Biometrika (2015), **xx**, x, *pp*. 1–7

Printed in Great Britain

Supplementary material for "Outlier detection for high dimensional data"

BY KWANGIL RO, CHANGLIANG ZOU, ZHAOJUN WANG
Institute of Statistics, Nankai University, Tianjin, China, 300071
rokwangil@yahoo.com.cn, nk.chlzou@gmail.com, zjwang@nankai.edu.cn

AND GUOSHENG YIN
Department of Statistics and Actuarial Science, The University of Hong Kong Pokfulam Road, Hong Kong
gyin@hku.hk

Proof of Theorem 1

Similar to the argument in Agulló et al. (2008), which is concerned with the least trimmed squares estimator for the multivariate regression, we first prove that $\varepsilon_n(\hat{\mu}_{\text{MDP}}, \mathcal{Y}) \geq \min\{n-h+1, h-m(\mathcal{Y})\}/n$. We show that there exists a value M, which only depends on $\mathcal{Y}$, such that for every $\mathcal{Y}'$ obtained by replacing at most $\min\{n-h+1, h-m(\mathcal{Y})\}-1$ observations in $\mathcal{Y}$ we have $\|\hat{\mu}'_{\text{MDP}}\| \leq M$, where $\hat{\mu}'_{\text{MDP}}$ is the MDP estimator based on $\mathcal{Y}'$.

Let J be a subset of size $m(\mathcal{Y})+1$. It follows from the definition of $m(\mathcal{Y})$ that for any $1 \leq k \leq p$,

$$c_k(J) = \frac{1}{m(\mathcal{Y})+1} \sum_{j \in J} \{y_{jk} - \hat{\mu}_k(J)\}^2 > 0.$$

Denote $c_{\min} = \min_{1 \leq k \leq p} \min_J c_k(J) > 0$. Let $N = \max_{1 \leq i \leq n, 1 \leq k \leq p} |y_{ik}|$ and

$$M = p^{1/2} \left[\left\{ \left(\frac{4N^2 h}{m(\mathcal{Y})+1} \right)^p c_{\min}^{1-p} \right\}^{1/2} + N \right].$$

If we take any dataset $\mathcal{Y}'$ by replacing $\min\{n-h+1, h-m(\mathcal{Y})\}-1$ observations in $\mathcal{Y}$, there exists a subset $H_1 \in \mathcal{H}$ containing indices only corresponding to the data points of the original dataset $\mathcal{Y}$. Then,

$$\begin{aligned}
\det[\text{diag}\{\hat{\Sigma}(H_1)\}] &= \prod_{k=1}^p \left[\frac{1}{h} \sum_{j \in H_1} \{y_{jk} - \hat{\mu}_k(H_1)\}^2 \right] \\
&\leq \left[\frac{1}{hp} \sum_{k=1}^p \sum_{j \in H_1} \{y_{jk} - \hat{\mu}_k(H_1)\}^2 \right]^p \\
&\leq (4N^2)^p.
\end{aligned}$$

Suppose that $\|\hat{\mu}'_{\text{MDP}}\| > M$ and let H_2 be the optimal subset corresponding to $\hat{\mu}'_{\text{MDP}}$ such that $\hat{\mu}'_{\text{MDP}} = \hat{\mu}'(H_2)$ where $\hat{\mu}'(H_2) = h^{-1} \sum_{j \in H_2} Y'_j$. Since $h - [\min\{n-h+1, h-m(\mathcal{Y})\}-1] \geq m(\mathcal{Y})+1$, the set H_2 contains a subset J_0 of size $m(\mathcal{Y})+1$ corresponding to the original observations of $\mathcal{Y}$. Thus we have

$$\begin{aligned}\det[\mathrm{diag}\{\hat{\Sigma}(H_2)\}] &= \prod_{k=1}^{p}\left[\frac{1}{h}\sum_{j\in H_2}\{y_{jk}-\hat{\mu}'_k(H_2)\}^2\right]\\ &\geq \prod_{k=1}^{p}\left[\frac{1}{h}\sum_{j\in J_0}\{y_{jk}-\hat{\mu}'_k(H_2)\}^2\right]\\ &\geq \prod_{k=1}^{p}\left[\frac{1}{h}\sum_{j\in J_0}\{y_{jk}-\hat{\mu}_k(J_0)\}^2+\frac{m(\mathcal{Y})+1}{h}\{|\hat{\mu}_k(J_0)|-|\hat{\mu}'_k(H_2)|\}^2\right].\end{aligned}$$

