Biometrika (2015), **94**, 2, *pp*. 1–8

Advance Access publication on 31 July 2012
Printed in Great Britain

Supplementary material for 'change point estimation: another look at multiple testing problems'

BY HONGYUAN CAO

Department of Statistics, University of Missouri-Columbia, Columbia, Missouri 65211, U.S.A.
caohong@missouri.edu

AND WEI BIAO WU

Department of Statistics, University of Chicago, Chicago, Illinois 60637, U.S.A.
wbwu@uchicago.edu

SUMMARY

This Supplementary Material includes the proofs of Theorems 1–2, testing and estimation of change-points under dependence, and simulation studies.

1. PROOF OF THEOREM 1

Let $S_i = \sum_{j=1}^{i} 12^{1/2}(p_j - 1/2)$. The summands $12^{1/2}(p_j - 1/2)$ are independent identically distributed with mean 0 and variance 1. We shall apply the strong invariance principle result (Komlós et al., 1975, 1976): There exists a richer probability space on which we can define a Brownian motion $B(\cdot)$ such that

$$\max_{i \le m} |S_i - B(i)| = o(\log m)$$

almost surely. Hence the increment process $S_i - S_{i-k}$ satisfies the Gaussian approximation

$$\max_{k \le i \le m} \left| k^{-1/2}(S_i - S_{i-k}) - k^{-1/2}\{B(i) - B(i-k)\} \right| = o_p(k^{-1/2} \log m). \quad \text{(S1)}$$

By Corollary A1 of Bickel & Rosenblatt (1973)

$$(2 \log g_m)^{1/2} \max_{0 \le u \le m-k} |B(u+1) - B(u)| - A_{g_m} \to E \quad \text{(S2)}$$

in distribution. By the scaling property of Brownian motion, the two incremental processes $[\{B(u+k) - B(u)\}k^{-1/2}, u \ge 0]$ and $[\{B(u+1) - B(u)\}, u \ge 0]$ have the same distribution. Hence

$$(2 \log g_m)^{1/2} \max_{0 \le u \le m-k} \left| k^{-1/2}\{B(u+k) - B(u)\} \right| - A_{g_m} \to E \quad \text{(S3)}$$

in distribution. By (S1) and (S3), we have (7) in view of the discretization approximation $\max_{0 \le u \le m} |B(u) - B(\lfloor u \rfloor)| = O_p\{(\log m)^{1/2}\}$. To see the latter, by the Bonferroni inequality

$$\begin{aligned} \mathrm{pr}\left\{ \max_{0 \le u \le m} |B(u) - B(\lfloor u \rfloor)| \ge 4(\log m)^{1/2} \right\} &\le m\mathrm{pr}\left\{ \max_{0 \le u < 1} |B(u)| \ge 4(\log m)^{1/2} \right\} \\ &\le 4m\mathrm{pr}\{B(1) \ge 2(\log m)^{1/2}\} \to 0 \end{aligned}$$

as $m \to \infty$. Here we have applied the reflection principle of Brownian motion. Hence the proof of (7) is completed.

2. PROOF OF THEOREM 2

Let $Q_i = k^{-1} \sum_{j=i-k}^{i-1} p_j$ and $G_i = \sum_{j=1}^{i} \{p_j - E(p_j)\}$. Let A be the event $\{\max_{1+k \le i \le m} |Q_i - E(Q_i)| \le \gamma\}$. By the triangle inequality and Freedman (1975)'s martingale inequality, let $t = k\gamma/2$, since $E(G_i^2) \le i/12$, we have

$$\begin{aligned} \mathrm{pr}\left\{\max_{1+k \le i \le 2k} |Q_i - E(Q_i)| \ge \gamma\right\} &\le \mathrm{pr}\left(\max_{i \le 2k} |G_i| \ge t\right) + \mathrm{pr}\left(\max_{i \le k} |G_i| \ge t\right) \\ &\le 4 \exp\{-(t^2/2)/(t/3 + 2k/12)\} \\ &\le 4 e^{-3k\gamma^2/(4+4\gamma)}. \end{aligned}$$

Hence we have

$$\mathrm{pr}(A) \ge 1 - 4k^{-1} m e^{-3k\gamma^2/(4+4\gamma)}. \quad \text{(S4)}$$

In the rest of the proof we shall restrict ourselves to the event A. Recall that the null hypotheses correspond to the indices $\mathcal{S}_0 \cup \mathcal{S}_2 \cup \cdots$ and the alternative hypotheses correspond to the indices $\mathcal{S}_1 \cup \mathcal{S}_3 \cup \cdots$. By (6) and condition (C1), we have $k = o(\tau_1)$. Under event A, if $i = \tau_0 + k, \ldots, \tau_1 - k$, we have $E(Q_i) = 1/2$. Hence $\{\tau_0 + k, \ldots, \tau_1 - k\} \subset W_0$. Similarly, under event A, $\{\tau_2 + k, \ldots, \tau_3 - k\} \subset W_0$, etc.

