Biometrika (2016), **xx**, x, *pp*. 1–13

Printed in Great Britain

Supplementary material for "Projection-based outlier detection in functional data"

BY HAOJIE REN

Institute of Statistics, Nankai University, No.94 Weijin Road, Tianjin, China, 300071

haojieren@gmail.com

NAN CHEN

Department of Industrial & Systems Engineering, National University of Singapore, 1 Engineering Drive 2, Singapore 117576

isecn@nus.edu.sg

AND CHANGLIANG ZOU

Institute of Statistics, Nankai University, No.94 Weijin Road, Tianjin, China, 300071

nk.chlzou@gmail.com

TECHNICAL PROOFS

Proof of Theorem 1

Similar to the argument in Ro et al. (2015), which is concerned with the outlier detection in high-dimensional observations, we first prove that $b_{N,\hat{v}_R}(\hat{\mu}_{H_{\mathrm{L}}}, \mathcal{X}) \geq \min(N-h+1, h)/N$. We show that there exists a value M, which only depends on $\hat{v}_R$, such that for every $\mathcal{X}'$ obtained by replacing at most $\min(N-h+1,h)-1$ observations in $\mathcal{X}$ we have $\|\hat{\mu}'_{H_{\mathrm{L}}}\|_{\hat{v}_R} \leq M$, where $\hat{\mu}'_{H_{\mathrm{L}}}$ is the least trimmed functional scores estimator based on $\mathcal{X}'$. If we take any dataset $\mathcal{X}'$ by replacing $\min(N-h+1,h)-1$ observations in $\mathcal{X}$, there exists a subset $H_1 \in \mathcal{H}$ containing indices only corresponding to the data points of the original dataset $\mathcal{X}$. Then

$$\begin{aligned}\sum_{i\in H_1} D_i(H_1) &= \sum_{i\in H_1}\sum_{1\leq k\leq d}\left[\int_a^b \{X_i(t)-\hat{\mu}_{H_1}(t)\}\hat{v}_{kR}(t)dt\right]^2/\hat{\lambda}_{kR}\\ &\leq \sum_{i\in H_1}\sum_{1\leq k\leq d} 2\left\{\int_a^b X_i^2(t)dt+\int_a^b \hat{\mu}_{H_1}^2(t)dt\right\}/\hat{\lambda}_{kR}\\ &\leq 4hM_1\sum_{1\leq k\leq d}\hat{\lambda}_{kR}^{-1},\end{aligned}$$

where $M_1 = \max_{1\leq i\leq N}\int_a^b X_i^2(t)dt$ and the last inequality holds by using Cauchy-Schwartz inequality and the fact $\int_a^b \hat{\mu}_{H_1}^2(t)dt \leq M_1$.

Suppose that $\|\hat{\mu}'_{H_{\mathrm{L}}}\|_{\hat{v}_{lR}} = M$ for some l and let H_2 be the optimal subset corresponding to $\hat{\mu}'_{H_{\mathrm{L}}}$ such that $\hat{\mu}'_{H_{\mathrm{L}}} = \hat{\mu}'(H_2)$ where $\hat{\mu}'(H_2) = h^{-1}\sum_{j\in H_2} X'_j$. Since $h-\{\min(N-h+1,h)-1\} \geq 1$, the set

H_2 contains a subset J_0 of size $|J_0| \geq 1$ corresponding to the original observations of $\mathcal{X}$. Thus we have

$$\begin{aligned}
\sum_{i\in H_2} D_i(H_2) &= \sum_{1\leq k\leq d} \hat{\lambda}_{kR}^{-1} \sum_{i\in H_2} \left[\int_a^b \{X_i(t) - \hat{\mu}'_{H_2}(t)\}\hat{v}_{kR}(t)dt\right]^2 \\
&\geq \sum_{1\leq k\leq d} \hat{\lambda}_{kR}^{-1} \sum_{i\in J_0} \left[\int_a^b \{X_i(t) - \hat{\mu}'_{H_2}(t)\}\hat{v}_{kR}(t)dt\right]^2 \\
&= \sum_{1\leq k\leq d} \hat{\lambda}_{kR}^{-1} \sum_{i\in J_0} \left[\int_a^b \{X_i(t) - \hat{\mu}_{J_0}(t)\}\hat{v}_{kR}(t)dt\right]^2 \\
&\quad + \sum_{1\leq k\leq d} \hat{\lambda}_{kR}^{-1}|J_0| \left[\int_a^b \{\hat{\mu}'_{H_2}(t) - \hat{\mu}_{J_0}(t)\}\hat{v}_{kR}(t)dt\right]^2 \\
&\geq |J_0|\hat{\lambda}_{lR}^{-1} \left\{\int_a^b \hat{\mu}'_{H_2}(t)\hat{v}_{lR}(t)dt\right\}^2 - 2|J_0|\hat{\lambda}_{lR}^{-1} \int_a^b \hat{\mu}'_{H_2}(t)\hat{v}_{lR}(t)dt \int_a^b \hat{\mu}_{J_0}(t)\hat{v}_{lR}(t)dt \\
&\geq \hat{\lambda}_{lR}^{-1}(M^2 - 2hMM_1^{1/2})
\end{aligned}$$

by the definition of M. This implies $\sum_{i\in H_2} D_i(H_2) > \sum_{i\in H_1} D_i(H_1)$ provided $M > hM_1^{1/2} + (h^2M_1 + h\hat{\lambda}_{lR}M_1\sum_{k=1}^d \hat{\lambda}_{kR}^{-1})^{1/2}$, which contradicts the definition of $\hat{\mu}'_{H_L}$. So we conclude that $\|\hat{\mu}'_{H_L}\|_{\hat{v}_{lR}} \leq hM_1^{1/2} + (h^2M_1 + h\hat{\lambda}_{lR}M_1\sum_{k=1}^d \hat{\lambda}_{kR}^{-1})^{1/2}$.

On the other hand, to show $b_{N,\hat{v}_R}(\hat{\mu}_{H_L}, \mathcal{X}) \leq \min(N-h+1, h)/N$, we first prove that $b_{N,\hat{v}_R}(\hat{\mu}_{H_L}, \mathcal{X}) \leq (N-h+1)/N$. If we replace $N-h+1$ data points of $\mathcal{X}$, then the optimal subset H_2 of $\mathcal{X}'$ would contain at least one outlier, but the least squares method breaks down even with one single outlier. It then follows that $\|\hat{\mu}_{H_L}\|_{\hat{v}_R}$ is not bounded because we can simple replace the observation $X_i(t)$ by $X'_i(t) = K\hat{v}_{kR}(t)$ so that $|\int_a^b X'_i(t)\hat{v}_{kR}(t)dt| = |K|$ which can be arbitrarily large. Similarly, we can easily observe that $b_{N,\hat{v}_R}(\hat{\mu}_{H_L}, \mathcal{X}) \leq h/N$. □

Proof of Theorem 2

First, we have $\sum_{i=1}^h D_{(i)}(H_2) \leq \sum_{i\in H_2} D_i(H_2)$, as $D_{(i)}(H_2)$ $(i = 1, \ldots, h)$ is the h smallest distances based on H_2. Then, we can easily find that

$$\begin{aligned}
\sum_{i\in H_2} D_i(H_2) &= \sum_{i\in H_2} \sum_{1\leq k\leq d} \hat{\eta}_{ik}^2(H_2)/\hat{\lambda}_{kR} \\
&= \sum_{i\in H_2} \sum_{1\leq k\leq d} \left[\int_a^b \{X_i(t) - \hat{\mu}_{H_1}(t) + \hat{\mu}_{H_1}(t) - \hat{\mu}_{H_2}(t)\}\hat{v}_{kR}(t)dt\right]^2 /\hat{\lambda}_{kR} \\
&= \sum_{i\in H_2} D_i(H_1) - h\sum_{1\leq k\leq d} \left[\int_a^b \{\hat{\mu}_{H_1}(t) - \hat{\mu}_{H_2}(t)\}\hat{v}_{kR}(t)dt\right]^2 /\hat{\lambda}_{kR} \\
&\leq \sum_{i\in H_2} D_i(H_1) = \sum_{i=1}^h D_{(i)}(H_1),
\end{aligned}$$

from which the theorem follows immediately. □

Before proving Theorem 3 and Proposition 1, we present the following lemma.