It follows from $\|\hat{\mu}'(H_2)\| > M$ that there exists k_0 such that $|\hat{\mu}'_{k_0}(H_2)| > M/p^{1/2}$. Thus,

$$\Big|\,|\hat{\mu}_{k_0}(J_0)| - |\hat{\mu}'_{k_0}(H_2)|\,\Big| > \frac{M}{p^{1/2}} - N > 0,$$

and

$$\begin{aligned}\det[\mathrm{diag}\{\hat{\Sigma}(H_2)\}] &> \left\{\frac{m(\mathcal{Y})+1}{h}\right\}^p \prod_{\substack{1\leq k\leq p\\ k\neq k_0}}\left[\frac{1}{m(\mathcal{Y})+1}\sum_{j\in J_0}\{y_{jk}-\hat{\mu}_k(J_0)\}^2\right]\left(\frac{M}{p^{1/2}}-N\right)^2\\ &\geq \left\{\frac{m(\mathcal{Y})+1}{h}\right\}^p c_{\min}^{p-1}\left(\frac{M}{p^{1/2}}-N\right)^2\\ &= (4N^2)^p\end{aligned}$$

by the definition of M. This implies $\det[\mathrm{diag}\{\hat{\Sigma}(H_2)\}] > \det[\mathrm{diag}\{\hat{\Sigma}(H_1)\}]$, which contradicts the definition of $\hat{\mu}'_{\mathrm{MDP}}$, so we conclude that $\|\hat{\mu}'_{\mathrm{MDP}}\| \leq M$.

On the other hand, to show $\varepsilon_n(\hat{\mu}_{(\mathrm{MDP})}, \mathcal{Y}) \leq \min\{n-h+1, h-m(\mathcal{Y})\}/n$, we first prove that $\varepsilon_n(\hat{\mu}_{\mathrm{MDP}}, \mathcal{Y}) \leq (n-h+1)/n$. If we replace $n-h+1$ data points of $\mathcal{Y}$, then the optimal subset H_2 of $\mathcal{Y}'$ would contain at least one outlier, but the least squares method breaks down even with one single outlier. It then follows that $\|\hat{\mu}'(H_2)\|$ is not bounded.

To show that $\varepsilon_n(\hat{\mu}_{\mathrm{MDP}}, \mathcal{Y}) \leq (h-m(\mathcal{Y}))/n$, let $\bar{J} \subset \{1, 2, \ldots, n\}$ be the set of indices corresponding to the largest subset of $\mathcal{Y}$ satisfying that all the elements are the same with respect to at least one component. Thus, we have $|\bar{J}| = m(\mathcal{Y})$. Without loss of generality, suppose the first component y_{j1}, $j \in \bar{J}$, equals to a constant B, and we replace $h - m(\mathcal{Y})$ other observations of $\mathcal{Y}$ by those with the first component B. Denote H_2 as the set of indices corresponding to the observations of $\mathcal{Y}'$ of which the first component is B. Then, we have $\det[\mathrm{diag}\{\hat{\Sigma}'(H_2)\}] = 0$ but $\|\hat{\mu}'(H_2)\|$ is not bounded, because the $h - m(\mathcal{Y})$ contaminated data points in H_2 may have arbitrarily large norms.

Proof of Theorem 2

If $\det(D_2) = 0$, Theorem 2 clearly holds. We consider $\det(D_2) > 0$, and compute the distance based on T_2 and D_2, $d_i(T_2, D_2)$, for $i = 1, \ldots, n$. Then we have

$$\begin{aligned}\frac{1}{hp}\sum_{i\in H_2} d_i^2(T_2, D_2) &= (hp)^{-1}\mathrm{tr}\left\{\sum_{i\in H_2}(Y_i - T_2)^{\mathsf{T}} D_2^{-1}(Y_i - T_2)\right\}\\ &= \frac{1}{hp}\mathrm{tr}\left\{\sum_{i\in H_2} D_2^{-1/2}(Y_i - T_2)(Y_i - T_2)^{\mathsf{T}} D_2^{-1/2}\right\}\\ &= \frac{1}{p}\mathrm{tr}\left(D_2^{-1/2} S_2 D_2^{-1/2}\right)\\ &= \frac{1}{p}\mathrm{tr}(R_2)\\ &= 1. \end{aligned} \tag{S.1}$$

