Vietnam Journal of Mathematics **37**:4(2009) 443-456

Vietnam Journal
of
MATHEMATICS

On Bootstrapping Regression and Correlation Models with Random Resample Size

Nguyen Van Toan

University of Technical Education, Ho Chi Minh city, Vietnam

Received February 22, 2008
Revised September 02, 2009

Abstract. The regression and correlation models are considered. Under some fixed assumptions on random resample size we show that the bootstrap approximation with random resample size to the distribution of the least squares estimates works as bootstrap.

1991 Mathematics Subject Classification: Primary 62E20; Secondary 62G05, 62G15.

Key words: Regression, correlation, bootstrap, Mallows metric, random resample size.

1. Introduction

Efron [2] introduced a very general resampling procedure, called the bootstrap, for estimating the distributions of statistics based on independent observations. The procedure is more widely applicable and perhaps, sounds more basis than the popular Quenouille-Tukey jackknife. Efron considered a number of statistical problems and demonstrated the feasibility of the bootstrap method. Freedman [3] has developed some asymptotic theory for applications of Efron's bootstrap to regression with the resample size m to differ from the number n of data points. The purpose of this paper is to study bootstrap regression models with a random resample size.

Consider a linear model of the form

$$Y(n) = X(n)\beta + \varepsilon(n), \tag{1}$$

where $Y(n)$ is an $n \times 1$ data vector; $X(n)$ is an $n \times p$ data matrix, of full rank $p \leq n$; β is a $p \times 1$ vector of unknown parameters, to be estimated from the data; $\varepsilon(n)$ is an $n \times 1$ vector of unobservables. The models differ in the stochastic mechanism supposed to have generated the data. However, in all cases ε is supposed random. Data are observed in the form $(X(n),\ Y(n))$, and we call $X(n)$ the set of design points of the model. We say that (1) is a regression model if analysis is carried out conditional on the design points. The design points could be genuinely fixed (for example, they could be regularly spaced) or they could have a random source but be conditioned upon. In a regression model the components $\varepsilon_1,\ \varepsilon_2 \ldots,\ \varepsilon_n$ of $\varepsilon(n)$ would be regarded as independent random variables with zero mean, usually assumed to be identically distributed. The case where the rows $(X_i,\ Y_i),\ 1 \leq i \leq n$, of $(X(n),\ Y(n))$ are independent and identically distributed pairs of random vectors is, in our nomenclature, a correlation model.

Resampling in Regression and Correlation Models. The idea behind the bootstrap is to replace the true distribution function F, in a formula for an unknown quantity, by its empirical estimate $\hat{F}$. Under the regression model, F is the distribution function of the errors ε_i, and $\hat{F}$ is the empirical distribution function of the sequence of residuals

$$\hat{\varepsilon}_i = Y_i - X_i\hat{\beta}, \quad 1 \leq i \leq n,$$

where $\hat{\beta}$ is the usual least-squares estimates of β. In the case of the correlation model, F is the distribution function of the pairs $(X_i,\ Y_i),\ 1 \leq i \leq n$.

First we treat the case of the regression model. Here, the residuals $\hat{\varepsilon}_i$ are resampled; note that the residuals are centred, in the sense that $\Sigma\hat{\varepsilon}_i = 0$. We take $\{\hat{\varepsilon}_1^*, \ldots,\ \hat{\varepsilon}_n^*\}$ to be a resample drawn randomly, with replacement, from the set $\{\hat{\varepsilon}_1, \ldots,\ \hat{\varepsilon}_n\}$ and define

$$Y_i^* = X_i\hat{\beta} + \hat{\varepsilon}_i^*, \quad 1 \leq i \leq m.$$

The joint distribution of $\{Y_i^*,\ 1 \leq i \leq m\}$, conditional on

$$\mathcal{X} = \{(X_1,\ Y_1), \ldots,\ (X_n,\ Y_n)\},$$

is the bootstrap estimate of the joint distribution of $\{Y_i,\ 1 \leq i \leq n\}$, conditional on $X_1, \ldots,\ X_n$.

In the case of the correlation model, the pairs $(X_i,\ Y_i)$ are resampled directly. That is, we draw a resample

$$\mathcal{X}^* = \{(X_1^*,\ Y_1^*), \ldots,\ (X_m^*,\ Y_m^*)\}$$

randomly, with replacement, from

$$\mathcal{X} = \{(X_1,\ Y_1), \ldots,\ (X_n,\ Y_n)\}.$$

The joint distribution of $\mathcal{X}^*$ conditional on $\mathcal{X}$ is the bootstrap estimate of the unconditional joint distribution of $\mathcal{X}$.

The regression model. We continue to assume the regression model (1), with the following conditions.

(A1) The matrix $X(n)$ is not random.

