Moving Block Bootstrap for Analyzing Longitudinal Data

Hyunsu Ju

Department of Preventive Medicine and Community Health, The University of Texas Medical Branch, Galveston, Texas, USA

Abstract

In a longitudinal study subjects are followed over time. I focus on a case where the number of replications over time is large relative to the number of subjects in the study. I investigate the use of moving block bootstrap methods for analyzing such data. Asymptotic properties of the bootstrap methods in this setting are derived. The effectiveness of these resampling methods is also demonstrated through a simulation study.

Keywords

Longitudinal study; Moving block bootstrap

1. Introduction

Many longitudinal designs are the case that the number of subjects n is large and the number of replications m_i is bounded. Liang and Zeger (1986) proved that the generalized estimating equation (GEE) estimator is consistent and asymptotically normal with misspecification of covariance parameters. That asymptotic property is considered when the number of subjects n goes to infinity and m_i is bounded. Xie and Yang (2003) proved the almost sure existence and strong consistency of GEE estimators.

Alternatively in this article, I will consider the case of the longitudinal design in which the number of subjects n is bounded and the number of replications m, the same number for all subjects, is large. The model on which we focus is given by

$$y_i = x_i \beta + e_i, \quad (1)$$

where $i = 1, n, \ldots, x_i$ is $m \times p$ design matrix, β is $p \times 1$ vector of unknown parameters, $y_i = (y_{i1}, \ldots, y_{im})'$, and $e_i = (e_{i1}, \ldots, e_{im})'$.

I focus on the following estimating equation. The estimator $\hat{\beta}$ is called a regression estimator if it solves

Address correspondence to Hyunsu Ju, The University of Texas Medical Branch, 301 University Blvd, Galveston, TX 77573, hyju@utmb.edu.

$$\sum_{i=1}^{n_0} x_i' V_i^{-1}(y_i - x_i\beta)=0. \quad (2)$$

The repeated observations are correlated for each subject. Since bootstrap methods that resample small subjects or resample observations independently may not work well, we will investigate the moving block bootstrap method developed for time-series-correlated data.

The remaining part of this paper is organized as follows. In Sec. 2, we present some literature review. In Sec. 3, we explore the use of moving block bootstrap for analyzing longitudinal data. In Sec. 4, we demonstrate the moving block bootstrap justification in longitudinal data theoretically and empirically. In Sec. 5, we conclude the paper with a brief discussion.

2. Literature Review

Efron (1979) introduced the bootstrap procedure for estimating sampling distributions of statistics based on independent and identically distributed (i.i.d. observations. It is well known, in the i.i.d. setup, that the bootstrap often offers more accurate approximations than classical large sample approximations, e.g., Singh (1981) and Babu (1986). However, when the observations are not necessarily independent, the classical bootstrap no longer succeeds, as shown by Singh (1981).

In a time series case, Lahiri (1996) applied the moving block bootstrap method to multiple linear regression models

$$y_j=x_j'\beta+\epsilon_j, j=1,\dots,m, \quad (3)$$

Where x_j's are known $p \times 1$ vectors, β is a $p \times 1$ vector of parameters, and ϵ_1, ϵ_2, ..., ϵ_m are stationary, strongly mixing random variables. If $\hat{\beta_m}$ is an M-estimator of β corresponding to some score function φ, under some conditions, a two-term Edgeworth expansion for studentized multivariate M-estimator was observed. Also, Lahiri (1996) showed that the block bootstrap has a second-order correctness for some suitable bootstrap analogs of studentized $\hat{\beta_m}$.

Hall et al. (1995) showed that the optimal asymptotic rate of the block size for the moving blocks method depends significantly on context, being equal to $m^{1/3}$, $m^{1/4}$, and $m^{1/5}$ in the cases of variance or bias estimation, estimation of a one-sided distribution function, and estimation of a two-sided distribution function, respectively. The latter two quantities are needed for the construction of equal-tailed and symmetric confidence intervals, respectively. Hall et al. (1995) present a practical rule for selecting the block size empirically. It is based on the fact that the asymptotic formula is $b \sim Cm^{1/k}$, where $k = 3$, 4, or 5 is known, and C is a constant that depends on the underlying process. The rule suggested provides a way for estimating the optimal block for a time series of smaller length than the original.

Paparoditis and Politis (2002) presented a new block bootstrap variation, the tapered block bootstrap, which is applicable in the general time series case of approximately linear statistics. The asymptotic validity and the favorable bias properties of the tapered block bootstrap are shown in two important cases: smooth function of means and M-estimators.

3. Moving Block Bootstrap for Longitudinal Data

The major drawback with model-based resampling is that in practice not only the parameters of a model, but also its structure, must be identified from the data. If the chosen structure is incorrect, the resampled series will be generated for the wrong model, and they will not have the same statistical properties as the original data. The model-based approach is inconsistent if the model used for resampling is misspecified.

The moving bootstrap involves resampling possibly overlapping blocks. The mixed block bootstrep (MBB) does not force one to select a model and the only parameter required is the block length. If the block is long enough the original dependence will be reasonably preserved in the resampled series. This approximation is better if the dependence is weak and the blocks are as long as possible, thus preserving the dependence more faithfully. But the distinct values of the statistics must be as numerous as possible to provide a good estimate of the distribution of the statistics and this points toward short blocks.

Unless the length of the series is considerable to accommodate longer and more number of blocks, the preservation of the dependence structure may be difficult, especially for complex, long-range dependent structures. In such cases, the block resampling scheme tends to generate resampled series that are less dependent than the original ones. Furthermore, the resampled series often exhibits artifacts that are caused by joining randomly selected blocks. Then, the asymptotic variance–covariance matrices of the estimators based on the original series and those based on the bootstrap series are different and a modification of the original scheme is needed. This suggests a strategy intermediate between model-based and block resampling. The idea comes from a procedure of pre-whitening the original dependent series by fitting a model and is intended to remove much of the dependence between the original observations. A resampling series is generated by block resampling of residuals from the simple fitted model, and the innovation series is then post-blackened by applying the simple estimated model to the resampled innovations. The post-blackened version works more consistently in practice (Davison and Hinkley, 1997).