Moreover, we have

$$\frac{1}{hp}\sum_{i\in H_2} d_i^2(T_1, D_1) = \frac{1}{hp}\sum_{i=1}^{h} d_{(i)}^2(T_1, D_1) \le \frac{1}{hp}\sum_{i\in H_1} d_i^2(T_1, D_1) = 1. \tag{S.2}$$

If we take $\lambda = (hp)^{-1}\sum_{i\in H_2} d_i^2(T_1, D_1)$, then $\lambda > 0$ because $\det(D_2) > 0$. Considering the distance based on T_1 and λD_1, $d_i(T_1, \lambda D_1)$, it follows from (S.1) and (S.2) that

$$\begin{aligned}\frac{1}{hp}\sum_{i\in H_2} d_i^2(T_1, \lambda D_1) &= \frac{1}{hp}\mathrm{tr}\left\{\sum_{i\in H_2}(Y_i - T_1)^{\mathsf{T}}\frac{1}{\lambda}D_1^{-1}(Y_i - T_1)\right\}\\ &= \frac{1}{\lambda hp}\mathrm{tr}\left\{\sum_{i\in H_2} d_i^2(T_1, D_1)\right\}\\ &= 1.\end{aligned}$$

Similar to Grübel (1988), we can show that (T_2, D_2) is the unique minimizer of $\det(D)$ among all (T, D) for which $(hp)^{-1}\sum_{i\in H_2} d_i(T, D) = 1$. Therefore, we have

$$\det(D_2) \le \det(\lambda D_1).$$

On the other hand, from the inequality (S.2), we have

$$\det(D_2) \le \det(\lambda D_1) \le \det(D_1).$$

If $\det(D_2) = \det(D_1)$, then $\det(D_2) = \det(\lambda D_1)$, and it implies that $(T_2, D_2) = (T_1, \lambda D_1)$. From $\det(\lambda D_1) \le \det(D_1)$, we know $\lambda = 1$, and thus $(T_2, D_2) = (T_1, D_1)$.

Proof of Proposition 1

The assertion can be shown by verifying

$$A_1 = \max_{1\le i\le n}\left|(Y_i - \hat{\mu})^{\mathsf{T}}\hat{D}^{-1}(Y_i - \hat{\mu}) - (Y_i - \hat{\mu})^{\mathsf{T}} D^{-1}(Y_i - \hat{\mu})\right| = o_p[\{\mathrm{tr}(R^2)\}^{1/2}],$$

and

$$A_2 = \max_{1\le i\le n} \left|(Y_i - \hat{\mu})^{\mathsf{T}} D^{-1}(Y_i - \hat{\mu}) - (Y_i - \mu)^{\mathsf{T}} D^{-1}(Y_i - \mu)\right| = o_p[\{\mathrm{tr}(R^2)\}^{1/2}].$$

First, let $U = \mathrm{diag}(u_1, \ldots, u_p) = \mathrm{diag}\{n^{1/2}(s_{ii}/\sigma_{ii} - 1)\}$ and $V = (s_{ii}/\sigma_{ii} - 1)(\sigma_{ii}/s_{ii} - 1)$. Then we have

$$\begin{aligned} A_1 &= \max_{1\le i\le n} \left|(Y_i - \hat{\mu})^{\mathsf{T}} (\hat{D}^{-1} - D^{-1})(Y_i - \hat{\mu})\right| \\ &= \max_{1\le i\le n} \left|(Y_i - \hat{\mu})^{\mathsf{T}} D^{-1}(-n^{-1/2}U + V)(Y_i - \hat{\mu})\right| \\ &\le A_{11} + A_{12}, \end{aligned}$$

where

$$\begin{aligned} A_{11} &= \max_{1\le i\le n} \left|n^{-1/2}(Y_i - \hat{\mu})^{\mathsf{T}} D^{-1} U (Y_i - \hat{\mu})\right|, \\ A_{12} &= \max_{1\le i\le n} \left|(Y_i - \hat{\mu})^{\mathsf{T}} D^{-1} V (Y_i - \hat{\mu})\right|. \end{aligned}$$

