{F}_{k+n}^{\infty}) \to 0, \quad n \to \infty; \tag{4.1}$$

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

$$\varphi(n) := \sup_{k \in \mathbb{N}} \varphi(\mathcal{F}_1^k, \mathcal{F}_{k+n}^{\infty}) \to 0, \quad n \to \infty. \tag{4.2}$$

Uniformly strong mixing—introduced by Rosenblatt (1956)—implies strong mixing (Lin and Lu, 1997), which was presented by Ibragimov (1959). 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 (1958) comprehensively and intensively analyzed a class of m-dependent processes. These types of time series are φ-mixing, since are finite order ARMA processes with innovations satisfying *Doeblin's condition* (Billingsley (1968, p. 168), Doob (1953, p. 192)). Finite order processes, which do not satisfy Doeblin's condition, can be shown to be α-mixing (Ibragimov and Linnik, 1971, pp. 312–313). Rosenblatt (1971) provides general conditions under which stationary Markov processes are α-mixing. Since functions of mixing processes are themselves mixing (Bradley, 2005), time-varying functions of any of the processes just mentioned are mixing as well.

The *error structure* for independent observations proposed in Section 1.4 has to be suitably modified in order to cover the case of weakly dependent observations. On the *between-individual level*, the elements of rows $[\boldsymbol{\Theta}_{i,\bullet}, \varepsilon_i]$ suppose to form *weakly dependent* sequences with *zero mean*, e.g., 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 assumptions will be discussed in Discussion 4.2.1.

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

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 type of symmetry does not hold for φ-dependence. Indeed, Rosenblatt (1971, 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.

Strong consistency of the TLS estimate for independent errors is proved by Gleser (1981) and, moreover, *weak consistency*—again for independent errors, but with less restrictive assumptions—is widely discussed in Gallo (1982a). When a premise of independence cannot be assumed, a consistency of the TLS estimate under weak dependence of errors has to be explored (Pešta, 2009a). Similar situation occurs to the TLS estimate's asymptotic normality, which was proved by Gallo (1982b) for the case of independent errors. We will extend this result for the weakly dependent errors in this chapter.

4.2 Strong Consistency

First of all, a *strong law of large numbers* (SLLN) for α-dependent *non-identically distributed* variables needs to be recalled.

Lemma 4.1 (Strong law of large numbers for α-mixing). *Let $\{\xi_n\}_{n=1}^{\infty}$ be a sequence of α-mixing random variables satisfying*

$$\sup_{n\in\mathbb{N}} \mathbb{E}|\xi_n|^q < \infty \tag{4.3}$$

for some $q > 1$. Suppose that there exists $\delta > 0$ such that as $n \to \infty$,

$$\alpha(n) = \begin{cases} \mathcal{O}\left(n^{-\frac{q}{2q-2}-\delta}\right) & \text{if} \quad 1 < q < 2, \\ \mathcal{O}\left(n^{-\frac{2}{q}-\delta}\right) & \text{if} \quad q \geq 2. \end{cases} \tag{4.4}$$

Then

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n}(\xi_i - \mathbb{E}\xi_i)}{n} = 0 \quad a.s.$$

Proof. See Chen and Wu (1989, Theorem 1). □

Furthermore, a SLLN for φ-dependent non-identically distributed variables is desired as well.

Lemma 4.2 (Strong law of large numbers for φ-mixing). *Let $\{\xi_n\}_{n=1}^{\infty}$ be a sequence of zero mean φ-mixing random variables satisfying*

$$\sum_{n=1}^{\infty} \sqrt{\varphi(n)} < \infty \tag{4.5}$$

and let $\{b_n\}_{n=1}^{\infty}$ *be a non-decreasing unbounded sequence of positive numbers. Assume that*

$$\sum_{n=1}^{\infty} \frac{\mathbb{E}\xi_n^2}{b_n^2} < \infty, \tag{4.6}$$

then

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{n} \xi_i}{b_n} = 0 \quad a.s.$$

Proof. See Xuejun et al. (2009, Theorem 4.1). □

For a given random sequence $\xi_\circ \equiv \{\xi_n\}_{n=1}^{\infty}$ of random elements, the dependence coefficients $\alpha(n)$ will be denoted $\alpha(\xi_\circ, n)$. Analogous notation is used for φ-mixing sequences. Moreover, an auxiliary lemma for latter application of the SLLN for non-identically distributed random variables is stated.

The following lemma describes an asymptotic behavior of α- and φ-mixing coefficients of the corresponding random sequences after a transformation. More precisely, the Borel transformation preserves the property of α- and φ-mixing and, moreover, sustains the rate of the mixing coefficients.

Lemma 4.3. *Suppose that for each* $m = 1, 2, \ldots$, $\xi^{(m)} := \{\xi_k^{(m)}\}_{k\in\mathbb{Z}}$ *is a sequence of random variables. Suppose the sequences* $\xi^{(m)}$, $m = 1, 2, \ldots$ *are independent of each other. Suppose that for each* $k \in \mathbb{Z}$, $h_k : \mathbb{R} \times \mathbb{R} \times \ldots \to \mathbb{R}$ *is a Borel function. Define the sequence* $\xi := \{\xi_k\}_{k\in\mathbb{Z}}$ *of random variables by*

$$\xi_k := h_k\left(\xi_k^{(1)}, \xi_k^{(2)}, \ldots\right), \quad k \in \mathbb{Z}.$$

Then for each $n \geq 1$, *the following statements hold:*

(i) $\alpha(\xi, n) \leq \sum_{m=1}^{\infty} \alpha(\xi^{(m)}, n)$,

(ii) $\varphi(\xi, n) \leq \sum_{m=1}^{\infty} \varphi(\xi^{(m)}, n)$.

Proof. See Bradley (2005, Theorem 5.2). □

The preliminary statistical machinery is going to be used for a derivation of the main results of this section—strong consistency of the EIV estimate. Besides the main consistency results, an estimate of *nuisance* parameter σ^2 is defined as $\widehat{\sigma}^2 := \lambda/n$ and its strong consistency is proved as well.

Firstly, the EIV estimate is strongly consistent assuming α-mixing errors in the EIV model.

Theorem 4.4 (Strong consistency in EIV with α-mixing)**.** *Let the EIV model hold and assumption* (D) *be satisfied. Suppose*

$$\{\Theta_{n,1}\}_{n=1}^{\infty}, \ldots, \{\Theta_{n,p}\}_{n=1}^{\infty}, \quad \textit{and} \quad \{\varepsilon_n\}_{n=1}^{\infty} \tag{4.7}$$

are pairwise independent sequences of α-mixing random variables having

$$\alpha(\Theta_{\circ,j}, n) = \mathcal{O}(n^{-q_j/(2q_j-2)-\delta_j}), \quad j = 1, \ldots, p \tag{4.8}$$

and

$$\alpha(\varepsilon_\circ, n) = \mathcal{O}(n^{-q_{p+1}/(2q_{p+1}-2)-\delta_{p+1}}), \tag{4.9}$$

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

$$\sup_{n\in\mathbb{N}} Z_{n,j}^2 < \infty, \quad j \in \{1, \ldots, p\}, \tag{4.10}$$

$$\sup_{n\in\mathbb{N}} \mathbb{E}|\Theta_{n,j}|^{2q_j} < \infty, \quad j \in \{1, \ldots, p\}, \tag{4.11}$$

and

$$\sup_{n\in\mathbb{N}} \mathbb{E}|\varepsilon_n|^{2q_{p+1}} < \infty, \tag{4.12}$$

then

$$\lim_{n\to\infty} \widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} \quad a.s., \tag{4.13}$$

$$\lim_{n\to\infty} \frac{\lambda}{n} = \sigma^2 \quad a.s. \tag{4.14}$$

Proof. The estimate of $\boldsymbol{\beta}$ from (2.5) can be expressed as

$$\widehat{\boldsymbol{\beta}} = \left\{\mathbf{I} + (\mathbf{Z}^\top\mathbf{Z})^{-1}(\mathbf{Z}^\top\boldsymbol{\Theta} + \boldsymbol{\Theta}^\top\mathbf{Z} + \boldsymbol{\Theta}^\top\boldsymbol{\Theta} - n\sigma^2\mathbf{I} + (n\sigma^2 - \lambda)\mathbf{I})\right\}^{-1} \times (\mathbf{Z}^\top\mathbf{Z})^{-1}\left(\mathbf{Z}^\top\mathbf{Z}\boldsymbol{\beta} + \mathbf{Z}^\top\boldsymbol{\varepsilon} + \boldsymbol{\Theta}^\top\mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\Theta}^\top\boldsymbol{\varepsilon}\right). \tag{4.15}$$

If we want to prove (4.13), it is sufficient to show that

(i) $n^{-1}\mathbf{Z}^\top\boldsymbol{\Theta} \to \mathbf{0}$ a.s., $n \to \infty$;

(ii) $n^{-1}\boldsymbol{\Theta}^\top\mathbf{Z} \to \mathbf{0}$ a.s., $n \to \infty$;

(iii) $n^{-1}(\boldsymbol{\Theta}^\top\boldsymbol{\Theta} - n\sigma^2\mathbf{I}) \to \mathbf{0}$ a.s., $n \to \infty$;

(iv) $n^{-1}(n\sigma^2 - \lambda) \to 0$ a.s., $n \to \infty$;

(v) $n^{-1}\mathbf{Z}^\top\boldsymbol{\varepsilon} \to \mathbf{0}$ a.s., $n \to \infty$;

(vi) $n^{-1}\boldsymbol{\Theta}^\top\boldsymbol{\varepsilon} \to \mathbf{0}$ a.s., $n \to \infty$.

Note that

$$\sup_{n\in\mathbb{N}} \mathbb{E}|Z_{n,j}\Theta_{n,k}|^2 = \sigma^2 \sup_{n\in\mathbb{N}} Z_{n,j}^2 < \infty, \quad \forall j,k \in \{1,\ldots,p\}.$$

Moreover, Lemma 4.3(i) implies that $\alpha(Z_{\circ,j}\Theta_{\circ,k},n) = \mathcal{O}(n^{-q_k/(2q_k-2)-\delta_k})$, which implies $\alpha(Z_{\circ,j}\Theta_{\circ,k},n) = \mathcal{O}(n^{-1-\delta_k})$ for all $j,k \in \{1,\ldots,p\}$. Applying SLLN for α-mixing (Theorem 4.1), we have

$$n^{-1}\sum_{i=1}^{n} Z_{i,j}\Theta_{i,k} \xrightarrow{a.s.} 0, \quad n \to \infty, \quad \forall j,k \in \{1,\ldots,p\}.$$

Therefore, (i) holds and the similar arguments demonstrate (ii) and (v).

Again, it follows from Lemma 4.3(i) that $\alpha(\Theta_{\circ,j}\Theta_{\circ,k},n) = \mathcal{O}(n^{-1-\delta_j\wedge\delta_k})$ for all $j,k \in \{1,\ldots,p\}$ such that $j \neq k$. The supremum assumption of Theorem 4.1 is straightforwardly satisfied, because the pairwise independence from (4.7) provides

$$\sup_{n\in\mathbb{N}} \mathbb{E}|\Theta_{n,j}\Theta_{n,k}|^2 = \sup_{n\in\mathbb{N}} \mathbb{E}\Theta_{n,j}^2\mathbb{E}\Theta_{n,k}^2 = (\sigma^2)^2 < \infty$$

for all $j,k \in \{1,\ldots,p\}$, $j \neq k$. Hence, the SLLN for α-mixing yields

$$n^{-1}\sum_{i=1}^{n} \Theta_{i,j}\Theta_{i,k} \xrightarrow{a.s.} 0, \quad n \to \infty, \quad \forall j,k \in \{1,\ldots,p\},\ j \neq k.$$

Thus the off-diagonal part of (iii) is satisfied and, moreover, the analogous arguments demonstrate (vi).

Consequently, $\alpha(\Theta_{\circ,j}^2,n) = \mathcal{O}(n^{-q_j/(2q_j-2)-\delta_j})$ for all $j \in \{1,\ldots,p\}$ by Lemma 4.3(i). Since $\sup_{n\in\mathbb{N}} \mathbb{E}|\Theta_{n,j}^2|^{q_j} < \infty$ for all $j \in \{1,\ldots,p\}$, then the SLLN for α-mixing can be applied

$$n^{-1}\sum_{i=1}^{n} \Theta_{i,j}^2 \xrightarrow{a.s.} \sigma^2, \quad n \to \infty, \quad \forall j \in \{1,\ldots,p\},$$

and the "diagonal" part of (iii) holds as well.

Now,

$$n^{-1}(\lambda - n\sigma^2) = \lambda_{min}(n^{-1}[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}] - \sigma^2\mathbf{I})$$

due to the eigendecomposition property. Let $\mathbf{B} := n^{-1}[\mathbf{I},\boldsymbol{\beta}]^\top\mathbf{Z}^\top\mathbf{Z}[\mathbf{I},\boldsymbol{\beta}]$. For each $n \in \mathbb{N}$, $\mathbf{B}$ is a positive semidefinite matrix of rank p. Thus it has p positive eigenvalues and the smallest one being zero. Note that

$$n^{-1}([\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]-n\sigma^2\mathbf{I})-\mathbf{B}$$
$$=n^{-1}\left\{[\mathbf{I},\boldsymbol{\beta}]^\top\mathbf{Z}^\top[\boldsymbol{\Theta},\boldsymbol{\varepsilon}]+[\boldsymbol{\Theta},\boldsymbol{\varepsilon}]^\top\mathbf{Z}[\mathbf{I},\boldsymbol{\beta}]\right\}+n^{-1}\left\{[\boldsymbol{\Theta},\boldsymbol{\varepsilon}]^\top[\boldsymbol{\Theta},\boldsymbol{\varepsilon}]-n\sigma^2\mathbf{I}\right\} \quad (4.16)$$

The first summand on the right hand side of equation (4.16) converges almost surely to zero due to (i), (ii), and (v). The second one converges almost surely to zero as well, using similar arguments as in (iii) and (vi). Furthermore, it follows from Lemma 3.4 that

$$\lambda_{min}(n^{-1}[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]-\sigma^2\mathbf{I})\xrightarrow[n\to\infty]{a.s.}\lambda_{min}(\mathbf{B})=0,$$

which demonstrates (iv).

Finally, (iv) directly implies (4.14) and completes the proof. □

Similarly as above, φ-mixing errors yield the EIV estimate's strong consistency as well, but under slightly different assumptions.

Theorem 4.5 (Strong consistency in EIV with φ-mixing)**.** *Let the EIV model hold and assumption* (D) *be satisfied. Suppose*

$$\{\Theta_{n,1}\}_{n=1}^\infty,\ldots,\{\Theta_{n,p}\}_{n=1}^\infty,\quad \textit{and}\quad \{\varepsilon_n\}_{n=1}^\infty \quad (4.17)$$

are pairwise independent sequences of φ-mixing random variables such that

$$\sum_{n=1}^\infty\sqrt{\varphi(\Theta_{\circ,j},n)}<\infty,\quad j\in\{1,\ldots,p\} \quad (4.18)$$

and

$$\sum_{n=1}^\infty\sqrt{\varphi(\varepsilon_\circ,n)}<\infty. \quad (4.19)$$

If

$$\sum_{n=1}^\infty\frac{\mathbb{E}\Theta_{n,j}^4}{n^2}<\infty,\quad j\in\{1,\ldots,p\} \quad (4.20)$$

and

$$\sum_{n=1}^\infty\frac{\mathbb{E}\varepsilon_n^4}{n^2}<\infty, \quad (4.21)$$

then

$$\lim_{n\to\infty} \widehat{\boldsymbol{\beta}} = \boldsymbol{\beta} \quad a.s., \tag{4.22}$$

$$\lim_{n\to\infty} \frac{\lambda}{n} = \sigma^2 \quad a.s. \tag{4.23}$$

Proof. A process of proving this theorem is analogous to the proof of Theorem 4.4. The only difference is that the SLLN for φ-mixing is applied instead of the SLLN for α-mixing. Therefore, one does not have to take care about the supremum condition (4.3) and the dependence coefficient assumption (4.4) from Theorem 4.1. On the other hand, the convergence condition (4.5) on sum of the square roots of dependence coefficients $\varphi(n)$ and the convergence assumption (4.6) from Theorem 4.2 need to be fulfilled.

Let us consider six terms of (4.15) from the proof of Theorem 4.4. It follows from Lemma 4.3(ii) that $\{Z_{n,j}\Theta_{n,k}\}_{n=1}^{\infty}$ is also a φ-mixing sequence for all $j,k \in \{1,\ldots,p\}$ and, moreover,

$$\sum_{n=1}^{\infty} \sqrt{\varphi(Z_{\circ,j}\Theta_{\circ,k},n)} \leq \sum_{n=1}^{\infty} \sqrt{\varphi(\Theta_{\circ,k},n)} < \infty, \quad \forall j,k \in \{1,\ldots,p\}.$$

Assumption (D) implies

$$0 < n^{-1}\sum_{i=1}^{n} Z_{i,j}^2 \to \Delta_{j,j} < \infty, \quad n \to \infty, \quad \forall j \in \{1,\ldots,p\}. \tag{4.24}$$

Due to Lemma 3.7,

$$\sum_{n=1}^{\infty} \frac{\mathbb{E}\{Z_{n,j}\Theta_{n,k}\}^2}{n^2} = \sigma^2 \sum_{n=1}^{\infty} \frac{Z_{n,j}^2}{n^2} < \infty, \quad \forall j,k \in \{1,\ldots,p\},$$

which allows us to apply the SLLN for φ-mixing. Hence, (i) holds and the similar arguments provide (ii) and (v).

The rest of the proof is now pretty straightforward. In order to show (iii), (iv), and (vi), one has to realize that Lemma 4.3(ii) yields $\varphi(\xi_\circ^2,n) \leq \varphi(\xi_\circ,n)$, $\varphi(\xi_\circ\zeta_\circ,n) \leq \varphi(\xi_\circ,n) + \varphi(\zeta_\circ,n)$, and, furthermore,

$$\sum_{n=1}^{\infty} \sqrt{\varphi(\xi_\circ\zeta_\circ,n)} \leq \sum_{n=1}^{\infty} \sqrt{\varphi(\xi_\circ,n)} + \sum_{n=1}^{\infty} \sqrt{\varphi(\zeta_\circ,n)} < \infty$$

for $\xi_n,\zeta_n \in \{\Theta_{n,1},\ldots,\Theta_{n,p},\varepsilon_n\}$, $\xi_n \neq \zeta_n$. Moreover, (4.20)–(4.21) hold and, due to the pairwise independence from (4.17),

$$\sum_{n=1}^{\infty} \frac{\mathbb{E}\{\xi_n\zeta_n\}^2}{n^2} = \sum_{n=1}^{\infty} \frac{\mathbb{E}\xi_n^2\mathbb{E}\zeta_n^2}{n^2} = \frac{\pi^2\sigma^4}{6} < \infty,$$

for $\xi_n, \zeta_n \in \{\Theta_{n,1}, \ldots, \Theta_{n,p}, \varepsilon_n\}$, $\xi_n \neq \zeta_n$, which completes the proof. □

4.2.1 Discussion

The assumptions of *pairwise independence* (4.7) and (4.17) between mixing sequences of errors on the within-individual level are crucial and cannot be omitted. The reason for this is that the form of the EIV estimate depends on the product of errors $\xi_n \zeta_n$, where $\xi_n, \zeta_n \in \{\Theta_{n,1}, \ldots, \Theta_{n,p}, \varepsilon_n\}$. The assumption of pairwise independence preserves the property of being α- or φ-mixing for such a product of weakly dependent disturbances due to Lemma 4.3. It may be thought of incorporating an extra form of weak dependence on the within-individual error level as well, but this could unfortunately require very complicated additional assumptions.

Heteroscedastic covariance structure of the within-individual errors can even be estimated without possessing repeated observations for each "individual", but a *structure of the covariance matrix* has to be predefined in advance according to some prior knowledge about the data dependence. E.g., if there is no reason to suppose that the error structure is changing over particular covariates and response, Toeplitz or AR(1) covariance models are reasonable choices.

