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

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

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

$$H_0 : p_1(t) = p_2(t) \text{ for all } t \quad \text{vs.} \quad H_1 : p_1(t) \neq p_2(t) \text{ for some } t, \tag{2}$$

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

$$H_0 : A = 0 \quad \text{vs.} \quad H_1 : A > 0. \tag{3}$$

For detecting the differece between $p_1(t)$ and $p_2(t)$, we can use

$$S_n = S_n(D_1, \ldots, D_n) = \left[\sum_{j=0}^{n-1} \left(\sum_{t=j+1}^{j+h} D_t \right)^2 \right] \left[n \sum_{t=1}^{n-1} \frac{(D_{t+1} - D_t)^2}{2} \right]^{-1}, \tag{4}$$

where $D_t = Y_t - X_t$ for $t = 1, \ldots, n$, or $D_t = Y_{t-n} - X_{t-n}$ for $t = n+1, \ldots, n+h$, $Y_t = \sum_{i=1}^{q_1} Y_i(t)/q_1$, $X_t = \sum_{j=1}^{q_2} X_j(t)/q_2$, $h = [np]$ 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,

$$T_{1n} = T_{1n}(D_1, \ldots, D_n) = \sum_{t=1}^{n} |D_t|, \quad T_{2n} = T_{2n}(D_1, \ldots, D_n) = \sum_{t=1}^{n} D_t^2, \tag{5}$$

$$T_{3n} = T_{3n}(D_1, \ldots, D_n) = \frac{1}{2} \sum_{t=1}^{n-1} (|D_t| + |D_{t+1}|) I_+ + \frac{1}{2} \sum_{t=1}^{n-1} \frac{|D_t|^2 + |D_{t+1}|^2}{|D_t| + |D_{t+1}|} I_-, \tag{6}$$

are also available, where $I_+ = I\{D_t D_{t+1} \geq 0\}$, $I_- = I\{D_t D_{t+1} < 0\}$ 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 α, 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 $C_{y,t} = Y_t - \bar{Y}$, $C_{x,t} = X_t - \bar{X}$, $\bar{Y} = \sum_{t=1}^{n} Y_t/n$ and $\bar{X} = \sum_{t=1}^{n} X_t/n$. The combined blocks are denoted by ξ_{pooled}.
3. Draw blocks with replacement from ξ_{pooled} to generate two resamples, $Y^{*b} = \{Y_1^{*b}, \ldots, Y_n^{*b}\}$ and $X^{*b} = \{X_1^{*b}, \ldots, X_n^{*b}\}$ $(b = 1, \ldots, B)$, corresponding to two samples, Y and X.
4. Calculate $t^{*b} = T(Y^{*b}, X^{*b}) = T(D_1^{*b}, \ldots, D_n^{*b})$.
5. Repeating steps 3 and 4 an appropriate number of times B, calculate $t^{*1}, \ldots, t^{*B}$.
6. From steps 1 and 5, approximate the achieved significance level by $\widehat{\text{ASL}} = \sum_{b=1}^{B} I\{t^{*b} \geq t_{obs}\}/B$, and reject H_0 when $\widehat{\text{ASL}} \leq \alpha$.

Algorithm 2.2 (Mixed MBB test)

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

3. (a) Combine ξ_y and ξ_x, and put $\xi_{\text{pooled}}=\{\xi_{y,1},\dots,\xi_{y,k}\,\xi_{x,1},\dots,\xi_{x,k}\}$.
(b) Draw $\xi_y^{*b}=\{\xi_{y,1}^{*b},\dots,\xi_{y,m}^{*b}\}$ and $\xi_x^{*b}=\{\xi_{x,1}^{*b},\dots,\xi_{x,m}^{*b}\}$ with replacement from ξ_{pooled} to obtain resamples $Y^{*b}=\{Y_1^{*b},\dots,Y_n^{*b}\}$ and $X^{*b}=\{X_1^{*b},\dots,X_n^{*b}\}$, where $m=n/\ell$ (if $[n/\ell]$ is an integer) or $m=[n/\ell]+1$ (otherwise), and $[n/\ell]$ is the integer part of a real n/ℓ.

Algorithm 2.3 (Mixed CBB test)

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

3. (a) Combine ξ_y and ξ_x, and put $\xi_{\text{pooled}}=\{\xi_{y,1},\dots,\xi_{y,n}\,\xi_{x,1},\dots,\xi_{x,n}\}$.
(b) Draw $\xi_y^{*b}=\{\xi_{y,1}^{*b},\dots,\xi_{y,m}^{*b}\}$ and $\xi_x^{*b}=\{\xi_{x,1}^{*b},\dots,\xi_{x,m}^{*b}\}$ with replacement from ξ_{pooled} to obtain resamples $Y^{*b}=\{Y_1^{*b},\dots,Y_n^{*b}\}$ and $X^{*b}=\{X_1^{*b},\dots,X_n^{*b}\}$, where m is defined in Algorithm 2.2.