Note that

$$\begin{aligned} A_{11} &\le 2 \max_{1\le i\le n} \left|n^{-1/2}(Y_i - \mu)^{\mathsf{T}} D^{-1} U (Y_i - \mu) + n^{-1/2}(\hat{\mu} - \mu)^{\mathsf{T}} D^{-1} U (\hat{\mu} - \mu)\right| \\ &= 2 \max_{1\le i\le n} \left|n^{-1/2}(Y_i - \mu)^{\mathsf{T}} D^{-1} U (Y_i - \mu)\right| + 2\left|n^{-1/2}(\hat{\mu} - \mu)^{\mathsf{T}} D^{-1} U (\hat{\mu} - \mu)\right| \\ &= 2 \max_{1\le i\le n} \left|n^{-1/2}(Y_i - \mu)^{\mathsf{T}} D^{-1} U (Y_i - \mu)\right| + o_p[\{\mathrm{tr}(R^2)\}^{1/2}], \end{aligned}$$

where the last equality follows from equation (3.7) in Srivastava and Du (2008).

Let $x_i = D^{-1/2}(Y_i - \mu)$ $(i = 1, \ldots, n)$, and then $x_i \sim N_p(0, R)$ and

$$\begin{aligned} \max_{1\le i\le n} \left|n^{-1/2}(Y_i - \mu)^{\mathsf{T}} D^{-1} U (Y_i - \mu)\right| &= \max_{1\le i\le n} \left|n^{-1/2} x_i^{\mathsf{T}} U x_i\right| \\ &= \max_{1\le i\le n} \left|n^{-1/2} \sum_{k=1}^{p} u_k x_{ik}^2\right| \\ &\le n^{-1/2} \sum_{k=1}^{p} u_k \max_{1\le i\le n} x_{ik}^2 \\ &= n^{-1/2} O(\log n)\mathrm{tr}(U). \end{aligned}$$

It follows from equations (3.8) and (3.9) in Srivastava and Du (2008) that

$$E\{\mathrm{tr}(U)\} = 0, \quad \mathrm{var}\{\mathrm{tr}(U)\} = 2\mathrm{tr}(R^2).$$

Thus, we have

$$A_{11} \le n^{-1/2} O(\log n) O[\{\mathrm{tr}(R^2)\}^{1/2}] + o_p[\{\mathrm{tr}(R^2)\}^{1/2}] = o_p[\{\mathrm{tr}(R^2)\}^{1/2}],$$

and similarly, $A_{12} = o_p[\{\mathrm{tr}(R^2)\}^{1/2}]$. Therefore, $A_1 = o_p[\{\mathrm{tr}(R^2)\}^{1/2}]$.

Next, we show $A_2 = o_p[\{\mathrm{tr}(R^2)\}^{1/2}]$. Let $z_i = R^{-1/2}D^{-1/2}(Y_i - \mu)$ $(i = 1, \ldots, n)$, then $z_i = (z_{i1}, \ldots, z_{ip})^{\mathsf{T}} \sim N_p(0, I_p)$ and

$$\begin{aligned} A_2 &= \max_{1\le i\le n} \left|(z_i - \bar{z})^{\mathsf{T}} R(z_i - \bar{z}) - z_i^{\mathsf{T}} R z_i\right| \\ &= \max_{1\le i\le n} \left|\sum_{k=1}^{p} \lambda_k(\bar{z}_{\cdot k}^2 - 2z_{ik}\bar{z}_{\cdot k})\right| \\ &\le A_{21} + A_{22}, \end{aligned}$$

where $A_{21} = \sum_{k=1}^{p} \lambda_k \bar{z}_{\cdot k}^2$, $A_{22} = 2\sum_{k=1}^{p} \lambda_k |\bar{z}_{\cdot k}| \max_{1\le i\le n} |z_{ik}|$, $\bar{z} = n^{-1}\sum_{i=1}^{n} z_i$ and $\bar{z}_{\cdot k} = n^{-1}\sum_{i=1}^{n} z_{ik}$.