Moreover, if we *compare the assumptions* for α- and φ-mixing in our EIV model, α-mixing has weaker assumptions on dependence of the errors (every φ-mixing is α-mixing, see e.g., Bradley (2005)), but stronger on the design (α-mixing requires bounded moments of the errors). For φ-mixing, it is the other way around. Indeed, assumptions (4.18)–(4.19) imply

$$\varphi(\Theta_{\circ,j}, n) = o(n^{-2}), \quad n \to \infty, \quad j \in \{1, \ldots, p\}$$

and

$$\varphi(\varepsilon_\circ, n) = o(n^{-2}), \quad n \to \infty.$$

Taking into account $\alpha(n) \leq \varphi(n)$ and supposing

$$\varphi(\Theta_{\circ,j}, n) = \mathcal{O}(n^{-2-\delta_j}), \quad n \to \infty, \quad j \in \{1, \ldots, p\},$$

$$\varphi(\varepsilon_\circ, n) = \mathcal{O}(n^{-2-\delta_{p+1}}), \quad n \to \infty$$

(which imply assumptions (4.18)–(4.19)), assumptions (4.8) and (4.9) are satisfied for some $4/3 \leq q_j \leq 2$, $j \in \{1, \ldots, p+1\}$. On the other hand, assumptions (4.11)–(4.12) with $q_j = 2$, $j \in \{1, \ldots, p+1\}$ clearly imply assumptions (4.20)–(4.21). The choice of q_j is essential as well. Smaller q_js make assumptions (4.11)–(4.12) less restrictive, but then, assumptions (4.8) and (4.9) become less realizable.

Additional design assumption (4.10), which is necessary for proving strong consistency for α-mixing errors, may be viewed as a competitive one to the "basic" design assumption (D). These assumptions are not equivalent and neither of them implies the other one. On the other hand, assumption (4.10) can be considered as a *supplementary* assumption to assumption (D) in the following sense: (D) implies (4.24). Hence, Lemma 3.7 yields $Z_{n,j}^2 = o(n^2)$, $n \to \infty$ for all $j \in \{1, \dots, p\}$, which is a weaker condition than the equiboundedness of $Z_{n,j}^2$ over all $n \in \mathbb{N}$ for all $j \in \{1, \dots, p\}$ from (4.10).

Finally, if identically distributed rows of errors (between-individual level) with an existence of their suitable moments are taken into account, assumptions (4.11)–(4.12) and (4.20)–(4.21) are trivially satisfied. Then, a *strict stationarity* of the between-individual errors with an existence of the appropriate moments has to imply these assumptions as well. Hence, moment assumptions (4.11)–(4.12) and (4.20)–(4.21) cannot be considered as unattainable. Moreover, for strictly stationary errors even the supremum in definitions (4.1) and (4.2) can simply be avoided.

4.3 Asymptotic Normality

A *weak invariance principle* (WIP) (also known as a *functional central limit theorem*) is a functional convergence of the sum of variables to the standard Wiener process $\mathcal{W}$. This principle for α-mixing variables will be recalled.

Put $S_n := \sum_{i=1}^{n} \xi_i$ and $\varsigma_n^2 := \mathbb{V}\mathrm{ar}\, S_n$. Define random elements on Skorokhod space $D[0,1]$ as follows:

$$\mathcal{W}_n(t) := \frac{S_{[nt]}}{\varsigma_n}, \quad 0 \le t \le 1, \tag{4.25}$$

where $[\cdot]$ denotes the nearest integer function. The expression ς_n^2 is usually called *long-run variance*.

Lemma 4.6 (Weak invariance principle for α-mixing)**.** *Let $\{\xi_n\}_{n=1}^{\infty}$ be a sequence of zero mean α-mixing random variables with*

$$\sup_{n \in \mathbb{N}} \mathbb{E}|\xi_n|^{2+\omega} < \infty \tag{4.26}$$

and

$$\sum_{n=1}^{\infty} \alpha(n)^{\omega/(2+\omega)} < \infty \tag{4.27}$$

for some $\omega > 0$. Suppose that

$$\frac{\mathbb{E}S_n^2}{n} \to \varsigma^2 > 0, \quad n \to \infty \tag{4.28}$$

is satisfied. Then

$$\mathcal{W}_n \xrightarrow{D[0,1]} \mathcal{W}, \quad n \to \infty. \tag{4.29}$$

Proof. See Herrndorf (1985) or Lin and Lu (1997, Corollary 3.2.1). ☐

Since the central limit theorem is just a special case of the weak invariance principle, then a corollary of previous Lemma 4.6 can be stated.

Corollary 4.7 (Central limit theorem for α-mixing)**.** *Suppose that all the assumptions of Lemma 4.6 on a sequence of zero mean α-mixing random variables $\{\xi_n\}_{n=1}^{\infty}$ are satisfied. Then*

$$\frac{S_n}{\varsigma_n} \xrightarrow{\mathscr{D}} \mathcal{N}(0,1), \quad n \to \infty. \tag{4.30}$$

Proof. Since a functional distributional limit (a convergence in Skorokhod space) implies the pointwise distributional limit, this corollary is just a special case of Lemma 4.6, when $\mathcal{W}_n(1)$ is considered. ☐

Lemma 4.8 (Lindeberg central limit theorem for φ-mixing)**.** *Let $\{\xi_n\}_{n=1}^{\infty}$ be a sequence of zero mean φ-mixing random variables having finite variance. Suppose that the Lindeberg condition*

$$\forall \delta > 0: \quad \lim_{n\to\infty} \frac{1}{\varsigma_n^2} \sum_{i=1}^{n} \mathbb{E}\xi_i^2 \mathcal{I}\{|\xi_i| > \delta\varsigma_n\} = 0 \tag{4.31}$$

is satisfied. Then

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

Proof. See Utev (1990, Corollary 4). ☐

Lindeberg condition (4.31) can be replaced by a stronger type of Lyapunov condition. This fact leads into the following corollary, which is more comfortable for us from the point of applicability.

Corollary 4.9 (Central limit theorem for φ-mixing)**.** *Let $\{\xi_n\}_{n=1}^{\infty}$ be a sequence of zero mean φ-mixing random variables such that*

$$\sup_{n\in\mathbb{N}} \mathbb{E}|\xi_n|^{2+\omega} < \infty \tag{4.32}$$

for some $\omega > 0$ and

$$\frac{\mathbb{E}S_n^2}{n} \to \varsigma^2 > 0, \quad n \to \infty. \tag{4.33}$$

Then

$$\frac{S_n}{\varsigma_n} \xrightarrow{\mathscr{D}} \mathcal{N}(0,1), \quad n \to \infty. \tag{4.34}$$

Proof. We show that assumptions (4.32) and (4.33) implies Lindeberg condition (4.31) from Lemma 4.8.

The first step is to show that conditions (4.32) and (4.33) 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

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

□

Assumption (4.33) may even be replaced by a weaker one:

$$\liminf_{n\to\infty} \frac{\mathbb{E}S_n^2}{n} = \varsigma^2 > 0,$$

where the limit inferior is used instead of the original limit.

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

Theorem 4.10 (Asymptotic normality in EIV with α-mixing)**.** *Let the EIV model hold and assumption* (D) *be satisfied. Suppose*

$$\{\Theta_{n,1}\}_{n=1}^{\infty}, \ldots, \{\Theta_{n,p}\}_{n=1}^{\infty}, \quad \textit{and} \quad \{\varepsilon_n\}_{n=1}^{\infty} \tag{4.35}$$

are pairwise independent sequences of α-mixing random variables having

$$\alpha(\Theta_{\circ,j}, n) = \mathcal{O}(n^{-1-\delta_j}), \quad j = 1, \ldots, p \tag{4.36}$$

and

$$\alpha(\varepsilon_\circ, n) = \mathcal{O}(n^{-1-\delta_{p+1}}), \tag{4.37}$$

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

$$\sup_{n\in\mathbb{N}} Z_{n,j}^2 < \infty, \quad j \in \{1, \ldots, p\}, \tag{4.38}$$

$$\sup_{n\in\mathbb{N}} \mathbb{E}|\Theta_{n,j}|^{4+\omega_j} < \infty, \quad j \in \{1, \ldots, p\}, \tag{4.39}$$

and

$$\sup_{n\in\mathbb{N}} \mathbb{E}|\varepsilon_n|^{4+\omega_{p+1}} < \infty \tag{4.40}$$

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

$$\frac{2}{\min_{j=1,\ldots,p+1} \omega_j} < \min_{j=1,\ldots,p+1} \delta_j. \tag{4.41}$$

If there exists a positive definite matrix $\beth$ *such that*

$$n^{-1}\mathbb{V}ar[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}] \begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix} \to \beth > \mathbf{0}, \quad n \to \infty; \tag{4.42}$$

then

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{\mathscr{D}} \mathcal{N}(\mathbf{0}, \cdot), \quad n \to \infty. \tag{4.43}$$

Proof. Assumptions of Theorem 4.10 imply the assumptions of Theorem 4.4. Indeed, the assumptions of pairwise independence (4.7) and (4.35) coincide. Similarly for assumptions (4.10) and (4.38). Assumptions on α-mixing rates (4.36) and (4.37) clearly imply assumptions (4.8) and (4.9) for any $\delta_j > 0$ and $1 < q_j \leq 2$, $j \in \{1, \ldots, p+1\}$. Supremum assumptions (4.39)–(4.40) imply (4.11)–(4.12) 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

$$(\mathbb{E}|\xi|^r)^{1/r} \leq (\mathbb{E}|\xi|^s)^{1/s},\ 0 < r < s < \infty. \tag{4.44}$$

Let us recall (3.41):

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

Assumptions of this theorem imply the assumptions of Theorem 4.4 and, hence, its consistency results can be used. With respect to (1.27), we have

$$\mathbf{\Delta}_n^{-1}\left([\mathbf{I},\boldsymbol{\beta}]\left[\mathbf{I},\widehat{\boldsymbol{\beta}}\right]^\top\right)^{-1}[\mathbf{I},\boldsymbol{\beta}]\xrightarrow{a.s.}\mathbf{\Delta}^{-1}\left([\mathbf{I},\boldsymbol{\beta}]\,[\mathbf{I},\boldsymbol{\beta}]^\top\right)^{-1}[\mathbf{I},\boldsymbol{\beta}]\,,\quad n\to\infty. \tag{4.45}$$

The inverse $\mathbf{\Delta}_n$ exists with probability tending to one due to (D) and (1.27) and the inverse of $[\mathbf{I},\boldsymbol{\beta}]\left[\mathbf{I},\widehat{\boldsymbol{\beta}}\right]^\top$ due to (3.42). 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 (4.45) and Slutsky's theorem (see Appendix A.2, Theorem A.2) reduce the problem of finding a limiting distribution for $\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})$ to that of finding a limiting distribution for

$$n^{-1/2}\left([\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]-\mathbb{E}[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]\right)\begin{bmatrix}\boldsymbol{\beta}\\-1\end{bmatrix}$$
$$=n^{-1/2}\left([\mathbf{I},\boldsymbol{\beta}]^\top\mathbf{Z}^\top[\mathbf{\Theta},\boldsymbol{\varepsilon}]+[\mathbf{\Theta},\boldsymbol{\varepsilon}]^\top[\mathbf{\Theta},\boldsymbol{\varepsilon}]-n\sigma^2\mathbf{I}\right)\begin{bmatrix}\boldsymbol{\beta}\\-1\end{bmatrix}. \tag{4.46}$$

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

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

and apply the Cramér-Wold theorem (see Appendix A.2, Theorem A.3). If

$$\mathbf{t}=c\begin{bmatrix}\boldsymbol{\beta}\\-1\end{bmatrix}$$

for some real constant $c\neq 0$, then each $\mathbf{t}^\top[\mathbf{Z},\mathbf{Z}\boldsymbol{\beta}]_{i,\bullet}^\top=0$. We have a sum of zero mean α-mixing random variables

$$\varpi_i:=c[\boldsymbol{\beta}^\top,-1]\left([\mathbf{\Theta},\boldsymbol{\varepsilon}]_{i,\bullet}^\top[\mathbf{\Theta},\boldsymbol{\varepsilon}]_{i,\bullet}-\sigma^2\mathbf{I}\right)\begin{bmatrix}\boldsymbol{\beta}\\-1\end{bmatrix}=c\left\{(\mathbf{\Theta}_{i,\bullet}\boldsymbol{\beta}-\varepsilon_i)^2-\sigma^2(1+\boldsymbol{\beta}^\top\boldsymbol{\beta})\right\}, \tag{4.47}$$

which satisfies all the assumptions of Corollary 4.7. In fact, assumption (4.26) holds, because

Lemma 3.5 (used twice) and (4.39)–(4.40) provide

$$\begin{aligned}
&\sup_{n\in\mathbb{N}} \mathbb{E}\left|c\left\{(\boldsymbol{\Theta}_{n,\bullet}\boldsymbol{\beta}-\varepsilon_n)^2-\sigma^2(1+\boldsymbol{\beta}^\top\boldsymbol{\beta})\right\}\right|^{2+\omega}\\
&\leq 2^{1+\omega}\left|c\sigma^2(1+\boldsymbol{\beta}^\top\boldsymbol{\beta})\right|^{2+\omega}+2^{1+\omega}c^{2+\omega}\sup_{n\in\mathbb{N}}\mathbb{E}\left(|\varepsilon_n|+\sum_{j=1}^{p}|\beta_j||\Theta_{n,j}|\right)^{4+2\omega}\\
&\leq 2^{1+\omega}\left|c\sigma^2(1+\boldsymbol{\beta}^\top\boldsymbol{\beta})\right|^{2+\omega}\\
&\quad+2^{1+\omega}c^{2+\omega}(p+1)^{3+2\omega}\left\{\sup_{n\in\mathbb{N}}\mathbb{E}|\varepsilon_n|^{4+2\omega}+\sum_{j=1}^{p}|\beta_j|^{4+2\omega}\sup_{n\in\mathbb{N}}\mathbb{E}|\Theta_{n,j}|^{4+2\omega}\right\}<\infty,
\end{aligned}$$

where it is sufficient to consider $\omega = 1/2\min_{j=1,\ldots,p+1}\omega_j$ and realize (4.44).

Since Lemma 4.3(i) yields

$$\alpha\left(c\left\{(\boldsymbol{\Theta}_{\circ,\bullet}\boldsymbol{\beta}-\varepsilon_\circ)^2-\sigma^2(1+\boldsymbol{\beta}^\top\boldsymbol{\beta})\right\},n\right)\leq\alpha(\varepsilon_\circ,n)+\sum_{j=1}^{p}\alpha(\boldsymbol{\Theta}_{\circ,j},n),$$

assumption (4.27) holds due to the concavity of function $u\mapsto u^{\omega/(2+\omega)}$, $\omega>0$ in $u\geq 0$:

$$\begin{aligned}
\sum_{n=1}^{\infty}\alpha\left(\varpi_\circ,n\right)^{\omega/(2+\omega)}&=\sum_{n=1}^{\infty}\alpha\left(c\left\{(\boldsymbol{\Theta}_{\circ,\bullet}\boldsymbol{\beta}-\varepsilon_\circ)^2-\sigma^2(1+\boldsymbol{\beta}^\top\boldsymbol{\beta})\right\},n\right)^{\omega/(2+\omega)}\\
&\leq\sum_{n=1}^{\infty}\alpha(\varepsilon_\circ,n)^{\omega/(2+\omega)}+\sum_{j=1}^{p}\sum_{n=1}^{\infty}\alpha(\boldsymbol{\Theta}_{\circ,j},n)^{\omega/(2+\omega)}<\infty,\quad\omega>0;
\end{aligned}$$

which is true because of (4.41) and the fact that

$$\alpha(\boldsymbol{\Theta}_{\circ,j},n)^{\omega/(2+\omega)}=\mathcal{O}\left(n^{-1-\frac{\delta_j\omega-2}{2+\omega}}\right),\quad\delta_j>2/\omega>0,\ j\in\{1,\ldots,p\}\tag{4.48}$$

and

$$\alpha(\varepsilon_\circ,n)^{\omega/(2+\omega)}=\mathcal{O}\left(n^{-1-\frac{\delta_j\omega-2}{2+\omega}}\right),\quad\delta_{p+1}>2/\omega>0.\tag{4.49}$$

Using (4.42) and (4.46), let us elaborate

$$\frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n}\varpi_i\right)^2=\frac{1}{n}\mathbb{E}\left\{c[\boldsymbol{\beta}^\top,-1]\left([\boldsymbol{\Theta},\boldsymbol{\varepsilon}]^\top[\boldsymbol{\Theta},\boldsymbol{\varepsilon}]-n\sigma^2\mathbf{I}\right)\left[\begin{array}{c}\boldsymbol{\beta}\\-1\end{array}\right]\right\}^2$$

$$= \frac{1}{n}\mathbb{E}[\boldsymbol{\beta}^\top, -1]\left([\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}] - \mathbb{E}[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]\right)\begin{bmatrix}\boldsymbol{\beta}\\ -1\end{bmatrix}$$
$$[\boldsymbol{\beta}^\top, -1]\left([\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}] - \mathbb{E}[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]\right)\begin{bmatrix}\boldsymbol{\beta}\\ -1\end{bmatrix}$$
$$= [\boldsymbol{\beta}^\top, -1]\left\{\frac{1}{n}\mathbb{V}\mathrm{ar}\,[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]\begin{bmatrix}\boldsymbol{\beta}\\ -1\end{bmatrix}\right\}\begin{bmatrix}\boldsymbol{\beta}\\ -1\end{bmatrix}$$
$$\to [\boldsymbol{\beta}^\top, -1]\sqsupset\begin{bmatrix}\boldsymbol{\beta}\\ -1\end{bmatrix} > 0, \quad n\to\infty; \tag{4.50}$$

and, hence, (4.28) is satisfied. Therefore, Corollary 4.7 provides the asymptotic normality of $n^{-1/2}\sum_{i=1}^n \varpi_i$.

