Necessary and Sufficient Conditions for Consistency of Resampling

Yuri K. Belyaev

Research Report
Centre of Biostochastics

Swedish University of Agricultural Sciences

Report 2003:1
ISSN 1651-8543

Necessary and Sufficient Conditions for Consistency of Resampling

Yuri K. Belyaev

Centre of Biostochastics,
Swedish University of Agricultural Sciences,
SE-901 83 Umeå

Abstract

A triangular array of independent non-identically distributed random variables is considered. The distribution functions of centered and rescaled sums of the random variables are estimated by resampling from the lists of their observed values. The estimators of distributions are called consistent (in probability) if they are weakly approaching the estimated distributions in probability, as the number of observations increases to infinity. Under some additional assumptions this type of consistency implies convergence in several metrics, e.g. in the uniform metric. A necessary and sufficient condition for consistency is given. In addition a new formulation of the Central Limit Theorem for triangular arrays, related to the notion of weakly approaching distribution functions, is stated. These results can be applied to justify the possibility of using resampling (bootstrap) techniques in many statistical applications, e.g. to justify the method of resampling from the list of weighted residuals in the case of a linear heteroscedastic regression.

Key words: triangular array, non-identically distributed random variables, central limit theorem, resampling, consistent estimation of distribution functions

1991 AMS subject classification: 60 F 05, 62 G 09, 62 E 20.

1 Introduction

The following statistical problem will be considered. There is a process of collecting statistical data $\{x_{1n}, ..., x_{nn}\}$, $n \to \infty$. The components $\{x_{hn}\}$ are values of independent non-identically distributed random variables (r.v.s). One can consider $\frac{1}{\rho_n}\sum_{h=1}^{n}(x_{hn} - u_{hn})$ as the observed value of the sum of the r.v.s, centered by $\{u_{hn}\}$ and rescaled by ρ_n. In many applications, it is useful to find estimators of the distributions of the centered and rescaled sums of r.v.s. The estimators have to be in some sense consistent. We will say that the estimators of distributions are (weakly) consistent if they are weakly approaching in probability to the estimated distributions. This type of consistency has been used in a paper by Belyaev (1995). An exact definition of this notion is given in the Appendix. Here, we say only that the notion of weakly approaching distributions is a convenient extension of weak convergence and that under some additional assumptions it implies convergence in the uniform metric and some other metrics, see Belyaev and Sjöstedt-de Luna (2000). We will find appropriate estimators by using resampling from the data $\{x_{1n}, ..., x_{nn}\}$. In this paper, we restrict ourselves to the case where the centered and rescaled sums of r.v.s are weakly approaching to the family of normal distributions. Our assumptions sufficient for consistency of considered estimators, are rather general, e.g. we avoid stating direct assumptions on existence of moments of random variables.

Theorems 1, 2 and 3 are the main results of this paper. Assumptions $\mathbf{A}_1$, $\mathbf{A}_2(\tau)$ and $\mathbf{A}_3(\tau)$ related to behavior of individual r.v.s, their sums, and expectations, respectively, are stated in the next section. Theorem 1 is an upgraded version of the well-known Central Limit Theorem where we use rather general assumptions $\mathbf{A}_1$ and $\mathbf{A}_2(\tau)$, which imply that the distributions of centered and rescaled sums of r.v.s are weakly approaching the family of normal distributions. Theorem 2 states that if assumptions $\mathbf{A}_1$ and $\mathbf{A}_2(\tau)$ hold, then $\mathbf{A}_3(\tau)$ is a necessary and sufficient condition that distributions of resampled, rescaled and centered sums based on the data $\{x_{1n}, ..., x_{nn}\}$, are weakly approaching in probability the distributions of related, centered and rescaled sums of r.v.s. In short, Theorem 2 says that resampling is valid for consistent estimation of the distribution of interest if and only if the assumption $\mathbf{A}_3(\tau)$ holds. Theorem 3 contains rather general assumptions which are sufficient for the usage of resampling in linear heteroscedastic regression.

The validity of resampling in the case of independent and identically distributed (i.i.d.) r.v.s has been carefully and exhaustively considered in a series of papers and books, for references see Hall (1992). In the case of i.i.d.

r.v.s the necessary condition for consistency of resampling has been obtained in Giné and Zinn (1989). The case of i.i.d. r.v.s with the infinite variance is quite different. It has been considered in the pioneer paper Athreya (1987). In the case of non-identically distributed independent r.v.s the situation is more complex. The most interesting results have been obtained in a paper by Mammen (1992) where the necessary and sufficient condition for validity of resampling has been found, in the case of a triangular array of independent real-valued r.v.s, and where the uniform distance between distributions has been used. The proofs in Mammen (1992) are rather long and many special results are used there. In our paper, quite different methods are used in the proofs related to the necessary and sufficient assumption $\mathbf{A}_3(\tau)$, which is close to that obtained by Mammen (1992). Instead of distributions convergent in the uniform metric we consider distributions weakly approaching in probability. Convergence in the uniform metric is obtained as a special case in Corollary 1 and Corollary 2. The methods suggested in this paper can easily be extended to the case of vector-valued r.v.s. One can encounter similar problems related to resampling if statistical data contain realisations of random processes, see e.g. Belyaev and Seleznjev (2000). The methods are rather simple because the notion used of weakly approaching distributions is directly connected with the pointwise convergence of the corresponding characteristic functions. Proofs of Theorems 1 and 2 are based on the analysis of characteristics functions.

The paper is organised as follows: In the next section we introduce most of the notation used and state the basic results. An application to heteroscedastic linear regression is given in Section 3. After that we present proofs of the stated results. The proofs are split into a series of lemmas. Several necessary definitions and propositions are presented in the Appendix. They are referred to the text as Definition 1, Proposition 1 etc. The notations $\overset{wa}{\leftrightarrow}$ and $\overset{wa(\mathrm{P})}{\longleftrightarrow}$ are also explained in the Appendix.

2 Basic results

We use the following notation: Capital letters denote r.v.s, E and P are symbols of expectation and probability, respectively, $\overset{\mathrm{P}}{\to}$ means convergence in probability, $\mathcal{L}(X)$ is the distribution law (d.l.) of an r.v. X. $\mathrm{I}(\mathcal{A})$ is the indicator of an event $\mathcal{A}$. By $\{\rho_n\}_{n\geq 1}$ we denote a sequence of positive non-random values which will be used to norm (rescale) r.v.s. Let $\mathbb{X} = \{X_{hn} : \{h, n\} \in \mathcal{T}\}$ be a triangular array of real r.v.s $\{X_{hn}\}$ which are independent for each $n = 2, 3, ...$, $\mathcal{T} = \{\{h, n\} : n = 1, 2, ..., h = 1, 2, ..., n\}$, $\mathbf{X}_n = \{X_{1n}, ..., X_{nn}\}$.

The r.v.s $X_{hn},\ h=1,...,n$, can be non-identically distributed. We consider the normed and shifted or centered r.v.s $Y_{hn} = X_{hn}/\rho_n$.
Let $Y_{hn}(\tau) = Y_{hn}\mathrm{I}(\mid Y_{hn} \mid \leq \tau),\ Y^{\ominus}_{hn}(\tau) = Y_{hn} - \mathrm{E}[Y_{hn}(\tau)]$,
$Y^{\ominus}_{\cdot n}(\tau) = \sum_{h=1}^{n} Y^{\ominus}_{hn}(\tau),\ Y^{\circ}_{hn}(\tau) = Y_{hn}(\tau) - \mathrm{E}[Y_{hn}(\tau)],\ Y^{\circ}_{\cdot n}(\tau) = \sum_{h=1}^{n} Y^{\circ}_{hn}(\tau)$, $\sigma^2_{\cdot n}(\tau) = \mathrm{Var}\,[Y^{\circ}_{\cdot n}(\tau)] = \sum_{h=1}^{n} \mathrm{E}[(Y^{\circ}_{hn}(\tau))^2]$. We will use notation $Y^{\circ}_{hn} = Y_{hn} - \mathrm{E}[Y_{hn}],\ Y^{\circ}_{\cdot n} = \sum_{h=1}^{n} Y^{\circ}_{hn}$ and $\sigma^2_{\cdot n} = \sum_{h=1}^{n} \mathrm{E}[(Y^{\circ}_{hn})^2]$ if $\mathrm{E}[Y^2_{hn}] < \infty,\ \{h,n\} \in \mathcal{T}$. We will also consider two lists of r.v.s

$$\mathbf{Y}^{\odot}_n = \{Y_{1n} - \bar{Y}_{\cdot n}, ..., Y_{nn} - \bar{Y}_{\cdot n}\},$$

and

$$\mathbf{Y}^{\odot}_n(\tau) = \{Y_{1n}(\tau) - \bar{Y}_{\cdot n}(\tau), ..., Y_{nn}(\tau) - \bar{Y}_{\cdot n}(\tau)\},$$

where

$$Y_{\cdot n} = \sum_{h=1}^{n} Y_{hn},\ \bar{Y}_{\cdot n} = Y_{\cdot n}/n,\ Y_{\cdot n}(\tau) = \sum_{h=1}^{n} Y_{hn}(\tau),\ \bar{Y}_{\cdot n}(\tau) = Y_{\cdot n}(\tau)/n,\ \tau > 0.$$

Let $\mathbf{J}^{\star}_n = \{J^{\star}_{1n}, ..., J^{\star}_{nn}\}$, be n i.i.d. r.v.s uniformly distributed on $\{1, 2, ..., n\}$, i.e. $\mathrm{P}^{\star}[J^{\star}_{in} = h] = 1/n,\ h = 1, ..., n$. We use the mark "$\star$" to show that r.v.s, probabilities and expectations are related to $\mathbf{J}^{\star}_n$. Values of r.v.s $J^{\star}_{in}$ can be obtained by simulation. Resampling copies of lists $\mathbf{Y}^{\odot}_n$ and $\mathbf{Y}^{\odot}_n(\tau)$, obtained via simulation of the r.v.s $\mathbf{J}^{\star}_n$, will be denoted as follows

$$\mathbf{Y}^{\star\odot}_n = \{Y^{\star\odot}_{1n}, ..., Y^{\star\odot}_{nn}\}, \quad \mathbf{Y}^{\star\odot}_n(\tau) = \{Y^{\star\odot}_{1n}(\tau), ..., Y^{\star\odot}_{nn}(\tau),\}$$

where $Y^{\star\odot}_{hn} = Y^{\star}_{hn} - \bar{Y}_{\cdot n},\ Y^{\star}_{hn} = Y_{J^{\star}_{hn},n},\ Y^{\star\odot}_{hn}(\tau) = Y^{\star}_{hn}(\tau) - \bar{Y}_{\cdot n}(\tau)$,
$Y^{\star}_{hn}(\tau) = Y_{J^{\star}_{hn},n}(\tau)$.

Let $N^{\star}_{hn} = \sum_{i=1}^{n} \mathrm{I}(J^{\star}_{in} = h)$. Then, we can write the sums of components in $\mathbf{Y}^{\star\odot}_n$ and $\mathbf{Y}^{\star\odot}_n(\tau)$ as follows

$$Y^{\star\odot}_{\cdot n} = \sum_{h=1}^{n} (N^{\star}_{hn} - 1) Y_{hn}, \quad Y^{\star\odot}_{\cdot n}(\tau) = \sum_{h=1}^{n} (N^{\star}_{hn} - 1) Y_{hn}(\tau). \tag{1}$$

We introduce the following assumptions:
There exists a sequence $\{\rho_n\}_{n\geq 1}$, *where every* $\rho_n > 0$ *and, a* $\tau > 0$ *such that*

$\mathbf{A}_1 :\quad \max_{1\leq h\leq n} \mid Y_{hn} \mid \xrightarrow{\mathrm{P}} 0, \quad n \to \infty;$

$\mathbf{A}_2(\tau) :\quad$ *the sequence of d.l.s* $\{\mathcal{L}(Y^{\ominus}_{\cdot n}(\tau))\}_{n\geq 1}$ *is tight;*

$\mathbf{A}_3(\tau) :\quad \sum_{h=1}^{n} (\mathrm{E}[Y_{hn}(\tau)])^2 - \frac{1}{n}\Big(\sum_{h=1}^{n} \mathrm{E}[Y_{hn}(\tau)]\Big)^2 \to 0,\ \ n \to \infty.$

$\mathbf{A}_1$ holds if and only if for every $\varepsilon > 0$

$$\sum_{h=1}^{n} \mathrm{P}[\mid Y_{hn} \mid > \varepsilon] \to 0, \quad n \to \infty.$$

Theorem 1. (The Central Limit Theorem (CLT) for triangular array) *Suppose that* $\mathbf{A}_1$ *holds. Then*

$$\mathcal{L}(Y_{\cdot n}^{\ominus}(\tau)) \overset{wa}{\leftrightarrow} \mathcal{N}_1(0, \sigma_n^2), \quad n \to \infty, \tag{2}$$

where $\{\sigma_n^2\}_{n \geq 1}$ *is a sequence of positive numbers,* $\sigma_+^2 = \sup_n \sigma_n^2 < \infty$, *if and only if* $\mathbf{A}_2(\tau)$ *holds.*

Theorem 1 says that if $\mathbf{A}_1$ holds , then the d.l.s of the sums $Y_{\cdot n}^{\ominus}(\tau)$ are weakly approaching (see Definition 3) the family of normal d.l.s $\mathfrak{N}_0(\sigma_+^2) = \{\mathcal{N}_1(0, \sigma^2): \ 0 \leq \sigma^2 \leq \sigma_+^2 < \infty\}$ if and only if $\mathbf{A}_2(\tau)$ holds, $\mathcal{N}_1(0,0) = \mathcal{L}(0)$.