(A2) The components $\varepsilon_1,\ \varepsilon_2 \ldots,\ \varepsilon_n$ of $\varepsilon(n)$ are independent, with common distribution F having mean 0 and finite variance σ^2; both F and σ^2 are unknown.

(A3) $\frac{1}{n}\{X(n)^T X(n)\} \to V$ which is positive definite.

Consider $X(n)$ as the first n of an infinite sequence of rows. Likewise, consider the disturbances $\varepsilon_1, \ldots,\ \varepsilon_n$ as the first n of an infinite sequence of independent random variables with common distribution function F. Least-squares estimate of β is given by

$$\hat{\beta}(n) = (X(n)^T X(n))^{-1} X(n)^T Y(n).$$

The observable column n-vector $\hat{\varepsilon}(n)$ of residuals is given by

$$\hat{\varepsilon}(n) = Y(n) - X(n)\hat{\beta}.$$

Let $\hat{F}_n$ be the empirical distribution of $\hat{\varepsilon}(n)$, centered at the mean, so $\hat{F}_n$ puts mass $1/n$ at $\hat{\varepsilon}_i(n) - \hat{\mu}_n$ and $\int x d\hat{F}_n^x = 0$, where $\hat{\mu}(n) = \frac{1}{n}\sum_{i=1}^n \hat{\varepsilon}_i(n)$. Given $Y(n)$, let $\varepsilon_1^*, \ldots,\ \varepsilon_m^*$ be conditionally independent, with common distribution $\hat{F}_n$; let $\varepsilon^*(m)$ be the m-vector whose i-th component is ε_i^*; and let

$$Y^*(m) = X(m)\hat{\beta}(n) + \varepsilon^*(m).$$

Bootstrap version of $\hat{\beta}(n)$ is

$$\hat{\beta}^*(m) = [X(m)^T X(m)]^{-1} X(m)^T Y^*(m).$$

The bootstrap principle is that the distribution of $\sqrt{m}\left\{\hat{\beta}^*(m) - \hat{\beta}(n)\right\}$, which can be computed directly from the data, approximates the distribution of $\sqrt{n}(\hat{\beta}(n) - \beta)$. Freedman [3] shown that this approximation is likely to be very good, provided n and m are large; and $\sigma^2 p.\text{trace}(X^T X)^{-1}$ is small. Theorem 2.2 of Freedman [3] is the a.s. justification of the bootstrap asymptotics as m and n tend to ∞.

In the case where the bootstrap resample size N is in itself a random variable, Mammen [4] has considered bootstrap with a Poisson random resample size which is independent of the original sample. Stemming from Efron's observation that the information content of a bootstrap sample is based on approximately $(1 - \mathrm{e}^{-1})100\% \approx 63\%$ of the original sample, Rao, Pathak and Koltchinskii [5] have introduced a sequential resampling method in which sampling is carried out one-by-one (with replacement) until $(m+1)$ distinct original observation appear, where m denotes the largest integer not exceeding $(1 - \mathrm{e}^{-1})n$. It has been shown that the empirical characteristics of this sequential bootstrap are within a distance $O(n^{-3/4})$ from the usual bootstrap. The authors provide a heuristic argument in favor of their sampling scheme and establish the consistency of the sequential bootstrap. Our work on this problem is limited to [6]-[14]. In these references we consider bootstrap for means with a random resample size which is independent of the original sample and find sufficient conditions for random resample size that random sample size bootstrap distribution can be used to approximate the sampling distribution. The purpose of this paper is to study bootstrap regression models with a random resample size and arrived at results very similar to those obtained for means in the one-sample problem.

In Sec. 2, we show that the bootstrap works for the regression model if the random bootstrap resample size N_n is a positive integer-valued random variable independent of $Y_1, Y_2, \ldots,\ Y_n$ such that

$$N_n \to_p \infty \quad \text{as } n \to \infty, \tag{2}$$

where $\to_p$ denotes convergence in probability.

The correlation model. In the case of the usual correlation model we assume that

(B1) The vectors (X_i, Y_i) are independent, with common (unknown) distribution μ in $\mathbb{R}^{p+1}$; and $E\{\|(X_i, Y_i)\|^4\} < \infty$, where $\| \, . \, \|$ is Euclidean length,

(B2) $\Sigma = E\{X_i^T X_i\}$, the $p \times p$ variance-covariance matrix of rows of X, is positive definite.

Let μ_n be the empirical distribution of the $(X_i,\ Y_i)$ for $i = 1, \ldots, n$. Given $\{X(n),\ Y(n)\}$, let $(X_i^*,\ Y_i^*)$ be independent, with common distribution μ_n, for $i = 1, \ldots, m$. Informally, this amounts to taking a resample of size m from the n observed vectors.