The case of $\mathbf{t} = \mathbf{0}$ is trivial. On the other hand, if

$$\mathbf{0} \neq \mathbf{t} \neq c\begin{bmatrix}\boldsymbol{\beta}\\ -1\end{bmatrix}$$

for any real constant $c \neq 0$, then we have a sum of zero mean α-mixing random variables

$$\rho_i := \mathbf{t}^\top\left([\mathbf{Z},\mathbf{Z}\boldsymbol{\beta}]_{i,\bullet}^\top[\boldsymbol{\Theta},\boldsymbol{\varepsilon}]_{i,\bullet} + [\boldsymbol{\Theta},\boldsymbol{\varepsilon}]_{i,\bullet}^\top[\boldsymbol{\Theta},\boldsymbol{\varepsilon}]_{i,\bullet} - \sigma^2\mathbf{I}\right)\begin{bmatrix}\boldsymbol{\beta}\\ -1\end{bmatrix} \tag{4.51}$$
$$= \mathbf{Z}_{i,\bullet}\mathbf{t}_{-(p+1)}\boldsymbol{\Theta}_{i,\bullet}\boldsymbol{\beta} + \mathbf{Z}_{i,\bullet}\boldsymbol{\beta}t_{p+1}\boldsymbol{\Theta}_{i,\bullet}\boldsymbol{\beta} - \mathbf{Z}_{i,\bullet}\mathbf{t}_{-(p+1)}\varepsilon_i - \mathbf{Z}_{i,\bullet}\boldsymbol{\beta}t_{p+1}\varepsilon_i$$
$$+ \boldsymbol{\Theta}_{i,\bullet}\mathbf{t}_{-(p+1)}\boldsymbol{\Theta}_{i,\bullet}\boldsymbol{\beta} + t_{p+1}\varepsilon_i\boldsymbol{\Theta}_{i,\bullet}\boldsymbol{\beta} - \boldsymbol{\Theta}_{i,\bullet}\mathbf{t}_{-(p+1)}\varepsilon_i - t_{p+1}\varepsilon_i^2 - \sigma^2\mathbf{t}^\top\begin{bmatrix}\boldsymbol{\beta}\\ -1\end{bmatrix},$$

which again satisfies all the assumptions of Corollary 4.7. In fact, assumption (4.26) holds for $\omega = 1/2\min_{j=1,\ldots,p+1}\omega_j$ realizing Lemma 3.5, (4.39), (4.40), and (4.44) together with the Cauchy-Schwarz inequality:

$$\sup_{n\in\mathbb{N}}\mathbb{E}|\rho_n|^{2+\omega} \leq 9^{1+\omega}\sup_{n\in\mathbb{N}}\mathbb{E}\Big\{|\mathbf{Z}_{n,\bullet}\mathbf{t}_{-(p+1)}\boldsymbol{\Theta}_{n,\bullet}\boldsymbol{\beta}|^{2+\omega} + |\mathbf{Z}_{n,\bullet}\boldsymbol{\beta}t_{p+1}\boldsymbol{\Theta}_{n,\bullet}\boldsymbol{\beta}|^{2+\omega}$$
$$+ |\mathbf{Z}_{n,\bullet}\mathbf{t}_{-(p+1)}\varepsilon_n|^{2+\omega} + |\mathbf{Z}_{n,\bullet}\boldsymbol{\beta}t_{p+1}\varepsilon_n|^{2+\omega} + |\boldsymbol{\Theta}_{n,\bullet}\mathbf{t}_{-(p+1)}\boldsymbol{\Theta}_{n,\bullet}\boldsymbol{\beta}|^{2+\omega}$$
$$+ |t_{p+1}\varepsilon_n\boldsymbol{\Theta}_{n,\bullet}\boldsymbol{\beta}|^{2+\omega} + |\boldsymbol{\Theta}_{n,\bullet}\mathbf{t}_{-(p+1)}\varepsilon_n|^{2+\omega} + |t_{p+1}\varepsilon_n^2|^{2+\omega}$$
$$+ |\sigma^2(\mathbf{t}_{-(p+1)}^\top\boldsymbol{\beta} - t_{p+1})|^{2+\omega}\Big\}$$
$$\leq 9^{1+\omega}\Big\{\sup_{n\in\mathbb{N}}|\mathbf{Z}_{n,\bullet}\mathbf{t}_{-(p+1)}|^{2+\omega}\sup_{n\in\mathbb{N}}\mathbb{E}|\boldsymbol{\Theta}_{n,\bullet}\boldsymbol{\beta}|^{2+\omega}$$

$$
\begin{aligned}
&+ \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} \\
&+ \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} \\
&+ \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} \\
&+ \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} \\
&+ \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} \\
&+ |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}\Bigg\} \\
\leq 9^{1+\omega}\Bigg\{ & p^{2+2\omega} \max_{j=1,\dots,p} |t_j|^{2+\omega} \sum_{j=1}^p \sup_{n\in\mathbb{N}} |Z_{n,j}|^{2+\omega} \max_{j=1,\dots,p} |\beta_j|^{2+\omega} \sum_{j=1}^p \sup_{n\in\mathbb{N}} \mathbb{E}|\Theta_{n,j}|^{2+\omega} \\
&+ p^{2+2\omega}|t_{p+1}| \max_{j=1,\dots,p} |\beta_j|^{2+\omega} \sum_{j=1}^p \sup_{n\in\mathbb{N}} |Z_{n,j}|^{2+\omega} \max_{j=1,\dots,p} |\beta_j|^{2+\omega} \sum_{j=1}^p \sup_{n\in\mathbb{N}} \mathbb{E}|\Theta_{n,j}|^{2+\omega} \\
&+ p^{1+\omega} \max_{j=1,\dots,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} \\
&+ p^{1+\omega}|t_{p+1}| \max_{j=1,\dots,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} \\
&+ p^{3+2\omega} \max_{j=1,\dots,p} |t_j|^{2+\omega} \max_{j=1,\dots,p} |\beta_j|^{2+\omega} \sum_{j=1}^p \sup_{n\in\mathbb{N}} \mathbb{E}|\Theta_{n,j}|^{4+2\omega} \\
&+ p^{3/2+\omega}|t_{p+1}|^{2+\omega} \max_{j=1,\dots,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} \\
&+ p^{3/2+\omega} \max_{j=1,\dots,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} \\
&+ |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}\Bigg\} < \infty,
\end{aligned}
$$

because assumption (4.38) and $\sup_{n\in\mathbb{N}} |Z_{n,j}|^{2+\omega} < \infty$, $\omega > 0$, $j \in \{1,\dots,p+1\}$ are equivalent.

Lemma 4.3(i) provides

$$
\alpha(\rho_\circ, n) \leq \alpha(\varepsilon_\circ, n) + \sum_{j=1}^p \alpha(\boldsymbol{\Theta}_{\circ,j}, n).
$$

Consequently, assumption (4.27) holds due to the concavity of function $u \mapsto u^{\omega/(2+\omega)}$, $\omega > 0$ in $u \geq 0$:

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

and due to (4.48) and (4.49).

Let us calculate

$$\begin{aligned}
\frac{1}{n}\mathbb{E}\left(\sum_{i=1}^{n} \rho_i\right)^2 &= \frac{1}{n}\mathbb{E}\left\{\mathbf{t}^\top \left([\mathbf{Z}, \mathbf{Z}\boldsymbol{\beta}]^\top [\boldsymbol{\Theta}, \boldsymbol{\varepsilon}] + [\boldsymbol{\Theta}, \boldsymbol{\varepsilon}]^\top [\boldsymbol{\Theta}, \boldsymbol{\varepsilon}] - n\sigma^2 \mathbf{I}\right) \begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix}\right\}^2 \\
&= \frac{1}{n}\mathbb{E}\mathbf{t}^\top \left([\mathbf{X}, \mathbf{Y}]^\top [\mathbf{X}, \mathbf{Y}] - \mathbb{E}[\mathbf{X}, \mathbf{Y}]^\top [\mathbf{X}, \mathbf{Y}]\right) \begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix} \\
&\quad [\boldsymbol{\beta}^\top, -1] \left([\mathbf{X}, \mathbf{Y}]^\top [\mathbf{X}, \mathbf{Y}] - \mathbb{E}[\mathbf{X}, \mathbf{Y}]^\top [\mathbf{X}, \mathbf{Y}]\right) \mathbf{t} \\
&= \mathbf{t}^\top \left\{\frac{1}{n}\mathbb{V}\mathrm{ar}\, [\mathbf{X}, \mathbf{Y}]^\top [\mathbf{X}, \mathbf{Y}] \begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix}\right\} \mathbf{t} \\
&\to \mathbf{t}^\top \beth \mathbf{t} > 0, \quad n \to \infty;
\end{aligned} \tag{4.52}$$

and assumption (4.28) is satisfied as well. Thus, Corollary 4.7 implies that $n^{-1/2} \sum_{i=1}^{n} \rho_i$ has asymptotically zero mean normal distribution. □

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

Theorem 4.11 (Asymptotic normality in EIV with φ-mixing). *Let the EIV model hold and assumption* (D) *be satisfied. Suppose*

$$\{\Theta_{n,1}\}_{n=1}^{\infty}, \ldots, \{\Theta_{n,p}\}_{n=1}^{\infty}, \quad \textit{and} \quad \{\varepsilon_n\}_{n=1}^{\infty} \tag{4.53}$$

are pairwise independent sequences of φ-mixing random variables such that

$$\sum_{n=1}^{\infty} \sqrt{\varphi(\Theta_{\circ,j}, n)} < \infty, \quad j \in \{1, \ldots, p\} \tag{4.54}$$

and

$$\sum_{n=1}^{\infty} \sqrt{\varphi(\varepsilon_\circ, n)} < \infty. \tag{4.55}$$

Moreover, assume that

$$\sup_{n\in\mathbb{N}} Z_{n,j}^2 < \infty, \quad j \in \{1,\ldots,p\}, \tag{4.56}$$

$$\sup_{n\in\mathbb{N}} \mathbb{E}|\Theta_{n,j}|^{4+\omega_j} < \infty, \quad j \in \{1,\ldots,p\}, \tag{4.57}$$

and

$$\sup_{n\in\mathbb{N}} \mathbb{E}|\varepsilon_n|^{4+\omega_{p+1}} < \infty \tag{4.58}$$

for some $\omega_j > 0$*,* $j \in \{1,\ldots,p+1\}$*. If there exists a positive definite matrix* $\daleth$ *such that*

$$n^{-1}\mathbb{V}ar[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}] \begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix} \to \daleth > \mathbf{0}, \quad n \to \infty; \tag{4.59}$$

then

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{\mathscr{D}} \mathcal{N}(\mathbf{0}, \cdot), \quad n \to \infty. \tag{4.60}$$

Proof. This proof contains very similar ideas as the proof of Theorem 4.10.

Assumptions of Theorem 4.11 imply the assumptions of Theorem 4.5. Indeed, assumptions of pairwise independence (4.17) and (4.53) coincide. Similarly for assumptions (4.18)–(4.19) and (4.54)–(4.55). Assumptions (4.20)–(4.21) follow directly from (4.57)–(4.58), because (4.44), (4.57), and (4.58) yield

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

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

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

Therefore, 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 4.10, it is necessary to find the limiting distribution of $\{\varpi_n\}_{n=1}^\infty$ defined in (4.47) and $\{\rho_n\}_{n=1}^\infty$ from (4.51) for

$$\mathbf{0} \ne \mathbf{t} \ne c\begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix}, \quad c \ne 0.$$

In light of Corollary 4.9, we only need to check whether sequences $\{\varpi_n\}_{n=1}^\infty$ and $\{\rho_n\}_{n=1}^\infty$ are

φ-mixing sequences, i.e., $\varphi(\varpi_\circ, n) \to 0$ and $\varphi(\rho_\circ, n) \to 0$ as $n \to \infty$. This follows directly from Lemma 4.3(ii) and assumptions $\varphi(\Theta_{\circ,j}, n) \to 0$ for $j = 1, \dots, p$ and $\varphi(\varepsilon_\circ, n) \to 0$ as $n \to \infty$. The rest of the assumptions of Corollary 4.9 is included in the assumptions of Corollary 4.7 and has been completely checked on sequences $\{\varpi_n\}_{n=1}^\infty$ and $\{\rho_n\}_{n=1}^\infty$ in the proof of Theorem 4.10. □

4.3.1 Discussion

Assuming that a sequence of random variables is φ-mixing implies that this sequence is α-mixing. On the other hand, the central limit theorem for φ-mixing (Corollary 4.9) has weaker assumptions than the central limit theorem for α-mixing (Corollary 4.7). Indeed, Corollary 4.9 does not require any assumption on *mixing rate* $\varphi(n)$ such as assumption (4.27) on α-mixing rates. Therefore in Theorem 4.11, we do not have to deal with mixing rate assumptions like (4.36)–(4.37) nor a restriction on the *moment order* like (4.41) as in Theorem 4.10. On the other hand, Theorem 4.9 requires mixing rate assumptions (4.54)–(4.55), which are inherited from the assumptions for the strong consistency of the EIV estimate. Both asymptotic normality results of the EIV estimate (Theorem 4.10 and Theorem 4.11) require all the assumptions from the strong mixing results (Theorem 4.4 and Theorem 4.5), because these convergences almost surely were used in the proofs of the asymptotic normality.

Assumptions (4.42) and (4.59) concerning the long-run variance of the EIV estimate are requisite and cannot be omitted, because they assure that the variance of the EIV estimate is bounded away from zero and, simultaneously, does not explode into infinity. These assumptions straightforwardly allow to apply the appropriate CLT in order to prove the asymptotic normality of $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$.

Assumptions and remarks regarding the error structure, the pairwise independence, or the strict stationarity of errors have been already discussed in Subsection 4.2.1.

Zero mean errors are implicitly assumed through the whole chapter and are not explicitly specified in every theorem, because this assumption is a part of the EIV model's definition valid for the whole thesis.

4.4 Conclusions

An error structure of the EIV model with *weakly dependent errors* is introduced in this chapter. Strong laws of large numbers for strong mixing and uniformly strong mixing are summarized. They allow us to derive and prove a *strong consistency* of the EIV estimate under both forms of errors' *asymptotic independence*. Furthermore, any form of stationarity does not have to be imposed on the errors. In these settings, the strong consistency of the nuisance variance parameter is proved as well.

Secondly, suitable central limit theorems for strong mixing and uniformly strong mixing are postulated. Consequently, an asymptotic normality of $\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})$ under α- and φ-mixing (generally non-stationary) errors of the EIV model is shown. Similar situation as in the independent errors' case would occur with the calculation of variance of the EIV estimate. Since the variance is already computationally useless when the errors are independent, another approach, that provides the approximate (estimated) asymptotic variance of the EIV estimate, needs to be proposed.

Chapter 5

Block Bootstrap

If you want to inspire confidence, give plenty of statistics. It does not matter that they should be accurate, or even intelligible, as long as there is enough of them.

LEWIS CARROLL

The asymptotic variance of the EIV estimate is very complicated and, moreover, computationally useless, when the errors in the EIV model are independent. This serious issue was overcome by the nonparametric bootstrapping.

Intuitively, a similar problem will arise when considering weakly dependent errors of the EIV model, because the independence is a special case of every type of the weak dependence. Therefore, it would be a waste of effort to try calculating the asymptotic variance of the EIV estimate, when the errors are weakly dependent.

A generalization of the nonparametric bootstrap (case sampling with replacement)—*block bootstrap*—is used. Instead of sampling individual cases, the *blocks of adjacent observations* are resampled with replacement. Stacking individual adjacent cases together into one solid block partly preserves the dependence between consecutive observations. Since the weak dependence can be seen as an asymptotic independence, the blocks can be resampled independently. It is a way how to achieve that the dependence between faraway observations is vanishing.

5.1 Moving Block Bootstrap

Various types of the block bootstrap procedures were proposed. Lahiri (2003) provided a comprehensive summary. The main difference between the block bootstrap types is the way of drawing the blocks of observations. *Non-overlapping block bootstrap* refers to resampling blocks, which do not overlap. This approach is less efficient for estimation, because some observations are not allowed to be joined into the same block. We consider *moving block bootstrap* (MBB), where a consecutive block is formed from the previous one via shifting the "stacking window" by one observation ahead. The exact algorithm of the MBB for a sample mean of univariate observations is precisely described in Procedure 5.1.

Procedure 5.1 Moving block bootstrap for the sample mean.

Input: Data consisting of n observations ξ_i and $n = mb$.

Output: Empirical bootstrap distribution of sample mean $\bar{\xi} := n^{-1}\sum_{i=1}^{n}\xi_i$, i.e., the empirical distribution where probability mass $1/D$ concentrates at each of ${}_{(1)}\bar{\xi}^*, \ldots, {}_{(D)}\bar{\xi}^*$.

define B_j as the block of b consecutive ξ_i's starting from ξ_j, that is $B_j = [\xi_j, \ldots, \xi_{j+b-1}]^\top$ for $j = 1, \ldots, q$, where $q := n - b + 1$

for $d = 1$ to D **do** // repeat in order to obtain empirical distribution of $\bar{\xi}$

  resample with replacement ${}_{(d)}C_1, \ldots, {}_{(d)}C_m$ independently from $\{B_1, \ldots, B_q\}$ with equal probability $1/q$, where each ${}_{(d)}C_i$, $i = 1, \ldots, m$, is a block of size b with ${}_{(d)}C_i = [{}_{(d)}c_{i1}, \ldots, {}_{(d)}c_{ib}]^\top$ // Let $\mathbb{P}^*$ be the (bootstrap) distribution of ${}_{(d)}C_i$ conditional on the sample $\{\xi_1, \ldots, \xi_n\}$. So, given $\xi_1, \ldots, \xi_n$, the m random blocks, ${}_{(d)}C_1, \ldots, {}_{(d)}C_m$, are *iid* distributed according to $\mathbb{P}^*$.

  the MBB resample of size n, denoted by ${}_{(d)}\xi_1^*, \ldots, {}_{(d)}\xi_n^*$, is formed by joining the ${}_{(d)}C_1, \ldots, {}_{(d)}C_m$ to one big block, i.e., ${}_{(d)}\xi_i^* = {}_{(d)}c_{\tau\nu}$ for $\tau = [(i-1)/b] + 1$, $\nu = i - b\tau$, and $i = 1, \ldots, n$ // ${}_{(d)}\boldsymbol{\xi}^* \equiv [{}_{(d)}\xi_1^*, \ldots, {}_{(d)}\xi_n^*]^\top$ is called the MBB version of $\boldsymbol{\xi} \equiv [\xi_1, \ldots, \xi_n]^\top$.

  let the resample average be ${}_{(d)}\bar{\xi}^* \leftarrow n^{-1}\sum_{i=1}^{n} {}_{(d)}\xi_i^*$

end for

The length of the blocks—*blocksize*—is denoted by $b \in \mathbb{N}$. Without loss of generality to the asymptotical properties of the EIV estimate, let us suppose that $b \mid n$, i.e., there exist $m \in \mathbb{N}$ such that $n = mb$. In other words, we just neglect an integer division problem. For practical and computational purposes, if $b \nmid n$, then we can truncate the quotient n/b to an integer value.

An extension of the MBB is a *circular block bootstrap*, where the observations are not ordered on a single line, but they are put into a circle. The order of the observations is preserved with the only exception that the last observation on the circle is followed by the first one. Hence, the stacking window can join the first and the last observations into one block. The application of the circular block bootstrap as an extension of the MBB is postponed for some further work and is not considered in this thesis.

Until this moment, we have considered the length of the blocks as a constant, but the whole idea of block bootstrapping can be generalized for a varying size of blocks. This extension is not taken into account in this thesis due to its huge complexity, which would bring us far beyond the borders of problems that are dealt here.

The MBB procedure was independently suggested by Künsch (1989) and Liu and Singh (1992) for the case of sample mean. Lahiri (1992) and Politis and Romano (1992) extended these results, but still considered only strictly stationary processes. Fitzenberger (1997) generalized previous approaches for non-stationary processes and applied the MBB approach for the linear regression setup.

5.2 Asymptotic Properties for Moving Block Bootstrap

Asymptotic properties for the MBB need to be postulated, which will be used for proving a correctness of the MBB procedure for the EIV estimate.

5.2.1 Law of Large Numbers for Moving Block Bootstrap

First of all, the *weak law of large numbers* for the bootstrapped sampled mean from a generally non-stationary α-mixing is shown.

Lemma 5.1 (Bootstrap weak law of large numbers for α-mixing). *Let $\{\xi_n\}_{n=1}^{\infty}$ be a sequence of zero mean α-mixing random variables satisfying*

$$\sup_{n\in\mathbb{N}} \mathbb{E}\xi_n^2 < \infty. \tag{5.1}$$

Assume that there exists $\delta > 0$ such that

$$\alpha(n) = \mathcal{O}\left(n^{-1-\delta}\right), \quad n \to \infty. \tag{5.2}$$

If $b \to \infty$ and $b = o(n^{1/2})$, $n \to \infty$, then under MBB Procedure 5.1

$$n^{-1}\sum_{i=1}^{n}\xi_i^* - n^{-1}\sum_{i=1}^{n}\xi_i \xrightarrow[n\to\infty]{\mathbb{P}^*_{\boldsymbol{\xi}}(\mathbb{P})} 0,$$

where $\boldsymbol{\xi}^* \equiv [\xi_1^*, \ldots, \xi_n^*]^\top$ is the MBB version of $\boldsymbol{\xi} \equiv [\xi_1, \ldots, \xi_n]^\top$.

Proof. The idea of this proof is borrowed from Fitzenberger (1997, Lemma A.3). Sequence $\{\xi_n\}_{n=1}^{\infty}$ is uniformly bounded in probability $\mathbb{P}$, because (5.1) is assumed. By the assumptions for $\{\xi_n\}_{n=1}^{\infty}$ and Lemma 4.1, it is known that a SLLN holds for the sample mean $n^{-1}\sum_{i=1}^{n}\xi_i$. By Corollary A.2 from Fitzenberger (1997), it follows that

$$\mathbb{V}\text{ar}_{\mathbb{P}^*_{\boldsymbol{\xi}}}\left(n^{-1}\sum_{i=1}^{n}\xi_i^*\right) = \mathcal{O}_{\mathbb{P}}(n^{-1}), \quad n \to \infty.$$

Thus the claim holds due to Lemma A.1 by Fitzenberger (1997), where

$$\mathbb{E}_{\mathbb{P}^*_{\boldsymbol{\xi}}} n^{-1} \sum_{i=1}^{n} \xi_i^* = n^{-1} \sum_{i=1}^{n} \xi_i + o_{\mathbb{P}}(n^{-1/2}), \quad n \to \infty.$$

□

Similarly as above, the weak law of large numbers for the bootstrapped sample mean is recalled, but a φ-mixing sequence is considered this time.