Bühlmann (1997) suggested the sieve bootstrap that is model based. The autoregressive (AR)(p) model is just used to filter the residual series. If the model used in the sieve bootstrap is not appropriate, the resulting residuals cannot be treated as i.i.d. A hybrid approach between the model based method and moving block bootstrap, named post-blacken bootstrap, was suggested by Davison and Hinkley (1997). The procedure is similar to the sieve bootstrap, but the residuals from AR(p) model are not resampled in an iid manner but by using the MBB bootstrap. If some residual dependent structure is present in the AR residuals, this is kept from the blockwise bootstrap. The linear model is used to pre-whiten the series by fitting the model that is intended to remove much of the dependence present the observations. A series of innovations is then generated by block resampling of

residuals obtained from the fitted model, and the innovation series is then post-blackened by applying the estimated model to the resampled innovations.

A block bootstrap algorithm in a longitudinal model is given as follows. We continue to assume (1) as our longitudinal model under consideration.

1. Let $\hat{e}_{ij}$, $i = 1, \ldots, n_0$, $j = 1, \ldots, m$ be the residuals from the model fit.

$$\hat{e}_{ij} = y_{ij} - x_{ij}\hat{\beta}, \quad (4)$$

where $\hat{\beta}$ is the ordinary least square estimate.

2. Now assuming that $m = bk$ with b and k integers: Let $B_1^*, \ldots, B_k^*$ denotes k uniform draws with replacement from the integers $\{0, 1, \ldots, m-b\}$. These represent the starting point for each block of length b. A block bootstrap resample of residuals, $(\hat{e}_{i1}^*, \ldots, \hat{e}_{im}^*)$, is defined by:

$$\hat{e}_{i,(j-1)b+s}^* = \hat{e}_{i, B_j^*+s}, (1 \leq j \leq k, 1 \leq s \leq b) \text{for} \quad \text{each } i. \quad (5)$$

3. The bootstrapped response, y_{ij}^*, are then generated from the estimated model with residuals $\hat{e}_{ij}^*$ and the original covariates:

$$y_{ij}^* = x_{ij}\hat{\beta} + \hat{e}_{ij}^*. \quad (6)$$

4. From the resampled responses, y_{ij}^*, and original covariates, we fit the model and obtain new parameter estimates.

5. Repeating steps (2) through (4) a large number, R, of times one obtains R bootstrap replicates from which features of the distribution of the parameter estimates can be estimated. In particular, the bootstrap variance estimates are simply variance of the B computed values for each parameter.

I consider the six different kinds of block bootstrap methods in a balanced longitudinal design in which the number of subjects is small and the number of replications is large:

Case 1, MBB1 (Within block bootstrap): For each i subject, we construct overlapping blocks with $m - b + 1$ blocks and block size b, i.e $B_1, \ldots, B_{m-b+1}$. Let us define $m/b = k$ which is assumed to be an integer for simplicity, in general $k = [m/b]$. We can add the k blocks with replacement among $B_1, \ldots, B_{m-b+1}$. We get the $B_1^*, \ldots, B_k^*$ with $kb = m$, and create $\hat{e}_{i1}^*, \ldots, \hat{e}_{im}^*$ from $\hat{e}_{i1}, \ldots, \hat{e}_{im}$, where $\hat{e}_{ij} = y_{ij} - \hat{\beta_0} - \hat{\beta_1}x_{ij}$. We can add up to n_0 individuals and plug this into the model and the results is a pseudo sample series $y_{11}^*, \ldots, y_{n_0m}^*$. From the model $y_{ij}^* = \hat{\beta}_0 + \hat{\beta}_1 x_{ij} + \hat{e}_{ij}^*$, we fit model and produce new parameters $\hat{\beta}_0^*$ and $\hat{\beta}_1^*$.

Case 2, MBB2 (Mixed block bootstrap): We have $m - b + 1$ blocks and block size b, i.e $B_1, \ldots, B_{m-b+1}$ and add up to n_0 subjects. We sample $n_0 k$ blocks with replacements among $B_1, \ldots, B_{n_0(m-b+1)}$. We construct $B_1^*, \ldots, B_{n_0 k}^*$ with $kb = m$, and plug this into the model and obtain a pseudo series $y_{11}^*, \ldots, y_{n_0 m}^*$. Similarly, from the model $y_{ij}^* = \hat{\beta}_0 + \hat{\beta}_1 x_{ij} + \hat{e}_{ij}^*$, we fit the model and produce new parameters $\hat{\beta}_0^*$ and $\hat{\beta}_1^*$.

Case 3, One-line moving block bootstrap: One can make up to one long series and perform the moving block bootstrap using a time series without splitting the different individual consecutive data.

Case 4, Standard bootstrap: This is a special case of $b = 1$ in MBB2.

Case 5, Resampling subject bootstrap: This is a special case of $b = m$ in MBB2.

Case 6, Stratified standard bootstrap: This is a special case of $b = 1$ in MBB1.

4. Justification of Moving Block Bootstrap in Longitudinal Data

I consider the justification of moving block bootstrap in longitudinal data. We focus on the within block bootstrap method (MBB1) in the six different kinds of scenario in previous section. I follow the Lahiri's (1996) assumptions. Let's consider the relationship between the GEE and M-estimators. The robust approach can be extended to the regression setup to analyze a predictor–outcome relationship. Suppose we have model (1) with $n = n_0$. The estimator $\hat{\beta}$ is called a robust regression estimator or an M-estimator if it solves

$$\sum_{i=1}^{n_0} \sum_{j=1}^{m} x_{ij}' \psi(y_{ij} - x_{ij}\beta) = 0, \quad (7)$$

for some choice of function $\psi(\cdot)$.