LEMMA 1. *Assume that $\hat{\mu}(\cdot)$, $\hat{\lambda}_k$ and $\hat{v}_k(\cdot)$ are $\sqrt{\zeta_N}$-consistent estimates of $\mu_0(\cdot)$, λ_k and $v_k(\cdot)$. Then $|T_i(\hat{\mu}, \hat{v}, \hat{\lambda}) - \sum_{k=1}^{d} \xi_{ik}^2| = O_p(\zeta_N^{-1/2} \log N)$, uniformly in i.*

Proof. Without loss of generality, assume that $\mu_0(t) = 0$. Recall that $\varepsilon_i(t) = \sum_{1 \le k < \infty} \lambda_k^{1/2} \xi_{ik} v_k(t)$ and $\xi_{ik} = \lambda_k^{-1/2} \int \varepsilon_i(t) v_k(t) dt$, $(i = 1, \ldots, N;\ k = 1, 2, \ldots)$ are independent identically distributed standard normal random variables. It follows that for each k, $\max_{1 \le i \le N} \xi_{ik}^2 = O_p(\log N)$. We denote

$$\phi_{ik} = \int_a^b \{X_i(t) - \hat{\mu}(t)\} v_k(t) dt,$$
$$\hat{\eta}_{ik} = \int_a^b \{X_i(t) - \hat{\mu}(t)\} \hat{v}_k(t) dt,\ (k = 1, \ldots, d).$$

Observe that

$$\max_i \hat{\eta}_{ik}^2 \le \max_i \|\varepsilon_i(t) - \hat{\mu}(t)\|^2$$
$$\le 2 \max_i \{\|\varepsilon_i(t)\|^2 + ||\hat{\mu}(t)||^2\} = 2 \sum_{k=1}^{\infty} \lambda_k \max_i \xi_{ik}^2 + 2||\hat{\mu}(t)||^2.$$

By assuming that $||\hat{\mu}(t) - \mu_0(t)||^2 = o_p(1)$, we have $||\hat{\mu}(t)||^2 = o_p(1)$. Also, $\sum_{k=1}^{\infty} \lambda_k = E(||\varepsilon_i(t)||^2) < \infty$, we can claim that $\max_i ||\varepsilon_i(t)||^2 = O_p(\log N)$ based on Markov inequality and then $\max_i \hat{\eta}_{ik}^2 = O_p(\log N)$. Combining the expressions above, we have

$$\left| \sum_{k=1}^{d} \max_i \frac{\hat{\eta}_{ik}^2}{\hat{\lambda}_k} - \sum_{k=1}^{d} \max_i \frac{\hat{\eta}_{ik}^2}{\lambda_k} \right| \le \sum_{k=1}^{d} \max_i \hat{\eta}_{ik}^2 \left| \hat{\lambda}_k^{-1} - \lambda_k^{-1} \right| = O_p(\zeta_N^{-1/2} \log N),$$

where we use the inequality $|\max_i |a_i| - \max_i |b_i|| \le \max_i |a_i - b_i|$.

Furthermore, we have

$$\max_i \left| \sum_{k=1}^{d} \frac{\hat{\eta}_{ik}^2}{\lambda_k} - \sum_{k=1}^{d} \frac{\phi_{ik}^2}{\lambda_k} \right| \le \max_i \sum_{k=1}^{d} \frac{1}{\lambda_k} |\hat{\eta}_{ik} - \hat{c}_k \phi_{ik}| (|\hat{\eta}_{ik}| + |\phi_{ik}|)$$
$$\le \sum_{k=1}^{d} \frac{1}{\lambda_k} \|\hat{v}_k(t) - \hat{c}_k v_k(t)\| \cdot 2 \max_i \|\varepsilon_i(t) - \hat{\mu}(t)\|^2$$
$$= O_p(\zeta_N^{-1/2} \log N),$$

by the consistency assumption of $\hat{v}_k(t)$, where $\hat{c}_k = \text{sgn}\left\{ \int_a^b v_k(t) \hat{v}_k(t) dt \right\}$.

Finally, we have

$$\max_i \left| \sum_{k=1}^{d} \frac{\phi_{ik}^2}{\lambda_k} - \sum_{k=1}^{d} \xi_{ik}^2 \right| \le 2 \max_i \sum_{k=1}^{d} \|\varepsilon_i(t)\|_{v_k} \|\hat{\mu}(t)\|_{v_k} + \sum_{k=1}^{d} \frac{1}{\lambda_k} \|\hat{\mu}(t)\|_{v_k}^2$$
$$\le 2\|\hat{\mu}(t)\| \sum_{k=1}^{d} \max_i \|\varepsilon_i(t)\|_{v_k} + \sum_{k=1}^{d} \frac{1}{\lambda_k} \|\hat{\mu}(t)\|^2 = o_p(\zeta_N^{-1/2} \log N),$$

from which the lemma follows. □

Proof of Theorem 3

(i) Denote $H = H_O \cup \bar{H}_O$, where H_O is the outlier subset of size m_N and $\bar{H}_O$ is its complement. Construct $H' = \bar{H}_O \cup J$, where J contains m_N normal observations which are arbitrarily taken from $\mathcal{X} \setminus \mathcal{O}_N \setminus \bar{H}_O$. We can assume that $\mu_0(t) \equiv 0$ without loss of generality. Let $\delta_{lL} \equiv \min_{1\le i\le n} \Delta_{il}^2 > 0$ for some l. Then

$$
\begin{aligned}
&\sum_{i\in H}\left[\int_a^b \{X_i(t)-\hat{\mu}_H(t)\}v_l(t)dt\right]^2 - \sum_{i\in H'}\left[\int_a^b \{X_i(t)-\hat{\mu}_{H'}(t)\}v_l(t)dt\right]^2 \\
&\ge \left[\sum_{i\in H_O}\left\{\int_a^b \mu_i(t)v_l(t)dt\right\}^2 - h^{-1}\left\{\int_a^b \sum_{i\in H_O}\mu_i(t)v_l(t)dt\right\}^2\right] \\
&\quad + \left[\sum_{i\in H_O}\left\{\int_a^b \varepsilon_i(t)v_l(t)dt\right\}^2 - \sum_{i\in J}\left\{\int_a^b \varepsilon_i(t)v_l(t)dt\right\}^2\right] \\
&\quad + 2\sum_{i\in H_O}\left\{\int_a^b \varepsilon_i(t)v_l(t)dt\right\}\left\{\int_a^b \mu_i(t)v_l(t)dt\right\} \\
&\quad - h^{-1}\left\{\int_a^b \sum_{i\in H}\varepsilon_i(t)v_l(t)dt\right\}^2 - 2h^{-1}\left\{\int_a^b \sum_{i\in H}\varepsilon_i(t)v_l(t)dt\right\}\left\{\int_a^b \sum_{i\in H_O}\mu_i(t)v_l(t)dt\right\} \\
&\equiv M_l + R_{1l} + R_{2l} - R_{3l} - R_{4l}.
\end{aligned}
$$

Using the fact that $h^{-1}m_N \le h^{-1}|\mathcal{O}_N| \approx 2|\mathcal{O}_N|/N \to 2\rho$ and

$$
h^{-1}\left\{\int_a^b \sum_{i\in H_O}\mu_i(t)v_l(t)dt\right\}^2 \le \frac{m_N}{h}\sum_{i\in H_O}\left\{\int_a^b \mu_i(t)v_l(t)dt\right\}^2,
$$

we can see $M_l \ge (1-m_N/h)\sum_{i\in H_O}\left\{\int_a^b \mu_i(t)v_l(t)dt\right\}^2 \ge m_N\delta_{lL}(1-2\rho)$.