The p-vector $\hat{\beta}(n)$ minimizes $\frac{1}{n}\sum_{i=1}^n \{Y_i - X_i\hat{\beta}(n)\}^2$. Thus, $\hat{\beta}(n)$ approaches to μ_n as β to the true law μ of $(X_i,\ Y_i)$. Let $\hat{\beta}^*(m)$ be the least squares estimate based on the resample

$$\hat{\beta}^*(m) = \{X^*(m)^T X^*(m)\}^{-1} X^*(m)^T Y^*(m). \tag{3}$$

Freedman [3] has shown that the conditional law of $\sqrt{m}\{\hat{\beta}^*(m) - \hat{\beta}(n)\}$ must be close to the unconditional law of $\sqrt{n}\{\hat{\beta}(n) - \beta\}$, i.e., the bootstrap approximation is valid. In Sec. 3, it will be shown that if the random bootstrap resample size N_n is a positive integer-valued random variable independent of $\{X(n),\ Y(n)\}$ such that (2) holds, the bootstrap works for the correlation model.

2. The Regression Model

Now return to the regression model described in Sec. 1. Let F_n be the empirical distribution function of $\varepsilon_1, \ldots, \varepsilon_n$; let $\tilde{F}_n$ be the empirical distribution function of the residuals $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n$ from the original regression on n data vectors, and let $\hat{F}_n$ be $\tilde{F}_n$ centered at its mean $\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^n \hat{\varepsilon}_i$. Let $\Psi_n(F)$ be the distribution of $\sqrt{n}\{\hat{\beta}(n) - \beta\}$, where F is the law of the ε's. So $\Psi_n(F)$ is a probability in $\mathbb{R}^p$. Then $\Psi_m(\hat{F}_n)$ is the law of $\sqrt{m}\{\hat{\beta}^*(m) - \hat{\beta}(n)\}$.

Let d_i^p be the Mallows metric for probabilities in $\mathbb{R}^p$, relative to the Euclidean norm $\| \,.\, \|$. For details, see Section 8 of Bickel and Freedman [1]. Only $i = 1$ or 2 are of present interest. Let $N(0,\ \sigma^2 V^{-1})$ be the normal distribution with mean 0 and variance-covariance matrix $\sigma^2 V^{-1}$.

Lemma 2.1. *Assume the regression model , with* (1), (*A*1)$-$(*A*3)*. Along almost all sample sequences, given* $Y_1, \dots,\ Y_n$*, as* m *and* n *tend to infinity,*

$$d_2^p\{\Psi_m(\hat{F}_n),\ N(0,\ \sigma^2 V^{-1})\} \to 0.$$

Proof. Use Theorem 2.2 of Freedman [3] and Lemma 8.3 of Bickel and Freedman [1]. ■

Let $\hat{\varepsilon}^*(m)$ be the bootstrap version of $\hat{\varepsilon}(n) : \hat{\varepsilon}^*(m) = Y^*(m) - X(m)\hat{\beta}^*(m)$. The theoretical variance $\sigma^2 = E(\varepsilon_i^2)$ is estimated from the n original data vectors by

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i^2(n) - \mu_n^2, \qquad \text{where} \qquad \mu_n = \frac{1}{n}\sum_{i=1}^{n} \hat{\varepsilon}_i(n),$$

and the bootstrap estimate is

$$\hat{\sigma}_m^{*^2} = \frac{1}{m}\sum_{i=1}^{n} \hat{\varepsilon}_i^{*^2}(m) - \mu_m^{*^2}, \qquad \text{where} \qquad \mu_m^* = \frac{1}{m}\sum_{i=1}^{m} \hat{\varepsilon}_i^*(m).$$

Let

$$\sigma_m^{*2} = \frac{1}{m}\sum_{i=1}^{m} \varepsilon_i^{*2} - \left(\frac{1}{m}\sum_{i=1}^{m} \varepsilon_i^*\right)^2.$$

Lemma 2.2. *Assume the regression model, with* (1), (*A*1) $-$ (*A*3), (2)*. Along almost all sample sequences, given* $Y_1, \dots,\ Y_n$*, as* n *tends to infinity, the conditional distribution of* $\sigma_{N_n}^*$ *converges to point mass at* σ*.*

Proof. Let $f_m^*(t) = E^*\left(e^{it\sigma_m^{*2}}\right)$, $f_{N_n}^*(t) = E^*\left(e^{it\sigma_{N_n}^{*2}}\right)$ and $h(t) = e^{i\sigma^2 t}$, where $E^*(\dots)$ is the conditional expectation $E(\dots|Y_1,\ \dots,\ Y_n)$.