It is straightforward to see that $A_{21} = o_p[\{\mathrm{tr}(R^2)\}^{1/2}]$ under Condition 3. Under Condition 4, we have

$$A_{22} = 2\sum_{k=1}^{p} \lambda_k |\bar{z}_{\cdot k}| O\{(\log n)^{1/2}\} \le 2\sum_{k=1}^{p} |\bar{z}_{\cdot k}| \max_{1\le k\le p} \lambda_k O\{(\log n)^{1/2}\} = o_p[\{\mathrm{tr}(R^2)\}^{1/2}].$$

Finally, we have that $A_2 = o_p[\{\mathrm{tr}(R^2)\}^{1/2}]$ and this completes the proof of Proposition 1.

Proof of Proposition 2

We first consider the moment generating function,

$$M(t) = E(e^{t^{\mathsf{T}} Y_1} \mid w_1 = 1). \tag{S.3}$$

Assume that $R = P^{\mathsf{T}}\Lambda P$, where $P^{\mathsf{T}}P = I$ and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$, then we have

$$\begin{aligned} M(t) &= \frac{1}{1-\delta} E\{e^{t^{\mathsf{T}} Y_1} I(w_1 = 1)\} \\ &= \frac{1}{1-\delta}\frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \int_{\{(Y-\mu)^{\mathsf{T}}D^{-1}(Y-\mu)\le a_\delta\}} \exp\{t^{\mathsf{T}}Y - (Y-\mu)^{\mathsf{T}}\Sigma^{-1}(Y-\mu)/2\} dY \\ &= \frac{1}{1-\delta}\frac{1}{(2\pi)^{p/2}} e^{t^{\mathsf{T}}\mu + t^{\mathsf{T}}\Sigma t/2} \int_{\{z^{\mathsf{T}}\Lambda z\le a_\delta\}} \exp\{-(z - PR^{1/2}D^{1/2}t)^{\mathsf{T}}(z - PR^{1/2}D^{1/2}t)/2\} dz \\ &= \frac{1}{1-\delta} e^{t^{\mathsf{T}}\mu + t^{\mathsf{T}}\Sigma t/2} F_t(a_\delta), \end{aligned} \tag{S.4}$$

where $z = PR^{-1/2}D^{-1/2}(Y - \mu)$ and $F_t(a)$ is the cumulative distribution function of the non-negative definite quadratic form in non-central normal variables, that is

$$F_t(a) = \mathrm{pr}(Z_\upsilon^{\mathsf{T}}\Lambda Z_\upsilon \le a), \quad Z_\upsilon \sim N(\upsilon, I_p), \quad \upsilon = PR^{1/2}D^{1/2}t.$$

Without loss of generality, we prove the proposition for $y_{11} \mid w_1 = 1$, whose moment generating function is

$$m_1(t_1) = E(e^{t_1 y_{11}} \mid w_1 = 1).$$

In (S.3), let $t = (t_1, 0, \ldots, 0)^{\mathsf{T}}$ with $p - 1$ components of 0, then it follows from (S.4) that

$$m_1(t_1) = \frac{1}{1-\delta} e^{t_1\mu_1 + \sigma_1^2 t_1^2/2} F_{t_1}(a_\delta),$$

where

$$F_{t_1}(a) = \frac{1}{(2\pi)^{p/2}} \int_{\{z^\mathsf{T}\Lambda z \le a\}} \exp\{-(z - t_1 \upsilon_1)^\mathsf{T}(z - t_1\upsilon_1)/2\} dz, \quad \upsilon_1^\mathsf{T}\upsilon_1 = \sigma_{11}.$$

It follows from the Berry-Esseen inequality that

$$\frac{a_\delta - p}{\{2\mathrm{tr}(R^2)\}^{1/2}} = z_\delta + o(1).$$

Then it is straightforward to show that

$$F_{t_1}(a_\delta)\big|_{t_1=0} = \frac{1}{(2\pi)^{p/2}} \int_{\{z^\mathsf{T}\Lambda z \le a_\delta\}} \exp(-z^\mathsf{T}z/2) dz = \mathrm{pr}\{d_i^2(\mu, D) < a_\delta\},$$

$$\begin{aligned}
&\left.\frac{\partial F_{t_1}(a_\delta)}{\partial t_1}\right|_{t_1=0} \\
=&\left.\frac{1}{(2\pi)^{p/2}} \int_{\{z^\mathsf{T}\Lambda z \le a_\delta\}} (\upsilon_1^\mathsf{T} z - t_1 \upsilon_1^\mathsf{T}\upsilon_1) \exp\{-(z - t_1\upsilon_1)^\mathsf{T}(z - t_1\upsilon_1)/2\} dz\right|_{t_1=0} \\
=&\frac{1}{(2\pi)^{p/2}} \int_{\{z^\mathsf{T}\Lambda z \le a_\delta\}} (\upsilon_1^\mathsf{T} z) \exp(-z^\mathsf{T}z/2) dz = 0,
\end{aligned}$$