For any $\epsilon > 0$, by Hoeffding inequality for Sub-Gaussian variables,

$$
\begin{aligned}
&\mathrm{pr}(R_{3l}/\lambda_l > \epsilon) = \mathrm{pr}\left\{\left(h^{-1/2}\sum_{i\in H}\xi_{il}\right)^2 > \epsilon\right\} \le 2\exp\left(-\epsilon^2/2\right), \\
&\mathrm{pr}(R_{2l}/\lambda_l > \epsilon) = \mathrm{pr}\left(2\sum_{i\in H_O}\xi_{il}\Delta_{il} > \lambda_l^{1/2}\epsilon\right) \le \exp\left(-\frac{\lambda_l\epsilon^2}{8\sum_{i\in H_O}\Delta_{il}^2}\right) \le \exp\left(-\frac{\lambda_l\epsilon^2}{8m_N\delta_U}\right), \\
&\mathrm{pr}(R_{4l}/\lambda_l > \epsilon) = \mathrm{pr}\left(2\sum_{i\in H}\xi_{il}\sum_{i\in H_O}\Delta_{il} > h\lambda_l^{1/2}\epsilon\right) \le \exp\left(-\frac{h\lambda_l\epsilon^2}{8(\sum_{i\in H_O}\Delta_{il})^2}\right) \le \exp\left(-\frac{\lambda_l\epsilon^2}{8m_N\delta_U}\right),
\end{aligned}
$$

where ξ_{il} $(i = 1, \ldots, N)$ are independent and identically distributed $N(0,1)$ variables. Now, let us deal with R_{1l}. Because R_{1l}/λ_l is equivalent to $\sum_{i=1}^{m_N}(U_{i1} - U_{i2})$, where U_{i1}'s and U_{i2}'s are independent and identically distributed random variables from χ_1^2. $U_{i1} - U_{i2}$ is thus sub-exponential with parameters

$(2\sqrt{2}, 4)$. By one-sided Bernstein's inequality for sub-exponential variables,

$$\mathrm{pr}(R_{1l}/\lambda_l > m_N\epsilon) \le \exp\left(-\frac{m_N\epsilon^2}{8+4\epsilon/3}\right).$$

Combining all the three exponential inequalities for R_{1l}, R_{2l} and R_{3l}, we have

$$\mathrm{pr}\left(|R_{1l}+R_{2l}-R_{3l}| > m_N\delta_{lL}\right) \le \exp\left(-c_1 m_N\right),$$

where c_1 is a constant depending only on δ_L, δ_U and λ_l. Accordingly,

$$\begin{aligned}
&\mathrm{pr}\left\{\sum_{i\in H} D_i(H) - \sum_{i\in H'} D_i(H') < 0\right\}\\
\le&\mathrm{pr}\left\{\sum_{k=1}^{d}\lambda_k^{-1}|R_{1k}+R_{2k}-R_{3k}| > \sum_{k=1}^{d} m_N\delta_{kL}(1-2\rho)/\lambda_k\right\}\\
\le&\sum_{k=1}^{d}\mathrm{pr}\left\{\lambda_k^{-1}|R_{1k}+R_{2k}-R_{3k}| > m_N\delta_{kL}(1-2\rho)/\lambda_k\right\}\\
\le&\exp(-c_2 m_N),
\end{aligned}$$

where c_2 is a constant depending on d, δ_L, δ_U and λ_k's.

(ii) For simplicity, we assume that $d=1$ because the proof for $d>1$ is similar.

$$\hat{\lambda}\sum_{i\in H} D_i(H) = \sum_{i\in \bar{H}_O}\left\{\int_a^b X_i(t)\hat{v}_l(t)dt\right\}^2 + \sum_{i\in H_O}\left\{\int_a^b X_i(t)\hat{v}_l(t)dt\right\}^2 - h\left\{\int_a^b \hat{\mu}_H(t)\hat{v}_l(t)dt\right\}^2.$$

Thus, by Lemma 1

$$\begin{aligned}
\hat{\lambda}\left\{\sum_{i\in H} D_i(H) - \sum_{i\in H'} D_i(H')\right\} =& \sum_{i\in H_O}\left\{\int_a^b X_i(t)\hat{v}_l(t)dt\right\}^2 - \sum_{i\in J}\left\{\int_a^b X_i(t)\hat{v}_l(t)dt\right\}^2\\
&- h\left\{\int_a^b \hat{\mu}_H(t)\hat{v}_l(t)dt\right\}^2 + h\left\{\int_a^b \hat{\mu}_{H'}(t)\hat{v}_l(t)dt\right\}^2\\
\ge& \sum_{i\in H_O}\left[\int_a^b \{\varepsilon_i(t)+\mu_i(t)\}v_l(t)dt\right]^2 - \sum_{i\in J}\left\{\int_a^b \varepsilon_i(t)v_l(t)dt\right\}^2\\
&- h\left\{\int_a^b \hat{\mu}_H(t)\hat{v}_l(t)dt\right\}^2 + O_p(\zeta_N^{-1/2} m_N\log N),
\end{aligned}$$

where we have

$$\begin{aligned}
h\left\{\int_a^b \hat{\mu}_H(t)\hat{v}_l(t)dt\right\}^2 =& h^{-1}\left\{\int_a^b \sum_{i\in H}\varepsilon_i(t)\hat{v}_l(t)dt\right\}^2 + h^{-1}\left\{\int_a^b \sum_{i\in H_O}\mu_i(t)\hat{v}_l(t)dt\right\}^2\\
&+ 2h^{-1}\left\{\int_a^b \sum_{i\in H}\varepsilon_i(t)\hat{v}_l(t)dt\right\}\left\{\int_a^b \sum_{i\in H_O}\mu_i(t)\hat{v}_l(t)dt\right\}.
\end{aligned}$$

It is easy to see

$$h^{-1}\left\{\int_a^b \sum_{i\in H}\varepsilon_i(t)\hat{v}_l(t)dt\right\}^2 = O_p(1),$$

$$h^{-1}\left\{\int_a^b \sum_{i\in H}\varepsilon_i(t)\hat{v}_l(t)dt\right\}\left\{\int_a^b \sum_{i\in H_O}\mu_i(t)\hat{v}_l(t)dt\right\} = O_p(m_N h^{-1/2}),$$

by the weak convergence in $\{D[0,1], \mathcal{L}^2(\mathcal{T})\}$ of the partial sum process $\sum_{i\in H}\varepsilon_i(t)$. Write

$$\begin{aligned}
&\sum_{i\in H_O}\left[\int_a^b \{\varepsilon_i(t)+\mu_i(t)\}v_l(t)dt\right]^2 \\
&\quad = \sum_{i\in H_O}\left\{\int_a^b \varepsilon_i(t)v_l(t)dt\right\}^2 + \sum_{i\in H_O}\left\{\int_a^b \mu_i(t)v_l(t)dt\right\}^2 \\
&\qquad + 2\sum_{i\in H_O}\left\{\int_a^b \varepsilon_i(t)v_l(t)dt\right\}\left\{\int_a^b \mu_i(t)v_l(t)dt\right\},
\end{aligned}$$

where we denote the last term as U. Note that since $E(U)=0$ and $\mathrm{var}(U) \le 4m_N\delta_U E(\|\varepsilon_i(t)\|^2) = O(m_N)$, we have $U = O_p(m_N^{1/2})$. Combining these results together,

$$\begin{aligned}
&\hat{\lambda}\left\{\sum_{i\in H} D_i(H) - \sum_{i\in H'} D_i(H')\right\} \\
&\ge \left[\sum_{i\in H_O}\left\{\int_a^b \varepsilon_i(t)v_R(t)dt\right\}^2 - \sum_{i\in J}\left\{\int_a^b \varepsilon_i(t)v_R(t)dt\right\}^2\right] \\
&\quad + m_N\delta_L(1-2\rho) + O_p(m_N^{1/2} + m_N h^{-1/2} + \zeta_N^{-1/2} m_N \log N) + O_p(1).
\end{aligned}$$

By the fact that the first term in the last inequality can be viewed as m_N independent and identically distributed variables with zero mean and bounded variances, we have as $N\to\infty$,

$$\begin{aligned}
&\mathrm{pr}\left\{\sum_{i=1}^h D_{(i)}(H') > \sum_{i=1}^h D_{(i)}(H)\right\} \\
&\le \mathrm{pr}\left\{\delta_L(1-2\rho) < O_p(m_N^{-1/2} + h^{-1/2} + \zeta_N^{-1/2}\log N) + O_p(m_N^{-1})\right\} \to 0,
\end{aligned}$$

provided that $m_N\to\infty$ and $\zeta_N^{-1/2}\log N \to 0$.