Since N_n is independent of $Y_1, Y_2, \dots, Y_n$,

$$\begin{aligned}|f_{N_n}^*(t) - h(t)| &= \sum_{m=1}^{\infty} P(N_n = m)|f_m^*(t) - h(t)| \\ &\le 2P(N_n \le m_0) + \sum_{m>m_0} P(N_n = m)|f_m^*(t) - h(t)|.\end{aligned}$$

By the proof of Theorem 2.2b in [3], the conditional distribution of σ_m^{*2} converges to point mass at σ^2 as m and n tend to infinity, it follows that with probability one for every $\epsilon > 0$ there exist $m_0,\ n_0 > 0$ such that

$$|f_m^*(t) - h(t)| < \epsilon \quad \forall m > m_0, \forall n > n_0.$$

Because of (2) there exists $n_1 > 0$ such that

$$P(N_n \le m_0) < \epsilon \quad \forall n > n_1.$$

Therefore

$$|f_{N_n}^*(t) - h(t)| < 3\epsilon \quad \forall n > \max\{n_0, n_1\}$$

that is

$$f_{N_n}^*(t) \to h(t) \ a.s.,$$

and the lemma follows. ■

We can now state the main result.

Theorem 2.3. *Assume the regression model, with* (1), *(A*1*)* − *(A*3*),* (2)*. Along almost all sample sequences, given* $Y_1, \ldots, Y_n$*, as* n *tends to infinity,*

a) *the conditional distribution of* $\sqrt{N_n}\{\hat{\beta}^*(N_n) - \hat{\beta}(n)\}$ *converges weakly to a normal distribution with mean 0 and variance-covariance matrix* $\sigma^2 V^{-1}$.

b) *the conditional distribution of* $\hat{\sigma}_{N_n}^*$ *converges to point mass at* σ.

c) *the conditional distribution of the pivot* $\{X(N_n)^T X(N_n)\}^{1/2}\{\hat{\beta}^*(N_n) - \hat{\beta}(n)\}/\hat{\sigma}_{N_n}^*$ *converges to standard normal in* $\mathbb{R}^p$.

Proof. Claim a) The bootstrap construction can be put into present notation as follows: conditionally, law of $\sqrt{N_n}\{\hat{\beta}^*(N_n) - \hat{\beta}(n)\}$ is $\Psi_{N_n}(F_n)$. Since N_n is independent of $Y_1, Y_2, \ldots, Y_n$,

$$\begin{aligned} & d_2^p \left\{ \Psi_{N_n}\left(\hat{F}_n\right),\ N\left(0,\ \sigma^2 V^{-1}\right) \right\} \\ &= \sum_{m=1}^{\infty} P\left(N_n = m\right) d_2^p \left\{ \Psi_m \left(\hat{F}_n\right),\ N\left(0,\ \sigma^2 V^{-1}\right) \right\} \\ &= A(k) + B(k), \end{aligned} \tag{4}$$

where

$$\begin{aligned} A(k) &= \sum_{m=1}^{k} P(N_n = m) d_2^p \left\{ \Psi_m\left(\hat{F}_n\right),\ N(0,\ \sigma^2 V^{-1}\right) \right\}, \\ B(k) &= \sum_{m>k} P(N_n = m) d_2^p \left\{ \Psi_m\left(\hat{F}_n\right),\ N(0,\ \sigma^2 V^{-1}\right) \right\}. \end{aligned}$$

Lemma 2.1 implies that with probability one for every $\delta > 0$ there exist $k,\ l > 0$ such that

$$d_2^p \left\{ \Psi_m(\hat{F}_n),\ N\left(0,\ \sigma^2 V^{-1}\right) \right\} < \delta \quad \forall m > k,\ \forall n > l. \tag{5}$$

Because of (2) there exists $q > 0$ such that

$$P(N_n \le k) < \delta \left[\max_{1\le m\le k} d_2^p \left\{ \Psi_m(\hat{F}_n),\ N\left(0,\ \sigma^2 V^{-1}\right)\right\}\right]^{-1} \quad \forall n > q. \tag{6}$$

(4)-(6) imply

$$d_2^p \left\{ \Psi_{N_n}(\hat{F}_n),\ N\left(0,\ \sigma^2 V^{-1}\right)\right\} < 2\delta \quad \forall n > \max\{l, q\}$$

that is

$$d_2^p \left\{ \Psi_{N_n}(\hat{F}_n),\ N\left(0,\ \sigma^2 V^{-1}\right)\right\} \to 0 \qquad \text{a.s.}$$

This shows Claim a).