If $i = \tau_1 + k, \ldots, \tau_2 - k$, we have $E(Q_i) \le \rho$. By our conditions, $\rho + \gamma < 1/2$. Then under A, $|Q_i - 1/2| > \gamma$, which implies $\{\tau_1 + k, \ldots, \tau_2 - k\} \subset W_2$. The latter interval can be slightly extended. Let $K = \lfloor 2k\gamma/(1/2 - \rho) \rfloor$. If $\tau_1 + K \le i \le \tau_1 + k$, under A, we also have $|Q_i - 1/2| > \gamma$ since $|Q_i - E(Q_i)| \le \gamma$ and $1/2 - E(Q_i) > 2\gamma$. The latter follows from

$$\begin{aligned} kE(Q_i) = \sum_{j=\tau_1}^{i} E(p_j) + \sum_{j=i-k+1}^{\tau_1 - 1} E(p_j) &\le (i - \tau_1 + 1)\rho + (\tau_1 - 1 - i + k)/2 \\ &\le 2^{-1}k - (K+1)(\rho - 1/2) < (1/2 - 2\gamma)k. \end{aligned} \quad \text{(S5)}$$

Similarly, if $i = \tau_2 - k, \ldots, \tau_2 - K - 1$, under A we also have $|R_i - 1/2| > \gamma$. Hence under A, $\{\tau_1 + K, \ldots, \tau_2 - K - 1\} \subset W_2$. Similarly, $\{\tau_3 + k, \ldots, \tau_4 - k - 1\} \subset W_2$, etc.

Under A, we have $\{\tau_h - k + K + 1, \ldots, \tau_h - 1\} \subset W_1$ for all odd indices h. Let $h = 1$ and $i = \tau_1 - k + K + 1, \ldots, \tau_1 - 1$. Similar to (S5),

$$\sum_{j=i}^{i+k-1} E(p_j) = \sum_{j=i}^{\tau_1 - 1} E(p_j) + \sum_{j=\tau_1}^{i+k-1} E(p_j) \le (\tau_1 - i)/2 + (k - \tau_1 + i)\rho,$$

which implies that $|k^{-1} \sum_{j=i}^{i+k-1} E(p_j) - 1/2| > 2\gamma$. Hence $R_i > \gamma$. By the same token, for even indices, we have $\{\tau_g, \ldots, \tau_g + k - K - 1\} \subset W_1$ for all even indices g.

For the connected component $\mathcal{M}_1$ of W_1 whose length is larger than $k/2$, under A, we have

$$\{\tau_1 - k + K + 1, \ldots, \tau_1 - 1\} \subset \mathcal{M}_1 \subset \{\tau_1 - k + 1, \ldots, \tau_1 + K - 1\}. \quad \text{(S6)}$$

Clearly $L_j < \gamma$ for $j \in \mathcal{M}_1$. We now consider the maximizer $\hat{\tau}_1 = \mathrm{argmax}_{j \in \mathcal{M}_1} R_j 1(R_j > \gamma)$. If $j = \tau_1 - k + 1, \ldots, \tau_1 - K - 1$, under A, we have $R_j < R_{\tau_1}$. To see the latter, let $F_i = \sum_{j=i}^{i+k-1} E(p_j)$. We have $E(F_j) - E(F_{\tau_1}) \ge (\tau_1 - j)(1/2 - \rho) \ge (K+1)(1/2 - \rho)$. Under