Now, let us prove the second part of (ii). To show the the probability tends to zero uniformly in H, the key step is to construct an appropriate J. Let $y_i = \sum_{k=1}^d \lambda_k^{-1}\left\{\int_a^b \varepsilon_i(t)v_k(t)dt\right\}^2$ for $i\in\mathcal{O}_N$ and $y_{(1)} \le \cdots \le y_{(|\mathcal{O}_N|)}$ be the order statistics of y_i's. Similarly, let $z_i = \sum_{k=1}^d \lambda_k^{-1}\left\{\int_a^b \varepsilon_i(t)v_k(t)dt\right\}^2$ for $i\in\mathcal{X}\setminus\mathcal{O}_N\setminus\bar{H}_O$ and $z_{(1)}\le\cdots\le z_{(k_N)}$ be the order statistics of z_i's, where we denote $k_N = |\mathcal{X}\setminus\mathcal{O}_N\setminus\bar{H}_O|$. Let $J = \{i\in\mathcal{X}\setminus\mathcal{O}_N\setminus\bar{H}_O : z_i \le z_{(m_N)}\}$. Construct $H'_{\mathrm{LTFS}} = \bar{H}_O\cup J$. By similar arguments used in

the proof of the last part and Lemma 1,

$$\begin{aligned}
&\sum_{i\in H_{\mathrm{LTFS}}} D_i(H_{\mathrm{LTFS}}) - \sum_{i\in H'_{\mathrm{LTFS}}} D_i(H'_{\mathrm{LTFS}}) \\
&\geq m_N(1-2\rho)\lambda_1^{-1}\delta_L + \sum_{k=1}^{d}\lambda_k^{-1}\left[\sum_{i\in H_O}\left\{\int_a^b \varepsilon_i(t)v_k(t)dt\right\}^2 - \sum_{i\in J}\left\{\int_a^b \varepsilon_i(t)v_k(t)dt\right\}^2\right] \\
&\quad + 2\sum_{k=1}^{d}\lambda_k^{-1}\sum_{i\in H_O}\left\{\int_a^b \varepsilon_i(t)v_k(t)dt\right\}\left\{\int_a^b \mu_i(t)v_k(t)dt\right\} \\
&\quad - h^{-1}\sum_{k=1}^{d}\lambda_k^{-1}\left\{\sum_{i\in H_{\mathrm{LTFS}}}\int_a^b \varepsilon_i(t)v_k(t)dt\right\}^2 \\
&\quad - 2h^{-1}\sum_{k=1}^{d}\lambda_k^{-1}\left\{\sum_{i\in H_{\mathrm{LTFS}}}\int_a^b \varepsilon_i(t)v_k(t)dt\right\}\left\{\sum_{i\in H_O}\int_a^b \mu_i(t)v_k(t)dt\right\} + O_p(\zeta_N^{-1/2}m_N\log N) \\
&\equiv m_N(1-2\rho)\lambda_1^{-1}\delta_L + R_1 + R_2 - R_3 - R_4 + O_p(\zeta_N^{-1/2}m_N\log N).
\end{aligned}$$

Unlike the first part, we will provide uniform bounds for the R_i's which always hold regardless of H_{LTFS}. First of all,

$$\left|\sum_{i\in H_{\mathrm{LTFS}}}\xi_{ik}\right| \leq \max_{H\in\mathcal{H}}\left|\sum_{i\in H}\xi_{ik}\right| \leq \max\left\{\left|\sum_{j=1}^{h}\xi_{(j)k}\right|, \left|\sum_{j=0}^{h-1}\xi_{(N-j)k}\right|\right\},$$

where $\xi_{(j)k}$'s are the order statistics of $(\xi_{1k},\ldots,\xi_{Nk})$. By the central limit theorem for L statistics (Chernoff et al., 1967),

$$N^{-1/2}\left\{\sum_{j=1}^{h}\xi_{(j)k} - \psi_\Phi(0.5)\right\} \to \mathrm{N}(0,\psi_\sigma^2(0.5))$$

in distribution, where $\psi_\Phi(q) = \int_0^q \Phi^{-1}(z)dz$, $\psi_\sigma^2(q) = \int_0^q \frac{1}{(1-u)^2}\left\{\int_u^q \Phi^{-1}(v)(1-v)dv\right\}^2 du$ and $\Phi(\cdot)$ is the standard normal distribution function (Serfling, 1980). By symmetry, $|\sum_{i\in H_{\mathrm{LTFS}}}\xi_{ik}| \leq O_p(N^{1/2})$, from which we can claim that $R_3 = O_p(1)$. Similarly,

$$R_4 \leq O(h^{-1})O_p(N^{1/2})m_N\delta_U^{1/2}.$$

Hence, with probability tending to one, $R_3 + R_4 < m_N(1-2\rho)\lambda_1^{-1}\delta_L$.

Now, let us deal with R_1 and R_2. We only discuss the case of $\rho > 0$ here but the case of $\rho = 0$ can be more easily handled. Two distinguishing cases are: the intermediate case $m_N/|\mathcal{O}_N| \to 0$ and the central case $m_N/|\mathcal{O}_N| \to q < 1$. In the first case,

$$\begin{aligned}
R_1 &\geq -\sum_{i\in J}\sum_{k=1}^{d}\left\{\int_a^b \lambda_k^{-1/2}\varepsilon_i(t)v_k(t)dt\right\}^2 \\
&\geq -m_N z_{(m_N)} \geq -m_N z_{(\lceil N\beta_1\rceil)},
\end{aligned}$$

for any $0 < \beta_1 < 1$ as long as N is sufficiently large. By Cauchy inequality,

$$\begin{aligned}|R_2| &\le 2\left(m_N\delta_U\lambda_d^{-1}\right)^{1/2}\sum_{k=1}^{d}\left(\sum_{i\in H_O}\xi_{ik}^2\right)^{1/2}\\ &\le 2m_N\left(\delta_U\lambda_d^{-1}\right)^{1/2}\sum_{k=1}^{d}|\xi|_{(m_N)k}\\ &\le 2m_N d\left(\delta_U\lambda_d^{-1}\right)^{1/2}|\xi|_{(\lceil N\beta_2\rceil)}, \end{aligned} \tag{S.1}$$

for any $0 < \beta_2 < 1$ as long as N is sufficiently large, where $|\xi|_{(j)k}$ are the order statistics of $(|\xi|_{1k},\ldots,|\xi|_{|\mathcal{O}_N|k})$. By the convergence of sample quantiles, we know that $z_{(\lceil N\beta_1\rceil)} \to c_{\beta_1}$ and $\xi_{\lceil N\beta_2\rceil} \to c_{\beta_2}$ in probability, where c_{β_1} and c_{β_2} are the β_1th and β_2th quantiles of χ_d^2 and absolute normal distributions, respectively. We can choose β_1 and β_2 so that $c_{\beta_1} + 2d(\delta_U\lambda_d^{-1})^{1/2}c_{\beta_2} < (1-2\rho)\lambda_1^{-1}\delta_L$. Thus, when $m_N/|\mathcal{O}_N| \to 0$, the assertion holds.

Finally, assume that $m_N/|\mathcal{O}_N| \to q < 1$. By the weak convergence of linear combination of order statistics again,

$$m_N^{-1}\sum_{k=1}^{d}\lambda_k^{-1}\sum_{i\in H_O}\left\{\int_a^b\varepsilon_i(t)v_k(t)dt\right\}^2 \le m_N^{-1}\sum_{i=1}^{m_N}y_{(i)} = \psi_G(q) + O_p(N^{-1/2}),$$

$$m_N^{-1}\sum_{k=1}^{d}\lambda_k^{-1}\sum_{i\in J}\left\{\int_a^b\varepsilon_i(t)v_k(t)dt\right\}^2 = \psi_G(q') + O_p(N^{-1/2}).$$

Similar to (S.1),

$$\left|m_N^{-1}R_2\right| \le 2d\left(\delta_U\lambda_d^{-1}\right)^{1/2} + o_p(1).$$

Combining all these results together, the assertion holds provided that

$$2d\left(\delta_U\lambda_d^{-1}\right)^{1/2} + \max_{0<q<1}\left\{\psi_G(q') - \psi_G(q)\right\} < (1-2\rho)\lambda_1^{-1}\delta_L,$$

i.e., Condition A4 holds. □

Proof of Proposition 1

This result is a direct corollary of Lemma 1, so the proof is omitted. □

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. Figure S1 shows boxplots of the numbers of outliers contained in H_{LTFS} and H_{MDP} under Case (II). Figure S2 presents boxplots of the numbers of outliers contained in H_{LTFS} using estimators of $\{v_k(\cdot), \lambda_k\}$ based on Ro et al. (2015), Filzmoser et al. (2008) and all samples. Tables S1–S3 report empirical false positive rates under Cases (I) and (II) when $N = 100, 200, 500, 1000$ with $p = 100$ or 500, respectively. Figure S3 shows empirical false positive rates and false negative rates under Case (II)

for various values of γ, when $\rho = 0.1$, $N = 500$ and $p = 500$. Figure S4 presents empirical false positive and false negative rates with various values of mixture ratio ω for generating outliers from the two classes of functions (a) and (b), under $\rho = 0.1$, $N = 500$, $\gamma = 2$ and $p = 500$. Similar conclusions to the paper can be drawn from these numerical studies: the initial subset H_{LTFS} is reasonably clean; The proposed refined least trimmed functional scores procedure is a promising alternative for functional outlier detection, considering its robustness in terms of the false positive and false negative rates.