and

$$\begin{aligned}
&\left.\frac{\partial^2 F_{t_1}(a_\delta)}{\partial t_1^2}\right|_{t_1=0} \\
=&\left.\frac{1}{(2\pi)^{p/2}} \int_{\{z^\mathsf{T}\Lambda z \le a_\delta\}} \{(\upsilon_1^\mathsf{T} z - t_1 \upsilon_1^\mathsf{T}\upsilon_1)^2 - \upsilon_1^\mathsf{T}\upsilon_1\} \exp\{-(z - t_1\upsilon_1)^\mathsf{T}(z - t_1\upsilon_1)/2\} dz\right|_{t_1=0} \\
=&\frac{1}{(2\pi)^{p/2}} \int_{\{z^\mathsf{T}\Lambda z \le a_\delta\}} \left(\sum_{k=1}^p \upsilon_{1k}^2 z_k^2\right) \exp(-z^\mathsf{T}z/2) dz - \sigma_{11}\mathrm{pr}(d_i^2(\mu, D) < a_\delta) \\
=&\sum_{k=1}^p \upsilon_{1k}^2 \left[\Phi\left\{\frac{a_\delta - p}{\sqrt{2\mathrm{tr}(R^2)}}\right\} - 2\phi\left\{\frac{a_\delta - p}{\sqrt{2\mathrm{tr}(R^2)}}\right\}\left\{\frac{\lambda_k}{\sqrt{2\mathrm{tr}(R^2)}} + \frac{a_\delta - p}{\sqrt{2\mathrm{tr}(R^2)}}\frac{\lambda_k^2}{2\mathrm{tr}(R^2)}\right\} + o(1)\right] \\
&- \sigma_{11}\mathrm{pr}\{d_i^2(\mu, D) < a_\delta\}.
\end{aligned}$$

Thus, we have

$$\begin{aligned}
E(y_{11} \mid w_1 = 1) &= \left.\frac{\partial m_1(t_1)}{\partial t_1}\right|_{t_1=0} \\
&= \frac{1}{1-\delta}\left\{\mu_1 F_{t_1}(a_\delta)|_{t_1=0} + \left.\frac{\partial F_{t_1}(a_\delta)}{\partial t_1}\right|_{t_1=0}\right\} \\
&= \mu_1
\end{aligned}$$

and

$$\mathrm{var}(y_{11} \mid w_1 = 1) = \left.\frac{\partial^2 m_1(t_1)}{\partial t_1^2}\right|_{t_1=0} - \mu_1^2 = \sigma_{11} + \frac{1}{1-\delta}\left.\frac{\partial^2 F_{t_1}(a_\delta)}{\partial t_1^2}\right|_{t_1=0}.$$

Table 1. *Average type I errors (%) under Case (I) for various values of p, n^* and α when $n = 200$*

Correlation	p	$n^* = 10$			$n^* = 20$		
		$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$	$\alpha = 1\%$	$\alpha = 5\%$	$\alpha = 10\%$
AR	50	1.7	5.6	10.0	1.4	4.8	8.7
	100	1.5	5.2	9.6	1.0	4.2	8.2
	200	1.2	4.8	9.1	0.8	3.7	7.5
	400	1.0	4.3	8.3	0.6	3.1	6.8
MA	50	1.7	5.4	9.5	1.1	4.1	7.6
	100	1.5	5.1	9.1	1.0	3.8	7.2
	200	1.4	4.8	8.7	0.9	3.6	7.0
	400	1.1	4.3	8.0	0.8	3.2	6.6

AR stands for autoregressive and MA for moving average.