4.1 Expansion for M-estimator

It is known that

$$\left(\sum_{i=1}^{n_0} \sum_{j=1}^{m} x_{ij}' x_{ij} \right)^{1/2} (\hat{\beta}_{n_0 m} - \beta) \sim N_p \left(0, \frac{\mathrm{E}\psi^2(e_{11})}{(\mathrm{E}\psi'(e_{11}))^2} I_p \right), \quad (8)$$

where I_p denotes the identity matrix of order p.

Let $\hat{e}_{ij} = y_{ij} - x_{ij}\hat{\beta}_{n_0 m}$ denote residuals. Define

$$\sigma_i(k) = \mathrm{E}\psi(e_{i1})\psi(e_{i(1+k)}), k \geq 0; \tau$$

$$= \tau_i = \mathrm{E}\psi'(e_{i1}) \text{for each } i \hat{\sigma}_{im}(k)$$

$$= (m-k)^{-1} \sum_{j=1}^{m-k} \psi(\hat{e}_{ij})\psi(\hat{e}_{i(j+k)}), 0 \leq k \leq m$$

$$-1, \hat{\tau}_{im}$$

$$= m^{-1} \sum_{j=1}^{m} \psi'(\hat{e}_{ij}).$$

Also, let $\sigma_i(k) = \sigma(k)$ and $\hat{\sigma_{im}}(k) = \hat{\sigma_m}(k)$

Assumption 1.

(A.1)

    i. ψ is twice differentiable, and ψ'' satisfies a Lipschitz condition of order $\delta_1 > 0$,

    ii. ψ, ψ', ψ'' are bounded.

(A.2)

    i.   for each i $\mathrm{E}\psi(e_{i1}) = 0$, $\tau \equiv \mathrm{E}\psi'(e_{i1})$ 0,

    ii. $\sigma_\infty \equiv \sigma(0) - 2\sum_{k=1}^{\infty} |\sigma(k)| > 0.$

(A.3) There exists $\rho > 0$ such that

    i. $\sup\{|P(A \cap B) - P(A)P(B)|: A \in \mathscr{F}_{-\infty}^r, B \in \mathscr{F}_{r+k}^\infty, r \geq 1\} \leq \rho^{-1}\exp(-\rho k)$ for all k 1,

    ii.   for all r 1, and all k ρ^{-1}, there exists a $\mathscr{F}_{r-k}^{r+k}$-measurable random variable $\tilde{e}_{ir,k}$ such that $\mathrm{E}|e_{ir} - \tilde{e}_{ir,k}|$ $\rho^{-1}\exp(-\rho k)$,

    iii. for all r, k, q ρ^{-1} and

        $A \in \mathscr{F}_{r-q}^{r+q}, \mathrm{E}|P(A|\mathscr{F}_j: j \neq r) - P(A|\mathscr{F}_j: 0 < |j-r| \leq q+k)| \leq \rho^{-1}\exp(-\rho k)$, and

    iv. for all r ρ^{-1}, k r and all $t_{r-k}, \ldots, t_{r+k} \in \mathbb{R}$ with $|t_r| > \rho$,

        $\mathrm{E}|\mathrm{E}(\exp(\sqrt{-1}\sum_{i=1}^{n_0}\sum_{j=r-k}^{r+k} t_j \psi(e_{ij}))|\mathscr{F}_j: j \neq r| < \exp(-\rho).$

(A.4) $\max\{\|x_{ij}\| : 1$ j $m\} = O(1)$ and $\lim \inf_{m \to \infty} m^{-1}\lambda_m \equiv \lambda > 0$, where λ_m denotes the smallest eigenvalue of $(\sum_{j=1}^{m} x'_{ij} x_{ij})$.

Let $D_{n_0m} = (\sum_{i=1}^{n_0}\sum_{j=1}^{m} x'_{ij} x_{ij})^{-1/2}$ and $d_{ij} = D_{n_0m} x'_{ij}, 1$ j m. When e_{ij} are weakly dependent for each i, the asymptotic covariance of $D_{n_0m}^{-1}(\hat{\beta}_{n_0m} - \beta)$ matrix is given by

$$\text{Cov}_{n_0m} \equiv (\text{E}\psi'(e_{11}))^{-2} \times \sum_{i=1}^{n_0}\sum_{k=0}^{m} L_{ikm}\text{E}\psi(e_{i1})\psi(e_{i(1+k)}), \quad (9)$$

where $L_{i0m} = I_p$ and $L_{ikm} = \sum_{j=1}^{m-k}(d_{ij}d'_{i(j+k)} + d_{i(j+k)}d'_{ij}), 1 \quad k \quad m-1.$

To define the studentized version of $\hat{\beta}_{n_0m}$, note that the asymptotic matrix $D_{n_0m}^{-1}(\hat{\beta}_{n_0m} - \beta)$ is given by

$$\sum_{n_0m} \equiv \text{Cov}\left(\sum_{i=1}^{n_0}\sum_{j=1}^{m}d_{ij}\psi(e_{ij})\right) = \sum_{i=1}^{n_0}\sum_{k=0}^{m-1}L_{ikm}\sigma(k). \quad (10)$$

Therefore, a natural estimator of Σ_{n_0m} is

$$\hat{\sum}_{n_om} = \sum_{i=1}^{n_0}\sum_{k=0}^{l}L_{ikm}\hat{\sigma}_m(k), \quad (11)$$

where $1 \quad l \equiv l_m \quad m-1$ is an integer. If $l \to \infty$ slowly with m, then $\|\hat{\Sigma}_{n_0m} - \Sigma_{n_0m}\| = o_p(1)$. $\hat{\Sigma}_{n_0m}$ is non singular with high probability for m large, and can be inverted to define the studentized statistic,

$$T_{n_0m} = \hat{\sum}_{n_0m}^{-1/2}D_{n_0m}^{-1}(\hat{\beta}_{n_0m} - \beta). \quad (12)$$

Next, I extend Lahiri's (1996) results for longitudinal case: assume that (A.1),(A.2),(A.3)(i), (ii), and (A.4) hold. Then, there exists a sequence of statistics $\{\hat{\beta}_m\}$ such that

$$P(\hat{\beta}_{n_0m} \text{ satisfies (7) and} \|D_{n_0m}^{-1}(\hat{\beta}_{n_0m} - \beta)\|^2 \leq C\log m) = 1 - o(m^{-1/2}). \quad (13)$$

If we have a unique solution $\hat{\beta}_{n_0m}$, then $\|D_{n_0m}^{-1}(\hat{\beta}_{n_0m} - \beta)\| = O_P((\log m)^{1/2})$. When (7) has a unique solution, one can obtain the strong consistency of $\hat{\beta}_{n_0m}$ as in Lahiri (1992). The next result gives a first-order Edgeworth expansion for the studentized M-estimator.