Table S4 compares empirical false positive and false negative rates under Case (II) with different values of total explained variation, when $\alpha = 5\%$, $\gamma = 2$, $N = 200$ and $p = 500$. It clearly shows that the performances are quite stable across all scenarios, especially for percentages between 85% and 95%.

Table S5 presents empirical false positive rates based on Yu et al. (2012) under Case (II) when $N = 200, 500, 1000$ and $p = 500$. From the results in Tables 2–3 in the paper, we observe that Yu et al. (2012)'s procedure is generally more conservative than the other competitors. This is not surprising to us because this method stepwisely deletes single observation, by comparing the largest measures with a threshold obtained from an extreme distribution. It is able to control the family-wise error rate when the dataset contains no outlier, say the proportion of good datasets that are wrongly declared to contain outliers. However, when the dataset contains some outliers, there is no guarantee to control the false positive rates. Table S5 tells us that its false positive rates are not consistent across various values of ρ and N and vary much with different correlation structures either.

Table S6 reports the average computing time of different methods under Case (I) and autoregressive structure with $\rho = 0.1$ and $\gamma = 2$. We can see that our procedure runs fast and our computing time is less than the refined minimum diagonal product (Ro et al., 2015) and depth-based functional outlier detection procedures (Febrero et al., 2008).

Table S1. *Empirical false positive rates (%) under Case (II) for different outlier ratios ρ with nominal size $\alpha = 1\%, 5\%, 10\%$, when $N = 100, 200, 500, 1000$, $\gamma = 2$ and $p = 500$*

$\varepsilon_i(t)$	N	$\rho = 0.04$			$\rho = 0.1$			$\rho = 0.2$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
AR	100	0.7	4.4	10.1	0.7	4.0	9.3	0.5	3.5	8.4
	200	0.9	4.8	10.6	0.8	4.7	10.1	0.7	4.3	9.2
	500	0.8	4.9	10.6	0.8	4.7	10.2	0.7	4.4	9.7
	1000	1.0	5.1	10.9	0.9	5.0	10.6	0.8	4.8	10.0
MA	100	0.8	4.6	9.9	0.7	4.3	9.3	0.5	3.6	8.4
	200	0.8	4.8	10.5	0.8	4.5	10.0	0.6	4.0	9.2
	500	0.9	4.9	10.7	0.8	4.7	10.3	0.7	4.3	9.5
	1000	0.9	5.1	10.8	0.9	4.8	10.4	0.7	4.4	9.7
BM	100	1.4	7.1	14.4	1.5	6.8	13.4	1.4	6.0	11.9
	200	1.2	6.4	13.1	1.1	6.0	12.3	1.1	5.3	11.2
	500	1.1	5.9	12.4	1.0	5.6	11.8	1.0	5.2	10.8
	1000	1.1	5.8	12.2	1.0	5.6	11.6	0.9	5.1	10.7

AR stands for autoregressive, MA for moving average and BM for Brownian motion.

CASE STUDY 1: OUTLIER DETECTION IN TONNAGE PROFILES

In this section, we apply the proposed methodology to a real dataset taken from an industrial multi-operation forging process. In this process, a forging machine shown in Fig. S5(a) is comprised of multiple dies, each assigned to perform one operation during a stroke. Tonnage forces exerted on all dies are measured by four strain sensors that are mounted on four columns of the press. In this sensing system, each sensor records the tonnage force profile at the predefined equal sampling interval of a rotational

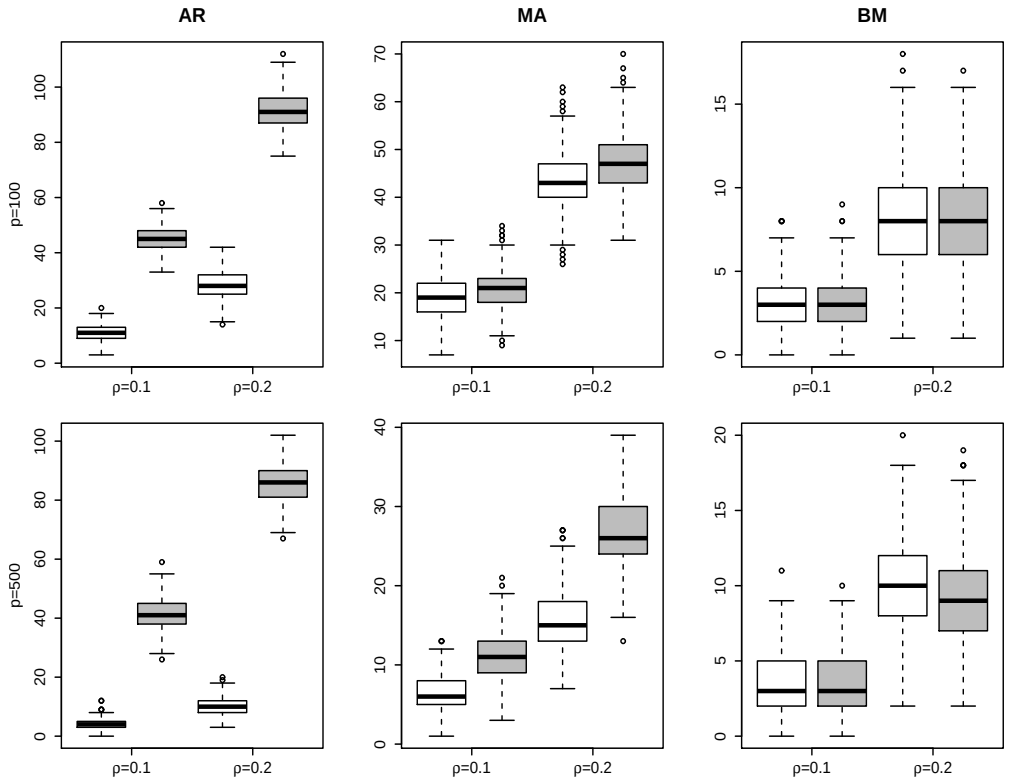

Fig. S1. Boxplots of the numbers of outliers contained in H_{LTFS} (white boxes) and H_{MDP} (grey boxes) under Case (II) when $N = 1000$ and $\gamma = 2$.

crank angle. This results in functional profile data shown in Fig. S5(b). More detailed discussion about this example can be found in Lei et al. (2010).

A sample of 496 profiles was collected under different experimental settings, containing 151 normal profiles collected under the normal production condition, and 5 groups of 69 abnormal profiles. For the illustration purpose, the average profiles of these six groups are shown in Fig. S5(c). In this example, the main goal is to set up an efficient classifier which heavily relies on the assumption that sufficient data from both normal and faulty operations are available that can be used for training classifiers. Here, we focus on constructing an outlier detection method, so we do not use the information of faulty classes. Using this dataset, we define five subsets, each of which includes the normal profiles and one of the five abnormal profile groups. We would like to automatically identify any abnormal profile observations from the whole dataset using our proposed detection procedures.