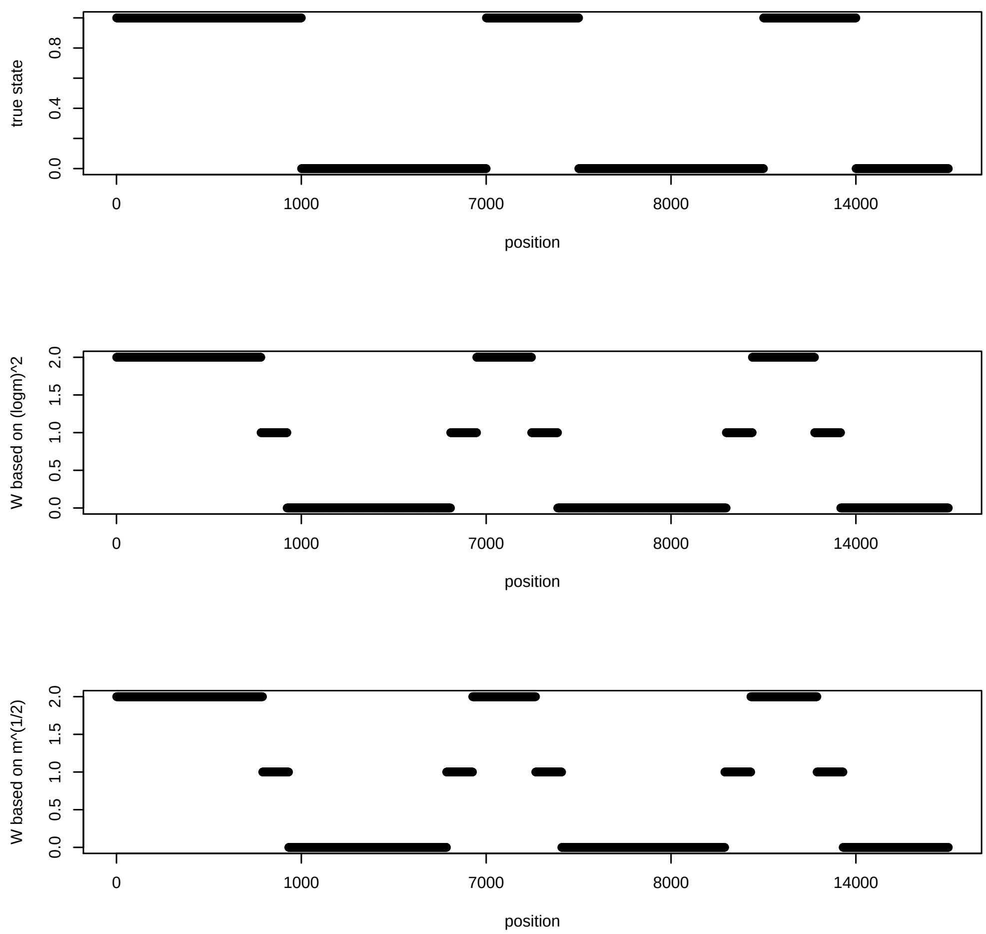

Fig. S1. Realization from one experiment. top panel: true state; middle panel: W based on $k = \{\log(m)\}^2$; bottom panel: W based on $k = m^{1/2}$.

A,

$$\begin{aligned} R_{\tau_1} - R_j &= |1/2 - k^{-1} F_{\tau_1}| - |1/2 - k^{-1} F_j| \\ &> |1/2 - k^{-1} E(F_{\tau_1})| - |1/2 - k^{-1} E(F_j)| - 2\gamma \\ &\geq (K+1)(1/2 - \rho) - 2\gamma > 0. \end{aligned}$$

So $\hat{\tau}_1 \geq \tau_1 - K$. By (S6), $\hat{\tau}_1 \leq \tau_1 + K - 1$. Hence event A implies the event in (9), which follows from (S4).

3. A SIMULATION STUDY WITH INDEPENDENT p-VALUES

In this section, we shall investigate the proposed procedure for the multiple testing problem with clustered signals through Monte Carlo simulation. Numbers of tests are set to be $m =$

$20,000$ and $m = 100,000$. The number of change-points is set to be 5. The z-value at each locus follows normal distribution with mean exhibit in Table 1 and variance 1. We change the signal strength at the first 2.5% loci of m tests. The p-values are calculated based on the standard

Table 1. *Signal and noise configuration*

Segment ($\%$ among m)	2.5	2.5	30	2.5	30	2.5	30
Signal strength (mean level)	μ	-1.5	0	1	0	-1.5 and 1 alternating	0

normal distribution. $1,000$ datasets are generated to do the experiment. A realization from one experiment is presented in Figure S1, showing the true state and W based on our algorithm with $k = \lfloor \{\log(m)\}^2 \rfloor$ and $k = \lfloor m^{1/2} \rfloor$, where $m = 20,000$ and the signal strength for the first 2.5% of m tests is $\mu = 1.8$.