Theorem 1. Assume that Assumptions (A.1)–(A.4) hold and that $\{\hat{\beta}_{n_0m}\}$ is a sequence of measurable solutions of (7). Then, there exist's a polynomial $p_m(\cdot)$ on $\mathbb{R}^p$ such that

$$\sup_{B \in \mathscr{B}}\left|P(T_{n_0m} \in B) - \int_B(1 + p_{n_0m}(x))d\Phi(x)\right| = o(m^{-1/2}) \quad (14)$$

for every class $\mathscr{B}$ of Borel subsets of $\mathbb{R}^p$ satisfying

$$\sup_{B \varepsilon \mathscr{B}} \Phi((\partial B)^\epsilon) = O(\epsilon) \ as \ \eta \downarrow 0. \quad (15)$$

Here $\|p_{n_0 m}\varphi\|_\infty = O(m^{-1/2})$ with sup norm $\|\|_\infty$, Φ denotes the standard normal distribution on $\mathbb{R}^p(p \quad 1)$, and the coefficient of $p_{n_0 m}(\cdot)$ are continuous functions of cross-product moments of $\psi(e_{ij})$, $\psi'(e_{ij})$, and $\psi''(e_{ij})$. Here $\quad B$ denote's the boundary of a set $B \subseteq \mathbb{R}^p$ and $(\quad B)^\epsilon = \{x : \|x - y\| < \epsilon$ for some $y \in \quad B\}$.

4.2. Expansion for Bootstrap M-estimator

Define the bootstrap M-estimator $\hat{\beta}^*_{n_o m}$ as a solution of the equation in $t \in \mathbb{R}^p$

$$\sum_{i=1}^{n_0} \left(\sum_{j=1}^{m} x'_{ij}(\psi(y^*_{ij} - x_{ij}t)) - \hat{\mu}_m \right) = 0, \quad (16)$$

where $\hat{\mu}_m = \frac{1}{b} E^*_{n_0 m}\{\psi(\hat{e}^*_{11}) + \ldots + \psi(\hat{e}^*_{1b})\}$ and y^*_{ij} is given in (6). The $\sum^*_{n_0 m}$ is the conditional covariance matrix of $\sum_{i=1}^{n_0}\sum_{j=1}^{m}d_{ij}\psi(\hat{e}^*_{ij})$ which is given by

$$\sum_{i=1}^{n_0}\sum_{u=1}^{k}\mathrm{Cov}_m\left(\sum_{j=1}^{b}d_{i,(u-1)b+j}\psi(\hat{e}^*_{ij})\right). \quad (17)$$

The natural estimator of $\sum^*_{n_0 m}$ is

$$\hat{\sum}^*_{n_0 m} = \sum_{i=1}^{n_0}\sum_{j=0}^{b-1}\sum_{u=1}^{k}\sum_{l=1}^{b-j}D^*_{i,luj}\hat{\sigma}^*_{n_0 m}(j), \quad (18)$$

where $D^*_{luj} = (1 - 2^{-1}I(j=0))(\tilde{D}^*_{luj} + \tilde{D}^{*'}_{luj})$, $\tilde{D}^*_{luj} = d_{(u-1)b+l}d'_{(u-1)b+l+j}$. The bootstrap version $T^*_{n_0 m}$ of $T_{n_0 m}$ is given by

$$T^*_{n_0 m} = (\hat{\sum}^*_{n_0 m})^{-1/2} D^{-1}_{n_0 m}(\hat{\beta}^*_{n_0 m} - \hat{\beta}_{n_0 m}). \quad (19)$$

By assumptions, there exists a sequence of statistics $\{\hat{\beta}^*_{n_0 m}\}$ such that

$$P(\hat{\beta}^*_{n_0 m}\text{ satisfies (16) and}\|D^{-1}_{n_0 m}(\hat{\beta}^*_{n_0 m} - \hat{\beta}_{n_0 m})\|^2 \leq C\log m) = 1 - o_p(m^{-1/2}). \quad (20)$$

Theorem 2. Assume that the conditions in Theorem 1. hold. Suppose that $T^*_{n_0 m}$ is defined for some measurable sequence $\{\hat{\beta}^*_{n_0 m}\}$ satisfying (20) and also suppose that $m^\delta b^{-1} = O(1)$ and $b = O(m^{(1-\kappa)/4})$ for some $\delta > 0$, and $\kappa > \max\{p + 3, 5\}\delta_0$. Then

$$\sup_{B \in \mathscr{B}} \left| P^*(T^*_{n_0m} \in B) - P(T_{n_0m} \in B) \right| = o_p(m^{-1/2}) \quad (21)$$

for any class $\mathscr{B}$ of Borel subset of $\mathbb{R}^p$.

Theorem 2 shows that the MBB indeed provides more accurate approximation for studentized multivariate M-estimator of the regression parameter vector β than normal approximation. Consequently, Theorem 2 is useful for constructing second-order correct multivariate inference procedures for β under multiple regression model. The studentized moving block bootstrap statistics obtain the second-order accuracy for the bounded $n = n_0$ case.