Using subsets 1 and 4, including normal data and corresponding to faults 1 and 4 data, we compare our method with the other benchmarks used in the simulation study, including those methods in Febrero et al. (2008), Yu et al. (2012), Ro et al. (2015) and Filzmoser et al. (2008). The reason for the use of subsets 1 and 4 is that profile samples corresponding to faults 1, 4 are very similar to the profiles under the normal operation, thus difficult to separate. The resulting false positive and false negative rates of the five methods with $\alpha = 5\%$ are reported in Table S7. For fault 1, our method, Ro et al. (2015)'s method and Filzmoser et al. (2008)'s method are equally effective as they can approximately maintain the desired level and identify all outliers. Febrero et al. (2008)'s method and Yu et al. (2012)'s method cannot

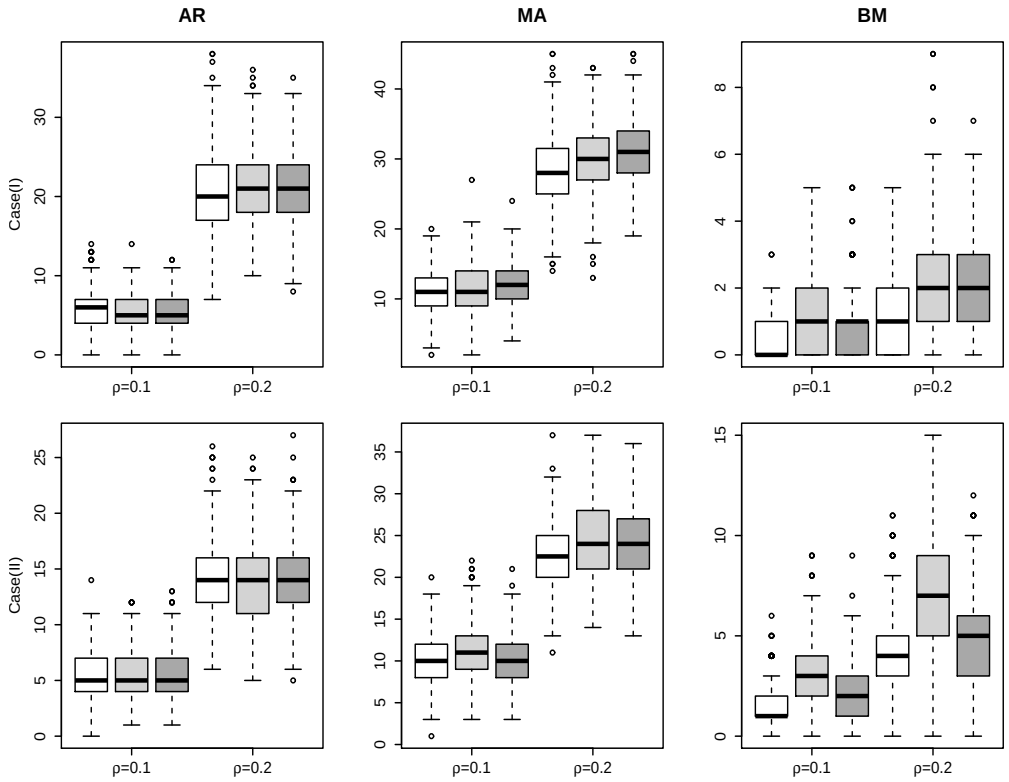

Fig. S2. Boxplots of the numbers of outliers contained in H_{LTFS} using estimators of $\{v_k(\cdot), \lambda_k\}$ based on Ro et al. (2015) (white boxes), Filzmoser et al. (2008) (lightgrey boxes) and all samples (darkgrey boxes) under Case (I) (top) and Case (II) (bottom) when $N = 500$, $\gamma = 2$ and $p = 100$.

deliver correct detection due to the masking effect. Regarding fault 4, all the methods fail to identify most outliers as the difference between the normal and faulty profiles in this subset is indeed slight, though our method can still control the false positive rate. To make this comparison more informative, we select 20 most outlying profiles out of all the 69 faulty profiles based on their minimum diagonal product distances (Ro et al., 2015), and then apply all the five methods to the new dataset, i.e. 151 normal profiles with 20 fault 4 profiles. The detection results are given in the last two columns of Table S7, corresponding to fault 4*. Now, our method is capable of identifying most outliers and performs better than the others in terms of false positive and false negative rates. We further zoom in the average profiles for faults 1, 4 and normal functional data. Figure S6 shows that there are indeed differences in the mean function in the time range of [40,100]. The results show that our method does have superior performance in this challenging case as it is able to capture fine differences.

CASE STUDY 2: OUTLIER DETECTION IN REAL-TIME TRAFFIC SPEED DATA

We also apply the proposed methodology to a real dataset taken from the traffic speed detectors for illustration. This data contains real-time traffic average speed information from some highway locations. In each location, the speed detector automatically records the speed every five minutes. This results in

Table S2. *Empirical false positive rates (%) under Case (I) for different outlier ratios ρ with nominal size $\alpha = 1\%, 5\%, 10\%$, when $N = 100, 200, 500, 1000$ $\gamma = 2$ and $p = 100$*

$\varepsilon_i(t)$	N	$\rho = 0.04$			$\rho = 0.1$			$\rho = 0.2$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
AR	100	0.7	4.6	9.9	0.7	4.0	9.2	0.6	3.7	8.5
	200	0.8	4.8	10.6	0.7	4.3	9.7	0.6	3.8	8.7
	500	0.9	4.9	10.6	0.8	4.4	9.8	0.6	4.0	8.8
	1000	0.9	5.0	10.8	0.8	4.5	9.9	0.7	4.1	8.9
MA	100	0.7	4.5	9.8	0.6	4.1	9.4	0.6	3.8	8.6
	200	0.8	4.8	10.4	0.8	4.6	10.1	0.6	4.1	9.1
	500	0.8	4.7	10.4	0.8	4.5	9.9	0.7	4.0	9.2
	1000	0.9	5.0	10.7	0.8	4.6	10.1	0.7	4.2	9.4
BM	100	1.6	7.1	14.1	1.5	6.6	13.2	1.4	5.8	11.6
	200	1.3	6.4	13.0	1.3	6.0	12.3	1.2	5.6	11.3
	500	1.2	6.1	12.5	1.1	5.7	11.7	1.0	5.3	10.9
	1000	1.1	5.9	12.2	1.1	5.6	11.6	1.0	5.2	10.8

Table S3. *Empirical false positive rates (%) under Case (II) for different outlier ratios ρ with nominal size $\alpha = 1\%, 5\%, 10\%$, when $N = 100, 200, 500, 1000$, $\gamma = 2$ and $p = 100$*

$\varepsilon_i(t)$	N	$\rho = 0.04$			$\rho = 0.1$			$\rho = 0.2$		
		1%	5%	10%	1%	5%	10%	1%	5%	10%
AR	100	0.8	4.7	10.5	0.7	4.3	9.6	0.6	3.6	8.6
	200	0.9	4.9	10.5	0.8	4.5	9.7	0.6	3.9	8.6
	500	0.9	5.0	10.8	0.8	4.6	10.1	0.6	4.1	9.1
	1000	0.9	5.1	10.8	0.8	4.7	10.2	0.6	4.1	9.2
MA	100	0.7	4.4	9.9	0.7	4.2	9.4	0.5	3.5	8.2
	200	0.8	4.8	10.5	0.7	4.4	9.9	0.6	3.8	8.8
	500	0.8	4.8	10.6	0.7	4.4	9.9	0.6	3.9	9.0
	1000	0.9	5.0	10.7	0.8	4.7	10.2	0.6	4.1	9.2
BM	100	1.5	6.9	13.7	1.4	6.4	13.0	1.5	6.0	11.6
	200	1.2	6.4	13.1	1.2	6.0	12.2	1.1	5.4	11.2
	500	1.1	6.0	12.3	1.1	5.7	11.7	1.0	5.1	10.7
	1000	1.1	5.9	12.3	1.0	5.6	11.6	0.9	5.1	10.7

time-dependent functional data. A sample of 31 original speed records was collected in May from one highway location. To reduce the measurement noise, firstly we smooth the data with the window size of half hour shown in the left panel of Fig. S7. In this case, we focus on detecting the abnormal speed curves. Here, the proposed method is applied to this dataset. The identified abnormal speed curves are corresponding to the days 1,10,11,17,18,19,28,29 and 31, which are either the weekends or the holidays in May. Meanwhile, we put the normal mean speed function and outlier mean speed function in the right panel of Fig. S7. Because we do not know the true outliers, we cannot compare the efficiency of our procedure with other related methods. However, from our result, our detection results appear reasonable and credible; the average speed curves in weekends or holidays are distinguished from those in weekdays.