Theorem 2. (The Central Limit Resampling Theorem (CLRT) for triangular array). *If* $\mathbf{A}_1$ *and* $\mathbf{A}_2(\tau)$ *hold, then* $\mathbf{A}_3(\tau)$ *is necessary and sufficient to give an asymptotically consistent estimation of* $\mathcal{L}(Y_{\cdot n}^{\ominus}(\tau))$, *as* $n \to \infty$, *based on resampling, i.e.*

$$\mathcal{L}(Y_{\cdot n}^{\star\odot} \mid \mathbf{X}_n) \overset{wa(\mathrm{P})}{\longleftrightarrow} \mathcal{L}(Y_{\cdot n}^{\ominus}(\tau)), \quad n \to \infty. \tag{3}$$

Theorem 2 states that the conditional d.l.s of the sums $Y_{\cdot n}^{\star\odot}$ of r.v.s in $\mathbf{Y}_n^{\star\odot}$, obtained by resampling from the list $\mathbf{Y}_n^{\odot}$, are weakly approaching in probability (see Definition 4) to the conditional d.l.s of the sums $Y_{\cdot n}^{\ominus}(\tau)$ given $\{\mathbf{X}_n\}_{n\geq 1}$ as $n \to \infty$.

Remark. We will prove Theorem 1 and Theorem 2 with $\sigma_n^2 = \sigma_{\cdot n}^2(\tau)$. It will be shown that $\sigma_{\cdot n}^2(\tau)$ does not essentially depend on τ, as $n \to \infty$, if $\mathbf{A}_1$ holds. If both $\mathbf{A}_2(\tau_1)$ and $\mathbf{A}_2(\tau_2)$ hold, $0 < \tau_1 < \tau_2 < \infty$, then $\mathcal{L}(Y_{\cdot n}^{\ominus}(\tau_1)) \overset{wa}{\leftrightarrow} \mathcal{L}(Y_{\cdot n}^{\ominus}(\tau_2))$, $n \to \infty$. If $\mathbf{A}_1$ holds and for a $\tau_0 > 0$, $\sum_{h=1}^{n}(\mathrm{E}[Y_{hn}(\tau_0)])^2 \to 0$, $n \to \infty$, then $\mathbf{A}_3(\tau)$ holds for any $\tau > 0$.

Let $F_n(x) = \mathrm{P}[Y_{\cdot n}^{\ominus}(\tau) \leq x]$ and $F_n^{\star}(x \mid \mathbf{X}_n) = \mathrm{P}[Y_{\cdot n}^{\star\odot} \leq x \mid \mathbf{X}_n]$ be the distribution functions (d.f.s) of the r.v.s $Y_{\cdot n}^{\ominus}(\tau)$, and the conditional d.f.s of $Y_{\cdot n}^{\star\odot}$ given $\mathbf{X}_n$, respectively, and $\Phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

Corollary 1. *If* $\mathbf{A}_1$ *and* $\mathbf{A}_2(\tau)$ *hold and* $\underline{\lim}_{n\to\infty}\sigma_n^2 > 0$, *then*

$$\sup_x \left| F_n(x) - \Phi\left(\frac{x}{\sigma_n}\right) \right| \to 0, \quad n \to \infty.$$

Corollary 2. *If* $\mathbf{A}_1$, $\mathbf{A}_2(\tau)$ *and* $\mathbf{A}_3(\tau)$ *hold and* $\underline{\lim}_{n\to\infty}\sigma_n^2 > 0$, *then*

$$\sup_x \left| F_n^\star(x \mid \mathbf{X}_n) - \Phi\left(\frac{x}{\sigma_n}\right)\right| \stackrel{\mathrm{P}}{\to} 0, \quad n \to \infty,$$

and correspondingly

$$\sup_x \mid F_n^\star(x \mid \mathbf{X}_n) - F_n(x) \mid \stackrel{\mathrm{P}}{\to} 0, \quad n \to \infty.$$

The case, where all r.v.s have the finite second moments is interesting in many applications.

Corollary 3. *Suppose that all second order moments* $\mathrm{E}[Y_{hn}^2] < \infty$, $\sup_n \sum_{h=1}^n \mathrm{E}[Y_{hn}^2] = C_2 < \infty$, $\{h, n\} \in \mathcal{T}$, $\sum_{h=1}^n (\mathrm{E}[Y_{hn}])^2 \to 0$, $n \to \infty$, *and for every* $\tau > 0$

$$\sum_{h=1}^n \mathrm{E}[(Y_{hn})^2 \mathrm{I}(\mid Y_{hn} \mid > \tau)] \to 0, \quad n \to \infty. \tag{4}$$

Then

$$\mathcal{L}(Y_{\cdot n}^{\circ}) \stackrel{wa}{\leftrightarrow} \mathcal{N}_1(0, \sigma_{\cdot n}^2), \quad n \to \infty, \tag{5}$$

and

$$\mathcal{L}(Y_{\cdot n}^{\star\odot} \mid \mathbf{X}_n) \stackrel{wa(\mathrm{P})}{\longleftrightarrow} \mathcal{L}(Y_{\cdot n}^{\circ}), \quad n \to \infty. \tag{6}$$

Here, the Lindeberg assumption (4) is used in order to have $\sigma_{\cdot n}^2$ in (5), instead of a σ_n^2 as it stated in Theorem 1. In the general case, σ_n^2 can essentially differ from $\sigma_{\cdot n}^2$ because, then we do not assume that the second moments are finite.

3 Application to heteroscedastic linear regression

Suppose that statistical data are the list of pairs $\{\{Y_1, \{\tilde{x}_{11}, ..., \tilde{x}_{r1}\}\}, ..., \{Y_n, \{\tilde{x}_{1n}, ..., \tilde{x}_{rn}\}\}\}$ with components satisfying the following relations of a linear heteroscedastic regression:

$$Y_h = \sum_{s=1}^r \tilde{x}_{sh}\beta_{s0} + W_h, \quad h = 1, ..., n, \tag{7}$$

where errors $\{W_h\}$ are independent r.v.s, $\{\tilde{x}_{sh}\}$ are explanatory variables (regressors), $\{\beta_{s0}\}$ are components of a parameter, and $\{Y_h\}$ are responses.

The errors can be non-identically distributed. We can rewrite (7) by using vector-matrix form

$$\mathbf{Y}_n = \tilde{\mathbb{X}}_n \boldsymbol{\beta}_0 + \mathbf{W}_n,$$

where $\mathbf{Y}_n = \{Y_1, ..., Y_n\}^T$, $\tilde{\mathbb{X}}_n = [x_{hs}]$, $x_{hs} = \tilde{x}_{sh}$, $\tilde{\mathbb{X}}_n^T = [\tilde{x}_{sh}]$, $s = 1, ..., r$, $h = 1, ..., n$, $\boldsymbol{\beta}_0 = \{\beta_{10}, ..., \beta_{r0}\}^T$, $\mathbf{W}_n = \{W_1, ..., W_n\}^T$, "^{T}" denotes transposition. The columns of the matrix $\tilde{\mathbb{X}}_n^T = [\tilde{\mathbf{x}}_1, ..., \tilde{\mathbf{x}}_n]$, can be considered as column-vectors $\tilde{\mathbf{x}}_h = \{\tilde{x}_{1h}, ..., \tilde{x}_{rh}\}^T$. Let $(\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+$ be the Moor-Penrose inverse matrix of $\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n$ and $\mathrm{tr}(\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+$ be its trace. If $\mathrm{rank}(\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n) = r$, then $(\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+ = (\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^{-1}$. Let $\mathcal{M}(\tilde{\mathbb{X}}_n^T) \subset \mathbf{R}^r$ be the linear space generated by vectors $\tilde{\mathbf{x}}_1, ..., \tilde{\mathbf{x}}_n$, and $\| \mathbf{a} \|^2 = \mathbf{a}^T \mathbf{a}$, $\mathbf{a} \in \mathbf{R}^r$.

Suppose that $\mathbf{c}^T \boldsymbol{\beta}_0$ is the parameter of interest and $\mathbf{c} \in \mathcal{M}(\tilde{\mathbb{X}}_n^T)$. Note, that

$$| \mathbf{a}_1^T (\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+ \mathbf{a}_2 | \leq \| \mathbf{a}_1 \| \| \mathbf{a}_2 \| \mathrm{tr}(\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+, \ \mathbf{a}_1, \mathbf{a}_2 \in \mathbf{R}^r. \tag{8}$$

We introduce the following assumptions:

$\mathbf{HR}_1$: *All values of explanatory variables are uniformly bounded,* $x_+ = \sup_{s,h} | \tilde{x}_{sh} | < \infty$, *and for every* $\varepsilon > 0$ *and a given* $\mathbf{c} \in \mathcal{M}(\tilde{\mathbb{X}}_n^T)$ *it holds*:

$$(i) \ \ t(\mathbf{c}, \tilde{\mathbb{X}}_n) = \frac{\mathrm{tr}(\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+}{\sqrt{\mathbf{c}^T (\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+ \mathbf{c}}} \to 0, \quad n \to \infty,$$

$$(ii) \ \ \sum_{h=1}^{n} \mathrm{E} \left[t^2(\mathbf{c}, \tilde{\mathbb{X}}_n) W_h^2 \mathrm{I}(t(\mathbf{c}, \tilde{\mathbb{X}}_n) \mid W_h \mid > \varepsilon) \right] \to 0, \quad n \to \infty;$$

$\mathbf{HR}_2$: *There are two constants* $0 < \sigma_-^2 \leq \sigma_+^2 < \infty$ *such that*

$$\sigma_-^2 \leq \sigma_h^2(\tilde{\mathbf{x}}_h) = \mathrm{E}[W_h^2] \leq \sigma_+^2, \ \text{ and } \ \mathrm{E}[W_h] = 0, \ h = 1, 2, ..., n;$$

$$\mathbf{HR}_3 : \qquad \frac{n(\mathrm{tr}(\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+)^3}{\mathbf{c}^T (\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+ \mathbf{c}} \to 0, \quad n \to \infty.$$

If $\mathbf{c}^T (\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+ \mathbf{c} \to 0$, $n \to \infty$, and $\mathbf{HR}_2$ holds, then the ordinary least-squares (OLS-)estimator $(\mathbf{c}^T \boldsymbol{\beta}_0)_n^\wedge = \mathbf{c}^T (\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+ \tilde{\mathbb{X}}_n^T \mathbf{Y}_n$ is unbiased and consistent. It is essential to know how accurate this estimator is. The problem has been thoroughly discussed in a paper by Wu (1986). It is interesting to estimate the d.l. of the OLS-estimator's deviations, e.g. the d.l. $\mathcal{L}\left(\frac{(\mathbf{c}^T \boldsymbol{\beta}_0)_n^\wedge - \mathbf{c}^T \boldsymbol{\beta}_0}{\sqrt{\mathbf{c}^T (\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+ \mathbf{c}}} \right)$. Here, the theory of linear regression with i.i.d