Lemma 5.2 (Bootstrap weak law of large numbers for φ-mixing)**.** *Let $\{\xi_n\}_{n=1}^{\infty}$ be a sequence of zero mean φ-mixing random variables satisfying*

$$\sum_{n=1}^{\infty} \frac{\mathbb{E}\xi_n^2}{n^2} < \infty$$

and

$$\sum_{n=1}^{\infty} \sqrt{\varphi(n)} < \infty. \tag{5.3}$$

If $b \to \infty$ and $b = o(n^{1/2})$, $n \to \infty$, then under MBB Procedure 5.1

$$n^{-1} \sum_{i=1}^{n} \xi_i^* - n^{-1} \sum_{i=1}^{n} \xi_i \xrightarrow[n\to\infty]{\mathbb{P}^*_{\boldsymbol{\xi}}(\mathbb{P})} 0,$$

where $\boldsymbol{\xi}^* \equiv [\xi_1^*, \ldots, \xi_n^*]^\top$ is the MBB version of $\boldsymbol{\xi} \equiv [\xi_1, \ldots, \xi_n]^\top$.

Proof. The proof is the same as the proof of previous Lemma 5.1 except one detail—the SLLN for φ-mixing (Lemma 4.2) has to be applied instead of Lemma 4.1 for α-mixing. □

5.2.2 Central Limit Theorem for Moving Block Bootstrap

Central limit theorems for the bootstrapped sample mean from non-stationary strong mixing or uniformly strong mixing sequences are stated.

Theorem 5.3 (Bootstrap central limit theorem for α-mixing)**.** *Let $\{\xi_n\}_{n=1}^{\infty}$ be a sequence of zero mean α-mixing random variables with*

$$\sup_{n\in\mathbb{N}} \mathbb{E}|\xi_n|^{4+\omega} < \infty \tag{5.4}$$

and

$$\alpha(n) = \mathcal{O}(n^{-1-\delta}), \quad n \to \infty \tag{5.5}$$

for some $\omega > 0$ *and* $\delta > 0$ *such that*

$$\frac{4}{\omega} < \delta. \tag{5.6}$$

Denote

$$\bar{\xi}_n := n^{-1}\sum_{i=1}^{n}\xi_i \quad \text{and} \quad \bar{\xi}_n^* := n^{-1}\sum_{i=1}^{n}\xi_i^*.$$

Suppose that

$$\liminf_{n\to\infty}\frac{\mathbb{E}S_n^2}{n} = \varsigma^2 > 0 \tag{5.7}$$

is satisfied. If $b \to \infty$ *and* $b = o(n^{1/2})$, $n \to \infty$, *then under MBB Procedure 5.1*

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}_{\boldsymbol{\xi}}^*\left(\frac{n}{\sqrt{\mathbb{E}S_n^2}}\left(\bar{\xi}_n^* - \bar{\xi}_n\right) \leq x\right) - \mathbb{P}\left(\frac{n}{\sqrt{\mathbb{E}S_n^2}}\bar{\xi}_n \leq x\right)\right| \xrightarrow{\mathbb{P}} 0, \quad n \to \infty, \tag{5.8}$$

where $\boldsymbol{\xi}^* \equiv [\xi_1^*, \ldots, \xi_n^*]^\top$ *is the MBB version of* $\boldsymbol{\xi} \equiv [\xi_1, \ldots, \xi_n]^\top$.

Proof. See Fitzenberger (1997, Theorem 3.1). □

Similar theorem for *independent and identically distributed* random variables was proved by Singh (1981). Consequently, Politis and Romano (1992) proved a version of this theorem for *strong stationary* α-mixing sequences.

Theorem 5.4 (Bootstrap central limit theorem for φ-mixing). *Let* $\{\xi_n\}_{n=1}^\infty$ *be a sequence of zero mean* φ*-mixing random variables with*

$$\sup_{n\in\mathbb{N}}\mathbb{E}|\xi_n|^4 < \infty \tag{5.9}$$

and

$$\varphi(n) = \mathcal{O}(n^{-2-\delta}), \quad n \to \infty \tag{5.10}$$

for some $\delta > 0$. *Denote*

$$\bar{\xi}_n := n^{-1}\sum_{i=1}^{n}\xi_i \quad \text{and} \quad \bar{\xi}_n^* := n^{-1}\sum_{i=1}^{n}\xi_i^*.$$

Suppose that

$$\liminf_{n\to\infty}\frac{\mathbb{E}S_n^2}{n} = \varsigma^2 > 0 \tag{5.11}$$

is satisfied. If $b \to \infty$ and $b = o(n^{1/2})$, $n \to \infty$, then under MBB Procedure 5.1

$$\sup_{x \in \mathbb{R}} \left| \mathbb{P}^*_{\boldsymbol{\xi}} \left(\frac{n}{\sqrt{\mathbb{E}S_n^2}} \left(\bar{\xi}_n^* - \bar{\xi}_n \right) \leq x \right) - \mathbb{P} \left(\frac{n}{\sqrt{\mathbb{E}S_n^2}} \bar{\xi}_n \leq x \right) \right| \xrightarrow{\mathbb{P}} 0, \quad n \to \infty, \tag{5.12}$$

where $\boldsymbol{\xi}^* \equiv [\xi_1^*, \ldots, \xi_n^*]^\top$ is the MBB version of $\boldsymbol{\xi} \equiv [\xi_1, \ldots, \xi_n]^\top$.

Proof. The proof of Theorem 5.4 remains almost the same as the proof of Theorem 3.1 by Fitzenberger (1997) except three facts: we use the CLT for φ-mixing observations, i.e., Corollary 4.9, instead of Theorem 5.3 by Gallant and White (1988); another covariance inequality, i.e., Lemma 1.2.8 by Lin and Lu (1997) instead of Theorem 3 by Doukhan (1994); and a suitable bootstrap WLLN, i.e., Lemma 5.2 instead of Lemma A.3 by Fitzenberger (1997). The proof of Theorem 3.1 provided by Fitzenberger (1997) is very long and nothing has to be changed in it except the usage of the CLT and the covariance inequality for the φ-mixing sequences. Therefore, we omit such a "copy-paste" exercise. □

Alternatively, a variation of Theorem 5.4 can be proved as a corollary of Theorem 5.3 realizing that for the mixing coefficients holds $\alpha(n) \leq \varphi(n)$, but in that case, additional assumption (5.6) needs to be assumed and the equiboundedness of higher moments (compare (5.4) against (5.9)) would be required. The property of being φ-mixing is more restrictive than being α-mixing. Moreover, the mixing rate assumption (5.10) for φ-mixing is stronger than (5.5) for α-mixing. On the other hand, the CLT for φ-mixing (Corollary 4.9) does not bind the mixing rate assumption and the moment order equiboundedness by (5.6) as the CLT for α-mixing (Corollary 4.7 or Theorem 5.3 by Gallant and White (1988)). Likewise, the moments' equiboundedness (5.9) for φ-mixing is not so restrictive than (5.4) for α-mixing. The reason for this is that the covariance inequality for φ-mixing only requires the uniformly bounded second moments. In the case of α-mixing, the covariance inequality needs uniformly bounded higher moments than the second ones, see Lin and Lu (1997, Section 1.2).

5.3 Asymptotic Validity for MBB in the EIV model

We propose moving block bootstrap Procedure 5.2 for the EIV estimate when the errors of the EIV model are α- or φ-mixing. Procedure 5.2 is an extension of the nonparametric bootstrap Procedure 3.1. Due to the dependence in the data—or better to say, in the error structure—residual type of bootstrap is not suitable for regression models (Fitzenberger, 1997).

A justification for the asymptotic validity of MBB Procedure 5.2 for the EIV estimate with weakly dependent errors will be provided. We prove that $\sqrt{n}(\widetilde{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})$ conditionally on the data converges to the asymptotic distribution of $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. The asymptotic results

Procedure 5.2 Moving block bootstrap for the EIV estimate.

Input: Data consisting of n row vectors of observations $[\mathbf{X}_{i,\bullet}, Y_i]$ and $n = mb$.

Output: Empirical bootstrap distribution of $\widehat{\boldsymbol{\beta}}$, i.e., the empirical distribution where probability mass $1/D$ concentrates at each of ${}_{(1)}\widetilde{\boldsymbol{\beta}}^*, \ldots, {}_{(D)}\widetilde{\boldsymbol{\beta}}^*$.

calculate the EIV estimate $\widehat{\boldsymbol{\beta}} \leftarrow (\mathbf{X}^\top\mathbf{X} - \lambda\mathbf{I})^{-1}\mathbf{X}^\top\mathbf{Y}$

define blocks of observation vectors $\mathbf{B}_j = [\mathbf{B}_j^x, \mathbf{B}_j^y]$ with $\mathbf{B}_j^y = [Y_j, \ldots, Y_{j+b-1}]^\top$, a $b \times 1$ vector, and $\mathbf{B}_j^x$, a $b \times p$ matrix with rows $\mathbf{X}_{j,\bullet}, \ldots, \mathbf{X}_{j+b-1,\bullet}$ for $j = 1, \ldots, q$, where $q := n - b + 1$

for $d = 1$ to D **do** // repeat in order to obtain empirical distribution of $\widehat{\boldsymbol{\beta}}$

resample with replacement $(b \times [p+1])$ blocks ${}_{(d)}\mathbf{C}_1, \ldots, {}_{(d)}\mathbf{C}_m$ independently from $\{{}_{(d)}\mathbf{B}_1, \ldots, {}_{(d)}\mathbf{B}_q\}$ with equal probability $1/q$, where each ${}_{(d)}\mathbf{C}_i$, $i = 1, \ldots, m$, is a block of size b with ${}_{(d)}\mathbf{C}_i = [{}_{(d)}\mathbf{c}_{i1}^\top, \ldots, {}_{(d)}\mathbf{c}_{ib}^\top]^\top$ // Let $\mathbb{P}^*$ be the (bootstrap) distribution of ${}_{(d)}\mathbf{C}_i$ conditional on the sample $\{[\mathbf{X}_{1,\bullet}, Y_1], \ldots, [\mathbf{X}_{n,\bullet}, Y_n]\}$. So, given $[\mathbf{X}_{1,\bullet}, Y_1], \ldots, [\mathbf{X}_{n,\bullet}, Y_n]$, the m random blocks, ${}_{(d)}\mathbf{C}_1, \ldots, {}_{(d)}\mathbf{C}_m$, are *iid* distributed according to $\mathbb{P}^*$.

the MBB resample of size n, denoted by $\{[{}_{(d)}\mathbf{X}_{1,\bullet}^*, {}_{(d)}Y_1^*], \ldots, [{}_{(d)}\mathbf{X}_{n,\bullet}^*, {}_{(d)}Y_n^*]\}$, is formed by joining (stacking) the resampled blocks ${}_{(d)}\mathbf{C}_1, \ldots, {}_{(d)}\mathbf{C}_m$ to one big block, i.e., $[{}_{(d)}\mathbf{X}_{i,\bullet}^*, {}_{(d)}Y_i^*] = {}_{(d)}\mathbf{c}_{\tau\nu}$ for $\tau = [(i-1)/b] + 1$, $\nu = i - b\tau$, and $i = 1, \ldots, n$

${}_{(d)}\lambda^*$ is the $(p+1)$-st eigenvalue of $[{}_{(d)}\mathbf{X}^*, {}_{(d)}\mathbf{Y}^*]^\top[{}_{(d)}\mathbf{X}^*, {}_{(d)}\mathbf{Y}^*]$

re-estimate ${}_{(d)}\widehat{\boldsymbol{\beta}}^* \leftarrow \left({}_{(d)}\mathbf{X}^{*\top}{}_{(d)}\mathbf{X}^* - {}_{(d)}\lambda^*\mathbf{I}\right)^{-1}{}_{(d)}\mathbf{X}^{*\top}{}_{(d)}\mathbf{Y}^*$

put

$$
\begin{aligned}
{}_{(d)}\widetilde{\boldsymbol{\beta}}^* \leftarrow & \widehat{\boldsymbol{\beta}} - \left({}_{(d)}\mathbf{X}^{*\top}{}_{(d)}\mathbf{X}^* - {}_{(d)}\lambda^*\mathbf{I}\right)^{-1}\left([\mathbf{I}, \widehat{\boldsymbol{\beta}}]\left[\mathbf{I}, {}_{(d)}\widehat{\boldsymbol{\beta}}^*\right]^\top\right)^{-1}[\mathbf{I}, \widehat{\boldsymbol{\beta}}] \\
& \left([{}_{(d)}\mathbf{X}^*, {}_{(d)}\mathbf{Y}^*]^\top[{}_{(d)}\mathbf{X}^*, {}_{(d)}\mathbf{Y}^*] - [\mathbf{X}, \mathbf{Y}]^\top[\mathbf{X}, \mathbf{Y}]\right)\begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ -1 \end{bmatrix}
\end{aligned}
$$

end for

from Chapter 4 provide crucial and basic steps for proving such a validity. Moreover, the main ideas of the forthcoming proofs will be inherited from the proof of Theorem 3.12.

Additionally, it will also be shown that the MBB procedure is *robust to autocorrelation of unknown type*. Doukhan (1994, Section 2.4) pointed out that all the ARMA processes with continuously distributed stationary innovations and bounded variance are strongly mixing such that $\alpha(n) = \mathcal{O}(n^{-\delta})$ for any $\delta > 0$, since the α-mixing coefficients of innovations are geometrically decreasing. However, Andrews (1984) constructed a stationary AR(1) process with Bernoulli innovations, which is not strongly mixing.

One has to realize that the theory regarding convergences' characterizations in the bootstrap world proved at the beginning of Chapter 3, e.g., Slutsky's extended Theorem 3.2, still holds, because they do not require the independence assumption whatsoever.

Firstly, let us consider α-mixing errors in the EIV model.

Theorem 5.5 (Validity of moving block bootstrap in EIV with α-mixing)**.** *Let the EIV model hold and assumption* (D) *be satisfied. Suppose*

$$\{\Theta_{n,1}\}_{n=1}^{\infty},\ldots,\{\Theta_{n,p}\}_{n=1}^{\infty}, \quad \text{and} \quad \{\varepsilon_n\}_{n=1}^{\infty} \tag{5.13}$$

are pairwise independent sequences of α-mixing random variables having

$$\alpha(\Theta_{\circ,j},n)=\mathcal{O}(n^{-1-\delta_j}), \quad j=1,\ldots,p \tag{5.14}$$

and

$$\alpha(\varepsilon_{\circ},n)=\mathcal{O}(n^{-1-\delta_{p+1}}), \tag{5.15}$$

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

$$\sup_{n\in\mathbb{N}} Z_{n,j}^2<\infty, \tag{5.16}$$

$$\sup_{n\in\mathbb{N}} \mathbb{E}|\Theta_{n,j}|^{8+2\omega_j}<\infty, \tag{5.17}$$

and

$$\sup_{n\in\mathbb{N}} \mathbb{E}|\varepsilon_n|^{8+2\omega_{p+1}}<\infty \tag{5.18}$$

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

$$\frac{4}{\min_{j=1,\ldots,p+1}\omega_j}<\min_{j=1,\ldots,p+1}\delta_j. \tag{5.19}$$

Moreover, assume that there exists a positive definite matrix $\beth$ such that

$$\lim_{n\to\infty}\frac{1}{n}\mathbb{V}ar_{\mathbb{P}}[\mathbf{X},\mathbf{Y}]^{\top}[\mathbf{X},\mathbf{Y}]\begin{bmatrix}\boldsymbol{\beta}\\-1\end{bmatrix}=\beth>\mathbf{0}. \tag{5.20}$$

If $b=o(n^{1/2})$, $n\to\infty$ and $b\to\infty$, then under MBB Procedure 5.2 holds

$$\sqrt{n}(\widetilde{\boldsymbol{\beta}}^*-\widehat{\boldsymbol{\beta}})\big|[\mathbf{X},\mathbf{Y}] \xleftrightarrow[n\to\infty]{\mathscr{D}(\mathbb{P})} \sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}).$$

Proof. The whole process of proving this Theorem 5.5 is totally the same as in Theorem 3.12 for independent variables. We only apply the newly developed asymptotic results for α-mixing and realize the sustainability of the property being α-mixing (see Lemma 4.3). The Cramér-Wold device helps us to derive a multivariate version of the bootstrap CLT from

Theorem 5.3 as in the proof of Theorem 3.11. Moreover, the strong consistency of $\widehat{\boldsymbol{\beta}}$ for α-mixing (Theorem 4.4), the asymptotic normality of $\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})$ for α-mixing (Theorem 4.10), and the WLLN for the MBB in α-mixing case (Lemma 5.1) are used as well.

The only dissimilarity that needs to be showed is a calculation of the conditional expectation

$$\begin{aligned}
&\mathbb{E}_{\mathbb{P}^*}[\mathbf{X}^*,\mathbf{Y}^*]^\top[\mathbf{X}^*,\mathbf{Y}^*]\\
&=\mathbb{E}_{\mathbb{P}^*}\sum_{i=1}^{n/b}[\mathbf{X}^*_{((i-1)b+1):ib,\bullet},\mathbf{Y}^*_{((i-1)b+1):ib}]^\top[\mathbf{X}^*_{((i-1)b+1):ib,\bullet},\mathbf{Y}^*_{((i-1)b+1):ib}]\\
&=\sum_{i=1}^{n/b}\sum_{j=1}^{n-b+1}\frac{1}{n-b+1}[\mathbf{X}_{j:(j+b-1),\bullet},\mathbf{Y}_{j:(j+b-1)}]^\top[\mathbf{X}_{j:(j+b-1),\bullet},\mathbf{Y}_{j:(j+b-1)}] \qquad (5.21)\\
&=\sum_{i=1}^{n/b}\sum_{j=1}^{n-b+1}\frac{1}{n-b+1}\sum_{k=j}^{j+b-1}[\mathbf{X}_{k,\bullet},Y_k]^\top[\mathbf{X}_{k,\bullet},Y_k]\\
&=\frac{n}{b(n-b+1)}\Bigg(b\sum_{i=1}^{n}[\mathbf{X}_{i,\bullet},Y_i]^\top[\mathbf{X}_{i,\bullet},Y_i]-\Big[(b-1)[\mathbf{X}_{1,\bullet},Y_1]^\top[\mathbf{X}_{1,\bullet},Y_1]\\
&\quad+(b-2)[\mathbf{X}_{2,\bullet},Y_2]^\top[\mathbf{X}_{2,\bullet},Y_2]+\ldots+[\mathbf{X}_{b-1,\bullet},Y_{b-1}]^\top[\mathbf{X}_{b-1,\bullet},Y_{b-1}]\\
&\quad+(b-1)[\mathbf{X}_{n,\bullet},Y_n]^\top[\mathbf{X}_{n,\bullet},Y_n]+(b-2)[\mathbf{X}_{n-1,\bullet},Y_{n-1}]^\top[\mathbf{X}_{n-1,\bullet},Y_{n-1}]\\
&\quad+\ldots+[\mathbf{X}_{n-b+2,\bullet},Y_{n-b+2}]^\top[\mathbf{X}_{n-b+2,\bullet},Y_{n-b+2}]\Big]\Bigg)\\
&=\frac{n}{n-b+1}[\mathbf{X},\mathbf{Y}]^\top[\mathbf{X},\mathbf{Y}]+\mathcal{O}_{\mathbb{P}}(b),\quad n\to\infty.
\end{aligned}$$

During the previous calculation, the process of stacking the random blocks—see Step 5 in Algorithm 5.2—needs to be considered. Adjustment (5.21) is the crucial step, where it is necessary to realize that $[\mathbf{X}^*_{((i-1)b+1):ib,\bullet},\mathbf{Y}^*_{((i-1)b+1):ib}]^\top[\mathbf{X}^*_{((i-1)b+1):ib,\bullet},\mathbf{Y}^*_{((i-1)b+1):ib}]$ has a discrete uniform distribution on support

$$\left\{[\mathbf{X}_{j:(j+b-1),\bullet},\mathbf{Y}_{j:(j+b-1)}]^\top[\mathbf{X}_{j:(j+b-1),\bullet},\mathbf{Y}_{j:(j+b-1)}]\right\}_{j=1}^{n-b+1}$$

for all $i=1,\ldots,n/b$, conditionally on $[\mathbf{X},\mathbf{Y}]$.

Under the circumstances ($b=o(n^{1/2})$ as $n\to\infty$), it has to be emphasized that the distributional closeness is considered in probability $\mathbb{P}$ and, in that event, all the negligible terms in probability $\mathbb{P}$ do not cause any harm.

In accordance with the assumptions of previously mentioned theorems and lemmas, we postulate sufficient assumptions for Theorem 5.5 in the way that all the required asymptotic results can be correctly applied. □

In other words, Theorem 5.5 affirms that $\sqrt{n}(\widetilde{\boldsymbol{\beta}}^*-\widehat{\boldsymbol{\beta}})$ under $\mathbb{P}^*$ and $\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})$ under $\mathbb{P}$ have the same limit distribution in the case of α-mixing errors.