Algorithm 2.4 (Mixed SB test)

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

$$\xi_y=\{\xi_y(1,L_1),\dots,\xi_y(n,L_n)\},\quad \xi_x=\{\xi_x(1,M_1),\dots,\xi_x(n,M_n)\},$$

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

3. (a) Combine ξ_y and ξ_x, and put $\xi_{\text{pooled}}=\{\xi_y(1,L_1),\dots,\xi_y(n,L_n),\ \xi_x(1,M_1),\dots,\xi_x(n,M_n)\}$.
(b) Draw $K_{y,b}$ and $K_{x,b}$ blocks with replacement from ξ_{pooled}, and put

$$\xi_y^{*b}=\{\xi(I_1^{*b},L_1^{*b}),\dots,\xi(I_{K_{y,b}}^{*b},L_{K_{y,b}}^{*b})\},\quad \xi_x^{*b}=\{\xi(J_1^{*b},M_1^{*b}),\dots,\xi(J_{K_{x,b}}^{*b},M_{K_{x,b}}^{*b})\},$$

where $\xi(t,\ell)=\xi_y(t,\ell)$ (if $1\leq t\leq n$), $\xi(t,\ell)=\xi_x(t,\ell)$ (otherwise), $I_1^{*b},\dots,I_{K_{y,b}}^{*b},\ J_1^{*b},\dots,J_{K_{x,b}}^{*b}$ are independent and identically distributed to a discrete uniform distribution on $\{1,\dots,n,\ n+1,\dots,2n\}$. A pair of random variables, (I_i^{*b},L_i^{*b}) or (J_i^{*b},M_i^{*b}), is one of $\{(1,L_1),\dots,(n,L_n),(1,M_1),\dots,(n,M_n)\}$, and $K_{y,b}=\min\{k:\sum_{i=1}^{k}L_i^{*b}\geq n\}$, $K_{x,b}=\min\{k:\sum_{j=1}^{k}M_j^{*b}\geq n\}$, respectively.
(c) Construct resamples, $Y^{*b}=\{Y_1^{*b},\dots,Y_n^{*b}\}$ and $X^{*b}=\{X_1^{*b},\dots,X_n^{*b}\}$, by putting the first n elements of ξ_y^{*b} and ξ_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

Table 1: Optimum ℓ in Mixed MBB, CBB and SB tests for $\alpha = 0.05$ and $\mathrm{V}(\varepsilon_i(t)) = 3$

			Mixed MBB				Mixed CBB				Mixed SB			
q_1	q_2	ϕ	T_{1n}	T_{2n}	T_{3n}	S_n	T_{1n}	T_{2n}	T_{3n}	S_n	T_{1n}	T_{2n}	T_{3n}	S_n
10	30	-0.2	6	7	7	2	1	1	6	2	1	1	8	2
		-0.1	5	5	4	2	1	1	8	2	1	1	8	1
		0	4	4	2	1	1	8	4	1	1	8	4	1
		0.1	2	1	2	1	3	1	9	1	3	1	9	1
		0.2	2	1	3	1	4	9	9	1	4	7	9	1
20	20	-0.2	6	4	4	2	9	9	8	2	1	9	5	2
		-0.1	4	3	3	2	9	9	4	2	9	9	3	2
		0	3	1	1	1	7	8	1	2	9	5	1	1
		0.1	1	1	1	1	3	4	1	1	4	1	1	1
		0.2	1	1	2	1	8	1	3	1	1	1	3	1

Table 2: Empirical levels for $\alpha = 0.05$ and $\mathrm{V}(\varepsilon_i(t)) = 3$