4.3. A Simulation Study

The block bootstrap captures the dependence in the series of residuals without the need to know the correlation structure. It can be simple and account for correlations in a regression model with correlated error. To obtain the bootstrap version of $\hat{\beta}_{n_0m}$, first form the observed blocks of residual length b as $\xi_{ih} = (\hat{e}_{ij}, \ldots, \hat{e}_{i(h+b-1)})$, $1 \le h \le q$, where $q = m - b + 1$ and $\hat{e}_{ij} = y_{ij} - x_{ij}\hat{\beta}_{n0m}$, $1 \le j \le m$, $1 \le i \le n$. Next draw $\xi^*_{i1}, \ldots, \xi^*_{ik}$ randomly, with replacements from $\xi_{i1}, \ldots, \xi_{iq}$, where $m/b = k$ is assumed to be an integer for simplicity. Note that each ξ^*_{ik} has b components. Denote the lth component of ξ^*_{ik}, $1 \le l \le b$ by ξ^*_{ikl}. Also, set $\hat{e}^*_{i((b-1)k+l)} = \xi^*_{ikl}$, $1 \le l \le b$, and we have the bootstrap pseudo-observations

$$y^*_{ij} = x_{ij}\hat{\beta}_{nm} + \hat{e}^*_{ij}, 1 \le i \le n_0, 1 \le j \le m. \quad (22)$$

Adapting Shorack's approach, we obtain the bootstrapped estimator $\hat{\beta}^*_{n_0m}$ as a solution of the equation $t \in R^p$,

$$g'_{n_0m} = \sum_{i=1}^{n_0} \sum_{j=1}^{m} x'_{ij} \left((y^*_{ij} - x_{ij}\beta) - \hat{\mu}_{nm} \right) = 0, \quad (23)$$

where $\hat{\mu}_{n_0m} = b^{-1} E_{n_0m}\{e^*_{11} + \ldots + e^*_{n_0b}\}$, and E_{n0m} denotes the conditional expectations under the MBB resampling scheme, given $\hat{e}_{11}, \ldots, \hat{e}_{nm}$. Centering the above equation by $\hat{\mu}_{n0m}$ makes the estimating equation conditionally unbiased at $\beta = \hat{\beta}_{n0m}$ and ensures the bootstrap analog. The bootstrap estimator is as follows:

$$\hat{\beta}^*_{n_0m} = \left(\sum_{i=1}^{n_0} \sum_{j=1}^{m} x'_{ij}x_{ij} \right)^{-1} \sum_{i=1}^{n_0} \sum_{j=1}^{m} x'_{ij}(y^*_{ij} - \hat{\mu}_{n_0m}). \quad (24)$$

Consider the following specific model in simulation work:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + \gamma_i + e_{ij}, i = 1, \ldots, n_0, j = 1, \ldots, m, \quad (25)$$

where $e_{ij} = \varphi e_{i(j-1)} + u_{ij}$,

$$\gamma_i \sim N(0, \sigma_\gamma^2) \text{ and } u_{ij} \sim N(0, \sigma_u^2).$$

In particular, let $n_0 = 5$, $m = 20$, and $x_{ij} = (1,\ldots, 20)'$.

$$\beta_0 = 10, \beta_1 = 1, b = 4, k = 5,$$

$$\sigma_\gamma^2 = 1, \sigma_u^2 = 1 \text{ and} \phi = 0.75.$$

4.4. Bootstrapping the Distribution of Statistics

Let R be the number of bootstrap simulations ($r = 1, \ldots, R$), and $\hat{\beta^*}$ be the bootstrap estimate of β for the r samples. The important result is that the distribution of $\hat{\beta^*}$, estimated by the empirical distribution function of the $\hat{\beta^{*,r}}(r = 1, \ldots, R)$, approximates the distribution of $\hat{\beta}$. Now the studentized statistic has the following form:

$$\hat{T}^* = \frac{\hat{\beta}^* - \hat{\beta}}{\hat{S}_{\hat{\beta}^*}^*}. \quad (26)$$

The difference between the distribution functions of $\hat{T}$ and $\hat{T^*}$ tends to 0, when the number of observations is large; thus we can use the quartiles of $\hat{T^*}$ instead of $\hat{T}$ to construct intervals or tests. Let $(T^{*,r}, r = 1, \ldots, R)$ be the rth sample of $\hat{T^*}$, where $\hat{T^*}$ is calculated in the same way as $\hat{T}$, replacing y_{ij} with y_{ij}^*. Let $\hat{q_a}$ be the percentile of the $T^{*,r}$. It can be shown that $P(\hat{T} \quad \hat{q_a})$ tends to a, when m tends to infinity. This gives a bootstrap confidence interval for β

$$\hat{I}_R = \left[\hat{\beta}^* - \hat{q}_{1-\frac{\alpha}{2}} \hat{S}_{\hat{\beta}^*}^*, \hat{\beta}^* - \hat{q}_{\frac{\alpha}{2}} \hat{S}_{\hat{\beta}^*}^* \right]. \quad (27)$$

For large m and R, the coverage probability of $\hat{I}_R$ is close to $1 - a$. The bootstrap estimation of the variance is calculated using the empirical variance of the R sample ($\hat{\beta^{*,r}}, r = 1, \ldots, R$):

$$\hat{S}^{*2} = \frac{1}{R-1} \sum_{r=1}^{R} (\hat{\beta}^{*,r} - \overline{\hat{\beta}^*})^2, \quad (28)$$

where $\overline{\hat{\beta}^*}$ is the sample mean $\overline{\hat{\beta}^*} = \sum_{r=1}^{R} \hat{\beta}^{*,r} / R$.

Coverage accuracy, where coverage is the probability that a confidence interval includes β, is the important property for a confidence interval procedure. Bootstrap confidence interval methods differ in their asymptotic properties. Our simulation results are given in Table 1.

MBB1 and MBB2 are similar to each other. Those two block bootstrap methods obtained correct coverage probability at the nominal level of 95%. The standard bootstrap or stratified ordinary bootstrap did not perform well in highly correlated longitudinal data with low coverage probabilities.