$W_h,\ h = 1, 2, ...$ can not be used. However, the consistent estimation of the d.l. can be realised by using the resampling (bootstrap) from weighted residuals. Let $\hat{Y}_h = \tilde{\mathbf{x}}_h^T(\tilde{\mathbb{X}}_n^T\tilde{\mathbb{X}}_n)^+\tilde{\mathbb{X}}_n^T\mathbf{Y}_n$ be the predicted value of $Y_h,\ h = 1, ..., n$. We can consider a resampling copy $\{\{\hat{Y}_1^\star - Y_1^\star, \tilde{\mathbf{x}}_1^\star\}, ..., \{\hat{Y}_n^\star - Y_n^\star, \tilde{\mathbf{x}}_n^\star\}\}$ of the list of residuals paired with vectors of explanatory variables $\{\{\hat{Y}_1 - Y_1, \tilde{\mathbf{x}}_1\}, ..., \{\hat{Y}_n - Y_n, \tilde{\mathbf{x}}_n\}\}$ where $Y_h^\star = Y_{J_{hn}^\star}$, $\hat{Y}_h^\star = \hat{Y}_{J_{hn}^\star}$, and, $\tilde{\mathbf{x}}_h^\star = \tilde{\mathbf{x}}_{J_{hn}^\star}$, $h = 1, ..., n$. Let

$$U_n^\star(\mathbf{c}, \mathbf{Y}_n, \tilde{\mathbb{X}}_n) = \sum_{h=1}^{n}(N_{hn}^\star - 1)\frac{\mathbf{c}^T(\tilde{\mathbb{X}}_n^T\mathbb{X}_n)^+\tilde{\mathbf{x}}_h}{(\mathbf{c}^T(\tilde{\mathbb{X}}_n^T\tilde{\mathbb{X}}_n)^+\mathbf{c})^{1/2}}(Y_h - \hat{Y}_h). \tag{9}$$

The r.v. $U_n^\star(\mathbf{c}, \mathbf{Y}_n, \tilde{\mathbb{X}}_n)$ is the centered sum of resampled and weighted residuals. Here, we justify usage of $\mathcal{L}(U_n^\star(\mathbf{c}, \mathbf{Y}_n, \tilde{\mathbb{X}}_n) \mid \mathbf{Y}_n)$ as a consistent estimator of the d.l. of normed deviations $(\mathbf{c}^T\boldsymbol{\beta}_0)_n^\wedge - \mathbf{c}^T\boldsymbol{\beta}_0$.

Theorem 3. *Suppose that errors $\{W_h\}_{h\geq 1}$ are independent, and that assumptions $\mathbf{HR}_1$ and $\mathbf{HR}_2$ hold. Then the normed d.l.s of deviations of the unbiased estimators $(\mathbf{c}^T\boldsymbol{\beta}_0)_n^\wedge$ are weakly approaching a tight family of normal d.l.s*

$$\mathcal{L}\left(\frac{(\mathbf{c}^T\boldsymbol{\beta}_0)_n^\wedge - \mathbf{c}^T\boldsymbol{\beta}_0}{\sqrt{\mathbf{c}^T(\tilde{\mathbb{X}}_n^T\tilde{\mathbb{X}}_n)^+\mathbf{c}}}\right) \overset{wa}{\leftrightarrow} \mathcal{N}_1(0, \sigma_n^2(\mathbf{c}, \tilde{\mathbb{X}}_n)), \quad n \to \infty, \tag{10}$$

where

$$\sigma_n^2(\mathbf{c}, \tilde{\mathbb{X}}_n) = \mathrm{E}\left[\left(\frac{(\mathbf{c}^T\boldsymbol{\beta}_0)_n^\wedge - \mathbf{c}^T\boldsymbol{\beta}_0}{\sqrt{\mathbf{c}^T(\tilde{\mathbb{X}}_n^T\tilde{\mathbb{X}}_n)^+\mathbf{c}}}\right)^2\right].$$

If in addition $\mathbf{HR}_3$ holds then

$$\mathcal{L}(U_n^\star(\mathbf{c}, \mathbf{Y}_n, \tilde{\mathbb{X}}_n) \mid \mathbf{Y}_n) \overset{wa(\mathrm{P})}{\longleftrightarrow} \mathcal{L}\left(\frac{(\mathbf{c}^T\boldsymbol{\beta}_0)_n^\wedge - \mathbf{c}^T\boldsymbol{\beta}_0}{\sqrt{\mathbf{c}^T(\tilde{\mathbb{X}}_n^T\tilde{\mathbb{X}}_n)^+\mathbf{c}}}\right), \quad n \to \infty. \tag{11}$$

Relation (11) means that under the stated assumptions, the resampling of weighted residuals gives consistency in asymptotic approximation of the d.l. of normed deviations of the OLS-estimator $(\mathbf{c}^T\boldsymbol{\beta}_0)_n^\wedge$ in the case of linear heteroscedastic regression. The assumptions stated in Theorem 3 also imply convergence in uniform metric.

Corollary 4. *If errors $\{W_n\}_{h\geq 1}$ are independent, and assumptions $\mathbf{HR}_1$, $\mathbf{HR}_2$ and $\mathbf{HR}_3$ hold then*

$$\sup_{z\in\mathbf{R}^1}\left|F_{U_n^\star}(z \mid \mathbf{Y}_n) - \Phi\left(\frac{z}{\sigma_n(\mathbf{c}, \tilde{\mathbb{X}}_n)}\right)\right| \xrightarrow{\mathrm{P}} 0, \ \ n \to \infty.$$

4 Proofs

Lemma 1. *If* $\mathbf{A}_1$ *holds, then for any* $\tau > 0$

$$(i)\ \max_{1\le h\le n} \mid Y_{hn}(\tau) \mid \xrightarrow{\mathrm{P}} 0,\ (ii)\ \max_{1\le h\le n} \mid \mathrm{E}[Y_{hn}(\tau)] \mid \to 0,$$
$$(iii)\ \max_{1\le h\le n} \mid Y^{\circ}_{hn}(\tau) \mid \xrightarrow{\mathrm{P}} 0,\ (iv)\ \max_{1\le h\le n} \mathrm{E}[(Y^{\circ}_{hn}(\tau))^2] \to 0,\ n\to\infty. \quad (12)$$

Proof. Relation (12)(i) is fulfilled because for any $\tau > 0$ $\mid Y_{hn}(\tau) \mid \le \mid Y_{hn} \mid$. For $\tau > 0$ and any $\varepsilon < \tau$

$$\mathrm{E}[Y_{hn}(\tau)] = \mathrm{E}[Y_{hn}(\varepsilon)] + \mathrm{E}[Y_{hn}\mathrm{I}(\varepsilon < \mid Y_{hn} \mid \le \tau)].$$

Therefore,

$$\mid \max_{1\le h\le n} \mathrm{E}[Y_{hn}(\tau)] \mid \le \varepsilon + \tau \mathrm{P}[\max_{1\le h\le n} \mid Y_{hn} \mid > \varepsilon].$$

By $\mathbf{A}_1$ for all sufficiently large $n \ge n(\varepsilon,\tau)$, we obtain $\mid \max_{1\le h\le n} \mathrm{E}[Y_{hn}(\tau)] \mid \le 2\varepsilon$. We can take an arbitrary small $\varepsilon > 0$. Hence, (12)(ii) is valid. Relations (12) (iii) and (iv) can be proved similarly. □

Lemma 2. *If* $\mathbf{A}_1$ *and* $\mathbf{A}_2(\tau)$ *hold, then the sequence of d.l.s* $\{\mathcal{L}(Y^{\circ}_{\cdot n}(\tau))\}_{n\ge 1}$ *is tight and*

$$\mathcal{L}(Y^{\circ}_{\cdot n}(\tau)) \overset{wa}{\leftrightarrow} \mathcal{L}(Y^{\ominus}_{\cdot n}(\tau)), \quad n\to\infty.$$

Proof. We have

$$Y^{\ominus}_{\cdot n}(\tau) - Y^{\circ}_{\cdot n}(\tau) = \sum_{h=1}^{n} Y_{hn}\mathrm{I}(\mid Y_{hn} \mid > \tau),$$

and for the following events it holds that

$$\{\max_{1\le h\le n} \mid Y_{hn} \mid \le \tau\} \subseteq \left\{ \sum_{h=1}^{n} Y_{hn}\mathrm{I}(\mid Y_{hn} \mid > \tau) = 0 \right\}. \quad (13)$$

Therefore, we have

$$\mathrm{P}\left[\sum_{h=1}^{n} Y_{hn}\mathrm{I}(\mid Y_{hn} \mid > \tau) = 0\right] \ge \mathrm{P}[\max_{1\le h\le n} \mid Y_{hn} \mid \le \tau] \to 1, \quad n\to\infty.$$

It implies that $Y^{\circ}_{\cdot n}(\tau) = Y^{\ominus}_{\cdot n}(\tau) + o_p(1),\ \ n\to\infty$. The desired result follows from Propositions 6 and 9. □

Lemma 3. *If* $\mathbf{A}_1$ *and* $\mathbf{A}_2(\tau)$ *hold, then*

$$\sup_n \sigma^2_{\cdot n}(\tau) < \infty. \tag{14}$$

Proof. Suppose on the contrary that $\overline{\lim}_{n\to\infty}\sigma^2_{\cdot n}(\tau) = \infty$, i.e. there exists a subsequence $\{n_k\}_{k\geq 1}$ such that $\sigma_{\cdot n_k}(\tau) \to \infty$. For each $\{h, n\}$ and $\sigma^2_{\cdot n}(\tau) \neq 0$, we consider the normed r.v.s $Z^{\circ}_{hn_k}(\tau) = Y^{\circ}_{hn_k}(\tau)/\sigma_{\cdot n_k}(\tau)$. We have from Lemma 1 that for all sufficiently large n_k

$$\max_{1\leq h\leq n_k} \mid Z^{\circ}_{hn_k}(\tau) \mid \leq \max_{1\leq h\leq n_k} \mid Y^{\circ}_{hn_k}(\tau) \mid \xrightarrow{\mathrm{P}} 0, \quad k \to \infty.$$

It follows that

(i) $\sum_{h=1}^{n_k} \mathrm{P}[\mid Z^{\circ}_{hn_k}(\tau) \mid > \varepsilon] \to 0, \;\; n \to \infty, \quad$ for every $\quad \varepsilon > 0;$

In addition, we have

(ii) $\sum_{h=1}^{n_k} \mathrm{E}[Z^{\circ}_{hn_k}(\tau)] = 0; \quad (iii) \sum_{h=1}^{n_k} \mathrm{E}[(Z^{\circ}_{hn_k}(\tau))^2] = 1, \;\; k = 1, 2, \ldots.$

Hence, by the Normal Convergence Criterion (Proposition 2) we have

$$\mathcal{L}(Z^{\circ}_{hn_k}(\tau)) \xrightarrow{w} \mathcal{N}_1(0, 1), \quad k \to \infty,$$

and for any k such that $\sigma^2_{n_k}(\tau) > 0$, it follows that

$$\mathrm{P}[\mid Y^{\circ}_{hn_k}(\tau) \mid > \sigma_{\cdot n_k}(\tau)] = \mathrm{P}[\mid Z^{\circ}_{hn_k}(\tau) \mid > 1] \to 1 - \frac{1}{\sqrt{2\pi}} \int_{-1}^{+1} e^{-x^2/2} dx > \frac{1}{4},$$

$k \to \infty$.