Claim b) Let $\bar{\varepsilon}^*(m) = \frac{1}{m}\sum_{i=1}^m \varepsilon_i^*$. Clearly

$$\sigma_m^* = \frac{1}{\sqrt{m}} \|\varepsilon^*(m) - \bar{\varepsilon}^*(m)\|$$

and

$$\hat{\sigma}_m^* = \frac{1}{\sqrt{m}} \|\hat{\varepsilon}^*(m) - \mu_m^*\|,$$

so

$$\begin{aligned}(\hat{\sigma}_m^* - \sigma_m^*)^2 &\le \frac{1}{m}\|(\hat{\varepsilon}^*(m) - \mu_m^*) - (\varepsilon^*(m) - \bar{\varepsilon}^*(m))\|^2 \\ &= \frac{1}{m}[\|\hat{\varepsilon}^*(m) - \varepsilon^*(m)\|^2 - (\mu_m^* - \bar{\varepsilon}^*(m))^2] \\ &= \frac{1}{m}\|\hat{\varepsilon}^*(m) - \varepsilon^*(m)\|^2\end{aligned}$$

and

$$\begin{aligned}(E^*|\hat{\sigma}_{N_n}^* - \sigma_{N_n}^*|)^2 &\le E^*[(\hat{\sigma}_{N_n}^* - \sigma_{N_n}^*)^2] \\ &= \sum_{m=1}^{\infty} P(N_n = m) E^*[(\hat{\sigma}_m^* - \sigma_m^*)^2] \\ &\le \sum_{m=1}^{\infty} P(N_n = m) E^* \left[\frac{1}{m}\|\hat{\varepsilon}^*(m) - \varepsilon^*(m)\|^2\right].\end{aligned} \tag{7}$$

Since

$$\hat{\varepsilon}^*(m) - \varepsilon^*(m) = -P(m)\varepsilon^*(m), \tag{8}$$

where $P(m) = X(m)\{X(m)^T X(m)\}^{-1} X(m)^T$ is the projection matrix onto the column space of $X(m)$, a routine computation starting from (8) shows that

$$E^*(\|\hat{\varepsilon}^*(m) - \varepsilon^*(m)\|^2) = \hat{\sigma}_n^2 p. \tag{9}$$

(7) and (9) imply

$$
\begin{aligned}
(E^*|\hat{\sigma}^*_{N_n} - \sigma^*_{N_n}|)^2 &\le \sum_{m=1}^{\infty} P(N_n = m)\frac{\hat{\sigma}^2_n p}{m} \\
&\le \hat{\sigma}^2_n p \left(\sum_{m=1}^{k} P(N_n = m) + \frac{1}{k} \sum_{m=k+1}^{\infty} P(N_n = m) \right) \qquad (10) \\
&\le \hat{\sigma}^2_n p \left(P(N_n \le k) + \frac{1}{k} \right).
\end{aligned}
$$

Since $\frac{1}{k} \to 0$ as $k \to \infty$, it follows that for every $\eta > 0$ there exists $k > 0$ such that

$$\frac{1}{k} < \eta. \qquad (11)$$

Because of (2) there exists $q > 0$ such that

$$P(N_n \le k) < \eta \quad \forall n > q. \qquad (12)$$

(2.10) of Freedman [3] and (10)-(12) imply

$$E^*|\hat{\sigma}^*_{N_n} - \sigma^*_{N_n}| \to 0 \qquad \text{a.s.} \qquad (13)$$

Claim b) follows from (13) and Lemma 2.2.

Claim c) This is immediate from a) and b) if we show that

$$\frac{1}{N_n}\{X(N_n)^T X(N_n)\} \to_p V. \qquad (14)$$

It is clear that

$$
\begin{aligned}
E\left[\left\|\frac{1}{N_n}\{X(N_n)^T X(N_n)\} - V\right\|\right] &= \sum_{m=1}^{\infty} P(N_n = m)\left\|\frac{1}{m}\{X(m)^T X(m)\} - V\right\| \\
&= M(k) + N(k),
\end{aligned}
$$

where

$$
\begin{aligned}
M(k) &= \sum_{m=1}^{k} P(N_n = m)\left\|\frac{1}{m}\left\{X(m)^T X(m)\right\} - V\right\|, \\
N(k) &= \sum_{m=k+1}^{\infty} P(N_n = m)\left\|\frac{1}{m}\left\{X(m)^T X(m)\right\} - V\right\|.
\end{aligned}
$$