5. Concluding Remarks and Discussion

The moving block bootstrap methods are used for analyzing longitudinal data in which a small number of subjects have a large number of replications over time by investigating the efficacy and utility of the methodology, theoretically and empirically, through a small simulation study. Those have second-order optimality in the case of dependent stationary data, under regular conditions.

Quasi-likelihood approaches such as the GEE of Liang and Zeger (1986) and quadratic inference function (QIF) of Qu et al. (2000) are useful for modeling longitudinal data in the form of large sample short time series. However, their advantages of simplicity and robustness against the misspecification of the correlation structure are offset by their loss of estimation efficiency and lack of procedure for model assessment and selection. In other words, these estimators have consistency properties, but are not fully efficient. To offset this inefficiency, subjects rather than observations can be resampled to obtain an estimation efficiency that takes correlation structure into account.

In a longitudinal study with a small sample and a long time series in which modeling the repeated time pattern is necessary, we assume a stationary process. The stationary process assumption represents a fundamentally different stochastic mechanism from other methods used to govern the structure and behavior of transitions over time. Using a moving block bootstrap estimation procedure is preferable in case of cluster high-correlated data with an equal large number of clusters or in spatial strong-correlated data collected from a large number of locations.

When both the number of subjects and the number of replications are large, the correlation structure is determined by a trade off between the number of nuisance parameters and the closeness of the mathematical model to the true underlying structure. The question is, which provides better estimation efficiency: a simpler correlation with small nuisance parameters or a model closer to the true structure with many nuisance parameters. Further simulations using the bootstrap procedure will be investigated in the future study.

Appendix

Proof of Theorem 1

Proof. We follow Lahiri (1996) notation and definitions. For a smooth function $h : \mathbb{R}^p \to \mathbb{R}$, let us $D_j h$ denote the partial derivative of $h(x)$ with respect to the jth coordinate of x, $1 \leq j \leq p$. For $p \times 1$ vectors $v = (v_1, \ldots, v_p)' \in \mathbb{Z}_+^p$ and $w = (w_1, \ldots, w_p)' \in \mathbb{R}^p$, let $|v| = v_1 + \ldots + v_p$, $v! = v_1! \ldots v_p!$, $w^v = \Pi_{i=1}^p (w_i)^{v_i}$, and $\|w\| = (w_1^2 + \ldots + w_p^2)^{1/2}$. Let D^v denote the differential

operator $D_1^{v_1} \ldots D_p^{v_p}$, namely, $D^v = \Pi_{j=1}^p \left(\frac{\partial}{\partial t^{(j)}}\right)^{v(j)}$. For $v \in \mathbb{Z}_+^p$ with $1 \quad |v| \quad s$, let χ_v denote the vth cumulant and μ_v is the vth moment of w. Note that $\sqrt{-1}^{|v|} \mu_v = D^v \hat{\Phi}(0)$, $D^\alpha \hat{\mu}(t) = (\sqrt{-1})^{|\alpha|} \int w^\alpha e^{\sqrt{-1}t'w} \mu(dw), t \in \mathbb{R}^p$, and $\sqrt{-1}^{|v|} \chi_v = (D^v \log \hat{\Phi})(0)$.

We consider w is a $\mathbb{R}^p$-valued random vector with $Ew = 0$ and $E\|w\|^s < \infty$ for some integer s 3.

Let $m_3 = [\log m \log \log(3 + m)]$, $v_{1m} = m^{-1/2}(\log m)^{1/2}$, $v_m = m^{-1/2}$, $v_{2m} = v_m(\log m)^{-1}$, and $v_{3m} = v_m(\log m)^{-3/2}$. Furthermore, define

$$G_{n_0 m} = D_{n_0 m}^{-1}(\hat{\beta}_{n_0 m} - \beta),$$
$$G_{1 n_0 m} = \sum_{i=1}^{n_0} \sum_{j=1}^m d_{ij} \psi(e_{ij})$$
$$d_{ij} = \left(\sum_{i=1}^{n_0} \sum_{j=1}^m x'_{ij} x_{ij}\right)^{-1/2} x_{ij},$$
$$D_{n_0 m} = \left(\sum_{i=1}^{n_0} \sum_{j=1}^m x'_{ij} x_{ij}\right)^{-1/2},$$
$$\tilde{\psi} = \psi(\cdot) - \hat{\mu}_{n_0 m},$$
$$A_{n_0 m} = \sum_{i=1}^{n_0} \sum_{j=1}^m d_{ij} d'_{ij} \psi'(e_{ij}),$$

$A = \tau I_p = E\psi'(e_{ij}) I_p$, $\tau = \tau_i = E\psi'(e_{i1})$ for each i, $w_{i1j} = \psi(e_{i1})$, $W_{i2j} = \psi'(e_{ij}) - \tau$, $W_{i3j} = \psi''(e_{ij}) - E\psi''(e_{i1})$, and $W_{i4j}(k) = \psi(e_{ij})\psi(e_{i(j+k)}) - \sigma_i(k)$. Also, write

$$\chi(U) = (-1)^{p/2} D_1 \ldots D_p E \exp(\sqrt{-1}t'U)|_{t=0}$$ for a random vector U in $\mathbb{R}^p$.

Let $\Delta = D_{n_0 m}^{-1}(t - \beta), t \in \mathbb{R}^p$. Then, by Taylor's expansion, one can get

$$\left[\sum_{i=1}^{n_0} \sum_{j=1}^m d_{ij} d'_{ij} \psi'(e_{ij})\right] \Delta = \sum_{i=1}^{n_0} \sum_{j=1}^m d_{ij} \psi(e_{ij}) + \frac{1}{2} \sum_{i=1}^{n_0} \sum_{j=1}^m d_{ij} (d'_{ij} \Delta)^2 \psi''(e_{ij}) + R_{n_0 m}(t), \quad (29)$$

Where $\|R_{n_0 m}(t)\| \leq C \sum_{i=1}^{n_0} \sum_{j=1}^m \|d_{ij}\|^{3+\delta_1} \|\Delta\|^{2+\delta}, t \in \mathbb{R}^p$.