For any large m there exists $k(m)$ such that for all $k > k(m)$, $\sigma_{\cdot n_k}(\tau) > m$. Therefore, we have $\mathrm{P}[\mid Y^{\circ}_{hn_k}(\tau) \mid > m] > 1/4$ for all sufficiently large k and thus $\{\mathcal{L}(Y^{\circ}_{hn_k}(\tau)\}_{k\geq 1}$ is not tight. This contradicts the result of Lemma 2. Hence, (14) holds. □

Lemma 4. *Suppose that* $\mathbf{A}_1$ *and* $\mathbf{A}_2(\tau_0)$ *hold, for some* $\tau_0 > 0$*. Then* $\mathbf{A}_2(\tau)$ *holds for any* $\tau > 0$*.*

Proof. We have

$$Y^{\ominus}_{\cdot n}(\tau) = Y^{\ominus}_{\cdot n}(\tau_0) + \sum_{h=1}^{n} \Big(\mathrm{E}[Y_{hn}\mathrm{I}(\mid Y_{hn} \mid \leq \tau_0)] - \mathrm{E}[Y_{hn}\mathrm{I}(\mid Y_{hn} \mid \leq \tau)] \Big). \tag{15}$$

From $\mathbf{A}_1$ it follows that if $\tau < \tau_0$, then

$$\left|\sum_{h=1}^{n} \mathrm{E}[Y_{hn}\mathrm{I}(\tau <\mid Y_{hn} \mid\leq \tau_0)]\right| \leq \tau_0 \sum_{h=1}^{n} \mathrm{P}[\mid Y_{hn} \mid > \tau] \to 0 \quad n \to \infty. \tag{16}$$

If $\tau > \tau_0$, then we need only to write τ instead of τ_0 and τ_0 instead of τ in (16). Therefore, $Y^{\ominus}_{\cdot n}(\tau) = Y^{\ominus}_{\cdot n}(\tau_0) + o(1)$, $n \to \infty$, which implies $\mathbf{A}_2(\tau)$. $\square$

Lemma 5. *If* $\mathbf{A}_1$ *holds, then for any* $0 < \tau_1 < \tau_2 < \infty$ *it follows that*

$$\mid \sigma^2_{\cdot n}(\tau_1) - \sigma^2_{\cdot n}(\tau_2) \mid \to 0, \quad n \to \infty. \tag{17}$$

Proof. We can write

$$\begin{aligned} & \sigma^2_{\cdot n}(\tau_1) - \sigma^2_{\cdot n}(\tau_2) \\ = & \sum_{h=1}^{n} \Big(\mathrm{E}[(Y^{\circ}_{hn}(\tau_1))^2] - \mathrm{E}[(Y^{\circ}_{hn}(\tau_2))^2] \Big) = \sum_{h=1}^{n} \mathrm{E}[Y^2_{hn}\mathrm{I}(\tau_1 <\mid Y_{hn} \mid\leq \tau_2)] \\ + & \sum_{h=1}^{n} \mathrm{E}[Y_{hn}\mathrm{I}(\tau_1 <\mid Y_{hn} \mid\leq \tau_2)]\mathrm{E}[Y_{hn}(\mathrm{I}(\mid Y_{hn} \mid\leq \tau_1) + \mathrm{I}(\mid Y_{hn} \mid\leq \tau_2))]. \end{aligned}$$

Here, we have

$$\left|\sum_{h=1}^{n} \mathrm{E}[Y^2_{hn}\mathrm{I}(\tau_1 <\mid Y_{hn} \mid\leq \tau_2)]\right| \leq \tau_2^2 \sum_{h=1}^{n} \mathrm{P}[\mid Y_{hn} \mid > \tau_1] \to 0,$$

$$\begin{aligned} & \left|\sum_{h=1}^{n} \mathrm{E}[Y_{hn}\mathrm{I}(\tau_1 <\mid Y_{hn} \mid\leq \tau_2)]\mathrm{E}[Y_{hn}(\mathrm{I}(\mid Y_{hn} \mid\leq \tau_1) + \mathrm{I}(\mid Y_{hn} \mid\leq \tau_2))]\right| \\ \leq & \quad \tau_2(\tau_1 + \tau_2) \sum_{h=1}^{n} \mathrm{P}[\mid Y_{hn} \mid > \tau_1] \to 0, \quad n \to \infty. \end{aligned}$$

Hence, (17) holds. $\square$

Proof of Theorem 1. (*Sufficiency of* $\mathbf{A}_2(\tau)$). As follows from Lemma 2, it is sufficient to prove that

$$\mathcal{L}(Y^{\circ}_{\cdot n}(\tau)) \overset{wa}{\leftrightarrow} \mathcal{N}_1(0, \sigma^2_{\cdot n}(\tau)), \quad n \to \infty. \tag{18}$$

Let $f^{\circ}_{\tau \cdot n}(t) = \mathrm{E}[e^{itY^{\circ}_{\cdot n}(\tau)}]$. From Lemma 2 we know that $\{\mathcal{L}(Y^{\circ}_{\cdot n}(\tau)\}_{n \geq 1}$ is tight. Then, the Continuity Theorem (Proposition 7) states that it is sufficient to check that for each $t \in \mathbf{R}^1$

$$f^{\circ}_{\tau \cdot n}(t) - e^{-t^2 \sigma^2_{\cdot n}(\tau)/2} \to 0, \quad n \to \infty. \tag{19}$$

We need the following elementary inequality

$$\left| e^{iz} - 1 - iz + \frac{1}{2}z^2 \right| \leq | z |^3, \quad z \in \mathbf{R}^1. \tag{20}$$

Note that (12) (iii) implies that for arbitrary small $\varepsilon > 0$

$$\sum_{h=1}^{n} \mathrm{P}[| Y_{hn}^{\circ}(\tau) |> \varepsilon] \to 0, \quad n \to \infty. \tag{21}$$

From (21) it follows that there is a sequence $\varepsilon(n) \to 0$ such that $\sum_{h=1}^{n} \mathrm{P}[| Y_{hn}(\tau) |> \varepsilon(n)] \to 0,\ n \to \infty$. By applying the inequality $| Y_{hn}^{\circ}(\tau) |\leq 2\tau$, we obtain

$$\mathrm{E}[| Y_{hn}^{\circ}(\tau) |^3] \leq \mathrm{E}[(Y_{hn}^{\circ}(\tau))^2]\varepsilon(n) + (2\tau)^3\mathrm{E}[\mathrm{I}(| Y_{hn}^{\circ}(\tau) |> \varepsilon(n))]$$

$$\leq \quad \sigma_{hn}^2(\tau)\varepsilon(n) + (2\tau)^3\mathrm{P}[| Y_{hn}^{\circ}(\tau) |> \varepsilon(n)]. \tag{22}$$

Therefore, from (14), (21) and (22) we have

$$\sum_{h=1}^{n} \mathrm{E} \mid Y_{hn}^{\circ}(\tau) \mid^3 = o(1), \quad n \to \infty. \tag{23}$$

Then inequalities (20) and (23) imply that

$$f_{\tau hn}^{\circ}(t) = \mathrm{E}[e^{itY_{hn}^{\circ}(\tau)}] = 1 - \frac{t^2}{2}\mathrm{E}[(Y_{hn}^{\circ}(\tau))^2] + r_{\tau hn}(t), \tag{24}$$

where $| r_{\tau hn}(t) |\leq t^3\mathrm{E}[| Y_{hn}^{\circ}(\tau) |^3]$. From (23), (24) and (14) for any $t \in \mathbf{R}^1$ we have

$$f_{\tau\cdot n}^{\circ}(t) = \prod_{h=1}^{n} f_{\tau hn}^{\circ}(t) = e^{-t^2\sigma_{\cdot n}^2(\tau)/2} + o(1), \quad n \to \infty. \tag{25}$$

We do not exclude the case when $\sigma_{\cdot n}^2(\tau) \to 0$, because then both $f_{\tau\cdot n}^{\circ}(t) \to 1$ and $e^{-t^2\sigma_{\cdot n}^2(\tau)/2} \to 1,\ n \to \infty$. Relation (19) follows from (25). Hence, (2) holds. Sufficiency of $\mathbf{A}_2(i)$ is proved.

(*Necessity of* $\mathbf{A}_2(\tau)$). The family of normal d.l.s. $\{\mathcal{N}_1(0, \sigma_n^2)\}_{n\geq 1}$ is tight because $\sigma_+^2 < \infty$. Hence, Proposition 6 and (2) imply that $\mathbf{A}_2(\tau)$ holds. □

Lemma 6. *If* $\mathbf{A}_1$ *and* $\mathbf{A}_2(\tau)$ *hold then*

$$\sum_{h=1}^{n} Y_{hn}^{\circ}(\tau)\mathrm{E}[Y_{hn}(\tau)] \xrightarrow{\mathrm{P}} 0, \quad n \to \infty. \tag{26}$$

Proof. The r.v.s $\tilde{Z}_{hn}(\tau) = Y^{\circ}_{hn}(\tau)\mathrm{E}[Y_{hn}(\tau)]$ satisfy the following assumptions of the Degenerate Convergence Criterion (Proposition 1):

(i) $\max_{1\le h\le n} \mid \tilde{Z}_{hn}(\tau) \mid \le \tau \max_{1\le h\le n} \mid Y^{\circ}_{hn}(\tau) \mid \xrightarrow{\mathrm{P}} 0, \quad n\to\infty;$

(ii) $\sum_{h=1}^{n} \mathrm{E}[\tilde{Z}_{hn}(\tau)] = 0;$

(iv) $\mathrm{Var}\left[\sum_{h=1}^{n} \tilde{Z}_{hn}(\tau)\right] = \sum_{h=1}^{n} \mathrm{E}[(Y^{\circ}_{hn}(\tau))^2](\mathrm{E}[Y_{hn}(\tau)])^2$

$\le \max_{1\le h\le n}(\mathrm{E}[Y_{hn}(\tau)])^2\sigma^2_{\cdot n}(\tau) \to 0, \quad n\to\infty,$
by Lemmas 1 and 3. Hence, (26) holds. □

Lemma 7. *If* $\mathbf{A}_1$ *and* $\mathbf{A}_2(\tau)$ *hold, then as* $n\to\infty$

$$\begin{aligned}\sum_{h=1}^{n}(Y_{hn}(\tau)-\bar{Y}_{\cdot n}(\tau))^2 &= \sum_{h=1}^{n}(Y^{\circ}_{hn}(\tau))^2+\sum_{h=1}^{n}(\mathrm{E}[Y_{hn}(\tau)])^2\\ &- \frac{1}{n}\left(\sum_{h=1}^{n}\mathrm{E}[Y_{hn}(\tau)]\right)^2+o_{\mathrm{P}}(1). \qquad (27)\end{aligned}$$

Proof. It is easy to check the following relation:

$$\begin{aligned}&\sum_{h=1}^{n}(Y_{hn}(\tau)-\bar{Y}_{\cdot n}(\tau))^2 = \sum_{h=1}^{n}(Y^{\circ}_{hn}(\tau))^2+\sum_{h=1}^{n}(\mathrm{E}[Y_{hn}(\tau)])^2\\ &- \frac{1}{n}\left(\mathrm{E}\left[\sum_{h=1}^{n}Y_{hn}(\tau)\right]\right)^2-\frac{1}{n}(Y^{\circ}_{\cdot n}(\tau))^2-\frac{2}{n}Y^{\circ}_{\cdot n}(\tau)\mathrm{E}[Y_{\cdot n}(\tau)]\\ &+ 2\sum_{h=1}^{n}Y^{\circ}_{hn}(\tau)\mathrm{E}[Y_{hn}(\tau)]. \qquad (28)\end{aligned}$$

From the tightness of $\{\mathcal{L}(Y^{\circ}_{\cdot n}(\tau)\}_{n\ge1}$ proved in Lemma 2, it follows that $\frac{1}{n}(Y^{\circ}_{\cdot n}(\tau))^2 \xrightarrow{\mathrm{P}} 0$, $n\to\infty$. From (12) (ii) we have that

$$\left|\frac{1}{n}Y^{\circ}_{\cdot n}(\tau)\mathrm{E}[Y_{\cdot n}(\tau)]\right| \le \frac{1}{n}\mid Y^{\circ}_{\cdot n}(\tau)\mid n\max_{1\le h\le n}\mid \mathrm{E}[Y_{hn}(\tau)]\mid \xrightarrow{\mathrm{P}} 0, \quad n\to\infty.$$

The last term in (28) also converges to zero as was shown in Lemma 6. □

Lemma 8. *If* $\mathbf{A}_1$ *and* $\mathbf{A}_2(\tau)$ *hold, then*

$$\sum_{h=1}^{n}(Y^{\circ}_{hn}(\tau))^2-\sum_{h=1}^{n}\mathrm{E}[(Y^{\circ}_{hn}(\tau))^2] \xrightarrow{\mathrm{P}} 0, \quad n\to\infty. \qquad (29)$$

Proof. We can consider $\{\sigma^2_{\cdot n}(\tau)\}_{n\geq 1}$ as a union of converging subsequences $\{\sigma^2_{\cdot n_k}(\tau)\}_{k\geq 1}$. We consider case 1 with a subsequence $\{n_k\}_{k\geq 1}$ such that $\lim_{k\to\infty} \sigma^2_{\cdot n_k}(\tau) \to 0,\ k\to\infty$, and case 2 with a subsequence $\{n_k\}_{k\geq 1}$ such that $\lim_{k\to\infty} \sigma^2_{\cdot n_k}(\tau) > 0$. Recall that the set of all limit points $\{\sigma^2_{\infty}(\tau) :\ \exists\ \{n_k\}_{k\geq 1},\ \sigma^2_{\infty}(\tau) = \lim_{k\to\infty} \sigma^2_{\cdot n_k}(\tau)\}$ is bounded, $\sigma^2_{\infty}(\tau) \leq \sigma^2_{+}$.