Finally,

$$\begin{aligned}
\mathrm{var}(y_{11} \mid w_1 = 1) &= \frac{1}{1-\delta}\sum_{k=1}^{p} \upsilon_{1k}^2 \left\{(1-\delta) - 2\phi(z_\delta)\left(\frac{\lambda_k}{\sqrt{2\mathrm{tr}(R^2)}} + z_\delta \frac{\lambda_k^2}{2\mathrm{tr}(R^2)}\right) + o(1)\right\} \\
&= \frac{1}{1-\delta}\sum_{k=1}^{p} \upsilon_{1k}^2 \left\{(1-\delta) - 2\phi(z_\delta)\frac{\lambda_k}{\sqrt{2\mathrm{tr}(R^2)}} + o(1)\right\} \\
&= \sigma_{11}\left\{1 - \frac{2\phi(z_\delta)}{1-\delta}\frac{(R^2)_{11}}{\sqrt{2\mathrm{tr}(R^2)}} + o(1)\right\} \\
&= \sigma_{11}\tau_{11},
\end{aligned}$$

which completes the proof.

Additional simulation results

We present additional simulation results based on different configurations to further examine the properties of the proposed method, and also make a comparison with other existing approaches for outlier detection. Similar conclusions can be drawn from these numerical studies: The proposed procedure maintains the test size and also possesses substantial power for outlier detection with high dimensional-data, while other methods often fail in the aspect of either type I error rates or type II error rates. The patterns of our findings are more prominent as p increases.

References

AGULLÓ, J., CROUX, C. & VAN AELST, S. (2008). The multivariate least-trimmed squares estimator. *J. Multivar. Anal.* **99**, 311–338.

GRÜBEL, R. (1988). A minimal characterization of the covariance matrix. *Metrika* **35**, 49–52.

SRIVASTAVA, M. S. & DU, M. (2008). A test for the mean vector with fewer observations than the dimension. *J. Multivar. Anal.* **99**, 386–402.

Table 2. *Average type I (α) and type II (β) errors under Cases (I)–(II) for various values of p with a nominal size of $\alpha = 0.05$, when $n = 100$ and $n^* = 20$*

Case	Correlation	p	R-MDP		PCOut		R-MCD		SDM	
			α	β	α	β	α	β	α	β
(I)	AR	50	5.4	0.4	2.8	1.1	1.5	21.2	6.7	0.1
		100	5.0	2.3	2.2	5.3	7.0	8.5	8.8	0.6
		200	4.8	10.4	2.4	10.0	10.9	18.8	6.1	3.5
		400	4.3	29.1	2.4	17.0	13.8	31.5	10.9	6.3
	MA	50	5.0	38.9	5.6	39.2	10.1	28.8	25.2	6.3
		100	4.6	26.3	4.5	39.3	7.9	9.3	16.4	2.8
		200	4.1	13.8	3.4	37.5	33.7	0.2	27.6	0.6
		400	3.6	6.1	2.6	18.5	30.4	0.0	34.8	0.0
(II)	AR	50	5.3	0.0	3.9	0.6	3.1	2.5	3.6	3.9
		100	4.9	0.6	5.1	55.3	9.4	0.0	4.0	8.0
		200	4.7	5.9	6.5	82.5	6.3	2.7	7.0	11.3
		400	4.3	24.7	7.4	88.5	15.3	6.7	8.7	16.7
	MA	50	5.2	30.2	4.7	3.1	5.1	8.2	10.2	0.6
		100	4.7	14.6	3.7	6.3	24.9	0.0	20.5	0.3
		200	4.5	4.7	3.5	33.4	32.4	0.0	5.8	0.1
		400	3.8	0.7	4.5	58.7	15.7	0.0	35.7	0.2

R-MDP: our refined minimum diagonal product method; PCOut: the principal component outlier detection procedure by Filzmoser et al. (2008); R-MCD: the regularized minimum covariance determinant method by Fritsch et al. (2011); and SDM: first constructing the initial subset based on the Stahel–Donoho outlyingness and then applying R-MCD.

Table 3. *Average type I (α) and type II (β) errors under Case (III) with $\psi = 2$, when $n = 100$ and $n^* = 20$*

p	R-MDP		PCOut		R-MCD		SDM	
	α	β	α	β	α	β	α	β
50	5.1	10.6	5.8	40.4	2.0	15.4	7.0	3.1
100	5.1	0.8	4.3	52.2	7.3	0.2	7.3	0.2
200	4.7	0.0	3.0	40.3	13.6	0.0	7.8	0.1
400	4.5	0.0	2.1	15.2	16.3	0.0	10.2	0.0