A similar theorem to the previous one can be restated, but this time for φ-mixing errors of the EIV model.

Theorem 5.6 (Validity of moving block bootstrap in EIV with φ-mixing)**.** *Let the EIV model hold and assumption* (D) *be satisfied. Suppose*

$$\{\Theta_{n,1}\}_{n=1}^{\infty}, \ldots, \{\Theta_{n,p}\}_{n=1}^{\infty}, \quad \textit{and} \quad \{\varepsilon_n\}_{n=1}^{\infty} \tag{5.22}$$

are pairwise independent sequences of φ-mixing random variables having

$$\varphi(\Theta_{\circ,j}, n) = \mathcal{O}(n^{-2-\delta_j}), \quad j = 1, \ldots, p \tag{5.23}$$

and

$$\alpha(\varepsilon_{\circ}, n) = \mathcal{O}(n^{-2-\delta_{p+1}}), \tag{5.24}$$

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

$$\sup_{n\in\mathbb{N}} Z_{n,j}^2 < \infty, \quad j = 1, \ldots, p, \tag{5.25}$$

$$\sup_{n\in\mathbb{N}} \mathbb{E}\Theta_{n,j}^8 < \infty, \quad j = 1, \ldots, p, \tag{5.26}$$

and

$$\sup_{n\in\mathbb{N}} \mathbb{E}\varepsilon_n^8 < \infty. \tag{5.27}$$

Moreover, assume that there exists a positive definite matrix $\daleth$ such that

$$\lim_{n\to\infty} \frac{1}{n} \mathbb{V}ar_{\mathbb{P}} [\mathbf{X}, \mathbf{Y}]^{\top} [\mathbf{X}, \mathbf{Y}] \begin{bmatrix} \boldsymbol{\beta} \\ -1 \end{bmatrix} = \daleth > \mathbf{0}. \tag{5.28}$$

If $b = o(n^{1/2})$, $n \to \infty$ and $b \to \infty$, then under MBB Procedure 5.2 holds

$$\sqrt{n}(\widetilde{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}) \big| [\mathbf{X}, \mathbf{Y}] \xleftrightarrow[n\to\infty]{\mathscr{D}(\mathbb{P})} \sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}).$$

Proof. See the proof of Theorem 5.5, where the appropriate asymptotic results for φ-mixing—the strong consistency of $\widehat{\boldsymbol{\beta}}$ (Theorem 4.5), the asymptotic normality of $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ (Theorem 4.11), the WLLN for the MBB (Lemma 5.2), and the CLT for the MBB (Theorem 5.4)—are used instead of the asymptotic properties for α-mixing. □

In other words again, Theorem 5.6 asserts that $\sqrt{n}(\widetilde{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})$ under $\mathbb{P}^*$ and $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ under $\mathbb{P}$ have the same limit distribution in the case of φ-mixing errors.

5.3.1 Choice of Blocksize

A choice of blocksize b in the moving block bootstrap procedure is an important decision. Indeed, it will affect the bootstrapped EIV estimate and the consequent empirical inference. Therefore, the blocksize value can be viewed as a *tuning parameter* in the MBB procedure.

From the previous theory, it is already known that $b = o(n^{1/2})$ as n tends to infinity. This result is slightly cumbersome, especially for practical computational purposes. It would be preferable to have at least the *Landau big $\mathcal{O}$ relation* with respect to the sample size in order to have a more "precise" choice of blocksize b for the MBB. One possibility, how to proceed such an optimality choice, is to minimize the asymptotic *mean square error* (MSE) of the MBB variance estimate. Taking the sample mean and its MBB procedure into account, Fitzenberger (1997, Theorem 3.4) proved that $b = \mathcal{O}(n^{1/3})$ as $n \to \infty$ by imposing quite strong and complicated assumptions. Since the elements of the EIV estimate are asymptotically equal to a specific mean ($\widehat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X} - \lambda\mathbf{I})^{-1}\mathbf{X}^\top\mathbf{Y}$), it may be concluded that the choice $b = \mathcal{O}(n^{1/3})$, $n \to \infty$ could be asymptotically optimal with respect to the variance's MSE for the EIV estimate as well.

Despite of this approximate choice of blocksize, we still do not know how to precisely choose b according to n. Therefore, a simulation study could enlighten us.

5.4 Simulation Study

A continuation of the simulation study from Section 3.4 was proceeded, but in this case for the moving block bootstrap procedure of the EIV estimate. The interest still lies in a construction of the confidence intervals and their *coverage level*.

The basic (design) setup from the simulations for the nonparametric bootstrap technique was mostly preserved, i.e., 5000 random samples were generated from a one-dimensional EIV model with the design points $\left\{(-1)^i\sqrt{1-1/i}\right\}_{i=1}^{n}$. We changed the design points in order to demonstrate the impact of non-monotonically ordered covariates as discussed in Section 3.4.

Actually, if we want to show a performance of the MBB technique for weak dependence, longer sequences of observations are needed to be generated in order to capture and demonstrate the effect, because any form of weak dependence generally features an overlap of information brought by the adjacent observations.

Percentile bootstrap 95%-confidence intervals were considered and $D = 5000$ block bootstrap replications were conducted.

The average lengths of confidence intervals based upon the MBB percentile method were computed for various setups of the weakly dependent errors from the EIV model.

Six errors' models (processes) for $\{\Theta_i\}_{i=1}^n$ and $\{\varepsilon_i\}_{i=1}^n$ were considered:

$$\xi_t - 0.2\xi_{t-1} = \zeta_t + 0.3\zeta_{t-1}, \quad \zeta_t \overset{iid}{\sim} \mathcal{N}(0,\sigma^2),\ \forall t \in \mathbb{Z}, \tag{5.29}$$

$$\xi_t = 0.2\xi_{t-1} + \zeta_t, \quad \zeta_t \overset{iid}{\sim} \mathcal{N}(0,\sigma^2),\ \forall t \in \mathbb{Z}, \tag{5.30}$$

$$\xi_t = \zeta_t + 0.3\zeta_{t-1}, \quad \zeta_t \overset{iid}{\sim} \mathcal{N}(0,\sigma^2),\ \forall t \in \mathbb{Z}, \tag{5.31}$$

$$\xi_t - 0.2\xi_{t-1} = \zeta_t + 0.3\zeta_{t-1}, \quad \zeta_t \overset{iid}{\sim} \sigma\sqrt{\frac{6.2}{8.2}} \times t_{8.2},\ \forall t \in \mathbb{Z}, \tag{5.32}$$

$$\xi_t = 0.2\xi_{t-1} + \zeta_t, \quad \zeta_t \overset{iid}{\sim} \sigma\sqrt{\frac{6.2}{8.2}} \times t_{8.2},\ \forall t \in \mathbb{Z}, \tag{5.33}$$

$$\xi_t = \zeta_t + 0.3\zeta_{t-1}, \quad \zeta_t \overset{iid}{\sim} \sigma\sqrt{\frac{6.2}{8.2}} \times t_{8.2},\ \forall t \in \mathbb{Z}; \tag{5.34}$$

where $\xi_t \in \{\Theta_t, \varepsilon_t\}$, $\forall t \in \mathbb{Z}$. Sequences of errors $\{\Theta_i\}_{i=1}^n$ and $\{\varepsilon_i\}_{i=1}^n$ are generated independently. The length of the "burn-in" period for the recursively generated sequences is set to 300, i.e., we throw the first 300 generated elements of the recursively defined sequences away. As it may be noticed from (5.29)–(5.34), we took two ARMA(1,1), two AR(1), and two MA(1) errors' models, where the innovations are *iid* having normal distribution or scaled Student t-distribution, respectively. Four different types of sequences of the errors' pairs for each errors' model were generated in the way that two different values of the standard deviation parameter σ for innovations were taken into account ($\sigma = 10^{-2}$ or $\sigma = 10^{-3}$) and combined with two different numbers of generated errors' pairs n ($n \approx 60$ or $n \approx 180$). The "unknown" (true) EIV parameter β is set to one.

It should be recalled that the number of degrees of freedom for the Student t-distribution has to be greater than eight in order to have finite required moments.

The simulation results for our six setups are summarized in Tables 5.1–5.6 for various blocksizes ($b = 3, \ldots, 11$).

After comparing the empirical coverages of the CIs, there is no striking difference in the CIs' empirical coverages between various types of the models for weakly dependent errors (ARMA, AR, and MA models with normally and t-distributed innovations) neither the choice of the blocksize, when the blocksize is roughly equivalent to the third root of the sample size. The most negligible differences in the empirical coverages are when changing the value of the standard deviation parameter σ for innovations from 10^{-2} to 10^{-3}, while blocksize b and number of generated errors' pairs n are kept fixed.

Let us go deeper into the details and compare the empirical coverages of the confidence intervals for two ARMA(1,1) errors' models (5.29) and (5.32) more precisely. In the case of normally distributed innovations, the most appropriate choice of blocksize b is 7 for $n \approx 60$ and 11 for $n \approx 180$, respectively. Similarly, if the innovations of ARMA(1,1) have Student t-distribution, the most appropriate choice of blocksize b is 7 for $n \approx 60$ and 10 for $n \approx 180$, respectively. When considering AR(1) errors' models (5.30) and (5.33), the optimal choice

Block size	Sample size	Standard deviation	Empirical CI coverage Moving block bootstrap (proper)
$b = 3$	$n = 60$	$\sigma = 10^{-2}$	98.46%
		$\sigma = 10^{-3}$	98.82%
	$n = 180$	$\sigma = 10^{-2}$	99.06%
		$\sigma = 10^{-3}$	98.84%
$b = 4$	$n = 60$	$\sigma = 10^{-2}$	97.30%
		$\sigma = 10^{-3}$	97.28%
	$n = 180$	$\sigma = 10^{-2}$	98.42%
		$\sigma = 10^{-3}$	98.10%
$b = 5$	$n = 60$	$\sigma = 10^{-2}$	96.26%
		$\sigma = 10^{-3}$	96.86%
	$n = 180$	$\sigma = 10^{-2}$	97.62%
		$\sigma = 10^{-3}$	97.72%
$b = 6$	$n = 60$	$\sigma = 10^{-2}$	95.70%
		$\sigma = 10^{-3}$	95.98%
	$n = 180$	$\sigma = 10^{-2}$	97.10%
		$\sigma = 10^{-3}$	97.16%
$b = 7$	$n = 56$	$\sigma = 10^{-2}$	95.30%
		$\sigma = 10^{-3}$	95.20%
	$n = 175$	$\sigma = 10^{-2}$	97.14%
		$\sigma = 10^{-3}$	96.88%
$b = 8$	$n = 56$	$\sigma = 10^{-2}$	94.56%
		$\sigma = 10^{-3}$	94.14%
	$n = 176$	$\sigma = 10^{-2}$	96.26%
		$\sigma = 10^{-3}$	97.04%
$b = 9$	$n = 54$	$\sigma = 10^{-2}$	94.38%
		$\sigma = 10^{-3}$	94.26%
	$n = 180$	$\sigma = 10^{-2}$	96.02%
		$\sigma = 10^{-3}$	96.28%
$b = 10$	$n = 60$	$\sigma = 10^{-2}$	91.84%
		$\sigma = 10^{-3}$	91.48%
	$n = 180$	$\sigma = 10^{-2}$	95.70%
		$\sigma = 10^{-3}$	95.64%
$b = 11$	$n = 55$	$\sigma = 10^{-2}$	92.48%
		$\sigma = 10^{-3}$	91.56%
	$n = 176$	$\sigma = 10^{-2}$	94.76%
		$\sigma = 10^{-3}$	95.08%

Table 5.1: Simulations of 95% confidence intervals for the moving block bootstrap when $\beta = 1$ and the errors are ARMA(1,1) with normally distributed innovations as stated in (5.29).

Block size	Sample size	Standard deviation	Empirical CI coverage Moving block bootstrap (proper)
$b = 3$	$n = 60$	$\sigma = 10^{-2}$	94.60%
		$\sigma = 10^{-3}$	95.16%
	$n = 180$	$\sigma = 10^{-2}$	95.84%
		$\sigma = 10^{-3}$	95.72%
$b = 4$	$n = 60$	$\sigma = 10^{-2}$	93.52%
		$\sigma = 10^{-3}$	93.74%
	$n = 180$	$\sigma = 10^{-2}$	95.72%
		$\sigma = 10^{-3}$	95.08%
$b = 5$	$n = 60$	$\sigma = 10^{-2}$	92.76%
		$\sigma = 10^{-3}$	93.04%
	$n = 180$	$\sigma = 10^{-2}$	95.00%
		$\sigma = 10^{-3}$	95.14%
$b = 6$	$n = 60$	$\sigma = 10^{-2}$	91.98%
		$\sigma = 10^{-3}$	92.40%
	$n = 180$	$\sigma = 10^{-2}$	94.68%
		$\sigma = 10^{-3}$	94.42%
$b = 7$	$n = 56$	$\sigma = 10^{-2}$	91.82%
		$\sigma = 10^{-3}$	91.94%
	$n = 175$	$\sigma = 10^{-2}$	94.92%
		$\sigma = 10^{-3}$	94.78%
$b = 8$	$n = 56$	$\sigma = 10^{-2}$	90.72%
		$\sigma = 10^{-3}$	90.80%
	$n = 176$	$\sigma = 10^{-2}$	94.06%
		$\sigma = 10^{-3}$	94.64%
$b = 9$	$n = 54$	$\sigma = 10^{-2}$	91.20%
		$\sigma = 10^{-3}$	90.56%
	$n = 180$	$\sigma = 10^{-2}$	93.26%
		$\sigma = 10^{-3}$	94.20%
$b = 10$	$n = 60$	$\sigma = 10^{-2}$	87.78%
		$\sigma = 10^{-3}$	87.28%
	$n = 180$	$\sigma = 10^{-2}$	93.42%
		$\sigma = 10^{-3}$	93.62%
$b = 11$	$n = 55$	$\sigma = 10^{-2}$	89.12%
		$\sigma = 10^{-3}$	88.06%
	$n = 176$	$\sigma = 10^{-2}$	92.98%
		$\sigma = 10^{-3}$	93.22%

Table 5.2: Simulations of 95% confidence intervals for the moving block bootstrap when $\beta = 1$ and the errors are AR(1) with normally distributed innovations as stated in (5.30).

Block size	Sample size	Standard deviation	Empirical CI coverage Moving block bootstrap (proper)
$b=3$	$n=60$	$\sigma=10^{-2}$	96.66%
		$\sigma=10^{-3}$	97.44%
	$n=180$	$\sigma=10^{-2}$	97.64%
		$\sigma=10^{-3}$	97.88%
$b=4$	$n=60$	$\sigma=10^{-2}$	95.64%
		$\sigma=10^{-3}$	95.44%
	$n=180$	$\sigma=10^{-2}$	97.34%
		$\sigma=10^{-3}$	96.90%
$b=5$	$n=60$	$\sigma=10^{-2}$	94.54%
		$\sigma=10^{-3}$	95.00%
	$n=180$	$\sigma=10^{-2}$	96.48%
		$\sigma=10^{-3}$	96.40%
$b=6$	$n=60$	$\sigma=10^{-2}$	93.68%
		$\sigma=10^{-3}$	94.04%
	$n=180$	$\sigma=10^{-2}$	95.78%
		$\sigma=10^{-3}$	95.96%
$b=7$	$n=56$	$\sigma=10^{-2}$	93.74%
		$\sigma=10^{-3}$	93.68%
	$n=175$	$\sigma=10^{-2}$	96.04%
		$\sigma=10^{-3}$	96.00%
$b=8$	$n=56$	$\sigma=10^{-2}$	92.60%
		$\sigma=10^{-3}$	92.54%
	$n=176$	$\sigma=10^{-2}$	95.04%
		$\sigma=10^{-3}$	95.76%
$b=9$	$n=54$	$\sigma=10^{-2}$	92.92%
		$\sigma=10^{-3}$	92.34%
	$n=180$	$\sigma=10^{-2}$	94.72%
		$\sigma=10^{-3}$	95.26%
$b=10$	$n=60$	$\sigma=10^{-2}$	89.76%
		$\sigma=10^{-3}$	89.34%
	$n=180$	$\sigma=10^{-2}$	94.52%
		$\sigma=10^{-3}$	94.64%
$b=11$	$n=55$	$\sigma=10^{-2}$	90.76%
		$\sigma=10^{-3}$	89.78%
	$n=176$	$\sigma=10^{-2}$	93.66%
		$\sigma=10^{-3}$	94.20%

Table 5.3: Simulations of 95% confidence intervals for the moving block bootstrap when $\beta=1$ and the errors are MA(1) with normally distributed innovations as stated in (5.31).

Block size	Sample size	Standard deviation	Empirical CI coverage Moving block bootstrap (proper)
$b = 3$	$n = 60$	$\sigma = 10^{-2}$	98.34%
		$\sigma = 10^{-3}$	98.64%
	$n = 180$	$\sigma = 10^{-2}$	98.70%
		$\sigma = 10^{-3}$	98.82%
$b = 4$	$n = 60$	$\sigma = 10^{-2}$	97.84%
		$\sigma = 10^{-3}$	97.50%
	$n = 180$	$\sigma = 10^{-2}$	98.24%
		$\sigma = 10^{-3}$	98.42%
$b = 5$	$n = 60$	$\sigma = 10^{-2}$	96.74%
		$\sigma = 10^{-3}$	96.90%
	$n = 180$	$\sigma = 10^{-2}$	97.50%
		$\sigma = 10^{-3}$	97.72%
$b = 6$	$n = 60$	$\sigma = 10^{-2}$	95.58%
		$\sigma = 10^{-3}$	95.18%
	$n = 180$	$\sigma = 10^{-2}$	97.32%
		$\sigma = 10^{-3}$	97.38%
$b = 7$	$n = 56$	$\sigma = 10^{-2}$	95.68%
		$\sigma = 10^{-3}$	95.20%
	$n = 175$	$\sigma = 10^{-2}$	97.32%
		$\sigma = 10^{-3}$	96.80%
$b = 8$	$n = 56$	$\sigma = 10^{-2}$	94.24%
		$\sigma = 10^{-3}$	94.80%
	$n = 176$	$\sigma = 10^{-2}$	96.68%
		$\sigma = 10^{-3}$	96.20%
$b = 9$	$n = 54$	$\sigma = 10^{-2}$	94.48%
		$\sigma = 10^{-3}$	94.00%
	$n = 180$	$\sigma = 10^{-2}$	96.04%
		$\sigma = 10^{-3}$	96.06%
$b = 10$	$n = 60$	$\sigma = 10^{-2}$	91.80%
		$\sigma = 10^{-3}$	91.28%
	$n = 180$	$\sigma = 10^{-2}$	95.54%
		$\sigma = 10^{-3}$	95.28%
$b = 11$	$n = 55$	$\sigma = 10^{-2}$	91.46%
		$\sigma = 10^{-3}$	91.46%
	$n = 176$	$\sigma = 10^{-2}$	95.58%
		$\sigma = 10^{-3}$	95.76%

Table 5.4: Simulations of 95% confidence intervals for the moving block bootstrap when $\beta = 1$ and the errors are ARMA(1,1) with innovations that have Student's t-distribution as stated in (5.32).