Case 1. We have $\mathrm{E}[\sum_{h=1}^{n_k}(Y^{\circ}_{hn_k}(\tau))^2] = \sigma^2_{\cdot n_k}(\tau) \to 0,\ n\to\infty$. Hence, $\sum_{h=1}^{n_k}(Y^{\circ}_{hn_k}(\tau))^2 \xrightarrow{\mathrm{P}} 0,\ n_k\to\infty$, and (29) holds with $n = n_k,\ k\to\infty$.

Case 2. Here, from Proposition 2 we have $\mathcal{L}(Y^{\circ}_{\cdot n_k}(\tau)/\sigma_{\cdot n_k}) \xrightarrow{w} \mathcal{N}_1(0,1)$. Therefore, by the Raikov theorem (Proposition 4) we have $\sum_{h=1}^{n_k}\left(\frac{Y^{\circ}_{hn_k}(\tau)}{\sigma_{\cdot n_k}(\tau)}\right)^2 - 1 \xrightarrow{\mathrm{P}} 0,\ \ n\to\infty$. Here, for a $\delta > 0$ and all sufficiently large k, $\ \delta < \sigma^2_{\cdot n_k}(\tau) \leq \sigma^2_{+}$. Therefore this relation is equivalent to (29). Hence, (29) holds for every limit point $\sigma^2_{\infty}(\tau)$. $\square$

Lemma 9. *If* $\mathbf{A}_1$ *and* $\mathbf{A}_2(\tau)$ *hold, then*

$$\mathcal{L}(Y^{\star\odot}_{\cdot n} \mid \mathbf{X}_n) \overset{wa(\mathrm{P})}{\longleftrightarrow} \mathcal{L}(Y^{\star\odot}_{\cdot n}(\tau) \mid \mathbf{X}_n), \quad n\to\infty. \tag{30}$$

Proof. From (1) we have

$$Y^{\star\odot}_{\cdot n} - Y^{\star\odot}_{\cdot n}(\tau) = \sum_{h=1}^{n}(N^{\star}_{hn}-1)(Y_{hn} - Y_{hn}(\tau)) = \sum_{h=1}^{n}(N^{\star}_{hn}-1)Y_{hn}\mathrm{I}(\mid Y_{hn}\mid > \tau). \tag{31}$$

We can evaluate (31) as follows:

$$\mid Y^{\star\odot}_{\cdot n} - Y^{\star\odot}_{\cdot n}(\tau)\mid \leq (n-1)\sum_{h=1}^{n}\mid Y_{hn}\mid \mathrm{I}(\mid Y_{hn}\mid > \tau).$$

From (13) we obtain that

$$\mathrm{P}[(n-1)\sum_{h=1}^{n}\mid Y_{hn}\mid \mathrm{I}(\mid Y_{hn}\mid > \tau) = 0] \geq \mathrm{P}[\max_{1\leq h\leq n}\mid Y_{hn}\mid \leq \tau] \to 1, \quad n\to\infty.$$

Hence, $\mathrm{P}[Y^{\star\odot}_{\cdot n} = Y^{\star\odot}_{\cdot n}(\tau)] \to 1,\ n\to\infty$. It implies that for every $\varepsilon > 0$ and each continuous and bounded $f(\cdot)$, i.e. $f(\cdot) \in \mathcal{C}_b(\mathbf{R}^1)$,

$$\mid \mathrm{E}^{\star}[f(Y^{\star\odot}_{\cdot n}) \mid \mathbf{X}_n] - \mathrm{E}^{\star}[f(Y^{\star\odot}_{\cdot n}(\tau)) \mid \mathbf{X}_n]\mid \leq 2\sup_{z\in\mathbf{R}^1}\mid f(z)\mid \mathrm{P}[Y^{\star\odot}_{\cdot n} \neq Y^{\star\odot}_{\cdot n}(\tau)] \to 0,$$

$n\to\infty$. Therefore, (30) holds as corresponding to Definition 4 of conditional d.l.s weakly approaching in probability. $\square$

Proof of Theorem 2. (*Sufficiency of* $\mathbf{A}_3(\tau)$). Suppose that $\mathbf{A}_1, \mathbf{A}_2(\tau)$ and $\mathbf{A}_3(\tau)$ hold. From Theorem 1 and Lemma 2 we have that $\mathcal{L}(Y^{\circ}_{\cdot n}(\tau)) \overset{wa}{\leftrightarrow} \mathcal{N}_1(0, \sigma^2_{\cdot n}(\tau))$, $n \to \infty$. The main idea of the proof is to show that

$$\mathcal{L}(Y^{\star\odot}_{\cdot n}(\tau) \mid \mathbf{X}_n) \overset{wa(\mathrm{P})}{\longleftrightarrow} \mathcal{N}_1(0, \sigma^2_{\cdot n}(\tau)), \quad n \to \infty. \tag{32}$$

Then, due to the transitivity of d.l.s weakly approaching in probability, it follows that

$$\mathcal{L}(Y^{\star\odot}_{\cdot n}(\tau) \mid \mathbf{X}_n) \overset{wa(\mathrm{P})}{\longleftrightarrow} \mathcal{L}(Y^{\circ}_{\cdot n}(\tau)), \quad n \to \infty, \tag{33}$$

and then (3) follows by Lemma 9. The sequences of d.l.s $\{\mathcal{L}(Y^{\circ}_{\cdot n}(\tau)\}_{n\geq 1}$ and $\{\mathcal{N}_1(0, \sigma^2_{\cdot n}(\tau))\}_{n\geq 1}$ are tight. Therefore, we can apply the Continuity Theorem for sequences of random d.l.s weakly approaching in probability (Proposition 8). We need to use the ch.f.s $f^{\star\odot}_{\tau\cdot n}(t \mid \mathbf{X}_n) = \mathrm{E}[e^{itY^{\star\odot}_{\cdot n}(\tau)} \mid \mathbf{X}_n]$. Each resampling copy of $\mathbf{Y}^{\odot}_n(\tau)$ is obtained by n independent samplings from the components of the list $\mathbf{Y}^{\odot}_n(\tau)$, where each of the n components is selected with probability $1/n$. Therefore, we have

$$\begin{aligned} f^{\star\odot}_{\tau\cdot n}(t \mid \mathbf{X}_n) &= \left(\frac{1}{n}\sum_{h=1}^{n} e^{it(Y_{hn}(\tau)-\bar{Y}_{\cdot n}(\tau))}\right)^n \\ &= \left(1 - \frac{t^2}{2n}\sum_{h=1}^{n}(Y_{hn}(\tau) - \bar{Y}_{\cdot n}(\tau))^2 + R_n(t,\tau)\right)^n, \end{aligned} \tag{34}$$

where

$$\begin{aligned} R_n(t,\tau) &= \frac{1}{n}\sum_{h=1}^{n}\Bigg(e^{it(Y_{hn}(\tau)-\bar{Y}_{\cdot n}(\tau))} - 1 - it(Y_{hn}(\tau) - \bar{Y}_{\cdot n}(\tau)) \\ &+ \frac{t^2}{2}(Y_{hn}(\tau) - \bar{Y}_{\cdot n}(\tau))^2\Bigg). \end{aligned} \tag{35}$$

We evaluate $R_n(t,\tau)$ with the help of inequality (20)

$$\begin{aligned} \mid R_n(t,\tau) \mid &\leq \frac{t^3}{n}\sum_{h=1}^{n} \mid Y_{hn}(\tau) - \bar{Y}_{\cdot n}(\tau) \mid^3 \\ &\leq \frac{t^3}{n}\sum_{h=1}^{n}(Y_{hn}(\tau) - \bar{Y}_{\cdot n}(\tau))^2 2 \max_{1\leq h\leq n} \mid Y_{hn}(\tau) \mid . \end{aligned} \tag{36}$$

Therefore, by Lemma 1 $| R_n(t,\tau) |= o_p\left(\frac{t^2}{2n}\sum_{h=1}^{n}(Y_{hn}(\tau)-\bar{Y}_{\cdot n}(\tau))^2\right)$, and additionally from Lemmas 7 and 8 it follows that

$$\sum_{h=1}^{n}(Y_{hn}(\tau)-\bar{Y}_{\cdot n}(\tau))^2 = \sigma^2_{\cdot n}(\tau)$$
$$+ \quad \sum_{h=1}^{n}(\mathrm{E}[Y_{hn}(\tau)])^2 - \frac{1}{n}\left(\sum_{h=1}^{n}\mathrm{E}[Y_{hn}(\tau)]\right)^2 + o_p(1), \quad n\to\infty. \quad (37)$$

Recall that $\sigma^2_{\cdot n}(\tau) \le \sigma^2_{+} < \infty$ and assume that $\mathbf{A}_3(\tau)$ holds. Hence, from (34) - (37) we obtain

$$f^{\star\odot}_{\tau\cdot n}(t \mid \mathbf{X}_n) - e^{-t^2\sigma^2_{\cdot n}(\tau)/2} \xrightarrow{\mathrm{P}} 0, \quad n\to\infty.$$

Proposition 8 implies (32). Then (33) holds. Therefore, $\mathbf{A}_3(\tau)$ is sufficient for (3).

(*Necessity of* $\mathbf{A}_3(\tau)$). Suppose that resampling gives us a consistent estimator of $\mathcal{L}(Y^{\ominus}_{\cdot n}(\tau))$, i.e. (3) holds. Then, from Lemmas 2 and 9 we have

$$\mathcal{L}(Y^{\star\odot}_{\cdot n}(\tau) \mid \mathbf{X}_n) \overset{wa(\mathrm{P})}{\longleftrightarrow} \mathcal{L}(Y^{\circ}_{\cdot n}(\tau)), \quad n\to\infty. \quad (38)$$

From (38), Theorem 1 and Lemma 2 we obtain

$$\mathcal{L}(Y^{\star\odot}_{\cdot n}(\tau) \mid \mathbf{X}_n) \overset{wa(\mathrm{P})}{\longleftrightarrow} \mathcal{N}_1(0,\sigma^2_{\cdot n}(\tau)), \quad n\to\infty. \quad (39)$$

We know that $\sigma^2_{\cdot n}(\tau) \le \sigma^2_{+} < \infty$. Hence, the sequence of d.l.s $\{\mathcal{N}_1(0,\sigma^2_{\cdot n}(\tau)\}_{n\ge 1}$ is tight. Then from (39), and by the Continuity Theorem for d.l.s weakly approaching in probability (Proposition 8), it follows that for every $t\in\mathbf{R}^2$

$$f^{\star\odot}_{\tau\cdot n}(t \mid \mathbf{X}_n) - e^{-t^2\sigma^2_{\cdot n}(\tau)/2} \xrightarrow{\mathrm{P}} 0, \quad n\to\infty. \quad (40)$$

From Lemma 8 for every $t\in\mathbf{R}^1$, we have

$$e^{-t^2\sigma^2_{\cdot n}(\tau)/2} - e^{-t^2\sum_{h=1}^{n}(Y^{\circ}_{hn}(\tau))^2/2} \xrightarrow{\mathrm{P}} 0, \quad n\to\infty. \quad (41)$$