Analysis similar to that in the proof of Claim a) shows that

$$E\left[\left\|\frac{1}{N_n}\{X(N_n)^T X(N_n)\} - V\right\|\right] \to 0,$$

and (14) is proved. ■

3. The Correlation Model

In this section, the object is to develop some asymptotic theory for applications of Efron's [2] bootstrap to correlation model with random resample size. Assume the correlation model, with conditions (B1), (B2). The original n data vectors are $(X_i,\ Y_i)$ for $i = 1, \ldots, n$; these are independent, with common distribution μ; their empirical distribution is μ_n. Given $\{X(n),\ Y(n)\}$, the resampled vectors $(X_i^*,\ Y_i^*)$ are independent, with common distribution μ_n, for $i = 1, \ldots, m$. Let $X^*(m)$ be the $m \times p$ matrix whose i-th row is X_i^* and $Y^*(m)$ be the $m \times 1$ column vector of Y_i^*'s. The least squares estimate based on the original data is $\hat{\beta}(n)$; on the resample, $\hat{\beta}^*(m)$, see (3). In the original data, the vector of unobservable disturbances is $\varepsilon(n)$, with $\varepsilon_i = Y_i - X_i\beta$; the observable residuals are

$$\hat{\varepsilon}(n) = Y(n) - X(n)\hat{\beta}(n). \tag{15}$$

In the resample, the $m \times 1$ column vector of disturbances is ε^*, with

$$\varepsilon_i^* = Y_i^* - X_i^*(m)\hat{\beta}(n). \tag{16}$$

The $m \times 1$ column vector of residuals is $\hat{\varepsilon}^*(m)$, with

$$\hat{\varepsilon}_i^*(m) = Y_i^* - X_i^*(m)\hat{\beta}^*(m). \tag{17}$$

Let

$$W^*(m) = \frac{1}{m} X^*(m)^T X^*(m) = \frac{1}{m}\sum_{i=1}^{m} {X_i^*}^T X_i^*,$$
$$Z^*(m) = \frac{1}{\sqrt{m}} X^*(m)^T \varepsilon^*(m) = \frac{1}{\sqrt{m}}\sum_{i=1}^{m} {X_i^*}^T \varepsilon_i^*$$

and $N(0, M)$ be the normal distribution with mean 0 and variance-covariance matrix M, where the nonnegative definite matrix M is defined by

$$M_{jk} = E(X_{ij} X_{ik} \varepsilon_i^2).$$

The proof of Theorem 3.1 in [3] implies the following result

Lemma 3.1. *Assume the correlation model, with conditions* $(B1), (B2)$. *Along almost all sample sequences, given* (X_i, Y_i) *for* $1 \le i \le n$, *as* m *and* n *go to infinity,*

a) $d_1^{p\times p}\{W^*(m),\ \Sigma\} \to 0.$

b) $d_2^p\{Z^*(m); N(0,M)\} \to 0.$

The next theorems show that the bootstrap works for the correlation model if the random bootstrap resample size N_n is a positive integer-valued random variable independent of $\{X(n), Y(n)\}$ such that

$$N_n \to_p \infty \quad \text{as } n \to \infty. \tag{18}$$

Theorem 3.2. *Assume the correlation model, with conditions* $(B1), (B2)$ *and* (18). *Along almost all sample sequences, given* $(X_i,\ Y_i)$ *for* $1 \le i \le n$, *as* n *tends to infinity,*

a) $\frac{1}{N_n} X^*(N_n)^T X^*(N_n)$ *converges in conditional probability to* Σ,

b) *the conditional law of* $\sqrt{N_n}\{\hat\beta^*(N_n) - \hat\beta(n)\}$ *goes weakly to normal with mean 0 and variance-covariance matrix* $\Sigma^{-1} M \Sigma^{-1}$.

Proof. Claim a) Since N_n is independent of $\{X(n), Y(n)\}$,

$$d_1^{p\times p}\{W^*(N_n), \Sigma\} = \sum_{m=1}^{\infty} P(N_n = m) d_1^{p\times p}\{W^*(m), \Sigma\} = A(k) + B(k), \tag{19}$$

where

$$A(k) = \sum_{m=1}^{k} P(N_n = m) d_1^{p\times p}\{W^*(m), \Sigma\},$$
$$B(k) = \sum_{m>k} P(N_n = m) d_1^{p\times p}\{W^*(m), \Sigma\}.$$

By Lemma 3.1a and (18), (19), we can now proceed analogously to the proof of Theorem 2.3a, and thus obtain

$$d_1^{p\times p}\{W^*(N_n), \Sigma\} \to 0 \quad \text{a.e.},$$

which proves Claim a).

Claim b). We first prove that

$$d_2^p\{Z^*(N_n), N(0,M)\} \to 0 \quad \text{a.e. as } n \to \infty. \tag{20}$$

This follows by the same method as in the proof of Claim a). Since N_n is independent of $\{X(n),\ Y(n)\}$,

$$\begin{aligned} d_2^p\{Z^*(N_n), N(0,M)\} &= \sum_{m=1}^{\infty} P(N_n = m) d_2^p\{Z^*(m), N(0,M)\} \\ &= C(k) + D(k), \end{aligned} \tag{21}$$

where

$$C(k) = \sum_{m=1}^{k} P(N_n = m) d_2^p \{Z^*(m), N(0, M)\},$$
$$D(k) = \sum_{m>k} P(N_n = m) d_2^p \{Z^*(m), N(0, M)\}.$$

Therefore (20) follows from Lemma 3.1b, (18) and (21).