Following Lahiri (1992), we obtain that

$$G_{n_0 m} = (A^{-1} + \tau^{-2}(A - A_{n_0 m})) G_{1 n_0 m} + (2\tau^3)^{-1} \sum_{i=1}^{n_0} \sum_{j=1}^m d_{ij} (d'_{ij} \theta_{1 n_0 m})^2 E\psi''(e_{ij}) + R'_{n_0 m}, \quad (30)$$

where $P(\|R'_{n_0 m}\| > C(\sigma_\infty) v_{2m}) = o(v_m)$,

$$T_{n_0m}=\sum\nolimits_{n_0m}^{-1/2}\left[G_{1n_0m}+\tau^{-1}G_{1n_0m}\left(m^{-1}\sum_{j=1}^{n_0}\sum_{j=1}^{m}W_{i2j}-(m\tau)^{-1}\sum_{i=1}^{n_0}\sum_{j=1}^{m}(d'_{ij}G_{1n_0m})\mathrm{E}\psi''(e_{i1})\right)+\tau^{-1}(A-A_{n_0m})G_{1n_0m}+(2\tau^2\right.$$

$$+\sum_{|\beta|=1}(\hat{\sum}\nolimits_{n_0m})^{\beta}D^{\beta}(\sum\nolimits_{n_0m}^{-1/2})G_{1n_0m}+R^{\dagger}_{n_0m}.$$

(31)

If we have $T_{n_0m}=T_{1n_0m}+R_{n_0ms}$, where R_{n_0ms} is the remainder term that under the moment condition $E\|y_{11}\|^s<\infty$ satisfies $P(\|R_{n_0ms}\|>\delta_{m,s})=\delta_{m,s}$ for some sequence $\delta_{m,s}=o(m^{-(s-2)/2})$, then the random variable T_{1n_0m} is called a $(s-2)$th -order stochastic approximation to T_{n_0m}. Note that the $(s-2)$th order Edgeworth expansions for T_{n_0m} and T_{1n_0m} coincide. The reason for T_{1n_0m} is that the first term is the same as T_{n_0m}, but the remaining terms consist of all independent variables for deriving a simpler expansion. The stochastic approximation T_{1n_0m} can be expressed in the form

$$T_{1n_0m}=\sum\nolimits_{n_0m}^{-1/2}G_{1n_0m}+\sum_{r=1}^{p}G'_{1n_0m}\Lambda_{n_0rm}G_{1n_0m}q_r+\sum_{r=1}^{p}\tilde{W}'_{2n_0m}\Lambda_{1n_0rm}G_{1n_0m}q_r+\sum_{|v|=1}(\hat{\sum}\nolimits_{1n_0m})^{v}\Lambda_{vm}G_{1n_0m},$$ (32)

where $q_1=(1,0,\ldots,0),\ldots,q_p=(0,0,\ldots,1)$ are the standard basis of $\mathbb{R}^p$,

$\tilde{W}_{2n_0m}=((A-A_{n_0m})':\sum_{i=1}^{n_0}m^{-1}\sum_{j=1}^{m}W_{i2j})'$ and $\Lambda_{n_0rm},\Lambda_{1n_0rm},\Lambda_{vm}$ are non random matrices satisfying $\max\{m^{1/2}\|\Lambda_{n_0rm}\|+\|\Lambda_{1n_0rm}\|+\|\Lambda_{vm}\|:1\quad r\quad p,|v|=1\}=o(1)$. In the following $C,C(\cdot)$ denotes pure constants which depend on each arguments, and the dependance of $C(\cdot)$ on p,a, and the finite moments of $\psi(e_{ij}),\psi'(e_{ij})$, and $\psi''(e_{ij})$ will be suppressed for notational simplicity. Using Lahiri's (1992, 1996) arguments, we can show that

$$P\left(\|\sum_{i=1}^{n_0}\sum_{j=1}^{m}d_{ij}\psi(e_{ij})\|>C(\log m)^{1/2}\right)=o(m^{-1/2}),$$ (33)

$$P(\|A_{n_0m}-A\|>Cm^{-1/4}(\log m)^{-2})=o(m^{-1/2}),$$ (34)

$$P\left(|\sum_{i=1}^{n_0}\sum_{j=1}^{m}d_{iju}d_{ijl}d_{ijz}|>Cm^{-5/8}\right)\leq Cm^{-3/4}$$ (35)

for all $1\quad u,l,z\quad p$, where d_{iju} denote the uth component of d_{ij}, we have $T_{n_0m}=T_{1n_0m}+R_{n_0m}$, where $P(\|R_{n_0m}\|>Cv_{2m})=o(v_m)$. We know that the first order Edgeworth expansions for T_{1n_0m} and T_{n_0m} coincide.

Let $G_{2n_0m}=\sum_{i=1}^{n_0}\sum_{j=a}^{m}d_{ij}W_{i1j}$ with $a=[m^{(1-2\delta_0)/2}]$,
$A_{2n_0m}=((\sum_{i=1}^{n_0}\sum_{j=a}^{m}d_{ijk}d_{ijl}W_{i2j}))_{p\times p}$,

$$\hat{W}_{2n_0m}=((A'_{2n_0m} - \mathrm{E}A'_{2n_0m}):\sum_{i=1}^{n_0}m^{-1}\sum_{j=1}^{m}W_{i2j})',$$

$$\hat{\sum}_{2n_0m}=\sum_{i=1}^{n_0}\sum_{k=0}^{l}(m-k)^{-1}L_{ikm}\sum_{j=a}^{m-k}[W_{i4j}-G'_{2n_0m}\gamma_{ijk}],\text{ and}$$

$$T_{2n_0m}=\sum_{n_0m}^{-1/2}G_{1n_0m}+\sum_{r=1}^{p}G'_{2n_0m}\Lambda_{n_0rm}G_{2n_0m}q_r+\sum_{r=1}^{p}\hat{W}'_{2n_0m}\Lambda_{1n_0rm}G_{2n_0m}q_r+\sum_{|v|=1}(\hat{\sum}_{2n_0m})^{v}\Lambda_{vm}G_{2n_0m}. \quad (36)$$

The reason for T_{2n_0m} is that the first term is the same as T_{1n_0m}, but the remaining terms consist of truncated independent variables for obtaining the simplified forms of expansion. Using an Edgeworth expansion under dependence for T_{1n_0m} (Lahiri 1994, 1996), we have

$$P(\|T_{1n_0m}-T_{2n_0m}\|>Cv_m)=o(v_m). \quad (37)$$

We have the same first-order Edgeworth expansions for T_{2n_0m}, T_{1n_0m}, and T_{n_0m}; namely, three statistics are close to each other.