From (27), (34), and (35), which hold under assumptions $\mathbf{A}_1$ and $\mathbf{A}_2(\tau)$, it follows that for every $t\in\mathbf{R}^1$

$$f^{\star\odot}_{\tau\cdot n}(t \mid \mathbf{X}_n) = \exp\left\{-\frac{t^2}{2}\left(\sum_{h=1}^{n}(Y^{\circ}_{hn}(\tau))^2 + W_n(\tau)\right)\right\} + o_p(1), \quad n\to\infty,$$

where

$$W_n(\tau) = \sum_{h=1}^{n} (\mathrm{E}[Y_{hn}(\tau)])^2 - \frac{1}{n}\left(\sum_{h=1}^{n} \mathrm{E}[Y_{hn}(\tau)]\right)^2 \geq 0. \tag{42}$$

From (29), (40), (41) and (42) we obtain

$$f^{\star\odot}_{\tau\cdot n}(t \mid \mathbf{X}_n) - e^{-t^2\sigma^2_{\cdot n}(\tau)/2} = e^{-t^2\sum_{h=1}^{n}(Y^{\bigcirc}_{hn}(\tau))^2/2}(e^{-t^2 W_n(\tau)/2} - 1) + o_p(1) \xrightarrow{\mathrm{P}} 0,$$

$n \to \infty$. Hence, $W_n(\tau) \to 0,\ n \to \infty$, i.e. $\mathbf{A}_3(\tau)$ holds. □

Proofs of Corollary 1 and Corollary 2. Proofs follow directly from Proposition 10, because $\underline{\lim}_{n\to\infty}\sigma^2_{\cdot n}(\tau) > 0$, together with uniform boundness $\sigma^2_{\cdot n}(\tau) \leq \sigma^2_{+} < \infty$, imply that all $\mathcal{N}_1(0, \sigma^2_{\cdot n}(\tau))$ have uniformly bounded densities $\frac{1}{\sqrt{2\pi}\sigma_{\cdot n}(\tau)} e^{-x^2/(2\sigma^2_{\cdot n}(\tau))}$. Hence, the family of normal d.f.s is uniformly continuous and we can apply Proposition 10. □

Proof of Corollary 3. From the Chebyshev inequality and (4) it follows that for every $\tau > 0$

$$\sum_{h=1}^{n} \mathrm{P}[\mid Y_{hn} \mid > \tau] \leq \frac{1}{\tau^2}\sum_{h=1}^{n} \mathrm{E}[Y^2_{hn}\mathrm{I}(\mid Y_{hn} \mid > \tau)] \to 0, \quad n \to \infty.$$

Hence, $\mathbf{A}_1$ holds. We will use the following inequalities

$$\tau \mathrm{I}(\mid Y_{hn} \mid > \tau) \leq \mid Y_{hn} \mid, \tag{43}$$

$$\mid \mathrm{E}[Y_{hn}\mathrm{I}(\mid Y_{hn} \mid \leq \tau)]\mathrm{E}[Y_{hn}\mathrm{I}(\mid Y_{hn} \mid > \tau)] \mid \leq \mathrm{E}[Y^2_{hn}\mathrm{I}(\mid Y_{hn} \mid > \tau)]. \tag{44}$$

From (4) and (43) we obtain

$$\left|\sum_{h=1}^{n} \mathrm{E}[Y_{hn}\mathrm{I}(\mid Y_{hn} \mid > \tau)]\right| \leq \frac{1}{\tau}\sum_{h=1}^{n} \mathrm{E}[Y^2_{hn}\mathrm{I}(\mid Y_{hn} \mid > \tau)] \to 0, \quad n \to \infty. \tag{45}$$

We have

$$\begin{aligned} \sigma^2_{\cdot n} - \sigma^2_{\cdot n}(\tau) &= \sum_{h=1}^{n} \Big(\mathrm{E}[Y^2_{hn}\mathrm{I}(\mid Y_{hn} \mid > \tau)] - (\mathrm{E}[Y_{hn}\mathrm{I}(\mid Y_{hn} \mid > \tau)])^2 \\ &- 2\mathrm{E}[Y_{hn}\mathrm{I}(\mid Y_{hn} \mid \leq \tau)]\mathrm{E}[Y_{hn}\mathrm{I}(\mid Y_{hn} \mid > \tau)] \Big). \end{aligned} \tag{46}$$

The Jensen inequality, (4), (44) and (46) imply that

$$\mid \sigma^2_{\cdot n} - \sigma^2_{\cdot n}(\tau) \mid \leq 3\sum_{h=1}^{n} \mathrm{E}[Y^2_{hn}\mathrm{I}(\mid Y_{hn} \mid > \tau)] \to 0, \quad n \to \infty. \tag{47}$$

From (45) we have

$$\begin{aligned} Y_{\cdot n}^{\circ} &= Y_{\cdot n}^{\ominus}(\tau) - \sum_{h=1}^{n} \mathrm{E}[Y_{hn}\mathrm{I}(\mid Y_{hn} \mid > \tau)] = Y_{\cdot n}^{\circ}(\tau) + \sum_{h=1}^{n} Y_{hn}\mathrm{I}(\mid Y_{hn} \mid > \tau) \\ &- \sum_{h=1}^{n} \mathrm{E}[Y_{hn}\mathrm{I}(\mid Y_{hn} \mid > \tau)] = Y_{\cdot n}^{\circ}(\tau) + o_{\mathrm{P}}(1), \quad n \to \infty. \end{aligned} \tag{48}$$

From assumption $\sum_{h=1}^{n} \mathrm{E}[Y_{hn}^2] \leq C_2$ and (48) we have that the sequences of d.l.s $\{\mathcal{L}(Y_{\cdot n}^{\circ})\}_{n\geq 1}$, $\{\mathcal{L}(Y_{\cdot n}^{\ominus}(\tau))\}_{n\geq 1}$, and $\{\mathcal{L}(Y_{\cdot n}^{\circ}(\tau))\}_{n\geq 1}$, are tight. Hence, $\mathbf{A}_2(\tau)$ holds. From (48), Proposition 9, (18) and (47) we obtain (5).
For every $\tau > 0$ we have

$$\begin{aligned} &\sum_{h=1}^{n}(\mathrm{E}[Y_{hn}(\tau)])^2 = \sum_{h=1}^{n}(\mathrm{E}[Y_{hn}] - \mathrm{E}[Y_{hn}\mathrm{I}(\mid Y_{hn} \mid > \tau)])^2 \\ \leq\ & 2\sum_{h=1}^{n}(\mathrm{E}[Y_{hn}])^2 + 2\sum_{h=1}^{n} \mathrm{E}[Y_{hn}^2]\mathrm{P}[\max_{1\leq h\leq n} \mid Y_{hn} \mid > \tau] \\ \leq\ & 2\sum_{h=1}^{n}(\mathrm{E}[Y_{hn}])^2 + 2C_2\mathrm{P}[\max_{1\leq h\leq n} \mid Y_{hn} \mid > \tau] \to 0, \quad n \to \infty. \end{aligned}$$

Therefore, we have that $\mathbf{A}_2(\tau)$ and $\mathbf{A}_3(\tau)$ hold for every τ.
Then (33) and (48) imply

$$\mathcal{L}(Y_{\cdot n}^{\star\odot}(\tau)) \overset{wa(\mathrm{P})}{\longleftrightarrow} \mathcal{N}(0, \sigma_{\cdot n}^2), \quad n \to \infty. \tag{49}$$

We can write the following identity

$$Y_{\cdot n}^{\star\odot} = Y_{\cdot n}^{\star\odot}(\tau) + \sum_{h=1}^{n}(N_{hn}^{\star} - 1)Y_{hn}\mathrm{I}(\mid Y_{hn} \mid > \tau). \tag{50}$$

We can evaluate the second moment of the sum in (50) as follows

$$\begin{aligned} &\mathrm{E}\left[\mathrm{E}^{\star}\left[\left(\sum_{h=1}^{n}(N_{hn}^{\star} - 1)Y_{hn}\mathrm{I}(\mid Y_{hn} \mid > \tau)\right)^2 \mid \mathbf{X}_n\right]\right] \\ =\ & \mathrm{E}\left[\sum_{h=1}^{n}\left(1 - \frac{1}{n}\right)Y_{hn}^2\mathrm{I}(\mid Y_{hn} \mid > \tau)\right. \end{aligned}$$

$$- \quad \frac{1}{n}\sum_{h_1=1}^{n}\sum_{h_2=1_{h_1\neq h_2}}^{n}\prod_{i=1,2} Y_{h_i n}\mathrm{I}(\mid Y_{h_i n}\mid > \tau)\Bigg]$$

$$= \quad \mathrm{E}\left[\sum_{h=1}^{n} Y_{hn}^2 \mathrm{I}(\mid Y_{hn}\mid > \tau) - \frac{1}{n}\left(\sum_{h=1}^{n} Y_{hn}\mathrm{I}(\mid Y_{hn}\mid > \tau)\right)^2\right]$$

$$\leq \quad \mathrm{E}\left[\sum_{h=1}^{n} Y_{hn}^2 \mathrm{I}(\mid Y_{hn}\mid > \tau)\right] \to 0, \quad n \to \infty.$$

Therefore, we have

$$Y_{\cdot n}^{\star\odot} = Y_{\cdot n}^{\star\odot}(\tau) + o_{\mathrm{P}}(1), \quad n \to \infty. \tag{51}$$

The desired result (6) follow from (51), Proposition 9, (49) and (50). □

Proof of Theorem 3. We can consider the triangular array of r.v.s $X_{hn} = \mathbf{c}^T(\tilde{\mathbb{X}}_n^T\tilde{\mathbb{X}}_n)^+\tilde{\mathbf{x}}_h W_h,\ \mathbf{c}^T \in \mathcal{M}(\tilde{\mathbb{X}}_n^T)$, where $W_h,\ h = 1,...,n$, are (unobserved) errors in (7). Let $\rho_n = (\mathbf{c}^T(\tilde{\mathbb{X}}_n^T\tilde{\mathbb{X}}_n)^+\mathbf{c})^{1/2}$ and $Y_{hn} = \frac{1}{\rho_n}X_{hn},\ \{h,n\} \in \mathcal{T}$. Then the normed deviations of OLS-estimators can be written as follows

$$S_n(\mathbf{c}, \mathbf{W}_n, \tilde{\mathbb{X}}_n) = \frac{1}{\rho_n}((\mathbf{c}^T\beta_0)_n^{\wedge} - \mathbf{c}^T\beta_0) = \sum_{h=1}^{n} Y_{hn}. \tag{52}$$

From $\mathbf{HR}_2$ it follows that $\mathrm{E}[Y_{hn}] = 0$, i.e. $Y_{hn}^{\circ} = Y_{hn}$. Relation (4) follows from $\mathbf{HR}_1$. Assumption $\mathbf{HR}_2$ and (8) imply that

$$\begin{aligned} \sigma_n^2(\mathbf{c}, \tilde{\mathbb{X}}_n) &= \mathrm{E}[(S_n(\mathbf{c}, \mathbf{W}_n, \tilde{\mathbb{X}}_n))^2] = \sum_{h=1}^{n}\mathrm{E}[Y_{hn}^2] \\ &= \sum_{h=1}^{n}\frac{\sigma_h^2(\tilde{\mathbf{x}}_h)\mathbf{c}^T(\tilde{\mathbb{X}}_n^T\tilde{\mathbb{X}}_n)^+\tilde{\mathbf{x}}_h\tilde{\mathbf{x}}_h^T(\tilde{\mathbb{X}}_n^T\tilde{\mathbb{X}}_n)^+\mathbf{c}}{\mathbf{c}^T(\tilde{\mathbb{X}}_n^T\tilde{\mathbb{X}}_n)^+\mathbf{c}} \leq \sigma_+^2. \end{aligned} \tag{53}$$

Therefore, all assumptions, stated in Corollary 3, are fulfilled for the r.v.s $\{Y_{hn}\}$. Hence, (5) is valid and it is equivalent to (10).

In order to prove (11) we consider

$$S_n^\star(\mathbf{c}, \mathbf{W}_n, \tilde{\mathbb{X}}_n) = \sum_{h=1}^{n}(N_{hn}^\star - 1)\frac{\mathbf{c}^T(\tilde{\mathbb{X}}_n^T\tilde{\mathbb{X}}_n)^+\tilde{\mathbf{x}}_h}{\sqrt{\mathbf{c}^T(\tilde{\mathbb{X}}_n^T\tilde{\mathbb{X}}_n)^+\mathbf{c}}}W_h. \tag{54}$$

Relations (5) and (6) in Corollary 3 imply

$$\mathcal{L}(S_n^\star(\mathbf{c}, \mathbf{W}_n, \tilde{\mathbb{X}}_n) \mid \mathbf{W}_n) \overset{wa(\mathrm{P})}{\longleftrightarrow} \mathcal{N}_1(0, \sigma_n^2(\mathbf{c}, \tilde{\mathbb{X}}_n)), \quad n \to \infty. \tag{55}$$

We are not able to use (54) and (55) because the errors $W_1, ..., W_n$ are not observable.

