Comparison of block bootstrap testing methods of mean difference for longitudinal data

Hirohito Sakurai*
National Center for University Entrance Examinations, Tokyo, Japan - sakurai@rd.dnc.ac.jp

Masaaki Taguri
National Center for University Entrance Examinations, Tokyo, Japan - tagurimm@yahoo.co.jp

Abstract
In this paper, we focus on a two-sample problem, and compare three block bootstrap testing methods for detecting the difference of two means in longitudinal data when the data of two groups are not paired. For the detection of mean difference of two groups, we here consider the following four types of test statistics: (i) sum of absolute values of difference between two mean sequences, (ii) sum of squares of difference between two mean sequences, (iii) estimator of area-difference between two mean curves, and (iv) difference of kernel estimators based on two mean sequences. The block resampling techniques considered here include moving block bootstrap, circular block bootstrap and stationary bootstrap. These are used to approximate the null distributions of test statistics. Monte Carlo simulations are carried out in order to examine the sizes and powers of the testing methods.

Keywords: two-sample problem; resampling; sizes and powers of tests.

1. Introduction
One of the important topics in statistical inference is comparison of two means or regression curves of two samples. Suppose now that there are two samples given by \( \{ Y_i(t) \}_{i=1}^{q_1} \) and \( \{ X_j(t) \}_{j=1}^{q_2} \) for \( t = 1, \ldots, n \), and assume that they are mutually independent, where \( q_1 \) and \( q_2 \) are numbers of subjects, and \( n \) is the number of observed points. We also assume that, for fixed \( t \), \( Y_i(t) \) and \( X_j(t) \) are independent over \( q_1 \) and \( q_2 \) subjects, respectively. Then we consider the model

\[
\begin{cases}
Y_i(t) = p_1(t) + \varepsilon_i(t), & i = 1, \ldots, q_1, \\
X_j(t) = p_2(t) + \eta_j(t), & j = 1, \ldots, q_2,
\end{cases}
\]

where \( p_1(t) \) and \( p_2(t) \) are unknown regression functions, and \( \varepsilon_i(t) \) and \( \eta_j(t) \) are the error terms having means 0 and finite variances, respectively. Then, we are interested in a testing problem

\[ H_0 : p_1(t) = p_2(t) \text{ for all } t \quad \text{vs.} \quad H_1 : p_1(t) \neq p_2(t) \text{ for some } t, \]

where \( H_0 \) and \( H_1 \) denote the null and alternative hypotheses.

In this paper, we compare the testing methods for detecting the difference of two mean curves in two longitudinal data when two samples are not paired. Our approaches to the problem (2) are based on Moving Block Bootstrap (MBB), Circular Block Bootstrap (CBB) and Stationary Bootstrap (SB), which are proposed by Künsch (1989) and Politis & Romano (1992, 1994), respectively.

2. Testing methods using block bootstrap
In this section, we review the testing methods for (2) using block bootstrap proposed by Sakurai and Taguri (2005, 2010, 2013). Note that the area-difference given by \( A = \int |p_1(t) - p_2(t)| \, dt \) is 0 under \( H_0 \) and positive under \( H_1 \). Then, the hypothesis of our interest reduces to testing

\[ H_0 : A = 0 \quad \text{vs.} \quad H_1 : A > 0. \]
For detecting the difference between \( p_1(t) \) and \( p_2(t) \), we can use

\[
S_n = S_n(D_1, \ldots, D_n) = \left[ \sum_{j=0}^{n-1} \left( \sum_{t=j+1}^{j+h} D_t \right)^2 \right] \left[ n \sum_{t=1}^{n-1} \frac{(D_{t+1} - D_t)^2}{2} \right]^{-1},
\]

where \( D_t = Y_t - X_t \) for \( t = 1, \ldots, n \), or \( D_t = Y_{t-n} - X_{t-n} \) for \( t = n+1, \ldots, n+h \), \( Y_t = \sum_{i=1}^{q_1} Y_i(t)/q_1 \), \( X_t = \sum_{j=1}^{q_2} X_j(t)/q_2 \), \( h = \lfloor np \rfloor \) is the integer part of \( np \), and \( p \) is a tuning constant satisfying \( 0 < p < 1 \) which is determined by the fully data-driven approach; the second approach described in Hall and Hart (1990, pp.1043–1044). As another type of test statistics,