Block size	Sample size	Standard deviation	Empirical CI coverage Moving block bootstrap (proper)
$b = 3$	$n = 60$	$\sigma = 10^{-2}$	94.70%
		$\sigma = 10^{-3}$	95.52%
	$n = 180$	$\sigma = 10^{-2}$	95.80%
		$\sigma = 10^{-3}$	95.80%
$b = 4$	$n = 60$	$\sigma = 10^{-2}$	94.08%
		$\sigma = 10^{-3}$	93.92%
	$n = 180$	$\sigma = 10^{-2}$	94.72%
		$\sigma = 10^{-3}$	95.34%
$b = 5$	$n = 60$	$\sigma = 10^{-2}$	93.00%
		$\sigma = 10^{-3}$	93.02%
	$n = 180$	$\sigma = 10^{-2}$	94.96%
		$\sigma = 10^{-3}$	95.00%
$b = 6$	$n = 60$	$\sigma = 10^{-2}$	91.60%
		$\sigma = 10^{-3}$	91.08%
	$n = 180$	$\sigma = 10^{-2}$	94.52%
		$\sigma = 10^{-3}$	94.52%
$b = 7$	$n = 56$	$\sigma = 10^{-2}$	91.72%
		$\sigma = 10^{-3}$	92.04%
	$n = 175$	$\sigma = 10^{-2}$	95.04%
		$\sigma = 10^{-3}$	94.38%
$b = 8$	$n = 56$	$\sigma = 10^{-2}$	90.74%
		$\sigma = 10^{-3}$	91.50%
	$n = 176$	$\sigma = 10^{-2}$	94.12%
		$\sigma = 10^{-3}$	93.84%
$b = 9$	$n = 54$	$\sigma = 10^{-2}$	91.04%
		$\sigma = 10^{-3}$	90.24%
	$n = 180$	$\sigma = 10^{-2}$	93.52%
		$\sigma = 10^{-3}$	93.46%
$b = 10$	$n = 60$	$\sigma = 10^{-2}$	87.86%
		$\sigma = 10^{-3}$	87.24%
	$n = 180$	$\sigma = 10^{-2}$	93.24%
		$\sigma = 10^{-3}$	93.08%
$b = 11$	$n = 55$	$\sigma = 10^{-2}$	88.40%
		$\sigma = 10^{-3}$	87.42%
	$n = 176$	$\sigma = 10^{-2}$	93.64%
		$\sigma = 10^{-3}$	93.86%

Table 5.5: Simulations of 95% confidence intervals for the moving block bootstrap when $\beta = 1$ and the errors are AR(1) with innovations that have Student's t-distribution as stated in (5.33).

Block size	Sample size	Standard deviation	Empirical CI coverage Moving block bootstrap (proper)
$b = 3$	$n = 60$	$\sigma = 10^{-2}$	96.92%
		$\sigma = 10^{-3}$	97.40%
	$n = 180$	$\sigma = 10^{-2}$	97.54%
		$\sigma = 10^{-3}$	97.50%
$b = 4$	$n = 60$	$\sigma = 10^{-2}$	96.14%
		$\sigma = 10^{-3}$	95.78%
	$n = 180$	$\sigma = 10^{-2}$	96.60%
		$\sigma = 10^{-3}$	96.92%
$b = 5$	$n = 60$	$\sigma = 10^{-2}$	94.78%
		$\sigma = 10^{-3}$	94.96%
	$n = 180$	$\sigma = 10^{-2}$	96.32%
		$\sigma = 10^{-3}$	96.42%
$b = 6$	$n = 60$	$\sigma = 10^{-2}$	93.66%
		$\sigma = 10^{-3}$	93.12%
	$n = 180$	$\sigma = 10^{-2}$	95.90%
		$\sigma = 10^{-3}$	95.88%
$b = 7$	$n = 56$	$\sigma = 10^{-2}$	93.52%
		$\sigma = 10^{-3}$	93.92%
	$n = 175$	$\sigma = 10^{-2}$	96.18%
		$\sigma = 10^{-3}$	95.68%
$b = 8$	$n = 56$	$\sigma = 10^{-2}$	92.44%
		$\sigma = 10^{-3}$	93.02%
	$n = 176$	$\sigma = 10^{-2}$	95.50%
		$\sigma = 10^{-3}$	94.82%
$b = 9$	$n = 54$	$\sigma = 10^{-2}$	92.80%
		$\sigma = 10^{-3}$	92.08%
	$n = 180$	$\sigma = 10^{-2}$	94.62%
		$\sigma = 10^{-3}$	94.84%
$b = 10$	$n = 60$	$\sigma = 10^{-2}$	89.94%
		$\sigma = 10^{-3}$	89.18%
	$n = 180$	$\sigma = 10^{-2}$	94.50%
		$\sigma = 10^{-3}$	94.18%
$b = 11$	$n = 55$	$\sigma = 10^{-2}$	89.72%
		$\sigma = 10^{-3}$	89.52%
	$n = 176$	$\sigma = 10^{-2}$	94.52%
		$\sigma = 10^{-3}$	94.76%

Table 5.6: Simulations of 95% confidence intervals for the moving block bootstrap when $\beta = 1$ and the errors are MA(1) with innovations that have Student's t-distribution as stated in (5.34).

of b for $n \approx 60$ is 3 and $b = 5$ seems to be a suitable one for $n \approx 180$. MA(1) errors' models (5.31) and (5.34) prefer blocksize $b = 5$ for $n \approx 60$ and b equal to 8 or 9 for $n \approx 180$, respectively.

The fact that $\sqrt[3]{180/60} \approx 1.44$ corresponds to the approximate choice of blocksize $b = \mathcal{O}(n^{1/3})$, $n \to \infty$, because the most suitable choice of b in each of the errors' models (5.29)–(5.34) for the case of $n \approx 180$ is roughly 1.5-times the most appropriate choice of b in the case of $n \approx 60$.

When comparing errors' models (5.29)–(5.31) among each other with the same (fixed) "simulation" size n and standard deviation of innovations σ, the empirical coverages slightly differ for the same b. Or in other words, the optimal choice for various dependent errors' models slightly varies. Indeed, the optimal value of blocksize b depends on the variability of the errors and the random errors from errors' models (5.29)–(5.31) do not have the same variability, despite the fact that they have common innovations. A similar conclusion can be made when we move from normally distributed innovations onto innovations that have Student t-distributions, i.e., errors' models (5.32)–(5.34).

In order to provide a final *decision rule of thumb*, it seems that there is no big harm, if we simply take $b \approx n^{1/3}$.

However, the choice of the blocksize should not be too arbitrary. The blocksize, that is too small, would suppress the dependence between errors, which would lead to an incorrectly higher estimated variance of the EIV estimate. Contrariwise, immoderately high values of the blocksize would not mimic the empirical distribution of the EIV estimate, because they would underestimate the estimate's variance.

The above precaution and suggestions can be illustrated by the simulations from the ARMA (1,1) error model (5.29) with $\beta = 1$. We choose "bad" values of blocksize b. The empirical coverages of the confidence intervals are shown in Table 5.7. When $b = 1$ (too small), the moving block bootstrap becomes the nonparametric bootstrap, which ignores the dependence between data. The estimated variance is too high, which results into completely incorrect empirical CI's coverages (over 99.9%). When b is set to 15 (too high value of b for sample size of 60) or even 30, the empirical CI's coverages rapidly decrease and the theoretical (correct) level of the 95% cannot be reached.

We performed much more simulations than presented here, especially for higher values of n. It generally turns out that the larger the sample size is, the better empirical coverage is achieved, when the asymptotically optimal (appropriate) blocksize b is chosen according to n. The precision may also be increased (or, in other words, the noise decreased) by setting the variance smaller and, therefore, smaller deviations of the errors' innovations provide empirical coverages of the CIs closer to the theoretical 95% level, but the choice of b needs to be "approximately correct" as well.

Block size	Sample size	Standard deviation	Empirical CI coverage Moving block bootstrap (proper)
$b = 1$	$n = 60$	$\sigma = 10^{-2}$	99.92%
		$\sigma = 10^{-3}$	99.96%
	$n = 180$	$\sigma = 10^{-2}$	99.98%
		$\sigma = 10^{-3}$	99.98%
$b = 15$	$n = 60$	$\sigma = 10^{-2}$	87.16%
		$\sigma = 10^{-3}$	86.14%
	$n = 180$	$\sigma = 10^{-2}$	94.22%
		$\sigma = 10^{-3}$	94.02%
$b = 30$	$n = 60$	$\sigma = 10^{-2}$	66.86%
		$\sigma = 10^{-3}$	68.62%
	$n = 180$	$\sigma = 10^{-2}$	89.04%
		$\sigma = 10^{-3}$	88.88%

Table 5.7: Inappropriate choices of blocksize b in the simulations of 95% confidence intervals for the moving block bootstrap when $\beta = 1$ and the errors are ARMA(1,1) with normally distributed innovations as stated in (5.29).

5.5 Data Analysis – Brown Trout

A group of ecologists from the Water Research Institute of T. G. Masaryk—a public research institution in the Czech Republic—is interested in the relationship between the length and the weight of brown trouts (*Salmo trutta* morpha *fario*) from a small mountain stream of Šumava National Park located on the south border of the Czech Republic. The data set contains 59 length-weight measurement pairs of the adult brown trouts. The ecologists were catching the trouts from the spring of the stream to the junction with another slightly larger river, i.e., from the highest altitude of the observed stream to its lowest part. Consequently, the fish were released back to the stream.

The main ecologists' question is whether there is a *simple linear relationship* between the lengths and the weights of trouts. E.g., does an increase in the trout's length by one millimeter have an impact of the increase in the trout's weight by 2.1 grams (or vice-versa)?

If we want to model the previous situation and provide a reliable answer to the above postulated question, we need to realize some important aspects, which we are dealing with:

- Both brown trout's characteristics—length and weight—are measured with errors and, moreover, both of them are encumbered with the disturbances of nature.
- Measured characteristics of the trouts are not independent due to the several reasons proposed by the ecologists. The depth of the stream is rapidly changing from the spring to the junction, which limits the movement of bigger trouts. On the other hand,

a higher water level provides a living environment for some predators and, therefore, the small trouts keep away from the lower parts of the stream. The chemistry of the water and the source of a nutriment vary a lot as the stream flows. So, the trouts from the higher altitudes have different conditions for their growth than the trouts from the lower altitudes. Fish, that live nearby and occupy a similar environment, influence each other and, moreover, a trout has a stronger impact on the neighboring trouts than on the faraway ones, i.e., a presence of pheromones.

- A linear relationship between measured characteristics is of the scientific interest due to its parsimonious interpretation for the ecologists.

According to the above aspects and demands, our EIV model with weakly dependent errors seems to be a plausible choice for modeling such a realistic situation.

As it can be easily noticed, the EIV model cannot be directly applied to the trout's data, because we need to take into account a "baseline" shift present between the lengths and the weights. Mathematically speaking, an intercept should be incorporated in the linear EIV model. Indeed, we have already theoretically discussed such a setup in Section 1.5 and *Partial Errors-in-Variables Model* (1.20) is the appropriate extension of the classical EIV model that should be used for the trouts data analysis.

The brown trouts data are displayed in Figure 5.1 (left) with the corresponding *least squares-total least squares* (LS-TLS) estimate (a solution to the PEIV model) and the *ordinary least squares* (OLS) estimate. As it was already discussed in Section 1.5, the data needs to be transformed as in (1.21) at first, see Figure 5.1 (right). Consequently, *slope estimate* $\widehat{\boldsymbol{\beta}}_{LS-TLS}$ can be obtained as in (1.22) from the transformed data.

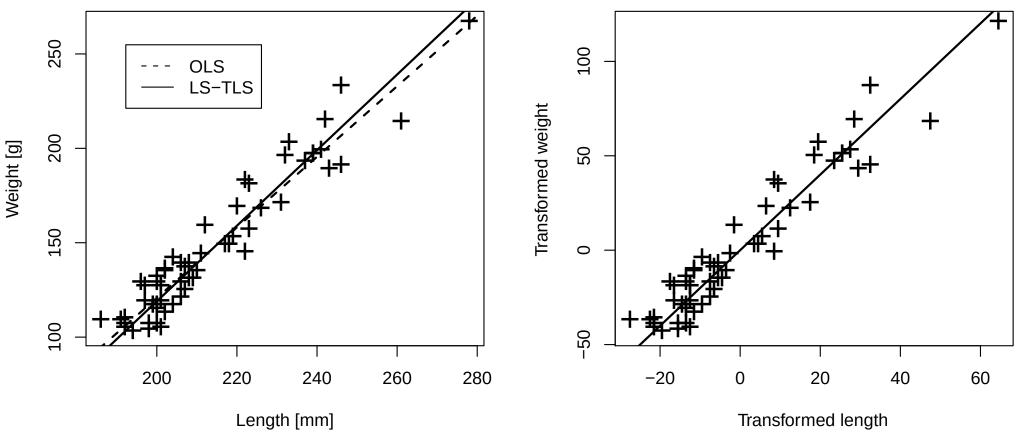

Figure 5.1: Brown trouts' lengths and weights.

The LS-TLS and the OLS slope estimates are shown in Table 5.8. Moreover, the corresponding 95% confidence intervals are constructed. For the OLS slope estimate, the CI based upon the asymptotic normality (blindly assuming independent errors only in the weighs and their homoscedasticity) is calculated. For the LS-TLS slope estimate, three types of the CIs are computed, i.e., the inappropriate CI based upon the approximate asymptotic normality, the CI based upon the proper nonparametric bootstrap (blindly assuming independent errors), and the proper CI based upon the correct MBB with the choice of blocksize $b = 4(\approx \sqrt[3]{59})$.

Method	Estimate of slope	Confidence interval	CI length
OLS (independence)	1.8687	[1.7228;2.0145]	0.2917
Approximate AN (independence)	2.0020	[1.8484;2.1556]	0.3072
NB (independence)	2.0020	[1.8314;2.2189]	0.3874
MBB (proper)	2.0020	[1.8714;2.1693]	0.2979

Table 5.8: Brown trout data set, $\widehat{\beta} = 2.0020$.

The OLS estimate dramatically differs from the LS-TLS one (see Table 5.8) as it can be expected due to the fact that the OLS estimate does not "allow" for the error disturbances in lengths.

If someone wants to test whether the slope between the lengths and the weights is equal 2.1 against the complementary alternative on the level of 5%, this hypothesis would be incorrectly rejected according to the classical regression setup (the OLS approach).

If we compare the lengths of the CIs for $\widehat{\boldsymbol{\beta}}_{LS-TLS}$ based upon the approximate asymptotic normality (assuming the independence of errors) and upon the proper MBB procedure, we figure out that the MBB confidence interval is slightly narrower. One of the main reasons can be that the approximate asymptotic normality gives larger amount of variability due to its inadequate independence assumption.

Applying the classical proper nonparametric bootstrap (NB) technique (without stacking observations into blocks and, hence, ignoring data dependence) provides confidence interval of length 0.3874, which differs a lot from the CI based on the proper MBB approach. This CI's length differs a lot from the length of the CI based on the asymptotic normality as well, which hints that some of the assumptions (e.g., independence of the errors) is probably violated.

Moreover, the lengths of the incorrect and the proper MBB confidence intervals coincide, which should not be a surprise, because the empirical distribution of $\widehat{\boldsymbol{\beta}}^*$ and $\widetilde{\boldsymbol{\beta}}^*$ differs only by correction (3.57) and the CIs are constructed by a percentile method. This correction is a constant for each fixed bootstrap sample, which is very small ($\approx 10^{-5}$) and numerically negligible. Therefore, the incorrect and the proper MBB CIs seem to coincide as well, but there is still a small difference between them.

The EIV estimate behaves robustly against *leverage observations*, which was already mentioned in Chapter 2. Fierro and Bunch (1994) and Van Huffel and Vandewalle (1991, Chapter 9) concentrate on the *multicollinearity* (nearly linear dependence) problem, which often occurs in the OLS estimation. There is a (di)similarity between the form of the EIV estimate and the ridge regression's estimate, where both estimation methods change the diagonal of matrix $\mathbf{X}^\top\mathbf{X}$ in the estimator's forms. Therefore, the EIV estimate can be viewed as a regularization technique (which may become ill conditioned for some data).

In our brown trouts data set, we can find one possible leverage observation—top-right corner of the left panel in Figure 5.1. As it can be seen, the OLS estimate is "persuaded" by this leverage point to "go" nearby and, hence, it is easily influenced by just one measurement. On the other hand, the LS-TLS estimate behaves more robust to that potential influential observation. Omitting such an observation provides only a small change in the LS-TLS estimate, but a larger one in the OLS estimate.

5.6 Conclusions

The EIV estimate for our EIV model with strong mixing and uniformly strong mixing errors is of main interest in this chapter. Since its asymptotic normality was derived in Chapter 4 for the case of weakly dependent errors, it is wanted to use this nice theoretical property for the practical (computational) inference. Nevertheless, all the pitfalls and problems about the EIV estimate's asymptotic variance still remain present as firstly pointed out in Chapter 3.

Indeed, the approximate asymptotic normality concerning the EIV estimate is computationally useless. Therefore, a proper *moving block bootstrap procedure* is proposed in order to *mimic the asymptotical distribution* of the EIV estimate. Its *validity* was proved for the case of α-mixing and φ-mixing errors.

A simulation study was conducted to demonstrate the theoretical results. It showed up that the moving block bootstrap is a *distributional-free method*, which depends on the choice of blocksize. A naive approximate choice of the blocksize was proposed. On the contrary, we remarked that no big harm is made when the blocksize is chosen close enough to that naive rule of thumb.

The real data set containing the lengths and the weights of brown trouts was described. The appropriateness of the *partial EIV model* was discussed and, consequently, the 95% confidence interval for the slope parameter using the proper MBB procedure was computed. On the top of that, the robustness to multicollinearity and leverage observations was illustrated.

5.6.1 Discussion and Remarks

Two justifications for the MBB procedure were provided in this chapter—one for the case of α-mixing errors and the other one for φ-mixing errors. These two theorems have *slightly different assumptions* on the errors. The *approximate coincidence* of the conditional MBB distribution of $\sqrt{n}(\widetilde{\boldsymbol{\beta}}^{*} - \widehat{\boldsymbol{\beta}})$ and the original distribution of $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ with α-mixing errors requires weaker assumptions on the mixing coefficients' rates than in the case of φ-mixing errors, i.e., φ-mixing implies α-mixing and (5.23)–(5.24) is more restrictive than (5.14)–(5.15). On the other hand, Theorem 5.6 contains less restrictive assumptions on the equiboundedness of the errors' moments than Theorem 5.5 (compare (5.26)–(5.27) with (5.17)–(5.18)). Moreover, the correctness of the MBB for the EIV estimate with α-mixing errors has the additional restriction (5.19), which *ties up* two contending assumptions: the mixing errors' rates and the order of errors' moments.

Finally, all the issues, remarks, and (dis)advantages already mentioned in the discussions of Chapter 3 and Chapter 4 retain relevant for weakly dependent errors and do not have to be repeated. Additionally, different block bootstrap procedures (e.g., non-overlapping or circular) can be shown to be valid for the EIV model in some future work.

Chapter 6

Nonparametric Estimation

Essentially, all models are wrong, but some are useful.

GEORGE E. P. BOX

A natural extension of the linear EIV model is to go for a *nonlinear* one. Amemiya (1997) proposed a way of the first order linearization of the nonlinear relations. Fazekas and Kukush (1997) investigated the asymptotic properties in the nonlinear EIV model. The TLS estimate in nonlinear EIV model is generally inconsistent as it was thoroughly discussed in, e.g., Fazekas et al. (2004). A correction to remove the bias was suggested by Kukush and Zwanzig (2002). Due to the inconsistency issue, the nonlinear EIV model will be not considered in this thesis anymore. We directly move on a *nonparametric estimation* in the errors-in-variables.

Since there are many ways and approaches in the nonparametric regression techniques, we concentrate on one specific, but very general way of nonparametric EIV setup.

We propose a class of *nonparametric* estimates for the EIV models over the sets of sufficiently *smooth* functions. The estimation takes place over the balls of functions which are elements of a suitable *Sobolev space*—special type of Hilbert spaces that facilitate calculation of the (total) least squares projection. The Hilbertnesss allows us to take projections and hence to decompose spaces into mutually orthogonal complements. Then we transform the problem of searching for the best fitting function in an infinite dimensional space into a finite dimensional optimization problem.

The regression setup proposed by Yatchew and Bos (1997) will be extended and combined with the total least squares approach introduced by Golub and Van Loan (1980). Our main interest lies only in the process of estimation. The inference part for our forthcoming nonparametric estimate is postponed for further research and exploration, because it would exceed the range of this thesis.

6.1 Introduction

Let us consider the simplest one-dimensional situation when one observes *input data*

$$[\mathbf{x}, \mathbf{y}] \equiv [(x_1, \ldots, x_n)^\top, (y_1, \ldots, y_n)^\top].$$

Moreover, these *observations* are considered to be measured with additive random *errors* $[\boldsymbol{\theta}, \boldsymbol{\varepsilon}]$. *Unobservable true* values $[\mathbf{x} + \boldsymbol{\theta}, \mathbf{y} + \boldsymbol{\varepsilon}]$ satisfy an *unknown* functional relationship, i.e., *regression*

$$y_i + \varepsilon_i = f(x_i + \theta_i), \quad i = 1, \ldots, n.$$

Unknown function f is thought to be *smooth*. The smoothness needs to be properly defined somehow. Suppose that our "smoothness" is the only assumption and, thereby, we want a modeling technique to be applicable on *various* types (large number) of data. Finally, we are searching for a suitable estimate $\widehat{f}$, where the misfit needs to be "*as small as possible*".