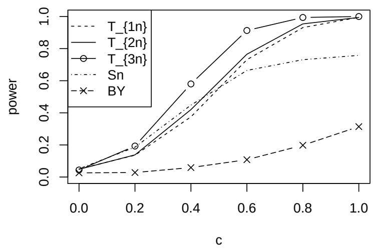
			Mixed MBB				Mixed CBB				Mixed SB				
q_1	q_2	ϕ	T_{1n}	T_{2n}	T_{3n}	S_n	T_{1n}	T_{2n}	T_{3n}	S_n	T_{1n}	T_{2n}	T_{3n}	S_n	BY
10	30	-0.2	0.048	0.046	0.053	0.051	0.026	0.018	0.033	0.050	0.024	0.019	0.037	0.048	0.043
		-0.1	0.050	0.049	0.051	0.052	0.032	0.028	0.045	0.052	0.030	0.026	0.051	0.049	0.036
		0	0.046	0.052	0.059	0.058	0.038	0.036	0.054	0.057	0.037	0.038	0.052	0.054	0.036
		0.1	0.054	0.049	0.083	0.057	0.047	0.049	0.067	0.054	0.047	0.049	0.072	0.058	0.037
		0.2	0.076	0.072	0.112	0.070	0.050	0.055	0.086	0.071	0.051	0.059	0.092	0.070	0.037
20	20	-0.2	0.050	0.048	0.044	0.051	0.035	0.045	0.049	0.051	0.032	0.042	0.050	0.050	0.026
		-0.1	0.044	0.044	0.050	0.055	0.042	0.055	0.051	0.053	0.043	0.054	0.050	0.056	0.025
		0	0.049	0.044	0.051	0.045	0.049	0.050	0.052	0.055	0.051	0.050	0.051	0.043	0.025
		0.1	0.053	0.055	0.080	0.058	0.049	0.055	0.081	0.059	0.051	0.055	0.080	0.061	0.025
		0.2	0.067	0.075	0.117	0.065	0.065	0.076	0.117	0.066	0.068	0.076	0.121	0.065	0.024

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: $\varepsilon_i(t) = \phi\varepsilon_i(t-1) + z_i(t)$ and $\eta_j(t) = \phi\eta_j(t-1) + z_j(t)$, where $z_i(t) \overset{i.i.d.}{\sim} N(0, \tau_1^2)$, $z_j(t) \overset{i.i.d.}{\sim} N(0, \tau_2^2)$, $\phi = 0, \pm 0.1, \pm 0.2$, $\tau_1^2 = \tau_2^2 = (1 - \phi^2)\mathrm{V}(\varepsilon_i(t))$, and $\mathrm{V}(\varepsilon_i(t)) = 1, 3, 5$. 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 $\mathrm{V}(\varepsilon_i(t)) = 3$ in this paper.

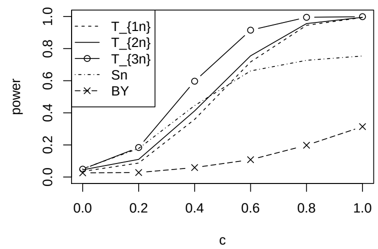
Since it is preferable that the empirical level is nearly equal to the nominal level α, our choice of ℓ in Mixed MBB, CBB and SB tests is made so that the empirical level is close to α. If there are some candidates which have the same level errors, we make the conservative choice, viz., we choose ℓ such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

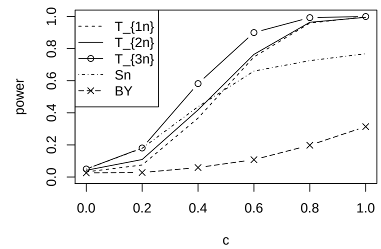
Next, we discuss the power studies based on Figure 1. The vertical and horizontal axes of Fig-



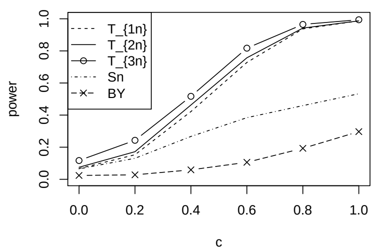
(a) Mixed MBB ($\phi = -0.2$)



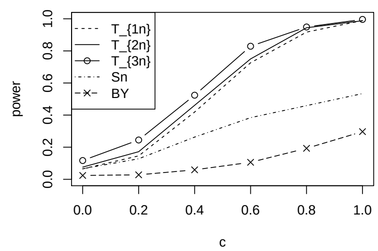
(b) Mixed CBB ($\phi = -0.2$)



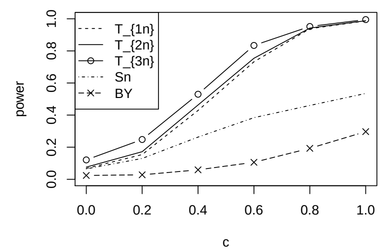
(c) Mixed SB ($\phi = -0.2$)



(d) Mixed MBB ($\phi = 0.2$)



(e) Mixed CBB ($\phi = 0.2$)



(f) Mixed SB ($\phi = 0.2$)

Figure 1: Comparison of T_{1n}, T_{2n}, T_{3n}, S_n and BY for $(q_1, q_2) = (20, 20)$ and $\mathrm{V}(\varepsilon_i(t)) = 3$