\[
T_{1n} = T_{1n}(D_1, \ldots, D_n) = \sum_{t=1}^{n} |D_t|, \quad T_{2n} = T_{2n}(D_1, \ldots, D_n) = \sum_{t=1}^{n} D_t^2, \quad T_{3n} = T_{3n}(D_1, \ldots, D_n) = \frac{1}{2} \sum_{t=1}^{n-1} (|D_t| + |D_{t+1}|)I_+ + \frac{1}{2} \sum_{t=1}^{n-1} \frac{|D_t|^2 + |D_{t+1}|^2}{|D_t| + |D_{t+1}|}I_-, \quad \text{(4)}
\]

are also available, where \( I_+ = I\{D_tD_{t+1} \geq 0\} \), \( I_- = I\{D_tD_{t+1} < 0\} \) and \( I\{\cdot\} \) is the indicator function, respectively. \( T_{3n} \) seems to have a complicate form, however it is a naive estimator of \( A \) in (3). The values (4), (5) and (6) will be small when \( H_0 \) is true, while they are large when \( H_0 \) is false. Therefore, the difference between \( p_1(t) \) and \( p_2(t) \) can be measured by \( T_{1n}, T_{2n}, T_{3n} \) and \( S_n \).

Next we briefly explain three testing methods for the problem (2) or (3) using MBB, CBB and SB; they are called Mixed MBB, Mixed CBB and Mixed SB tests (Sakurai and Taguri, 2005, 2010, 2013). The main ideas of each method are to make blocks of observations in each sample according to MBB, CBB or SB, and to generate resamples corresponding to two samples by drawing blocks with replacement from the mixed (pooled) MBB, CBB or SB type of blocks. The latter is a technique that can reflect the null hypothesis.

For simplicity, let \( T \) be a generic notation for \( T_{1n}, T_{2n}, T_{3n} \) or \( S_n \). For a given significance level \( \alpha \), the common steps of Mixed MBB, CBB and SB tests are summarized in Algorithm 2.1. The detailed steps 2 and 3 in Algorithm 2.1 for Mixed MBB, CBB and SB tests are given by Algorithms 2.2, 2.3 and 2.4, respectively.

**Algorithm 2.1**

1. Calculate \( t_{obs} = T(Y, X) = T(D_1, \ldots, D_n) \).
2. Divide centered samples, \( \{C_{y,1}, \ldots, C_{y,n}\} \) and \( \{C_{x,1}, \ldots, C_{x,n}\} \), into several collections of blocks corresponding to MBB, CBB or SB, and combine the blocks of two samples, where \( C_{y,t} = Y_t - \bar{Y}, \ C_{x,t} = X_t - \bar{X}, \ \bar{Y} = \frac{1}{n} \sum_{t=1}^{n} Y_t/n \) and \( \bar{X} = \frac{1}{n} \sum_{t=1}^{n} X_t/n \). The combined blocks are denoted by \( \xi_{\text{pooled}} \).
3. Draw blocks with replacement from \( \xi_{\text{pooled}} \) to generate two resamples, \( Y^{sb} = \{Y_{1}^{sb}, \ldots, Y_{n}^{sb}\} \) and \( X^{sb} = \{X_{1}^{sb}, \ldots, X_{n}^{sb}\} \) (\( b = 1, \ldots, B \)), corresponding to two samples, \( Y \) and \( X \).
4. Calculate \( t^{sb} = T(Y^{sb}, X^{sb}) = T(D_{1}^{sb}, \ldots, D_{n}^{sb}) \).
5. Repeating steps 3 and 4 an appropriate number of times \( B \), calculate \( t^{s1}, \ldots, t^{sB} \).
6. From steps 1 and 5, approximate the achieved significance level by \( \overline{\text{ASL}} = \sum_{b=1}^{B} I\{t^{sb} \geq t_{obs}\}/B \), and reject \( H_0 \) when \( \overline{\text{ASL}} \leq \alpha \).
Algorithm 2.2 (Mixed MBB test)