We can realize resamplings from the list of residuals as written in (9). We have

$$Y_h - \hat{Y}_h = W_h - \tilde{\mathbf{x}}_h^T(\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+ \tilde{\mathbb{X}}_n^T \mathbf{W}_n. \tag{56}$$

From (9) and (56) we obtain the following relation:

$$U_n^\star(\mathbf{c}, \mathbf{Y}_n, \tilde{\mathbb{X}}_n) = S_n^\star(\mathbf{c}, \mathbf{W}_n, \tilde{\mathbb{X}}_n) - V_n^\star(\mathbf{c}, \mathbf{W}_n, \tilde{\mathbb{X}}_n), \tag{57}$$

where

$$V_n^\star(\mathbf{c}, \mathbf{W}_n, \tilde{\mathbb{X}}_n) = \sum_{h=1}^n (N_{hn}^\star - 1) \frac{\mathbf{c}^T(\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+ \tilde{\mathbf{x}}_h \tilde{\mathbf{x}}_h^T (\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+ \tilde{\mathbb{X}}_n^T \mathbf{W}_n}{\sqrt{\mathbf{c}^T(\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+ \mathbf{c}}}.$$

We will prove that $V_n^\star(\mathbf{c}, \mathbf{W}_n, \tilde{\mathbb{X}}_n)$ converges in probability to zero.

Note, that $\mathrm{E}^\star[(N_{hn}^\star - 1)^2] = 1 - 1/n$, $\mathrm{E}^\star[(N_{h_1 n}^\star - 1)(N_{h_2 n}^\star - 1)] = -1/n$ and $\tilde{\mathbf{x}}_h^T(\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+ \tilde{\mathbb{X}}_n^T \mathrm{E}[\mathbf{W}_n \mathbf{W}_n^T] \mathbb{X}_n (\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+ \tilde{\mathbf{x}}_h \leq \sigma_+^2 \tilde{\mathbf{x}}_h^T (\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+ \tilde{\mathbf{x}}_h$.

We can evaluate the second moment of $V_n^\star(\mathbf{c}, \mathbf{W}_n, \tilde{\mathbb{X}}_n)$ in two steps. We start evaluating the conditional expectation given $\mathbf{W}_n$ and after that we take the expectation related to $\mathbf{W}_n$, and then we use (8). We have

$$\begin{aligned}
& \mathrm{E}^\star[(V_n^\star(\mathbf{c}, \mathbf{W}_n, \tilde{\mathbb{X}}_n))^2 \mid \mathbf{W}_n] \\
= \; & \sum_{h=1}^n \mathrm{E}^\star[(N_{hn}^\star - 1)^2] \left(\frac{\mathbf{c}^T(\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+ \tilde{\mathbf{x}}_h \tilde{\mathbf{x}}_h^T (\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+ \tilde{\mathbb{X}}_n^T \mathbf{W}_n}{\sqrt{\mathbf{c}^T(\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+ \mathbf{c}}} \right)^2 \\
- \; & \sum_{h_1=1}^n \quad \sum_{h_2=1,\, h_1 \neq h_2}^n \mathrm{E}^\star[(N_{h_1 n}^\star - 1)(N_{h_2 n}^\star - 1)] \\
\cdot \; & \prod_{j=1,2} \frac{\mathbf{c}^T(\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+ \tilde{\mathbf{x}}_{h_j} \tilde{\mathbf{x}}_{h_j} (\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+ \tilde{\mathbb{X}}_n^T \mathbf{W}_n}{\sqrt{\mathbf{c}^T(\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+ \mathbf{c}}} \\
= \; & \sum_{h=1}^n \left(\frac{\mathbf{c}^T(\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+ \tilde{\mathbf{x}}_h \tilde{\mathbf{x}}_h^T (\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+ \tilde{\mathbb{X}}_n^T \mathbf{W}_n}{\sqrt{\mathbf{c}^T(\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+ \mathbf{c}}} \right)^2 \\
- \; & \frac{1}{n} \left(\sum_{h=1}^n \frac{\mathbf{c}^T(\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+ \tilde{\mathbf{x}}_h \tilde{\mathbf{x}}_h^T (\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+ \tilde{\mathbb{X}}_n^T \mathbf{W}_n}{\sqrt{\mathbf{c}^T(\tilde{\mathbb{X}}_n^T \tilde{\mathbb{X}}_n)^+ \mathbf{c}}} \right)^2 .
\end{aligned}$$

From (8) and $\mathbf{HR}_3$ it follows that

$$\mathrm{E}[\mathrm{E}^\star[(V_n^\star(\mathbf{c},\mathbf{W}_n,\tilde{\mathbb{X}}_n))^2 \mid \mathbf{W}_n]]$$

$$\leq \sum_{h=1}^{n} \frac{\mathbf{c}^T(\tilde{\mathbb{X}}_n^T\tilde{\mathbb{X}}_n)^+\tilde{\mathbf{x}}_h\tilde{\mathbf{x}}_h^T(\tilde{\mathbb{X}}_n^T\tilde{\mathbb{X}}_n)^+\tilde{\mathbb{X}}_n^T\mathrm{E}[\mathbf{W}_n\mathbf{W}_n^T]\tilde{\mathbb{X}}_n(\tilde{\mathbb{X}}_n^T\tilde{\mathbb{X}}_n)^+\tilde{\mathbf{x}}_h\tilde{\mathbf{x}}_h^T(\tilde{\mathbb{X}}_n^T\tilde{\mathbb{X}}_n)^+\mathbf{c}}{\mathbf{c}^T(\tilde{\mathbb{X}}_n^T\tilde{\mathbb{X}}_n)^+\mathbf{c}}$$

$$\leq \sigma_+^2\sum_{h=1}^{n} \frac{(\mathbf{c}^T(\tilde{\mathbb{X}}_n^T\tilde{\mathbb{X}}_n)^+\tilde{\mathbf{x}}_h)^2\tilde{\mathbf{x}}_h^T(\tilde{\mathbb{X}}_n^T\tilde{\mathbb{X}}_n)^+\tilde{\mathbf{x}}_h}{\mathbf{c}^T(\tilde{\mathbb{X}}_n^T\tilde{\mathbb{X}}_n)^+\mathbf{c}}$$

$$\leq r^2\sigma_+^2 n\frac{\|\mathbf{c}\|^2\, x_+^4(\mathrm{tr}(\tilde{\mathbb{X}}_n^T\tilde{\mathbb{X}}_n)^+)^3}{\mathbf{c}^T(\tilde{\mathbb{X}}_n^T\tilde{\mathbb{X}}_n)^+\mathbf{c}} \to 0, \quad n\to\infty.$$

From Proposition 9 and (57) it follows

$$\mathcal{L}(U_n^\star(\mathbf{c},\mathbf{Y}_n,\tilde{\mathbb{X}}_n)) \overset{wa(\mathrm{P})}{\longleftrightarrow} \mathcal{L}(S_n^\star(\mathbf{c},\mathbf{W}_n,\tilde{\mathbb{X}}_n)), \quad n\to\infty. \tag{58}$$

Relations (55) and (58) imply (11). □

Proof of Corollary 4. From $\mathbf{HR}_2$ we obtain the result that $\sigma_n^2(\mathbf{c},\tilde{\mathbb{X}}_n) \geq \sigma_-^2 > 0$. The desired statement follows from Corollary 2. □

5 Appendix

Here, for reader's convenience we bring together definitions and theoretical results related to weak convergence and weakly approaching sequences of d.l.s.

Let $\mathbb{X}' = \{X'_{hn} : \{h,n\} \in \mathcal{T}\}$ be a triangular array of independent r.v.s for each n. The r.v.s can be non-identically distributed. Together with the list of independent r.v.s $\mathbf{X}'_n = \{X'_{1n},...,X'_{nn}\}$, we consider a real-valued r.v. T_n. Both $\mathbf{X}'_n$ and T_n are defined on the same probability space. Let $\mathcal{L}(T_n)$ and $\mathcal{L}(T_n \mid \mathbf{X}'_n)$ be the d.l. of T_n and the conditional d.l. of T_n given $\mathbf{X}'_n$. By $\mathcal{L}(0)$, we define the degenerate d.l. concentrated at 0. We consider regular conditional d.l.s, Dudley (1998). The diversity of conditional d.l.s $\{\mathcal{L}(T_n \mid \mathbf{X}'_n)\}$ is rather rich. For example, if $T_n^\star = X'^\star_{1n}$ where $X'^\star_{1n} = X'_{hn}$ with probability $1/n$, $h = 1,...,n$, then $\mathcal{L}(T_n^\star \mid \mathbf{X}'_n)$ is the empirical distribution of $X'_{\cdot n}$. By $\mathcal{C}_b(\mathbf{R}^1)$ we denote the set of all continuous and bounded functions $f(\cdot) : \mathbf{R}^1 \to \mathbf{R}^1$.

Definition 1. *A sequence of d.l.s* $\{\mathcal{L}(T_n)\}_{n\geq 1}$ *is called tight if for any* $\varepsilon > 0$ *there exists a constant* $C_\varepsilon > 0$ *such that*

$$\sup_n \mathrm{P}[\mid T_n \mid > C_\varepsilon] < \varepsilon.$$

Definition 2. *Let* T_0 *be an r.v. such that for each* $f(\cdot) \in \mathcal{C}_b(\mathbf{R}^1)$

$$\mathrm{E}[f(T_n)] \to \mathrm{E}[f(T_0)], \quad n \to \infty.$$

Then, the sequence of d.l.s $\{\mathcal{L}(T_n)\}_{n\geq 1}$ *is called weakly converging to* $\mathcal{L}(T_0)$. *In short we write*

$$\mathcal{L}(T_n) \xrightarrow{w} \mathcal{L}(T_0), \quad n \to \infty.$$

Below we let $X'_{hn}(\tau) = X'_{hn}\mathrm{I}(\mid X'_{hn} \mid \leq \tau)$.

Remark. The relations $\sum_{h=1}^{n} \mathrm{P}[\mid X'_{hn} \mid > \varepsilon] \to 0$, and $\max_{1\leq h\leq n} \mid X'_{hn} \mid \xrightarrow{\mathrm{P}} 0$, $n \to \infty$, are equivalent for the independent r.v.s $X'_{1n}, ..., X'_{nn}$.

Proposition 1. (The Degenerate Convergence Criterion). *For every* $\varepsilon > 0$ $\max_{1\leq h\leq n} \mathrm{P}[\mid X'_{hn} \mid > \varepsilon] \to 0$, $n \to \infty$, *and* $\mathcal{L}(X'_{\cdot n}) \xrightarrow{w} \mathcal{L}(0)$, $n \to \infty$, *if and only if for every* $\varepsilon > 0$ *and a* $\tau > 0$

(*i*) $\sum_{h=1}^{n} \mathrm{P}[\mid X'_{hn} \mid > \varepsilon] \to 0$, (*ii*) $\sum_{h=1}^{n} \mathrm{E}[X'_{hn}(\tau)] \to 0$,

(*iii*) $\sum_{h=1}^{n}(\mathrm{E}[X'^2_{hn}(\tau)] - (\mathrm{E}[X'_{hn}(\tau)])^2) \to 0$, *as* $n \to \infty$.

Proposition 2. (The Normal Convergence Criterion). *For every* $\varepsilon > 0$ $\max_{1\leq h\leq n} \mathrm{P}[\mid X'_{hn} \mid > \varepsilon] \to 0$, $n \to \infty$, *and* $\mathcal{L}(X'_{\cdot n}) \xrightarrow{w} \mathcal{N}_1(\mu, \sigma^2)$, $n \to \infty$, *if and only if for every* $\varepsilon > 0$ *and a* τ

(*i*) $\sum_{h=1}^{n} \mathrm{P}[\mid X'_{hn} \mid > \varepsilon] \to 0$, (*ii*) $\sum_{h=1}^{n} \mathrm{E}[X'_{hn}(\tau)] \to \mu$,

(*iii*) $\sum_{h=1}^{n}(\mathrm{E}[X'^2_{hn}(\tau)] - (\mathrm{E}[X'_{hn}(\tau)])^2) \to \sigma^2$ *as* $n \to \infty$.