Combining Claim a) with (20) we obtain Claim b), and the proof is complete. ■

The conventional estimate for $\sigma^2 = E(\varepsilon_i^2)$ is the mean square of the residuals (see (15)), denoted here by

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i(n)^2.$$

The corresponding estimate based on the resample is

$$\hat{\sigma}_m^{*^2} = \frac{1}{m} \sum_{i=1}^{m} \hat{\varepsilon}_i^*(m)^2$$

where the residuals $\hat{\varepsilon}_1^*(m), \ldots, \hat{\varepsilon}_m^*(m)$ are defined in (17).

Theorem 3.3. *Assume the correlation model, with conditions* $(B1)$, $(B2)$ *and* (18). *Along almost all sample sequences, given* $(X_i,\ Y_i)$ *for* $1 \le i \le n$, *as* n *tends to infinity, the conditional law of* $\hat{\sigma}_{N_n}^*$ *converges weakly to point mass at* σ.

Proof. Let

$$\sigma_m^{*^2} = \frac{1}{m} \sum_{i=1}^{m} \varepsilon_i^{*^2},$$

where the disturbances $\varepsilon_1^*(m), \ldots, \varepsilon_m^*(m)$ are defined in (16).

It will be shown that

$$\sigma_{N_n}^* \to \sigma \quad \text{a.e.} \tag{22}$$

In view of Lemma 8.6 of Bickel and Freedman [1], since N_n is independent of $\{X(n),\ Y(n)\}$,

$$d_1\left(\frac{1}{N_n}\sum_{i=1}^{N_n} \varepsilon_i^{*^2}, \frac{1}{N_n}\sum_{i=1}^{N_n} \varepsilon_i^2\right)$$
$$= \sum_{m=1}^{\infty} P(N_n = m) d_1\left(\frac{1}{m}\sum_{i=1}^{m} \varepsilon_i^{*^2}, \frac{1}{m}\sum_{i=1}^{m} \varepsilon_i^2\right) \le d_1(\varepsilon_i^{*^2},\ \varepsilon_i^2).$$

But $d_1(\varepsilon_i^{*^2}, \varepsilon_i^2)$ tends to 0 a.e. by Lemmas 3.2 and Lemma 3.1c of Freedman [3]. It follows that the conditional law of $\frac{1}{N_n}\sum_{i=1}^{N_n} \varepsilon_i^{*^2}$ is close to the unconditional law of $\frac{1}{N_n}\sum_{i=1}^{N_n}\varepsilon_i^2$. Next, let $f_m(t)=E\left(e^{it\sigma_m^2}\right)$, $f_{N_n}(t)=E\left(e^{it\sigma_{N_n}^2}\right)$ and $h(t)=e^{i\sigma^2 t}$, where $\sigma_m^2=\frac{1}{m}\sum_{i=1}^{m}\varepsilon_i^2$. Since N_n is independent of $\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n,\ldots$

$$\begin{aligned}|f_{N_n}(t)-h(t)| &= \sum_{m=1}^{\infty}P(N_n=m)|f_m(t)-h(t)| \\ &\le 2P(N_n\le m_0)+\sum_{m>m_0}P(N_n=m)|f_m(t)-h(t)|. \qquad (23)\end{aligned}$$

Since the distribution of σ_m^2 converges to point mass at σ^2 as m tends to infinity, it follows that with probability one for every $\epsilon>0$ there exists $m_0>0$ such that $|f_m(t)-h(t)|<\epsilon$ for all $m>m_0$. Because of (18) there exists $n_0>0$ such that $P(N_n\le m_0)<\epsilon$ for all $n>n_0$. From (23) it may be concluded that $|f_{N_n}(t)-h(t)|<3\epsilon$ for all $n>n_0$, and (22) follows.

Now

$$(\hat\sigma_m^*-\sigma_m^*)^2\le\frac{1}{m}\sum_{i=1}^{m}\{\hat\varepsilon_i^*(m)-\varepsilon_i^*(m)\}^2$$

so

$$\left(\hat\sigma_{N_n}^*-\sigma_{N_n}^*\right)^2\le\frac{1}{N_n}\sum_{i=1}^{N_n}\{\hat\varepsilon_i^*(N_n)-\varepsilon_i^*(N_n)\}^2$$

and it remains only to show that

(C) The conditional law of $\frac{1}{N_n}\sum_{i=1}^{N_n}\{\hat\varepsilon_i^*(N_n)-\varepsilon_i^*(N_n)\}^2$ concentrates near 0.