6.2 Sobolev Spaces and Total Least Squares

A wide applicability of the method for finding a suitable estimate in our setup yields a *non-parametric* approach as an adequate technique. Smoothness of the estimator function $\widehat{f}$ needs to be ensured, but, e.g., kernels, splines, or wavelets can be *too restrictive*. Therefore, we fit a function from a *general class* of smooth functions—*Sobolev spaces* equipped with a corresponding *Sobolev norm*

$$\left(\mathcal{H}^m, \|\cdot\|_{Sob,m}\right) := \left\{ g \in \mathsf{L}^2 \, : \, \|g\|_{Sob,m} := \left(\sum_{i=0}^{m} \int |g^{(i)}(t)|^2 \mathrm{d}t \right)^{1/2} < +\infty \right\}.$$

Previous definition indicates that unknown function $f \in \mathcal{H}^m$ needs to have derivatives up to the order m and, hence, one may speak about the *order* of corresponding Sobolev space. In many physical or econometric relationships, the order $m = 2$ seems to be quite satisfactory, see Yatchew and Härdle (2006).

6.2.1 Graphical Illustration

The observed data should be "as close as possible" to the true unobservable values, or in other words, the errors $[\boldsymbol{\theta}, \boldsymbol{\varepsilon}]$ should be "as small as possible". This can be reached by measuring the misfit in the "shortest" way, i.e., taking the perpendicular distance into account as demonstrated in Figure 6.1.

Since one assumes that $m \geq 1$, a tangent can be constructed for function f in its each point and, hence, the orthogonal distance from input values can be measured. This

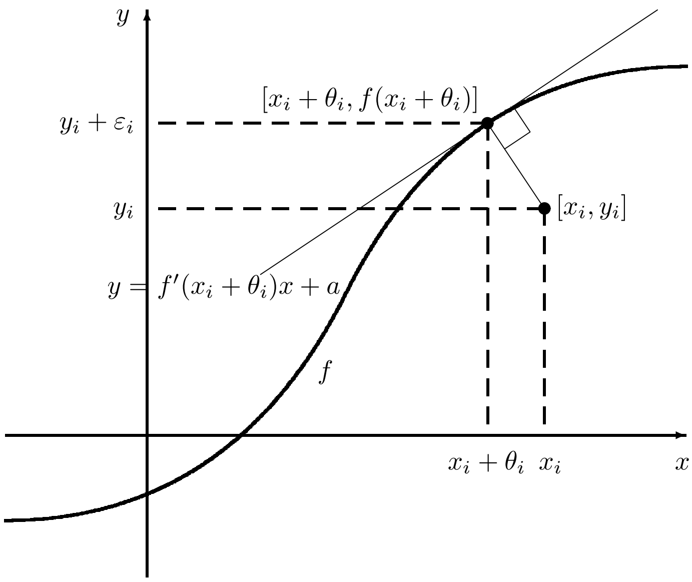

Figure 6.1: Orthogonal regression in the EIV setup with a smooth function.

orthogonal distance clearly corresponds to the Euclidean norm of the errors $[\boldsymbol{\theta}, \boldsymbol{\varepsilon}]$, i.e., total least squares. The TLS method is just another name for orthogonal regression in statistics.

On the other hand, the smoothness (or "wildness") of unknown function f is measured by its Sobolev norm. Hence, one should realize that the *better* the fit the *wilder* the function and vice versa. This can be written in an informal way

$$\frac{\text{small}}{\text{large}} \left\| \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\varepsilon} \end{bmatrix} \right\|_2 \quad \rightleftharpoons \quad \frac{\text{large}}{\text{small}} \|f\|_{Sob,m} .$$

6.2.2 Estimate

Searching for an estimate $\widehat{f}$ is simply nothing else than finding a reasonable *compromise* between misfit (Euclidean norm of the error vector) and smoothness (Sobolev norm of the estimated function). This compromise can be easily incorporated using so-called *smoothing parameter* $\chi > 0$:

$$\min_{f \in \mathcal{H}^m, \boldsymbol{\theta} \in \mathbb{R}^n, \boldsymbol{\varepsilon} \in \mathbb{R}^n} \left\{ \left\| \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\varepsilon} \end{bmatrix} \right\|_2^2 + \chi \|f\|_{Sob,m}^2 \right\}, \quad \text{s.t.} \;\; \mathbf{y} + \boldsymbol{\varepsilon} = \mathbf{f}(\mathbf{x} + \boldsymbol{\theta}). \tag{6.1}$$

Although, the optimizing problem (6.1) is very complicated to solved directly.

Let us consider for simplicity interval $[0,1]$ as a bounded domain, where our x-data lie. Sobolev space on this interval $\mathcal{H}^m[0,1]$ is a Banach space and one can define *Sobolev inner product* for each $g, h \in \mathcal{H}^m[0,1]$:

$$\langle g, h\rangle_{Sob,m} := \sum_{i=0}^{m} \int_0^1 g^{(i)}(t) h^{(i)}(t) \mathrm{d}t. \tag{6.2}$$

Hence, $\mathcal{H}^m[0,1]$ is also a Hilbert space. Using Riesz representation theorem and Arzelà-Ascoli theorem, Yatchew and Bos (1997) proved the following: for all $f \in \mathcal{H}^m[0,1]$ and for any $a \in [0,1]$, there exists $\psi_a \in \mathcal{H}^m[0,1]$ such that

$$f(a) = \langle \psi_a, f\rangle_{Sob,m} = \sum_{i=0}^{m} \int_0^1 \psi_a^{(i)}(t) f^{(i)}(t) \mathrm{d}t, \tag{6.3}$$

and ψ_a is called a *representor* at the point a. Hence, one may easily derive so-called *representor matrix* $\mathbf{\Psi}_{n\times n}(\mathbf{t})$ whose columns (and rows) equal the representors *evaluated* at $t_1, \ldots, t_n$

$$\Psi_{ij}(\mathbf{t}) = \left\langle \psi_{t_i}, \psi_{t_j} \right\rangle_{Sob,m} = \psi_{t_i}(t_j) = \psi_{t_j}(t_i), \quad \forall i, j. \tag{6.4}$$

The representor matrix is *symmetric* and *positive definite* as proved in Pešta (2006b).

A form of the representors was derived by Pešta (2006a):

$$\psi_a(t) = \sum_{k=1}^{2m} \exp\left[\Re\left(e^{i\delta_k}\right)t\right] \Big\{ \mathcal{I}_{[t\leq t_j]} \gamma_k(t_j) \cos\left[\Im\left(e^{i\delta_k}\right)t\right] + \mathcal{I}_{[t>t_j]} \gamma_{2m+k}(t_j) \sin\left[\Im\left(e^{i\delta_k}\right)t\right] \Big\}.$$

Here, the coefficients γ_ks and δ_ks are determined as a solution of the ordinary differential equation with some boundary conditions.

Let

$$\mathcal{M} := span\left\{\psi_{x_i+\theta_i} : i = 1, \ldots, n\right\}$$

and, afterwards, its orthogonal complement

$$\mathcal{M}^\perp = \{h \in \mathcal{H}^m[0,1] : \langle \psi_{x_i+\theta_i}, h\rangle_{Sob,m} = 0, \, i = 1, \ldots, n\}.$$

The Sobolev space can be written as a direct sum of its orthogonal subspaces, i.e., $\mathcal{H}^m[0,1] = \mathcal{M} \oplus \mathcal{M}^\perp$ since $\mathcal{H}^m[0,1]$ is a Hilbert space. Function $h \in \mathcal{M}^\perp$ takes on the value zero at

$x_1 + \theta_1, \ldots, x_n + \theta_n$ due to the property (6.3). Each $f \in \mathcal{H}^m[0,1]$ can be written in form

$$f = \sum_{i=1}^{n} c_i \psi_{x_i+\theta_i} + h, \quad h \in \mathcal{M}^{\perp}. \tag{6.5}$$

Then, one can rewrite the objective function from (6.1) incorporating the corresponding restriction simply by its substitution, applying the relation (6.5) and the representation (6.3), utilizing linearity of the Sobolev inner product (6.2), and using the definition of representor matrix (6.4) with its property of being symmetric

$$\begin{aligned}
\left\| \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\varepsilon} \end{bmatrix} \right\|_2^2 + \chi \|f\|_{Sob,m}^2 &= \|\boldsymbol{\theta}\|_2^2 + \left\| \mathbf{y} - \left\langle \boldsymbol{\psi}_{\mathbf{x}+\boldsymbol{\theta}}, \sum_{i=1}^{n} c_i \psi_{x_i+\theta_i} + h \right\rangle_{Sob,m} \right\|_2^2 \\
&+ \chi \left\langle \sum_{i=1}^{n} c_i \psi_{x_i+\theta_i} + h, \sum_{i=1}^{n} c_i \psi_{x_i+\theta_i} + h \right\rangle_{Sob,m}^2 \\
&= \|\boldsymbol{\theta}\|_2^2 + \|\mathbf{y} - \boldsymbol{\Psi}(\mathbf{x}+\boldsymbol{\theta})\mathbf{c}\|_2^2 \\
&+ \chi \mathbf{c}^\top \boldsymbol{\Psi}(\mathbf{x}+\boldsymbol{\theta})\mathbf{c} + \chi \|h\|_{Sob,m}^2
\end{aligned}$$

where for an arbitrary $g \in \mathcal{H}^m$ the following convention is used

$$\langle \boldsymbol{\psi}_{\mathbf{t}}, g \rangle_{Sob,m} = \left(\langle \psi_{t_1}, g \rangle_{Sob,m}, \ldots, \langle \psi_{t_n}, g \rangle_{Sob,m} \right)^\top.$$

Note further that $\sum_{i=1}^{n} c_i \psi_{x_i+\theta_i}$ minimizes (6.1) and, moreover, is the unique solution of that optimizing problem, because $\psi_{x_i+\theta_i}$ are the base elements of $\mathcal{M}$. Therefore, the *infinite dimensional* minimizing (6.1) is transformed into the *finite dimension*

$$\min_{\mathbf{c} \in \mathbb{R}^n, \boldsymbol{\theta} \in \mathbb{R}^n} \left\{ \|\boldsymbol{\theta}\|_2^2 + \|\mathbf{y} - \boldsymbol{\Psi}(\mathbf{x}+\boldsymbol{\theta})\mathbf{c}\|_2^2 + \chi \mathbf{c}^\top \boldsymbol{\Psi}(\mathbf{x}+\boldsymbol{\theta})\mathbf{c} \right\}. \tag{6.6}$$

A solution $\left\{ \widehat{\mathbf{c}}, \widehat{\boldsymbol{\theta}} \right\}$ of the finite optimizing problem (6.6) always exists and is unique, which can be proved similarly as in Pešta (2006a). A *derivative of representor matrix* $\boldsymbol{\Psi}_{n \times n}^{(1)}(\mathbf{t})$ needs to be defined as a matrix whose columns are equal to the first derivatives of the representors evaluated at $t_1, \ldots, t_n$; i.e.

$$\Psi_{i,j}^{(1)} = \psi_{t_j}'(t_i), \quad i,j = 1, \ldots, n.$$

Now, by setting all the partial derivatives of the objective function in (6.6) with respect to all elements of $\mathbf{c}$ and $\boldsymbol{\theta}$ equal zero, and taking into account the existence of the inverse of representor matrix (due to its positive definiteness), one can end up with a system of equations

$$\begin{aligned}\left[\boldsymbol{\Psi}(\mathbf{x}+\widehat{\boldsymbol{\theta}})+\chi\mathbf{I}\right]\widehat{\mathbf{c}} &= \mathbf{y},\\ \left[\mathbf{y}-\boldsymbol{\Psi}(\mathbf{x}+\widehat{\boldsymbol{\theta}})\widehat{\mathbf{c}}-\frac{\chi}{2}\widehat{\mathbf{c}}\right]\boldsymbol{\Psi}^{(1)}(\mathbf{x}+\widehat{\boldsymbol{\theta}}) &= \widehat{\boldsymbol{\theta}},\end{aligned}$$

which can be solved iteratively.

Once we find $\widehat{\mathbf{c}}$ and $\widehat{\boldsymbol{\theta}}$, a *unique* estimate $\widehat{f}$ can be obtained by

$$\widehat{f} = \sum_{i=1}^{n} \widehat{c}_i \psi_{x_i+\widehat{\theta}_i}.$$

6.3 Examples

Our technique will be demonstrated on two totally different real data sets. If we do not have any idea about the nature of our data, one cannot simply use a special technique. Here comes our method. Surely, our technique can behave worse on one concrete data set than an "appropriate" method for that kind of data. On the other hand, we do not lose as much as in a situation when an inappropriate method is chosen due to the lack of information about the data, e.g., a Pareto type model for estimating a probability density, which appears to be bimodal.

The first data set are the result of a National Institute of Standards and Technology (NIST) study involving the *thermal expansion of copper*. The response variable is the coefficient of thermal expansion and the predictor variable is temperature in kelvin. The data contain 236 observations and were firstly described by Hahn (1970). The precision of the thermometer used is surely not zero and that is why some disturbances in measured temperature should be taken into account in our model. Our fit can be seen in Figure 6.2.

The second data set are monthly averaged atmospheric pressure differences between Easter Island (Pacific) and Darwin (Australia) and can be found in Kahaner et al. (1989). This difference drives the trade winds in the southern hemisphere. Cycles in the pressure differences correspond to the *El Niño* and the *Southern Oscillation*. These data contain 168 observations and errors should be taken in the account in the explanatory variable (time) as well. The reason for this is very simple—one cannot know whether the data were collected weekly on the same day or daily at the same hour, and also simultaneously on both locations. The fitted curve for our technique is again shown in Figures 6.2 and 6.3.

6.4 Discussion and Conclusions

In this chapter, regression in Sobolev spaces using TLS is developed. Sobolev spaces provide the only general restriction—smoothness—on the unknown estimated function. Total least squares helps to incorporate (measurement) errors in the explanatory variable and in the

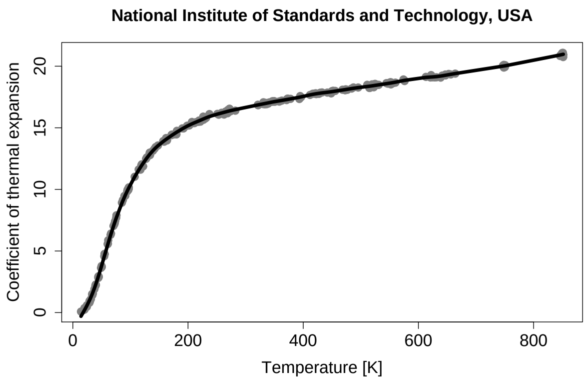

Figure 6.2: Thermal expansion of copper.

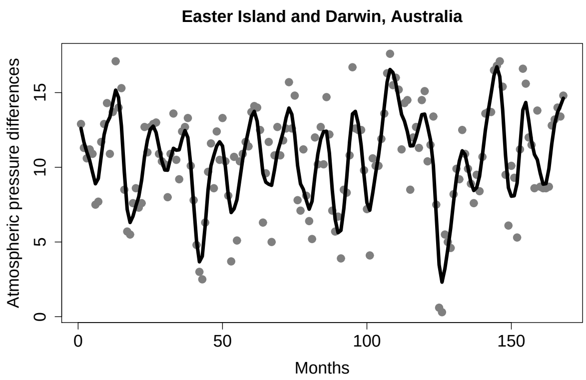

Figure 6.3: El Niño – southern oscillation.

predictor as well. Joining these two mathematical constructions together yields a method which easily provides an estimate with demanded properties as shown in Section 6.2. Thus one may conclude that it is a *very general* and *widely applicable* nonparametric smoothing technique as it was demonstrated in Section 6.3.

Moreover, our method *works without a prior knowledge* of functional relation or error distribution. This makes our technique *widely applicable* to various data relationships and quite robust with respect to the nature of data.

6.4.1 Remarks

It has to be remarked that regression in Sobolev spaces using TLS can be easily extended into a *multivariate* case, meaning more dimensions for covariates and response variables as well, but partial derivatives have to be taken into account.

A smoothing (tuning) parameter χ incorporated in (6.1), which controls the *trade-off* between the infidelity to the data versus the roughness of the estimated solution, can be chosen according, e.g., a *cross-validation* criterion

$$\mathcal{CV}(\chi) = \frac{1}{n}\sum_{i=1}^{n}\left[y_i - \widehat{f}_{-i}(x_i+\theta_i)\right]^2$$

where $\widehat{f}_{-i}$ is obtained by solving

$$\min_{f,\boldsymbol{\theta}_{-i},\boldsymbol{\varepsilon}_{-i}} \left\{ \left\| \begin{bmatrix} \boldsymbol{\theta} \\ \boldsymbol{\varepsilon} \end{bmatrix}_{-i} \right\|_2^2 + \chi \left\| f \right\|_{Sob,m}^2 \right\} \quad \text{s.t. } \mathbf{y}_{-i} + \boldsymbol{\varepsilon}_{-i} = \mathbf{f}(\mathbf{x}_{-i} + \boldsymbol{\theta}_{-i})$$

where the subscript $-i$ denotes omitted the i-th element of the corresponding vector. A starting value of χ can be chosen "ad hoc" by trying on values from the logarithmic lattice. But the cross-validation need not to be the only one which can provide a reasonable choice of the previously mentioned parameter, i.e., generalized cross-validation or various information criteria.

Regression in Sobolev spaces allows to add so-called *isotonic* restrictions (for more details see Pešta (2006a)) to the estimated function, e.g., monotonicity or convexity. Therefore, one can perform a cumulative density function or a probability density function estimation. Another application of isotonicity can be found in testing unimodality of a general regression curve.

Unfortunately, we have to admit that our method has a disadvantage. There is a problem of the invariant estimate with respect to a change of the variable magnitudes (scale). Let us consider the simplest one-dimensional setup discussed above. When the values of the explanatory variable are divided by a factor of two and the values of the response variable are kept unchanged, our technique provides a different estimated curve than the originally fitted curve shrunk twice in the horizontal way. This problem can be solved incorporating *scaled* total least squares (STLS) with a scaling parameter $\eta > 0$. Hence, optimizing problem of finding an estimate $\widehat{f}$ is shown as follows

$$\min_{f\in\mathcal{H}^m,\boldsymbol{\theta}\in\mathbb{R}^n,\boldsymbol{\varepsilon}\in\mathbb{R}^n} \left\{ \left\|\boldsymbol{\theta}\right\|_2^2 + \eta \left\|\boldsymbol{\varepsilon}\right\|_2^2 + \chi \left\| f \right\|_{Sob,m}^2 \right\}, \quad \text{s.t. } \mathbf{y} + \boldsymbol{\varepsilon} = \mathbf{f}(\mathbf{x} + \boldsymbol{\theta}). \tag{6.7}$$

On the other hand, an additional theoretical problem of the choice of scaling parameter appears when compared with previous simpler optimizing problem (6.1). Altogether, STLS approach (6.7) can be viewed as a skewed orthogonal regression, when one takes into account not the perpendicular distance to the unknown function, but the skewed one with a certain angle represented by parameter η. This scaling parameter can also serve as an *emphasizing parameter*, because it distributes emphasis on the errors corresponding to independent and dependent variables.

6.4.2 Further Research

For a further research, one may be interested in functional data. Hence, our method might be extended into this statistical branch. One of the theoretical reasons (except many practical demands) for the applicability of our approach into functional data analysis is that Hilbert-Schmidt operator nicely "shakes hands" with Sobolev spaces.

Our approach of regression in Sobolev spaces using TLS needs to be studied from the statistical point of view as well. Consistency and asymptotic normality of the estimate should be explored. Moreover, bootstrap techniques could be applied for confidence intervals and hypothesis testing.

Chapter 7

Conclusions

Je to hlavně celé o lidech ... někdy to třeba rozebereme u borovičky.
[It's mainly all about people ... sometimes maybe we will discuss it with gin.]

JAROMÍR ANTOCH

The linear errors-in-variables model is introduced with a corresponding total least squares estimate. A brief algebraic and statistical summary of the EIV problem and the TLS estimate is provided. General error structures with possible extensions are proposed in order to model some realistic situations. Several alternatives to the EIV model are suggested and reasonable ways of the estimation are carried out.