Theorem 2 only requires that $k/m \to 0$ and $\log(m)/k \to 0$ as $m \to \infty$. We choose the window size $k = \lfloor m^{1/2} \rfloor$ and $\lfloor \{\log(m)\}^2 \rfloor$. We obtain 10^4 independent realizations of $A_m = 12^{-1}\max_{k+1 \le i \le m} L_i$, based on m independent $\mathcal{U}(0,1)$ random variables. The critical value $\gamma_{m,\alpha}$ is estimated by the empirical 95% quantile of these 10^4 realizations of A_m. We choose $\alpha = 0.05$. This direct simulation-based approach performs better than the cutoff value given in (8). We implement our testing procedure following the change-point detection algorithm and evaluate its performance by the false discovery rate, the false non-discovery rate and the missed discovery rate. The false non-discovery rate is defined as the expected value of the ratio of falsely accepted hypotheses and total accepted hypotheses; and the missed discovery rate is defined as the expected value of the ratio of falsely accepted hypotheses and total alternative hypotheses. The false non-discovery and missed discovery rates can be used to describe the power of a multiple testing procedure, similar to the type II error rate in a single hypothesis testing setup.

We compare our methods with the smoothing method proposed by Zhang et al. (2011) and Benjamini & Hochberg (1995)'s procedure. At the realized false discovery rate level based on our procedure, we implement the smoothing method proposed by Zhang et al. (2011). Specifically, let $\hat{G}^*(t) = \{2\sum_{i=1}^m I(p_i^* > 0.5) + \sum_{i=1}^m I(p_i^* = 0.5)\}^{-1} \sum_{i=1}^m I\{p_i^* \ge 1 - t\}$ if $0 \le t \le 0.5$ and $\hat{G}^*(t) = 1 - \{2\sum_{i=1}^m I(p_i^* > 0.5) + \sum_{i=1}^m I(p_i^* = 0.5)\}^{-1} \sum_{i=1}^m I\{p_i^* \ge t\}$ if $0.5 < t \le 1$, where p_i^* is the median of the p-values in the k^*th neighbourhood of ith hypothesis. Following Zhang et al. (2011), the estimated false discovery rate is

$$\widehat{\mathrm{FDR}}(t) = \Big[\{R^*(t) \vee 1\}\{1 - \hat{G}^*(t)\}\Big]^{-1} W^*(\lambda)\hat{G}^*(t),$$

where $W^*(\lambda) = \sum_{i=1}^m I\{p_i^* > \lambda\}$ and λ is a tuning parameter. At false discovery rate level α, threshold $\hat{t}$ is chosen as the largest t such that $\widehat{\mathrm{FDR}}(\hat{t}) \le \alpha$. As in Zhang et al. (2011), we set the tuning parameter $\lambda = 0.1$ and the size of neighborhood k^* the same as our sliding window length k.

The results are summarized in Table 2. It suggests that, when the signal is moderate to large, our procedure performs similarly across a spectrum of bandwidths for a large number of tests. The false discovery rate and missed discovery rate are pretty small and get smaller with increased signal strength. With similar false discovery rates, our procedure performs uniformly better than Zhang et al. (2011)'s procedure and Benjamini & Hochberg (1995)'s procedure in terms of false non-discovery rate and missed discovery rate. Zhang et al. (2011)'s procedure takes into account the clustering structure and has improved performance with increased signal strength. In contrast, Benjamini & Hochberg (1995)'s procedure does not change much.

Table 2. *False discovery rate, false non-discovery rate and missed discovery rate with* $1,000$ *simulations for independent case*

	$m = 20000$				$m = 100000$			
μ	k	FDR	FNR	MDR	k	FDR	FNR	MDR
				Our procedure				
0.1	141	0.14	4.04	40.46	316	0.02	2.97	28.31
	98	0.08	4.60	45.47	132	0.02	3.25	30.64
0.8	141	0.12	2.75	27.25	316	0.02	0.59	5.52
	98	0.08	4.27	42.05	132	0.18	3.00	28.16
1.8	141	0.10	2.00	19.63	316	0.02	0.57	5.29
	98	0.06	2.39	23.09	132	0.01	0.69	6.33
				Zhang et al. (2011)'s procedure				
0.1	141	0.54	7.33	71.19	316	0.12	6.62	63.81
	98	0.53	7.87	76.91	132	0.14	8.14	79.73
0.8	141	0.49	7.28	70.67	316	0.12	6.55	63.08
	98	0.52	7.81	76.24	132	0.14	8.11	79.42
1.8	141	0.37	4.84	45.78	316	0.08	4.11	38.59
	98	0.29	5.64	53.88	132	0.07	5.98	57.28
				Benjamini & Hochberg (1995)'s procedure				
0.1		2.86	9.99	99.99		2.44	9.99	99.99
0.8		2.91	9.99	99.99		2.38	9.99	99.99
1.8		1.35	9.99	99.99		0.00	9.99	99.99