2. Divide \( \{C_{y,1}, \ldots, C_{y,n}\} \) and \( \{C_{x,1}, \ldots, C_{x,n}\} \) into \( k(=n-\ell+1) \) successive overlapping blocks with each length \( \ell \), and put the collection of blocks \( \xi_y = \{\xi_{y,1}, \ldots, \xi_{y,k}\} \) and \( \xi_x = \{\xi_{x,1}, \ldots, \xi_{x,k}\} \), where \( \xi_{y,t} = \{C_{y,t}, \ldots, C_{y,t+\ell-1}\} \) and \( \xi_{x,t} = \{C_{x,t}, \ldots, C_{x,t+\ell-1}\} \) (\( t = 1, \ldots, k \)), respectively.

3. (a) Combine \( \xi_y \) and \( \xi_x \), and put \( \xi_{\text{pooled}} = \{\xi_{y,1}, \ldots, \xi_{y,k}, \xi_{x,1}, \ldots, \xi_{x,k}\} \).

(b) Draw \( \xi_y^b = \{\xi_{y,1}^b, \ldots, \xi_{y,n}^b\} \) and \( \xi_x^b = \{\xi_{x,1}^b, \ldots, \xi_{x,m}^b\} \) with replacement from \( \xi_{\text{pooled}} \) to obtain resamples \( Y_1^b, \ldots, Y_n^b \) and \( X_1^b, \ldots, X_n^b \), where \( m = n/\ell \) (if \( n/\ell \) is an integer) or \( m = \lceil n/\ell \rceil + 1 \) (otherwise), and \( \lceil n/\ell \rceil \) is the integer part of a real \( n/\ell \).

Algorithm 2.3 (Mixed CBB test)

2. Divide \( \{C_{y,1}, \ldots, C_{y,n}\} \) and \( \{C_{x,1}, \ldots, C_{x,n}\} \) into \( n \) collections of blocks, \( \xi_y = \{\xi_{y,1}, \ldots, \xi_{y,n}\} \) and \( \xi_x = \{\xi_{x,1}, \ldots, \xi_{x,n}\} \), where \( \xi_{y,t} \) and \( \xi_{x,t} \) are blocks of length \( \ell \), obtained in the manner of Politis and Romano (1992).

3. (a) Combine \( \xi_y \) and \( \xi_x \), and put \( \xi_{\text{pooled}} = \{\xi_{y,1}, \ldots, \xi_{y,n}, \xi_{x,1}, \ldots, \xi_{x,n}\} \).

(b) Draw \( \xi_y^b = \{\xi_{y,1}^b, \ldots, \xi_{y,n}^b\} \) and \( \xi_x^b = \{\xi_{x,1}^b, \ldots, \xi_{x,m}^b\} \) with replacement from \( \xi_{\text{pooled}} \) to obtain resamples \( Y_1^b, \ldots, Y_n^b \) and \( X_1^b, \ldots, X_n^b \), where \( m \) is defined in Algorithm 2.2.

Algorithm 2.4 (Mixed SB test)

2. Divide \( \{C_{y,1}, \ldots, C_{y,n}\} \) and \( \{C_{x,1}, \ldots, C_{x,n}\} \) into \( n \) collections of blocks as follows:

\[
\xi_y = \{\xi_{y,1}(L_1), \ldots, \xi_{y,n}(L_n)\}, \quad \xi_x = \{\xi_{x,1}(M_1), \ldots, \xi_{x,n}(M_n)\},
\]

where \( \xi_{y,t} \) and \( \xi_{x,t} \) are the blocks starting from \( C_{y,t} \) and \( C_{x,t} \) with length \( \ell(\geq 1) \), \( L_1, \ldots, L_n \), \( M_1, \ldots, M_n \) are independent and identically distributed to a geometric distribution with parameter \( 1/\ell \).

3. (a) Combine \( \xi_y \) and \( \xi_x \), and put \( \xi_{\text{pooled}} = \{\xi_{y,1}(L_1), \ldots, \xi_{y,n}(L_n), \xi_{x,1}(M_1), \ldots, \xi_{x,n}(M_n)\} \).