A generalization of the TLS estimate—the EIV estimate—is derived. Surprisingly, the form of the estimate remains the same, but it solves much broader class of optimizing problems in the EIV setup. Several invariant and equivariant properties are shown. The EIV estimate can be viewed as a unitarily invariant and error-distance minimizing estimate. On the top of that, a robust behavior with respect to outliers and leverage observations of the EIV estimate is demonstrated.

Serious problems concerning asymptotic normality of the EIV estimate are pointed out. A solution to computational inefficiency of the normal approximation is brought by a bootstrap approach. Asymptotical validity of the proper nonparametric bootstrap procedure is shown and demonstrated in a simulation study and on real data as well.

In spite of the previous advantage of the bootstrap approach, the bootstrap inference cannot outperform the asymptotic normality. On the other hand, it can provide solid results even in the case when the normal approximation is computationally useless.

Independence of the errors does not have to be realistic every time. The error structure of the EIV model is extended by weakening the independence assumption. Weakly dependent errors—strong and uniformly strong mixing—are considered. Consequently, a strong consistency and an asymptotic normality of the EIV estimate are proved in such a setup.

The practical applicability of the previous asymptotic results remains a serious issue analogously as in the case when the errors were independent. A suitable block bootstrap method is proposed in order to overcome such a problem. A justification of the moving block bootstrap procedure is proved. A choice of the blocksize for the MBB is elaborated, which is illustrated via a simulation study. Real data, where the EIV model with weakly dependent errors seems to be plausible, is analyzed.

Finally, a nonlinear extension of the EIV model is commented. On the top of that, a nonparametric version of the EIV model is supposed. A way of estimation from a broad class of sufficiently smooth functions is suggested and, afterwards, demonstrated on various real data sets.

7.1 Overview

A skeleton of this thesis can be encapsulated in an ideological schema displayed in Figure (7.1).

Firstly, a linear EIV model is considered, which is generalized to a nonparametric one later on. Properties of the TLS estimate are summarized and its generalization, called the EIV estimate, is invented. Interesting and important properties of invariancy and equivariancy for the EIV estimate are derived.

Various error structures—from independent homoscedastic errors to weakly dependent heteroscedastic ones—are taken into account, while the design assumption (D) is still valid.

Afterwards, asymptotic properties like the strong consistency or the asymptotic normality is proved. Having troubles with normal approximations, proper and correct bootstrap methods are suggested. Their appropriateness and validity are shown theoretically and by simulations as well.

All important remarks and suggestions to the particular issues and problems are discussed immediately at the end of each chapter.

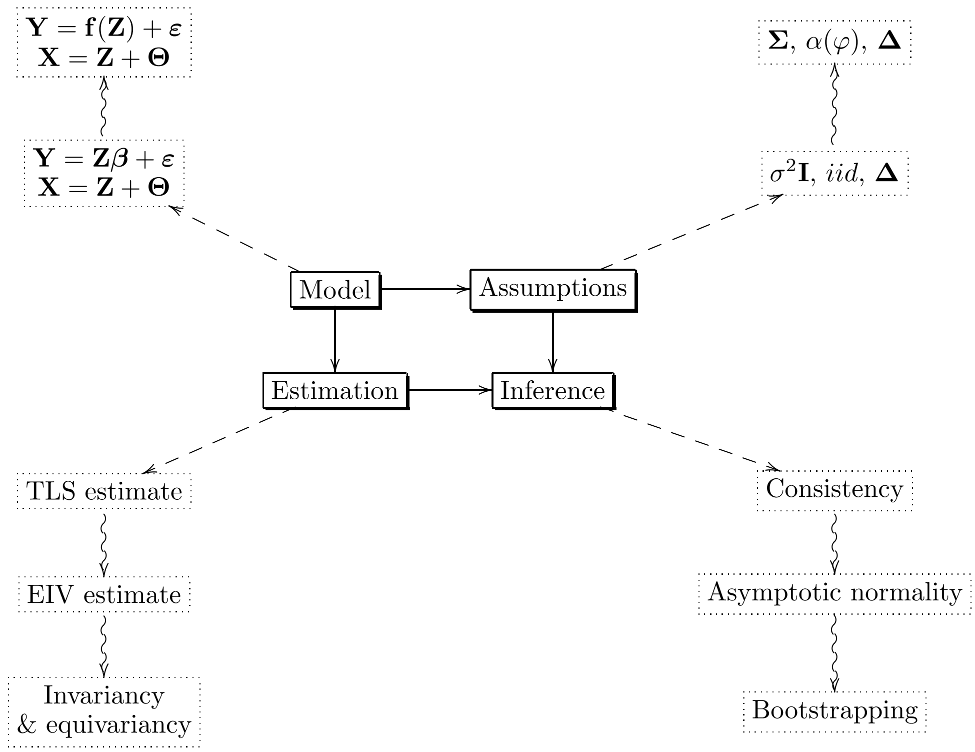

Figure 7.1: Ideological schema of statistical approaches in the doctoral thesis.

Appendix A

Useful Definitions and Theorems

Ja, maar niet te veel.
[Yes, but not too many.]

GERRIT ACHTERBERG

A.1 Additional Definitions

Definition A.1 (Unitary matrix)**.** A *unitary matrix* $\mathbf{A}$ is a square matrix satisfying $\mathbf{A}^\top\mathbf{A} = \mathbf{A}\mathbf{A}^\top = \mathbf{I}$.

Definition A.2 (Permutation matrix)**.** For a permutation $\pi : \{1, \ldots, p\} \to \{1, \ldots, p\}$, a *permutation matrix* is a square matrix $[\mathbf{e}_{\pi(1)}, \ldots, \mathbf{e}_{\pi(p)}]^\top$, where its rows are permuted canonical vectors.

Definition A.3 (Rotation matrix)**.** A *rotation matrix* is a unitary matrix whose determinant is equal to one.

Definition A.4 (Deterministic Landau symbols)**.** Let $\{a_n\}_{n=1}^\infty$ and $\{b_n\}_{n=1}^\infty$ be two sequences of real numbers. One writes

$$a_n = \mathcal{O}(b_n), \quad n \to \infty;$$

if and only if there exists a positive real number $M > 0$ and an integer $n_0 \in \mathbb{N}$ such that

$$|a_n| \leq M|b_n|, \quad \forall n \geq n_0.$$

One writes

$$a_n = o(b_n), \quad n \to \infty;$$

if and only if, for every positive real number $\tau > 0$, there exists an integer $n_0 \in \mathbb{N}$ such that

$$|a_n| \leq \tau |b_n|, \quad \forall n \geq n_0.$$

Definition A.5 (Stochastic Landau symbols)**.** Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of random variables and $\{a_n\}_{n=1}^{\infty}$ be a sequence of constants. One writes

$$X_n = \mathcal{O}_{\mathbb{P}}(a_n), \quad n \to \infty;$$

if and only if, for every positive real number $\epsilon > 0$, there exists a positive real number $M > 0$ and an integer $n_0 \in \mathbb{N}$ such that

$$\mathbb{P}\left[\left|\frac{X_n}{a_n}\right| > M\right] < \epsilon, \quad \forall n \geq n_0.$$

One writes

$$X_n = o_{\mathbb{P}}(a_n), \quad n \to \infty;$$

if and only if, for all positive real numbers $\epsilon > 0$ and $\tau > 0$, there exists an integer $n_0 \in \mathbb{N}$ such that

$$\mathbb{P}\left[\left|\frac{X_n}{a_n}\right| > \tau\right] < \epsilon, \quad \forall n \geq n_0.$$

A.2 Supplementary Theorems

Theorem A.1 (Eigen decomposition (spectral decomposition))**.** *Let* $\mathbf{P} \in \mathbb{R}^{n \times n}$ *be a matrix of eigenvectors of a given square matrix* $\mathbf{A} \in \mathbb{R}^{n \times n}$ *and* $\mathbf{W} \in \mathbb{R}^{n \times n}$ *be a diagonal matrix with the corresponding eigenvalues on the diagonal. Then, as long as* $\mathbf{P}$ *is a square matrix with full rank,* $\mathbf{A}$ *can be written as an eigen decomposition*

$$\mathbf{A} = \mathbf{P}\mathbf{W}\mathbf{P}^{-1}. \tag{A.1}$$

Furthermore, if $\mathbf{A}$ *is symmetric, then the columns of* $\mathbf{P}$ *are orthogonal vectors. If* $\mathbf{P}$ *is not a square matrix with full rank, then* $\mathbf{P}$ *cannot have a matrix inverse and* $\mathbf{A}$ *does not have an eigen decomposition.*

Theorem A.2 (Slutsky's)**.** *Let* $\{\boldsymbol{\xi}_n\}_{n=1}^{\infty}$ *and* $\{\boldsymbol{\zeta}_n\}_{n=1}^{\infty}$ *be sequences of scalar or vector or matrix random elements. If* $\boldsymbol{\xi}_n$ *converges in distribution to a random element* $\boldsymbol{\xi}$*, and* $\boldsymbol{\zeta}_n$ *converges in probability to a constant* $\mathbf{c}$*, then (for suitable dimensions)*

(i) $\boldsymbol{\xi}_n + \boldsymbol{\zeta}_n \xrightarrow[n\to\infty]{\mathscr{D}} \boldsymbol{\xi} + \mathbf{c}$*;*

(ii) $\boldsymbol{\zeta}_n \boldsymbol{\xi}_n \xrightarrow[n\to\infty]{\mathscr{D}} \mathbf{c}\boldsymbol{\xi}$;

(iii) $\boldsymbol{\zeta}_n^{-1} \boldsymbol{\xi}_n \xrightarrow[n\to\infty]{\mathscr{D}} \mathbf{c}^{-1}\boldsymbol{\xi}$, *provided that* $\boldsymbol{\zeta}_n$ *and* $\mathbf{c}$ *are invertible.*

Theorem A.3 (Cramér-Wold)**.** *Let*

$$\boldsymbol{\xi}_n = (\xi_{n,1}, \ldots, \xi_{n,k})^\top \quad \textit{and} \quad \boldsymbol{\xi} = (\xi_1, \ldots, \xi_k)^\top$$

be k*-dimensional random vectors. Then* $\boldsymbol{\xi}_n$ *converges to* $\boldsymbol{\xi}$ *in distribution if and only if*

$$\sum_{i=1}^{k} t_i \xi_{n,i} \xrightarrow{\mathscr{D}} \sum_{i=1}^{k} t_i \xi_i, \quad n \to \infty$$

for each $(t_1, \ldots, t_k)^\top \in \mathbb{R}^k$.

List of Procedures

List of Figures

List of Tables

Index

Bibliography

Amemiya, Y. (1997). Generalization of the TLS approach in the errors-in-variables problem. In Van Huffel, S., editor, *Proceedings of the Second International Workshop on Total Least Squares and Errors-in-Variables Modeling*, pages 77–86, Philadelphia, PA. SIAM.

Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis.* John Wiley & Sons, New York.

Andrews, D. W. K. (1984). Non-strong mixing autoregressive processes. *Journal of Applied Probability*, 21:930–934.

Belyaev, Y. K. (1995). Bootstrap, resampling and Mallows metric. Lecture Notes 1, Institute of Mathematical Statistics, Umeå University.

Belyaev, Y. K. and Sjöstedt-de Luna, S. (2000). Weakly approaching sequences of random distributions. *Journal of Applied Probability*, 37(3):807–822.

Bhatia, R. (1996). *Matrix Analysis. Volume 169 of Graduate Texts in Mathematics.* Springer-Verlag, New York.

Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Annals of Statistics*, 9(6):1196–1217.

Billingsley, P. (1968). *Convergence of Probability Measures.* John Wiley & Sons, New York, 1st edition.

Billingsley, P. (1999). *Convergence of Probability Measures.* John Wiley & Sons, New York, 2nd edition.

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

Chen, X. and Wu, Y. (1989). Strong law for mixing sequence. *Acta Mathematicae Applicatae Sinica*, 5(4):367 – 371.

Davidson, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press, New York.

Dobrushin, R. L. (1970). Prescribing a system of random variables by conditional distributions. *Theory of Probability and its Applications*, 15:458 – 486.

Doob, J. L. (1953). *Stochastic Processes*. John Wiley & Sons, New York.

Doukhan, P. (1994). *Mixing: Properties and Examples, Lecture Notes in Statistics, vol. 85*. Springer, New York.

Eckart, G. and Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrica*, 1:211 – 218.

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7(1):1 – 26.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*. SIAM, Philadelphia, PA.

Fazekas, I. and Kukush, A. (1997). Asymptotic properties of an estimator in nonlinear functional errors-in-variables models with dependent error terms. *Computers & Mathematics with Applications*, 34(10):23 – 39.

Fazekas, I., Kukush, A., and Zwanzig, S. (2004). Correction of nonlinear orthogonal regression estimator. *Ukrainian Mathematical Journal*, 56(8):1101 – 1118.

Fierro, R. D. and Bunch, J. R. (1994). Collinearity and total least squares. *SIAM Journal on Matrix Analysis and Applications*, 15(4):1167 – 1181.

Fitzenberger, B. (1997). The moving block bootstrap and robust inference for linear least squares and quantile regression. *Journal of Econometrics*, 82:235 – 287.

Gallant, A. R. and White, H. (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Basil Blackwell, New York.

Gallo, P. P. (1982a). Consistency of regression estimates when some variables are subject to error. *Communications in Statistics: Theory and Methods*, 11:973 – 983.

Gallo, P. P. (1982b). *Properties of Estimators in Errors-in-Variables Models*. PhD thesis, University of North Carolina, Chapel Hill, NC.

Gleser, L. J. (1981). Estimation in a multivariate “errors in variables” regression model: Large sample results. *Annals of Statistics*, 9:24 – 44.

Gleser, L. J. and Watson, G. S. (1973). Estimation of a linear transformation. *Biometrika*, 60:525 – 534.

Golub, G. H., Hoffman, A., and Stewart, G. W. (1987). A generalization of the Eckart-Young-Mirsky matrix approximation theorem. *Linear Algebra and Its Applications*, 88/89:322 – 327.

Golub, G. H. and Van Loan, C. F. (1980). An analysis of the total least squares problem. *SIAM Journal on Numerical Analysis*, 17(6):883 – 893.

Golub, G. H. and Van Loan, C. F. (1996). *Matrix Computation*. Johns Hopkins University Press, Baltimore, MD, 3rd edition.

Hahn, T. A. (1970). Thermal expansion of copper from 20 to 800K—standard reference material 736. *Journal of Applied Physics*, 41:5096 – 5101.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer, New York.

Healy, J. D. (1975). *Estimation and Tests for Unknown Linear Restrictions in Multivariate Linear Models*. PhD thesis, Purdue University.

Herrndorf, N. (1985). A functional central limit theorem for strongly mixing sequence of random variables. *Probability Theory and Related Fields*, 69(4):541 – 550.

Ibragimov, I. A. (1959). Some limit theorems for stochastic processes stationary in the strict sense. *Doklady Akademii Nauk SSSR (In Russian)*, 125:711 – 714.

Ibragimov, I. A. and Linnik, Y. V. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, The Netherlands.

Itō, K. (1993). *Encyclopedic Dictionary of Mathematics: The Mathematical Society of Japan*. MIT Press, Cambridge, MA, 2nd edition.

Kahaner, D., Moler, C., and Nash, S. (1989). *Numerical Methods and Software*. Englewood Cliffs, Prentice Hall, NJ.

Katz, M. L. (1963). Note on the Berry-Esseen theorem. *Annals Mathematical Statistics*, 34(3):1107 – 1108.

Kukush, A., Markovsky, I., and Van Huffel, S. (2002). On consistent estimators in linear and bilinear multivariate errors-in-variables models. In Van Huffel, S. and Lemmerling, P., editors, *Total Least Squares and Errors-in-Variables Modeling: Analysis, Algorithms and Applications*, pages 155 – 164, Philadelphia, PA. SIAM.

Kukush, A., Markovsky, I., and Van Huffel, S. (2005). Consistency of the structured total least squares estimator in a multivariate errors-in-variables model. *Journal of Statistical Planning and Inference*, 133(2):315 – 358.

Kukush, A. and Zwanzig, S. (2002). On consistent estimators in nonlinear functional errors-in-variables models. In Van Huffel, S. and Lemmerling, P., editors, *Total Least Squares and Errors-in-Variables Modeling: Analysis, Algorithms and Applications*, pages 145 – 154, Philadelphia, PA. SIAM.

Künsch, H. R. (1989). The jacknife and the bootstrap for general stationary observations. *Annals of Statistics*, 17:1217 – 1241.

Lahiri, S. N. (1992). Second order optimality of stationary bootstrap. In Lapage, R. and Billard, L., editors, *Exploring the Limits of Bootstrap*, pages 183 – 214, New York. Wiley.

Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*. Springer-Verlag, New York.

Lin, Z. and Lu, C. (1997). *Limit Theory for Mixing Dependent Random Variables*. Springer-Verlag, New York.

Liu, R. Y. and Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. In Lapage, R. and Billard, L., editors, *Exploring the Limits of Bootstrap*, pages 225 – 248, New York. Wiley.

Madansky, A. (1959). The fitting of straight lines when both variables are subject to error. *Journal of American Statistical Association*, 54:173 – 205.

Mirsky, L. (1960). Symmetric gauge functions and unitarily invariant norms. *Quarterly Journal of Mathematics*, 11:50 – 59.

Okamoto, M. (1973). Distinctness of the eigenvalues of a quadratic form in a multivariate sample. *Annals of Statistics*, 1:763 – 765.

Pešta, M. (2006a). Ideas and remarks on isotonic regression in Sobolev spaces. In Antoch, J. and Dohnal, G., editors, *Robust 2006 Conference Proceedings*, pages 253 – 260, Prague. JČMF.

Pešta, M. (2006b). Isotonic regression in Sobolev spaces. Master's thesis, Charles University in Prague, Faculty of Mathematics and Physics, Czech Republic.

Pešta, M. (2008). Total least squares approach in regression methods. In Šafránková, J. and Pavlů, J., editors, *WDS'08 Proceedings of Contributed Papers: Part I – Mathematics and Computer Sciences*, pages 88 – 93, Prague. MatfyzPress.

Pešta, M. (2009a). Strongly consistent estimation in dependent errors-in-variables. *Acta Universitatis Carolinae, Mathematica et Physica*. To appear.

Pešta, M. (2009b). Total least squares and bootstrapping with application in calibration. Submitted.

Politis, D. N. and Romano, J. P. (1992). A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation. *Annals of Statistics*, 20(4):1985 – 2007.

Rao, C. R. (1980). Matrix approximations and reduction of dimensionality in multivariate statistical analysis. In Krishnaiah, P. R., editor, *Multivariate Analysis*, volume 5, pages 3 – 22, Amsterdam. North-Holland.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences*, 42:43 – 47.

Rosenblatt, M. (1971). *Markov Processes: Structure and Asymptotic Behavior.* Springer-Verlag, Berlin.

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Annals of Statistics*, 9(6):1187 – 1195.

Thompson, R. C. (1972). Principal submatrices IX: Interlacing inequalities for singular values of submatrices. *Linear Algebra Applications*, 5:1 – 12.

Utev, S. A. (1990). The central limit theorem for φ-mixing arrays of random variables. *Theory of Probability and its Applications*, 35:131 – 139.

van der Vaart, A. W. (1998). *Asymptotic statistics.* Cambridge University Press, New York.

Van Huffel, S. and Vandewalle, J. (1991). *The Total Least Squares Problem: Computational Aspects and Analysis.* SIAM, Philadelphia, PA.

von Neumann, J. (1937). Some matrix inequalities and metrization of matrix-space. *Tomsk University Review*, 1:286 – 300.

Xuejun, W., Shuhe, H., Yan, S., and Wenzhi, Y. (2009). Moment inequalities for φ-mixing sequences and its applications. *Journal of Inequalities and Applications*, 2009(Article ID 379743):12 pages.

Yatchew, A. and Bos, L. (1997). Nonparametric least squares estimation and testing of economic models. *Journal of Quantitative Economics*, 13:81 – 131.

Yatchew, A. and Härdle, W. (2006). Nonparametric state price density estimation using constrained least squares and the bootstrap. *Journal of Econometrics*, 133(2):579 – 599.

Zagdaǹski, A. (2005). On the construction and properties of bootstrap-t prediction intervals for stationary time series. *Probability and Mathematical Statistics*, 25(1):133 – 153.