Note: All numbers are multiplied by 100; "μ" is the mean of signal strength at first 2.5% loci of m tests, "k" is bandwidth, "FDR" is false discovery rate, "FNR" is false non-discovery rate and "MDR" is missed discovery rate.

4. Testing and Estimation of Change-points under Dependence

In this section we shall generalize the results in Section 2 by allowing dependence in p-values. To generalize Theorem 1, we assume that the p-values $(p_1, \ldots, p_m)$ form a stationary process

$$p_i = G(\xi_i), \quad \xi_i = (\ldots, \varepsilon_{i-1}, \varepsilon_i, \varepsilon_{i+1}, \ldots), \tag{S7}$$

where the ε_i are independent identically distributed random variables, and G is a measurable function such that $p_i \sim \mathcal{U}(0,1)$ marginally for $i = 1, \ldots, m$. Note that (S7) defines a very general class of stationary processes. As in Wu (2005), we define the functional dependence measure of the sequence $(p_1, \ldots, p_m)$. Let $\varepsilon_i, \varepsilon'_j, i, j \in Z$, be independent identically distributed random variables. For $q > 2$, define the functional dependence measure:

$$\delta_{k,q} = \|G(\xi_i) - G(\xi_{i,\{i-k\}})\|_q, \tag{S8}$$

where $\xi_{i,\{k\}} = \{\ldots, \varepsilon_{i-1,(k)}, \varepsilon_{i,(k)}, \varepsilon_{i+1,(k)}, \ldots\}$, $\varepsilon_{j,(k)} = \varepsilon_j$ if $j \neq k$ and $\varepsilon_{j,(k)} = \varepsilon'_j$ if $j = k$. Then the sequence $(\delta_{k,q})_{k=-\infty}^{\infty}$ quantifies the dependence of $(p_{i+k})_{k=-\infty}^{\infty}$ on ε_i.

Theorem S1. *Assume the tail functional dependence measure*

$$\Delta_{m,q} = \sum_{|k| \geq m} \delta_{k,q} = O(m^{-\theta}), \tag{S9}$$

where $\theta > \max\left[1,\ (q-2)\{q + 2 + (q^2 + 20q + 4)^{-1}\}/(8q)\right]$, *and*

$$k_m^{-1/2} m^{1/q} (\log m)^{-1/2} + m^{-1} k_m \to 0$$

as $m \to \infty$. *Then under (S7) with* $p_i \sim \mathcal{U}(0,1)$, *(7) still holds with the constant* $1/12$ *therein replaced by the long-run variance* $\sigma^2 = \sum_{k \in Z} \mathrm{cov}(p_0, p_k)$.

Theorem S1 can be similarly proved by using the argument of Theorem 1. The only difference is that, instead of using the Gaussian approximation result in Komlós et al. (1975, 1976), we apply Corollary 2.1 in Berkes et al. (2014).

Example S1. Let $p_i = F(X_i)$, where $X_i = \sum_{j=0}^{\infty} a_j \varepsilon_{i-j}$ and F is the cumulative distribution function of X_i. A similar linear process model is considered in Clarkes & Hall (2009). Assume that the density $f(x) = dF(x)/dx$ is bounded and ε_i has a finite νth moment, $\nu > 0$. Then the functional dependence measure $\delta_{k,q} = O\Big([E\{\min(1, |a_k||\varepsilon_0 - \varepsilon_0'|)^q\}]^{1/q}\Big) = O\{|a_k|^{\min(1,\nu/q)}\}$. Hence, if $a_k = O(k^{-\beta})$, $\beta > 0$, then (S9) holds if $\beta > (1+\theta)/\min(1, \nu/q)$.