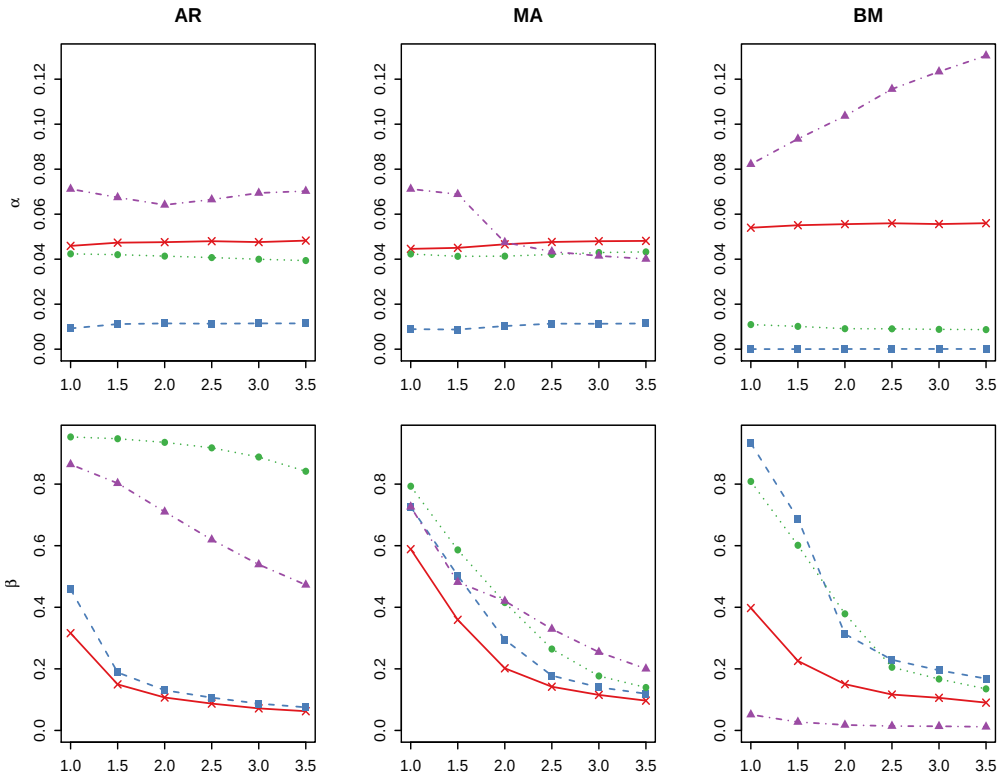

Fig. S3. Empirical false positive rates (α, top) and false negative rates (β, bottom) of four methods as follows: our refined least trimmed functional scores method (asterisks solid red line); Yu et al. (2012) (square dash blue line); Ro et al. (2015) (circle dot green line) and Filzmoser et al. (2008) (triangle dot-dash purple line) under Case (II) for various values of γ, when $\rho = 0.1$, $N = 500$ and $p = 500$.

REFERENCES

CHERNOFF, H., GASTWIRTH, J. L. & JOHNS, M. V. (1967). Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation. *Ann. Math. Stat.* **38**, 52–72.

FEBRERO, M., GALEANO, P. & GONZÁLEZ-MANTEIGA, W. (2008). Outlier detection in functional data by depth measures, with application to identify abnormal nox levels. *Environmetrics* **19**, 331–345.

FILZMOSER, P., MARONNA, R. & WERNER, M. (2008). Outlier identification in high dimensions. *Comput. Statist. Data Anal.* **52**, 1694–1711.

LEI, Y., ZHANG, Z. & JIN, J. (2010). Automatic tonnage monitoring for missing part detection in multi-operation forging processes. *J. Manuf. Sci. Eng.* **132**, 051010.1–10.

RO, K., ZOU, C., WANG, Z. & YIN, G. (2015). Outlier detection for high-dimensional data. *Biometrika* **102**, 589–599.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*, vol. 162. Wiley: New York.

YU, G., ZOU, C. & WANG, Z. (2012). Outlier detection in functional observations with applications to profile monitoring. *Technometrics* **54**, 308–318.

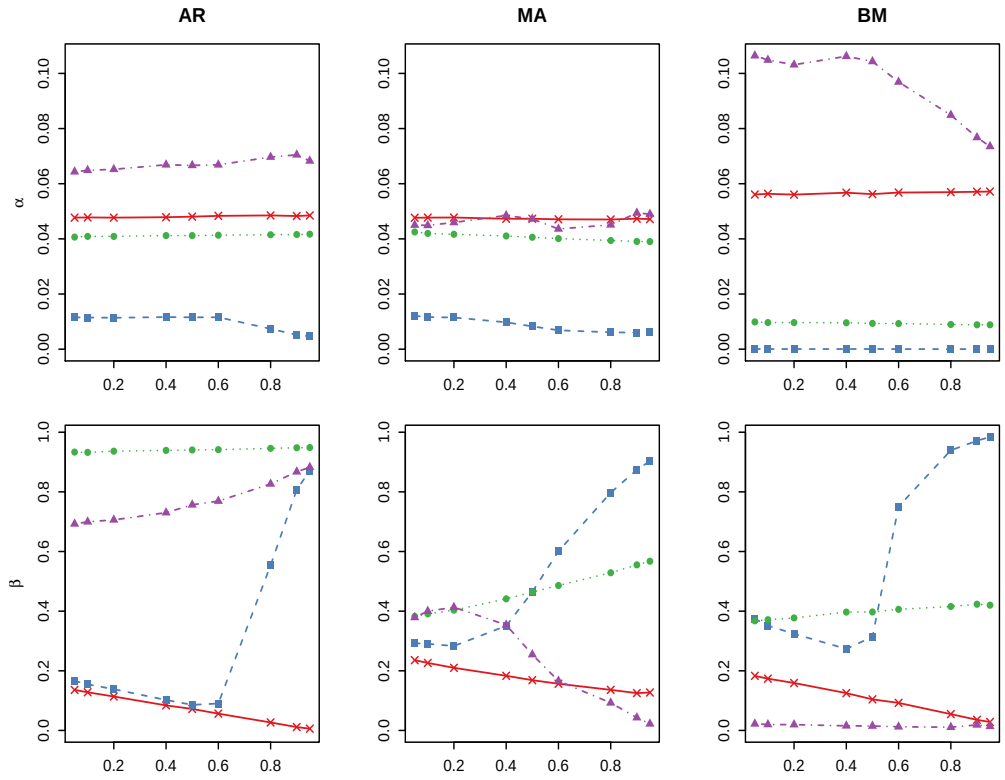

Fig. S4. Empirical false positive rates (α, top) and false negative rates (β, bottom) of four methods as follows: our refined least trimmed functional scores method (asterisks solid red line); Yu et al. (2012) (square dash blue line); Ro et al. (2015) (circle dot green line) and Filzmoser et al. (2008) (triangle dot-dash purple line) for various values of ω, when $\rho = 0.1$, $N = 500$, $\gamma = 2$ and $p = 500$.

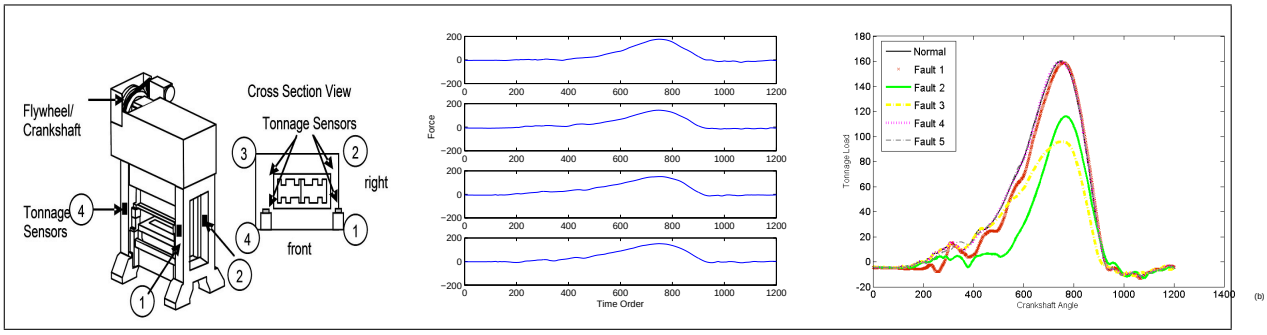

Fig. S5. (a)(left panel): A forging machine with four strain gages. (b)(middle panel): A sample of four-channel profiles. (c)(right panel) Average curves of aggregated tonnage profiles for normal and faulty operations.