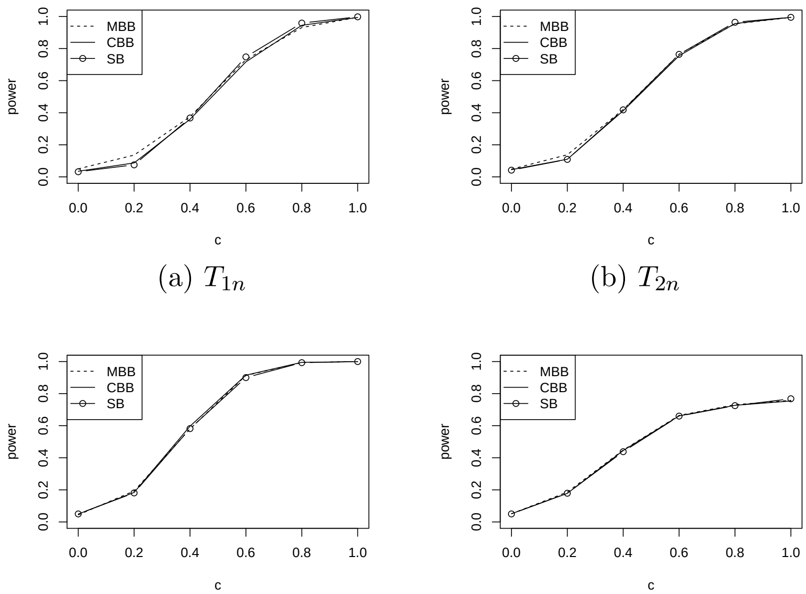
ure 1 are the empirical power of tests and c $(0 \leq c \leq 1)$ defined above. Since we found similar tendencies among the six cases of (q_1, q_2), we show the results for $(q_1, q_2) = (20, 20)$ with $\phi = \pm 0.2$. We can observe that the empirical power of T_{3n} is most powerful among those corresponding to five statistics, and that the overall relationship among powers in each bootstrap test together with BY test is given by $T_{3n} \geq T_{2n} \geq T_{1n} \geq S_n \geq$ BY. This indicates the numerical superiority of Mixed MBB, CBB and SB tests using T_{rn} $(r = 1, 2, 3)$ and S_n in power. Especially, the superiority of T_{3n} in power is confirmed from Figure 1. As the number of subjects increases, the empirical power was improved. For $0 \leq c \leq 1$, the empirical power of T_{1n} is nearly equal to that of T_{2n}, however the latter is slightly higher than the former for most cases. Figure 2 shows the comparison of empirical powers corresponding to Mixed MBB, CBB and SB tests when the test statistic is fixed. From this figure, we can observe no significant difference among the three tests.

4. Concluding remarks

In this paper we have compared three block bootstrap testing methods using MBB, CBB and SB for two means in longitudinal data. Our simulation results indicate the applicability of Mixed MBB, CBB and SB tests for weakly dependent data even when the sample size is very small. The problem on block length selection in the block resampling is very important, and the development of a fully data-driven approach to selecting (mean) block length in the above tests will be needed for practical data analyses. Further, extending the testing methods to several curves and/or grouped data problems would be required in the future.

References

Bowman, A. & Young, S. (1996). Graphical comparison of nonparametric curves. *Applied Statistics*, **45**, 83–98.



(a) T_{1n} (b) T_{2n}

(c) T_{3n} (d) S_n

Figure 2: Comparison of Mixed MBB, CBB and SB tests for $(q_1, q_2) = (20, 20)$, $\phi = -0.2$ and $\mathrm{V}(\varepsilon_i(t)) = 3$

Hall, P. & Hart, J. D. (1990). Bootstrap test for difference between means in nonparametric regression. *Journal of the American Statistical Association*, **85**, 1039–1049.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*, **17**, 1217–1241.

Politis, D. N. & Romano, J. P. (1992). A circular block-resampling procedure for stationary data. *Exploring the Limit of Bootstrap* (eds. LePage, R. & and Billard, L.), Wiley, 263–270.

Politis, D. N. & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, **89**, 1303–1313.

Sakurai, H. & Taguri, M. (2005). Test for difference of two means in longitudinal data using moving block bootstrap. *Proceedings of the 5th IASC Asian Conference on Statistical Computing*, 139–142.

Sakurai, H. & Taguri, M. (2010). Test of mean difference for longitudinal data using circular block bootstrap. *Proceedings of COMPSTAT'2010*, 1581–1588.

Sakurai, H. & Taguri, M. (2013). Testing methods of mean difference for longitudinal data based on stationary bootstrap. *Proceedings of the 59th World Statistics Congress of the International Statistical Institute, 2013*, 5303–5308.