To establish a version of Theorem 2 for dependent p-values, we shall apply Rosenblatt (1952)'s transformation and let

$$p_i = G_i(\xi_i), \tag{S10}$$

where G_i are measurable functions such that $G_i \sim \mathcal{U}(0,1)$ if H_i is a null hypothesis and otherwise if it is any alternative hypothesis. Note that $(p_1, \ldots, p_m)$ can be a non-stationary sequence. Extending (S8), we define the uniform functional dependence measure

$$\delta_{k,q} = \sup_i \|G_i(\xi_i) - G_i\{\xi_{i,(i-k)}\}\|_q.$$

Assume that there exists $0 < \zeta \leq 2$ such that

$$\overline{\lim_{q\to\infty}}\, q^{1/2-1/\zeta} \sum_{k\in Z} \delta_{k,q} < \infty. \tag{S11}$$

If $(p_1, \ldots, p_m)$ is ℓ-dependent, $\ell \geq 0$, in the sense that $G_i(\xi_i)$ only depends on $\varepsilon_{i-\ell}, \ldots, \varepsilon_i$, then $\delta_{k,q} = 0$ if $k \geq \ell$ and $k < 0$, and hence (S11) holds automatically with $\zeta = 2$. Under (S11), by the argument of Theorem 2 in Wu (2005), we have the following Hoeffding-type inequality for dependent random variables: there exist constants $C_1, C_2 > 0$, such that

$$\mathrm{pr}\left\{\max_{1\leq l\leq j} |S_{i,l} - E(S_{i,l})| \geq j^{-1/2}u\right\} \leq C_1 e^{-C_2 u^{\zeta}} \tag{S12}$$

for all $u > 0, i \geq 0$ and $j > 1$, where $S_{i,j} = \sum_{l=1+i}^{i+j} p_l$. Following the argument of Theorem 2, using inequality (S12), we obtain

THEOREM S2. *Assume (S10), (C1)–(C2) and (S11). Let $\gamma \asymp (k_m^{-1}\log m)^{1/2}$ and assume $\gamma + \rho < 1/2$. Then*

$$\mathrm{pr}\left\{\hat{l} = l, \max_{i\leq l} |\hat{\tau}_i - \tau_i| \leq (1/2-\rho)^{-1} 2k\gamma\right\} \geq 1 - C_3 k^{-1} m e^{-C_4 (k^{-1/2}\gamma)^{\zeta}},$$

as $m \to \infty$.

Under dependence, the convergence rate of our algorithm can be slower than under independence, as asserted by the bound given in Theorem S2. The primary impact of dependence on our testing procedure is that instead of using the marginal variance of p-values that follow $\mathcal{U}(0,1)$, we need to use the long-run variance $\sigma^2 = \sum_{k\in Z} \mathrm{cov}(p_0, p_k)$ to incorporate the dependence. In addition, the rate of convergence is slower, as $0 < \zeta \leq 2$.

5. A SIMULATION STUDY ON HOW DEPENDENCE AFFECTS TESTING PERFORMANCE

To investigate the impact of dependence on the procedure, we simulate data with AR(1) error structure. Specifically, the measurement errors now follow

$$e_i = \rho e_{i-1} + \epsilon_i, \quad \epsilon_i \sim N(0, 1), \quad i = 1, \ldots, m. \tag{S13}$$

The rest of simulation setup is the same as in the independent case. p-values are calculated using the standard normal distribution after standardization. Specifically, $p = 2\{1 - \Phi(|Z|)\}$, where $\Phi(\cdot)$ is the standard normal cumulative distribution function and $Z = \mu + \sqrt{1-\rho^2}e$, where μ is signal strength, and e is from (S13). To save space, we only show results based on our procedure and Zhang et al. (2011)'s procedure with correlation $\rho = 0.3$ and $\rho = 0.6$. As we can see from Table 3, the performances of both procedures deteriorate with dependent error term. The performances improve when sample size and signal strength increase.