(b) Draw \( K_{y,b} \) and \( K_{x,b} \) blocks with replacement from \( \xi_{\text{pooled}} \), and put

\[
\xi_y^b = \{\xi(I_{1}^{y,b}, L_1^{y,b}), \ldots, \xi(I_{K_{y,b}}^{y,b}, L_{K_{y,b}}^{y,b})\}, \quad \xi_x^b = \{\xi(J_{1}^{x,b}, M_1^{x,b}), \ldots, \xi(J_{K_{x,b}}^{x,b}, M_{K_{x,b}}^{x,b})\},
\]

where \( \xi(t, \ell) = \xi_y(t, \ell) \) (if \( 1 \leq t \leq n \)), \( \xi(t, \ell) = \xi_x(t, \ell) \) (otherwise), \( I_1^{y,b}, \ldots, I_{K_{y,b}}^{y,b}, J_1^{x,b}, \ldots, J_{K_{x,b}}^{x,b} \) are independent and identically distributed to a discrete uniform distribution on \( \{1, \ldots, n, n+1, \ldots, 2n\} \). A pair of random variables, \( (I_t^{y,b}, L_t^{y,b}) \) or \( (J_t^{x,b}, M_t^{x,b}) \), is one of \( \{(1, L_1), \ldots, (n, L_n), (1, M_1), \ldots, (n, M_n)\} \), and \( K_{y,b} = \min\{k : \sum_{i=1}^{k} L_i^{y,b} \geq n\} \), \( K_{x,b} = \min\{k : \sum_{j=1}^{k} M_j^{x,b} \geq n\} \), respectively.

(c) Construct resamples, \( Y_1^{x,b}, \ldots, Y_n^{x,b} \) and \( X_1^{x,b}, \ldots, X_n^{x,b} \), by putting the first \( n \) elements of \( \xi_y^b \) and \( \xi_x^b \).

3. Numerical study

In this section, we investigate finite sample behavior of the sizes and powers of Mixed MBB, CBB and SB tests by Monte Carlo simulations. Our numerical study also includes the comparison with Bowman and Young’s (1996) test for unpaired data (hereafter termed “BY” for short).

All our results are based on independent 2000 pairs of two samples, \( \{Y_i(t)\} \) and \( \{X_j(t)\} \), where \( B = 2000 \) replications of resampling in our tests are applied to every two samples, and the nominal
level of test is \( \alpha = 0.05 \). We generate initial two samples according to (1) whose means are specified by \( p_1(t) = 0 \) and \( p_2(t) = c \), where \( c = 0, 0.2, 0.4, 0.6, 0.8, 1.0; c = 0 \) or \( c \neq 0 \) corresponds to the null hypothesis or the alternative hypothesis being true. As for the error terms, \( \varepsilon_i(t) \) and \( \eta_j(t) \), we consider the following Gaussian AR(1) errors: 

\[
\varepsilon_i(t) = \phi \varepsilon_{i-1}(t) + z_i(t) \quad \text{and} \quad \eta_j(t) = \phi \eta_{j-1}(t) + z_j(t),
\]

where \( z_i(t) \sim i.i.d. \ N(0, \tau_i^2), \ z_j(t) \sim i.i.d. \ N(0, \tau_j^2), \phi = 0, \pm 0.1, \pm 0.2, \tau_i^2 = \tau_j^2 = (1 - \phi^2) V(\varepsilon_i(t)) \), and \( V(\varepsilon_i(t)) = 1, 3, 5 \). For \( n = 10 \) points, the cases of \( (q_1, q_2) = (10, 10), (10, 20), (10, 30), (20, 20), (20, 30), (30, 30) \) are examined. Due to limitations of space, however, we restrict ourselves to discussing the case of \( (q_1, q_2) = (10, 30), (20, 20) \) and \( V(\varepsilon_i(t)) = 3 \) in this paper.

Since it is preferable that the empirical level is nearly equal to the nominal level \( \alpha \), our choice of \( \ell \) in Mixed MBB, CBB and SB tests is made so that the empirical level is close to \( \alpha \). If there are some candidates which have the same level errors, we make the conservative choice, viz., we choose \( \ell \) such that the empirical level is less than the nominal level. Further if there are some candidates whose empirical levels are equal, we select \( \ell \) to maximize the empirical power among them. The resulting choices of such \( \ell \) are given in Table 1.