Let $P^*(\ldots)$ be the conditional probability $P(\ldots|\{X_1,Y_1\},\ldots,\{X_n,Y_n\})$. Since N_n is independent of $\{X(n),\ Y(n)\}$,

$$\begin{aligned}&P^*\left(\frac{1}{N_n}\sum_{i=1}^{N_n}\{\hat\varepsilon_i^*(N_n)-\varepsilon_i^*(N_n)\}^2>\delta\right)\\ &=\sum_{m=1}^{\infty}P(N_n=m)P^*\left(\frac{1}{m}\sum_{i=1}^{m}\{\hat\varepsilon_i^*(m)-\varepsilon_i^*(m)\}^2>\delta\right)\\ &\le P(N_n\le m_0)+\sum_{m>m_0}P(N_n=m)P^*\left(\frac{1}{m}\sum_{i=1}^{m}\{\hat\varepsilon_i^*(m)-\varepsilon_i^*(m)\}^2>\delta\right)\end{aligned}$$

for all $\delta>0$.

From (3.12) of Freedman [3] it follows that with probability one for every $\epsilon>0$ there exist $m_0,\ n_0>0$ such that

$$P^*\left(\frac{1}{m}\sum_{i=1}^{m}\{\hat\varepsilon_i^*(m)-\varepsilon_i^*(m)\}^2>\delta\right)<\epsilon\quad\forall m>m_0,\forall n>n_0.$$

Because of (18) there exists $n_1>0$ such that

$$P(N_n \leq m_0) < \epsilon \quad \forall n > n_1.$$

Therefore

$$P^* \left(\frac{1}{N_n} \sum_{i=1}^{N_n} \{\hat{\varepsilon}_i^*(N_n) - \varepsilon_i^*(N_n)\}^2 > \delta \right) < 2\epsilon \quad \forall n > \max\{n_0, n_1\},$$

and (C) follows. The theorem is proved. ■

Remark 3.4. In particular, as n tends to ∞, the conditional law of

$$\{X^*(N_n)^T X^*(N_n)\}^{1/2}\{\hat{\beta}^*(N_n) - \hat{\beta}(n)\}/\hat{\sigma}^*_{N_n}$$

converges to the appropriate limit: normal with mean 0 and variance-covariance matrix $\Sigma^{-1/2} M \Sigma^{-1/2}/\sigma^2$. In the case where the correlation model is "homoscedastic", which can be interpreted mathematically as follows:

$$E(\varepsilon_i^2 | X_i) = \sigma^2 \quad \text{a.e.},$$

this is just the $p \times p$ identity matrix $I_{p\times p}$.

Acknowledgements. The author is grateful to the referee for some helpful comments on an earlier version of the manuscript.

References

1. P. J. Bickel and D. A. Freedman, Some asymptotic theory for the bootstrap, *Ann. Statist.* **9** (1981), 1196–1217.
2. B. Efron, Bootstrap methods: Another look at the Jackknife, *Ann. Statist.* **7** (1979), 1–26.
3. D. A. Freedman, Bootstrapping regression models, *Ann. Statist.* **9** (1981), 1218–1228.
4. E. Mammen, Bootstrap, wild bootstrap, and asymptotic normality, *Prob. Theory Relat. Fields* **93** (1992), 439–455.
5. C. R. Rao, P. K. Pathak, and V. I. Koltchinskii, Bootstrap by sequential resampling, *J. Statist. Plann. Inference* **64** (1997), 257–281.
6. N. V. Toan, Wild bootstrap and asymptotic normality, *Bulletin, College of Science, Hue University* **10** (1) (1996), 48–52.
7. N. V. Toan, On the bootstrap estimate with random sample size, *Scientific Bulletin of Universities* **Mathematics - Informatics** (1998), 31-34
8. N. V. Toan, On the asymptotic accuracy of the bootstrap with random sample size, *Vietnam J. Math.* **26** (4) (1998), 351–356.
9. N. V. Toan, On the asymptotic accuracy of the bootstrap with random sample size, *Pakistan J. Statist.* **14** (3) (1998), 193–203.
10. N. V. Toan, Rate of convergence in bootstrap approximations with random sample size, *Acta Math. Vietnam.* **25** (2) (2000), 161–179.
11. N. V. Toan, On weak convergence of the bootstrap empirical process with random resample size, *Vietnam J. Math.* **28** (2) (2000), 153–158.

12. N. V. Toan, On the asymptotic distribution of the bootstrap estimate with random resample size, *Vietnam J. Math.* **33** (3) (2005), 261–270.
13. T. M. Tuan and N. V. Toan, On the asymptotic theory for the bootstrap with random sample size, *Proc. NCST VN,* **10** (2) (1998), 3–8.
14. T. M. Tuan and N. V. Toan, An asymptotic normality theorem of the bootstrap sample with random sample size, *VNU Journal of Science, Nat. Sci.* **t XIV** (1) (1998), 1–7.