Table 3. *False discovery rate, false non-discovery rate and missed discovery rate with* $1,000$ *simulations for dependent case*

		$m = 20000$				$m = 100000$			
ρ	μ	k	FDR	FNR	MDR	k	FDR	FNR	MDR
					Our procedure				
0.3	0.1	141	0.17	4.33	43.50	316	0.03	3.01	28.75
		98	0.12	4.92	48.74	132	0.02	3.61	34.09
	0.8	141	0.14	3.21	31.96	316	0.02	0.76	7.06
		98	0.11	4.59	45.39	132	0.02	3.41	32.14
	1.8	141	0.12	2.29	22.52	316	0.02	0.61	5.69
		98	0.08	2.70	26.20	132	00.02	1.06	9.73
0.6	0.1	141	0.18	5.40	54.91	316	0.04	3.27	31.29
		98	0.11	5.84	58.39	132	0.04	3.27	31.29
	0.8	141	0.16	4.80	48.50	316	0.04	1.92	18.16
		98	0.10	5.67	56.59	132	0.02	4.79	45.84
	1.8	141	0.13	3.35	33.33	316	0.03	0.86	8.07
		98	0.07	3.59	35.08	132	0.01	2.37	22.08
					Zhang et al. (2011)'s procedure				
0.3	0.1	141	1.32	7.34	71.29	316	0.26	6.69	64.59
		98	1.29	7.95	77.70	132	0.38	8.25	80.96
	0.8	141	1.20	7.24	70.31	316	0.24	6.61	63.69
		98	1.27	7.83	76.48	132	0.37	8.19	80.31
	1.8	141	0.82	4.99	47.28	316	0.19	4.17	39.25
		98	0.72	5.85	55.99	132	0.19	6.24	59.90
0.6	0.1	141	8.57	7.70	74.94	316	2.77	7.14	69.15
		98	10.27	8.18	79.98	132	5.07	8.43	82.76
	0.8	141	7.92	7.49	72.77	316	2.61	6.96	67.26
		98	0.10	5.67	56.59	132	4.78	8.30	81.43
	1.8	141	5.21	5.62	53.47	316	1.67	4.88	46.12
		98	0.07	3.59	35.08	132	2.59	6.66	64.22

Note: All numbers are multiplied by 100; "μ" is the mean of signal strength at first 2.5% loci of m tests, "k" is bandwidth, "FDR" is false discovery rate, "FNR" is false non-discovery rate and "MDR" is missed discovery rate.

All previous results are based on two-sided tests. We next look at one-sided tests with both positive and negative dependence. The setup is the same as in the dependent case and we use $k = \lfloor \{\log(m)\}^2 \rfloor$. It turns out that the type I error is not correct if we analyze dependent data under the independence assumption. For example, with $m = 20,000$ tests, if $\rho = 0.3$ and we treat as if the errors were independent, at significance level 0.05, the actual type I error is 0.8729; on the other hand, if $\rho = -0.3$ and we treat as if the errors were independent, at significance level 0.05, the actual type I error is 0.00001; both are far from the nominal level 0.05. Such examples illustrate that dependence must be accounted for in order to carry out correct inference.

REFERENCES

BERKES, I., LIU, W. & WU, W. B. (2014). Komlós-Major-Tusnády approximation under dependence. *Ann. Probab.* **42**, 794-817.

BENJAMINI, Y. & HOCHBERG, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* B **57**, 289–300.

BICKEL, P. & ROSENBLATT, M. (1973). On some global measures of the deviations of density function estimates. *Ann. Statist* **1**, 1071–1095.

BICKEL, P., BOLEY, N., BROWN, B., HUANG, H. & ZHANG, N. R. (2010). Non-parametric methods for genomic inference. *Ann. Appl. Stat.* **6**, 1660–1697.

CLARKES, S. & HALL, P. (2009). Robustness of multiple testing procedures against dependence. *Ann. Statist.* **37**, 332–358.

FREEDMAN, D. A. (1975). On tail probabilities for martingales. *Ann. Probab.* **3**, 100-118.

KOMLÓS, J., MAJOR, P. & TUSNÁDY, G. (1975). An approximation of partial sums of independent RV'-s, and the sample DF. I. *Probab. Theory Relat. Fields* **32**, 111-131.

KOMLÓS, J., MAJOR, P. & TUSNÁDY, G. (1976). An approximation of partial sums of independent RV'-s, and the sample DF. II. *Probab. Theory Relat. Fields* **34**, 33-58.

ROSENBLATT, M. (1952). Remarks on multivariate transformation. *Ann. Math. Statist.* **23**, 470-472.

WU, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proc. Natl. Acad. Sci.* **102**, 14150–14154.

ZHANG, C., FAN, J. & YU, T. (2011). Multiple testing via FDR_L for large-scale imaging data. *Ann. Statist.* **39**, 613–642.

ZHAO, Z. & WU, W. B. (2007). Inference of trends in time series. *J. R. Stat. Soc.* R **69**, 391-410.