Proposition 3. *Suppose that there is a limit d.l.* $\mathcal{L}(X'_0)$ *of an r.v.* X'_0, *such that* $\mathcal{L}(X'_{\cdot n}) \xrightarrow{w} \mathcal{L}(X'_0)$, $n \to \infty$. *Then* $\mathcal{L}(X'_0) = \mathcal{N}_1(\mu, \sigma^2)$ *and for every* $\varepsilon > 0$ $\max_{1\leq h\leq n} \mathrm{P}[\mid X'_{hn} \mid > \varepsilon] \to 0$, $n \to \infty$, *if and only if* $\max_{1\leq h\leq n} \mid X'_{hn} \mid \xrightarrow{\mathrm{P}} 0$, $n \to \infty$.

Proofs of Propositions 1, 2 and 3 are given in Loève (1977), pp. 328-329.

Proposition 4. (The Raikov theorem). *Let* $\mathbb{X}' = \{X'_{hn} : \{h, n\} \in \mathcal{T}\}$ *be a triangular array. If for every* $\varepsilon > 0$ $\max_{1\leq h\leq n} \mathrm{P}[\mid X'_{hn} \mid > \varepsilon] \to 0$ *as* $n \to \infty$, *then the following two statements are equivalent:*

(i) *there exists a sequence* $\{u_n\}_{n\geq 1}$ *such that*
$$\mathcal{L}(X'_{\cdot n} - u_n) \xrightarrow{w} \mathcal{N}_1(0,1), \quad n \to \infty;$$

(ii) $\sum_{h=1}^{n}(X'^{\ominus}_{hn}(\tau))^2 \xrightarrow{\mathrm{P}} 1, \quad n \to \infty,$
where $X'^{\ominus}_{hn} = X'_{hn} - \mathrm{E}[X'_{hn}\mathrm{I}(\mid X'_{hn} \mid \leq \tau)]$.

Proof. See Theorem 5 on p. 143 in Gnedenko and Kolmogorov (1968).

Proposition 5. (The Prokhorov theorem). *If a sequence of d.l.s* $\{\mathcal{L}(T_n)\}_{n\geq 1}$ *is tight, then there exists a subsequence* $\{\mathcal{L}(T_{n_k})\}_{k\geq 1}$ *and an r.v.* T_0 *such that* $\mathcal{L}(T_{n_k}) \xrightarrow{w} \mathcal{L}(T_0),\ n \to \infty$.

Proof is given in Shiryaev (1996).

The following definition eliminates the necessity to assume the existence of a limit d.l.

Definition 3. *Let* $\{\mathcal{L}(T'_n)\}_{n\geq 1}$ *and* $\{\mathcal{L}(T''_n)\}_{n\geq 1}$ *be two sequences of d.l.s. They are said to be weakly approaching (each other) if for every* $f(\cdot) \in \mathcal{C}_b(\mathbf{R}^1)$

$$\mathrm{E}[f(T'_n)] - \mathrm{E}[f(T''_n)] \to 0, \quad n \to \infty.$$

In short we write $\mathcal{L}(T'_n) \overset{wa}{\leftrightarrow} \mathcal{L}(T''_n),\ n \to \infty$.

Note that if $T''_n \equiv T_0$ then Definition 3 is reduced to weak convergence.

Proposition 6. *If* $\mathcal{L}(T'_n) \overset{wa}{\leftrightarrow} \mathcal{L}(T''_n),\ n \to \infty$, *and* $\{\mathcal{L}(T''_n)\}_{n\geq 1}$ *is tight, then* $\{\mathcal{L}(T'_n)\}_{n\geq 1}$ *is also tight.*

Proof. See Lemma 5, pp. 817 - 818 in Belyaev and Sjöstedt-de Luna (2000).

Definition 4. *Let* $\{T'_n, T''_n\}_{n\geq 1}$ *be r.v.s defined for each* n *on the same probability space as* $\mathbf{X}'_n$*. Then, the sequences* $\{\mathcal{L}(T'_n \mid \mathbf{X}'_n)\}_{n\geq 1}$ *and* $\{\mathcal{L}(T''_n \mid \mathbf{X}'_n)\}_{n\geq 1}$ *of conditional d.l.s, given* $\mathbf{X}'_n$*, are said to be weakly approaching (each other) in probability along* $\mathbf{X}'_n$ *if for every function* $f(\cdot) \in \mathcal{C}_b(\mathbf{R}^1)$

$$\mathrm{E}[f(T'_n) \mid \mathbf{X}'_n] - \mathrm{E}[f(T''_n) \mid \mathbf{X}'_n] \xrightarrow{\mathrm{P}} 0, \quad n \to \infty.$$

In short we write

$$\mathcal{L}(T'_n \mid \mathbf{X}'_n) \overset{wa(\mathrm{P})}{\longleftrightarrow} \mathcal{L}(T''_n \mid \mathbf{X}'_n), \quad n \to \infty. \tag{59}$$

In the case when, instead of $\mathcal{L}(T''_n \mid \mathbf{X}'_n)$ and $\mathrm{E}[f(T''_n) \mid \mathbf{X}'_n]$, we have $\mathcal{L}(T''_n)$ and $\mathrm{E}[f(T''_n)]$, respectively, then instead of (59), we write

$$\mathcal{L}(T'_n \mid \mathbf{X}'_n) \overset{wa(\mathrm{P})}{\longleftrightarrow} \mathcal{L}(T''_n), \quad n \to \infty.$$

Let $f_n(t) = \mathrm{E}[e^{itT'_n}]$, $f_n(t \mid \mathbf{X}'_n) = \mathrm{E}[e^{itT'_n} \mid \mathbf{X}'_n]$, and $g_n(t) := \mathrm{E}[e^{itT''_n}]$ be the characteristic functions of T'_n, T''_n given $\mathbf{X}'_n$, and T''_n.

Proposition 7. (The Continuity Theorem for weakly approaching sequences of d.l.s). *Let $\{\mathcal{L}(T'_n)\}_{n\geq 1}$ and $\{\mathcal{L}(T''_n)\}_{n\geq 1}$ be two sequences of d.l.s where $\{\mathcal{L}(T''_n)\}_{n\geq 1}$ is tight. Then*

$$\mathcal{L}(T'_n) \overset{wa}{\leftrightarrow} \mathcal{L}(T''_n), \quad n \to \infty$$

if and only if for each $t \in \mathbf{R}^1$

$$f_n(t) - g_n(t) \to 0, \quad n \to \infty.$$

Proposition 8. (The Continuity theorem for weakly approaching random d.l.s). *Let $\{\mathcal{L}(T'_n \mid \mathbf{X}'_n)\}_{n\geq 1}$ and $\{\mathcal{L}(T''_n)\}_{n\geq 1}$ be two sequences of d.l.s and $\{\mathcal{L}(T''_n)\}_{n\geq 1}$ be tight. Then*

$$\mathcal{L}(T'_n \mid \mathbf{X}'_n) \overset{wa(\mathrm{P})}{\longleftrightarrow} \mathcal{L}(T''_n), \quad n \to \infty,$$

if and only if for every $t \in \mathbf{R}^1$

$$f_n(t \mid \mathbf{X}'_n) - g_n(t) \overset{\mathrm{P}}{\to} 0, \quad n \to \infty.$$

Proof. See Theorems 1 and 2 in Belyaev and Sjöstedt-de Luna (2000).

Proposition 9. (The Stability Theorem for weakly approaching d.l.s). *Let $\{T'_n, U_n\}$ be a sequence of pairs of r.v.s defined on the same probability space as $\mathbf{X}'_n$ and let $U_n \overset{\mathrm{P}}{\to} 0$, $n \to \infty$. If $\{T''_n\}_{n\geq 1}$ is a tight sequence of r.v.s and $\mathcal{L}(T'_n) \overset{wa}{\leftrightarrow} \mathcal{L}(T''_n)$ $(\mathcal{L}(T'_n \mid \mathbf{X}'_n) \overset{wa(\mathrm{P})}{\longleftrightarrow} \mathcal{L}(T''_n))$, $n \to \infty$, then also*

$$\mathcal{L}(T'_n + U_n) \overset{wa}{\leftrightarrow} \mathcal{L}(T''_n) \quad (\mathcal{L}(T'_n + U_n \mid \mathbf{X}'_n) \overset{wa(\mathrm{P})}{\longleftrightarrow} \mathcal{L}(T''_n)) \quad \textit{as} \quad n \to \infty.$$

Proof. See Lemmas 7 and 8 in Belyaev and Sjöstedt-de Luna (2000).

Proposition 10. *Let $\{T'_n\}_{n\geq 1}$ be a sequence of real-valued r.v.s defined on the same probability space as $\mathbf{X}'_n$, $\{T''_n\}_{n\geq 1}$ be a tight sequence of r.v.s and*

$\mathcal{L}(T'_n \mid \mathbf{X}'_n) \overset{wa(\mathrm{P})}{\longleftrightarrow} \mathcal{L}(T''_n),\ n \to \infty$. *Suppose that the sequence of distribution functions (d.f.s)* $\{G_n(\cdot)\}_{n\geq 1}$ *of r.v.s* T''_n *is uniformly continuous. Then*

$$\sup_{z\in\mathbf{R}^1} \mid F_n(z \mid \mathbf{X}'_n) - G_n(x) \mid \xrightarrow{\mathrm{P}} 0 \quad as \quad n \to \infty,$$

where $F_n(\cdot \mid \mathbf{X}'_n)$ *is the conditional d.f. of* T'_n *given* $\mathbf{X}'_n$.

Proof. See Lemma 9 and Corollary 1 in Belyaev and Sjöstedt-de Luna (2000).

Acknowledgements

This work is a part of the Swedish research programme "Remote Sensing for the Environment", RESE (home page http://www.lantmateriet.se/rese/), financed by the Swedish Foundation for Strategic Environmental Research, MISTRA. It is also supported by the Bank of Sweden Tercentenary Foundation. The author is thankful to Prof. B. Ranneby, Drs. M. Ekström, O. Seleznjev and S. Sjöstedt-de Luna for several useful comments on the draft version of this report and to Dr. P. Haemig for improving the English language.

References

Athreya, K. B. (1987). Bootstrap of the main in the infinite variance case. *Ann. Statist.,* **15**, pp. 724-731.

Belyaev, Yu. K. (1995). *Bootstrap, Resampling and Mallows Metric.* Lecture Notes 1, Institute of Mathematical Statistics, Umeå University, Sweden.

Belyaev, Yu. K. and Seleznjev, O. V. (2000). Approaching in distribution with applications to resampling of stochastic processes. *Scand. Journ. of Statistics,* **27**, No 3, pp. 371-384.

Belyaev, Yu. K. and Sjöstedt, S. (1997). Resampling from independent heterogeneous random variables with varying mean values. *Theory of Stochastic Processes,* **3**, No 1-2, pp. 121-131.

Belyaev, Yu. K. and Sjöstedt-de Luna, S. (2000). Weakly approaching sequences of random distributions. *Journal of Applied Probability,* **37**, No 3, pp. 807-822.

Dudley, R. M. (1989). *Real Analysis and Probability.* Chapman & Hall, New York.

Giné, E. and Zinn, J. (1989). Necessary conditions for the bootstrap of the mean. *Ann. Stat.*, **17**, pp. 684-691.

Gnedenko, B. V. and Kolmogorov, A. N. (1968). *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley, London.

Hall, P. (1992). *The Bootstrap and the Edgeworth Expansion.* Springer-Verlag, New York.

Loève, M. (1977). *Probability Theory I,* v. I, 4th ed., Springer-Verlag, New York.

Mammen, E. (1992). Bootstrap, wild bootstrap and asymptotic normality. *Probability Theory and Related Fields,* **93**, pp. 439-455.

Shiryaev, A. N. (1996). *Probability*, 2nd ed., Springer, New York.

Wu, C. F. J. (1986). Jacknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics,* **14**, pp. 1261-1295.