We obtain that

$$\sup_{B\in\mathscr{B}}\left|P(T_{n_0m}\in B)-\int_{B}(1+p_{n_0m}(x))d\Phi(x)\right|=o(m^{-1/2}), \quad (38)$$

with $\|p_{n_0m}\varphi\|_{\infty})=O(m^{-1/2})$. The proof is then complete.

Proof of Theorem 2

Proof. Let $G^*_{n_0m}=D^{-1}_{n_0m}(\hat{\beta}^*_{n0m}-\hat{\beta}_{n_0m})$, $G^*_{1n_0m}=\sum_{i=1}^{n_0}\sum_{u=1}^{k}W^*_{i1u}$, $A^*_{n_0m}=\sum_{u=1}^{k}W^*_{i2u}$, $\hat{A}_{n_0m}=E_{n_0m}A^*_{n_0m}$, $\tau^*_{1n_0m}=m^{-1}\sum_{i=1}^{n_0}\sum_{j=1}^{m}\psi'(\hat{e}^*_{ij})$, $\hat{\tau}_{1n_0m}=E^*_{n_0m}(\tau^*_{1n_0m})$, $\xi^*_{uj}=j$th, component of B^*_u, for $u = 1, k,...,$

$$W^*_{i1u}=\sum_{j=1}^{b}d_{i((u-1)b+j)}\hat{\psi}(\xi^*_{uj}), W^*_{i2u}=\sum_{j=1}^{b}d_{i((u-1)b+j)}d'_{i((u-1b+j)}\psi'(\xi^*_{uj}), 1\leq k\leq m.$$

As in the proof of Theorem 4.5.1 and in Lahiri's (1996) result, we have

$$T^*_{n_0m}=T^*_{1n_0m}+R^{\dagger*}_{n_0m} \quad (39)$$

where $P_{n_0m}(\|R^{\dagger*}_{n_0m}\|>Cv_{2m})=O_p(v_{3m})$. We also use $T^*_{1n_0m}$ and $T^*_{2n_0m}$, which is the same definition of T_{1n_0m} and T_{2n_0m} in Theorem 4.5.1.

Now using the results of Bhattachararya and Ranga Rao (1986) and Lahiri (1996), We have the same first-order Edgeworth expansion forms for $T_{2n_0m}^*$, $T_{1n_0m}^*$, and $T_{n_0m}^*$, since those are close to each other. We obtain

$$\sup_{B \in \mathscr{B}} \left| P^*(T_{n_0m}^* \in B) - P(T_{n_0m} \in B) \right| = o_p(v_m). \quad (40)$$

References

Babu GJ. Bootstrapping statistics with linear combinations of chi-squares as weak limit. Sankhya Ser A. 1986; 56:85–93.

Bhattacharya, RN.; Ranga Rao, R. Normal Approximations and Asymptotic Expansions. Malabar, FL: Krieger; 1986.

Bühlmann P. Sieve bootstrap for time series. Bernoulli. 1997; 3:128–148.

Davison, AC.; Hinkley, DV. Bootstrap Methods and Their Application. Cambridge, UK: Cambridge University Press; 1997.

Efron B. Bootstrap methods: Another look at the jackknife. Annals Stat. 1979; 7:1–16.

Hall P, Horowitz JL, Jing BY. On blocking rules for the bootstrap withe dependent data. Biometrika. 1995; 82:561–574.

Lahiri SN. Bootstrapping m-estimators of a multiple linear regression parameter. Annals Stat. 1992; 20:1548–1570.

Lahiri SN. On two-term Edgeworth expansions and bootstrap approximations for studentized multivariate m-estimators. Sankhya Ser A. 1994; 56:201–226.

Lahiri SN. On Edgeworth and moving block bootstrap for studentized m-estimators in multiple linear regression models. J Multivar Anal. 1996; 56:42–59.

Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika. 1986; 73:13–22.

Paparoditis E, Politis DN. Tapered block bootstrap for general statistics from stationary sequences. Econometric J. 2002; 5:131–148.

Qu A, Lindsay BG, Li B. Improving generalized estimating equations using quadratic inference function. Biometrika. 2000; 87:823–836.

Singh K. On the asymptotic accuracy of the Efron's bootstrap. Annals Stat. 1981; 9:1187–1195.

Xie M, Yang Y. Asymptotics for generalized estimating equations with large cluster sizes. Annals Stat. 2003; 31:13–22.

Table 1

Coverage probability and length of CI $(\hat{\beta}_1^*)$: 500 replications; $\varphi = 0.75$, SOB is stratified ordinary bootstrap, SB is a standard bootstrap estimation, and $\hat{\beta_1}$ is a robust estimation with unknown covariance structure

Methods	$\mathbf{CI}(\hat{\beta}_1^*)$	Probability	Length
MBB1	(0.849, 1.168)	0.949	0.318
MBB2	(0.844, 1.163)	0.952	0.320
SOB	(0.805, 1.091)	0.768	0.286
SB	(0.804, 1.086)	0.747	0.282
$\hat{\beta_1}$	(0.840, 1.169)	0.950	0.329