We first summarize the results of the level studies in Table 2. The table shows that the empirical levels of Mixed MBB, CBB and SB tests with \( T_{rn} \) \( (r = 1, 2, 3) \) and \( S_n \) tend to keep the nominal level \( \alpha \) when \( \phi \leq 0 \), however it is not true for most cases of \( \phi = 0.2 \). When \( \phi > 0 \), the level error of \( T_{3n} \) and \( S_n \) seems to be slightly larger. \( S_n \) does not need longer \( \ell \) than \( T_{rn} \) \( (r = 1, 2, 3) \) to keep the nominal level as is shown in Table 1. On the other hand, BY test shows a tendency to underestimate the nominal level for all cases.

Next, we discuss the power studies based on Figure 1. The vertical and horizontal axes of Fig-

<table>
<thead>
<tr>
<th>( q_1 )</th>
<th>( q_2 )</th>
<th>( \phi )</th>
<th>( T_{1n} )</th>
<th>( T_{2n} )</th>
<th>( T_{3n} )</th>
<th>( S_n )</th>
<th>( T_{1n} )</th>
<th>( T_{2n} )</th>
<th>( T_{3n} )</th>
<th>( S_n )</th>
<th>( BY )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>30</td>
<td>-0.2</td>
<td>0.048</td>
<td>0.046</td>
<td>0.053</td>
<td>0.051</td>
<td>0.026</td>
<td>0.018</td>
<td>0.033</td>
<td>0.050</td>
<td>0.024</td>
</tr>
<tr>
<td>0.05</td>
<td>0.05</td>
<td>0.052</td>
<td>0.053</td>
<td>0.052</td>
<td>0.032</td>
<td>0.028</td>
<td>0.045</td>
<td>0.052</td>
<td>0.030</td>
<td>0.026</td>
<td>0.051</td>
</tr>
<tr>
<td>0.045</td>
<td>0.046</td>
<td>0.052</td>
<td>0.059</td>
<td>0.058</td>
<td>0.047</td>
<td>0.049</td>
<td>0.067</td>
<td>0.054</td>
<td>0.047</td>
<td>0.049</td>
<td>0.072</td>
</tr>
<tr>
<td>0.076</td>
<td>0.072</td>
<td>0.112</td>
<td>0.070</td>
<td>0.050</td>
<td>0.055</td>
<td>0.086</td>
<td>0.071</td>
<td>0.051</td>
<td>0.059</td>
<td>0.092</td>
<td>0.070</td>
</tr>
<tr>
<td>0.05</td>
<td>0.05</td>
<td>0.053</td>
<td>0.055</td>
<td>0.058</td>
<td>0.045</td>
<td>0.049</td>
<td>0.050</td>
<td>0.025</td>
<td>0.051</td>
<td>0.050</td>
<td>0.051</td>
</tr>
<tr>
<td>0.053</td>
<td>0.055</td>
<td>0.080</td>
<td>0.058</td>
<td>0.049</td>
<td>0.055</td>
<td>0.081</td>
<td>0.059</td>
<td>0.051</td>
<td>0.055</td>
<td>0.080</td>
<td>0.061</td>
</tr>
<tr>
<td>0.067</td>
<td>0.075</td>
<td>0.117</td>
<td>0.065</td>
<td>0.065</td>
<td>0.065</td>
<td>0.117</td>
<td>0.066</td>
<td>0.068</td>
<td>0.076</td>
<td>0.121</td>
<td>0.065</td>
</tr>
</tbody>
</table>
Figure 1: Comparison of $T_{1n}$, $T_{2n}$, $T_{3n}$, $S_n$ and BY for $(q_1, q_2) = (20, 20)$ and $V(\varepsilon_i(t)) = 3$.
Figure 2: Comparison of Mixed MBB, CBB and SB tests for \((q_1, q_2) = (20, 20), \phi = -0.2\) and \(V(\varepsilon_i(t)) = 3\)