Table S4. *Empirical false positive* (α) *and false negative* (β) *rates (%) under Case (II) with various values of total explained variation, when* $\alpha = 5\%$, $\gamma = 2$, $N = 200$ *and* $p = 500$

$\varepsilon_i(t)$	ρ	75%		85%		90%		95%	
		α	β	α	β	α	β	α	β
AR	0.02	4.4	8.1	4.5	7.7	4.9	6.7	5.9	6.4
	0.04	4.3	11.2	4.5	10.0	4.8	9.6	5.7	9.2
	0.1	4.2	11.2	4.3	10.4	4.5	10.1	5.3	9.6
	0.2	3.9	12.0	4.1	11.4	4.2	11.1	4.6	10.6
		$d = 9$		$d = 11$		$d = 12$		$d = 14$	
MA	0.02	4.3	20.0	4.5	16.3	4.9	15.7	6.1	14.0
	0.04	4.1	23.1	4.4	19.3	4.8	18.4	5.9	17.0
	0.1	3.9	23.5	4.1	21.3	4.5	20.0	5.5	18.8
	0.2	3.6	26.7	3.7	24.6	4.0	23.1	4.7	21.4
		$d = 9$		$d = 11$		$d = 12$		$d = 14$	
BM	0.02	6.9	51.8	6.8	17.3	6.6	10.3	6.5	6.1
	0.04	6.6	47.6	6.6	18.2	6.4	12.1	6.4	7.5
	0.1	6.4	40.6	6.3	19.9	6.0	14.0	5.9	8.9
	0.2	6.3	31.8	5.8	18.0	5.6	13.9	5.4	9.2
		$d = 3$		$d = 4$		$d = 5$		$d = 7$	

d: is the median of the empirical numbers of the eigenfunctions to explain the corresponding percentages of total variation

Table S5. *Yu et al. (2012)'s empirical false positive rates (%) under Case (II) for various values of outlier ratio* ρ *with nominal size* $\alpha = 10\%, 15\%, 20\%$, *when* $N = 200, 500, 1000$ *and* $p = 500$

$\varepsilon_i(t)$	N	$\rho = 0.04$			$\rho = 0.1$			$\rho = 0.2$		
		10%	15%	20%	10%	15%	20%	10%	15%	20%
AR	200	5.0	8.0	11.6	5.3	8.6	12.4	1.1	1.7	2.5
	500	2.1	3.0	3.8	2.0	3.0	3.9	0.5	0.8	1.1
	1000	1.0	1.4	1.8	0.6	1.2	1.8	0.2	0.3	0.4
MA	200	5.2	8.2	11.6	5.2	8.6	12.8	2.2	3.5	5.1
	500	2.1	3.1	4.0	1.2	2.1	3.1	1.0	1.4	1.9
	1000	0.9	1.3	1.7	0.5	0.8	1.2	0.4	0.5	0.7
BM	200	0.0	0.1	0.1	0.0	0.1	0.1	0.0	0.0	0.0
	500	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	1000	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table S6. *Computing time (sec.) for various sample size* N, *with the outlier ratio* $\rho = 0.1$, $p = 500$ *and* $\alpha = 5\%$.

N	ReLTFS	DFOD	SFOD	RMDP	PCOut
100	1.67	63.16	0.86	15.30	0.12
200	2.79	187.40	0.23	24.71	0.20
500	7.63	1087.43	1.18	55.96	0.74
1000	16.11	4043.30	1.95	94.02	1.29

ReLTFS: our refined least trimmed functional scores method; DFOD: the depth-based functional outlier detection procedure introduced by Febrero et al. (2008); SFOD: the stepwise functional outlier detection procedure proposed by Yu et al. (2012); RMDP: refined minimum diagonal product procedure by Ro et al. (2015); PCOut: principal component outlier detection procedure by Filzmoser et al. (2008)

Table S7. *False positive* (α) *and false negative* (β) *rates (%) of the five related methods for outlier detection of faults 1 and 4 in the forging example*

Method	Fault 1		Fault 4		Fault 4*	
	α	β	α	β	α	β
ReLTFS	6.7	0.0	4.0	91.3	4.0	5.0
DFOD	0.0	91.3	0.7	97.1	0.7	90.0
SFOD	0.0	100.0	0.0	100.0	0.0	100.0
RMDP	4.6	0.0	2.7	97.1	2.7	85.0
PCOut	8.8	0.0	22.5	89.9	12.6	100.0

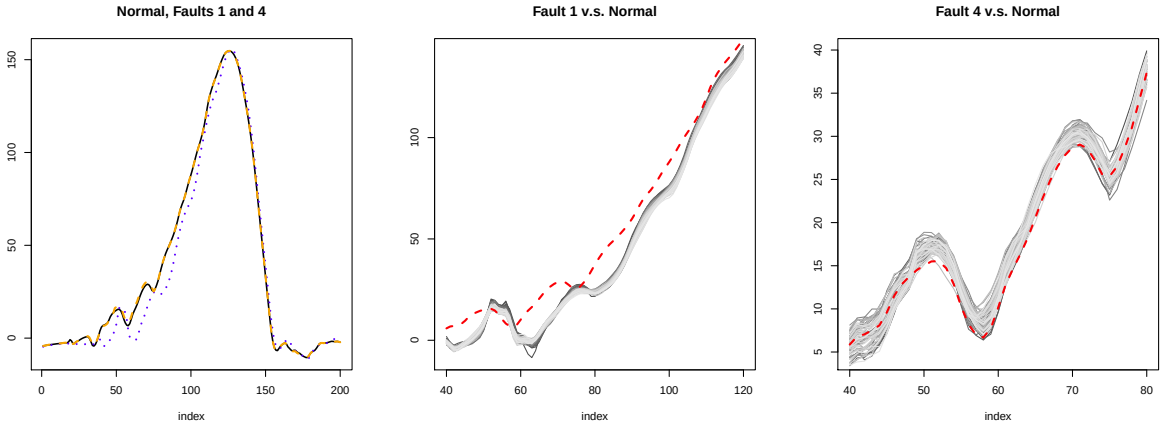

Fig. S6. Left panel: Average profiles of Normal (solid black curve), Faults 1 (dash orange curve) and 4 (dot blue curve). Middle panel: Average normal profile (dash red curve) and 69 fault 1 profiles zoomed-in between 40 and 100 (grey curves). Right panel: Average normal profile (dash red curve) and 69 fault 4 profiles zoomed-in between 40 and 80 (grey curves).

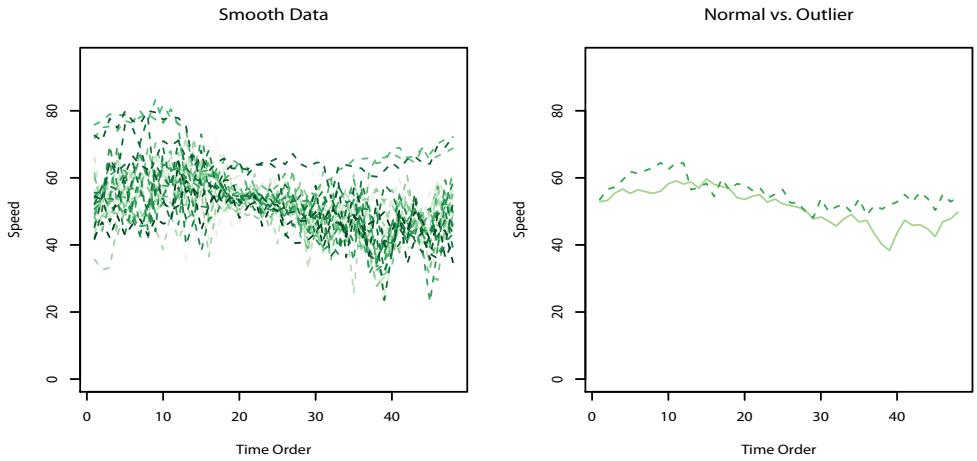

Fig. S7. Left panel: 31 smoothed speed curves in May; Right panel: Average speed curves from normal (solid curve) and outlier (dash curve) days. The detected outlier days are weekends and holidays.