AN ADAPTIVE WEIGHTED COMPONENT TEST FOR HIGH-DIMENSIONAL MEANS

Department of Statistics & Actuarial Science,HKU.

Faculty of Business Administration, University of Macau.

Department of Biostatistics, City University of Hong Kong.

Supplementary Material

This supplementary materials provides detailed proofs of Theorem 1 and 2 and Proposition 1-3 as well as the power simulation results under the heteroscedastic condition.

S1 Appendix: Proof of Main Theorems

S1.1 Lemmas for proof of theorems

Lemma A.1. (Lemma 2.1(i) inYang (2007)) Suppose that ξ and η are $\mathcal{F}_1^k$ - measurable and $\mathcal{F}_{k+n}^{\infty}$ - measurable random variables, respectively. If $E|\xi|^p < \infty, E|\eta|^p < \infty$ for some $p, q, s > 1$ with $1/p + 1/q + 1/s = 1$, then

$$|E(\xi\eta) - (E\xi)(E\eta)| \leq 10\alpha^{1/s}(n)(E|\xi|^p)^{1/p} \cdot (E|\eta|^q)^{1/q}.$$

Lemma A.2. Suppose $\{\xi_i, i \geq 1\}$ is a random sequence and $f : \mathbf{R} \to \mathbf{R}$ is a Borel function. If $\{\xi_i, i \geq 1\}$ is α-mixing, i.e., $\alpha_\xi(s) \to 0$ as $s \to \infty$, then

$\{\eta_i = f(\xi_i), i \geq 1\}$ is also α-mixing with $\alpha_\eta(s) = \alpha_\xi(s)$.

Proof. Let $\mathcal{F}_a^b = \sigma\{\xi_a, \xi_{a+1}, \ldots, \xi_b\}$ and $\mathcal{B}_a^b = \sigma\{\eta_a, \eta_{a+1}, \ldots, \eta_b\}$ for integers $a < b$, then by the definition of strong mixing coefficient,

$$\begin{aligned}
&\alpha_\eta(s)\\
=&\sup_{k\geq 1}\left\{|P\left(Y \in (A\cap B)\right) - P(Y\in A)P(Y\in B)| : A \in \mathcal{B}_1^k, B \in \mathcal{B}_{k+s}^\infty\right\}\\
=&\sup_{k\geq 1}\left\{\left|P\left(X \in f^{-1}(A\cap B)\right) - P(X\in f^{-1}(A))P(X\in f^{-1}(B))\right| : A \in \mathcal{B}_1^k, B \in \mathcal{B}_{k+s}^\infty\right\}\\
=&\sup_{k\geq 1}\left\{\left|P\left(X \in f^{-1}(A)\cap f^{-1}(B)\right) - P(X\in f^{-1}(A))P(X\in f^{-1}(B))\right| : A \in \mathcal{B}_1^k, B \in \mathcal{B}_{k+s}^\infty\right\}\\
=&\sup_{k\geq 1}\left\{|P\left(X \in C\cap D\right) - P(X\in C)P(X\in D)| : C \in \mathcal{F}_1^k, D \in \mathcal{F}_{k+s}^\infty\right\} = \alpha_\xi(s).
\end{aligned}$$

Lemma A.3. (Theorem 17.2.1 in Ibragimov (1975)) Suppose that ξ and η are $\mathcal{F}_1^k$ - measurable and $\mathcal{F}_{k+n}^\infty$ - measurable random variables, respectively. If $|\xi| \leq C_1, |\eta| \leq C_2$, then

$$|E(\xi\eta) - (E\xi)(E\eta)| \leq 4C_1C_2\alpha(n).$$

Lemma A.4. (Theorem 1 in Kim (1994)) Suppose $\{\xi_i, i \geq 1\}$ is a sequence of dependent random variables satisfying $E|\xi_i|^p < \infty$ for some $p \geq 1$. Assume that $E(\xi_i) = 0$ and $M_{2r+\epsilon} = \sup\left\{(E|\xi_i|^{(2r+\epsilon)})^{1/(2r+\epsilon)}, i \geq 1\right\}$ for some $\epsilon > 0$ and $r = 1, 2$. Further, suppose that $p(k) \leq (2r+\epsilon)/k, q(k) \leq (2r+\epsilon)/(2r-k)$ for $k = 1, \ldots, 2r-1$, let $f(r,\epsilon) = \min_{k=1,\ldots,2r-1} f[p(k), q(k)]$, then

$$\boldsymbol{E}\left(\sum_{i=1}^{n}\xi_i\right)^{2r} \leqq Cn^r\left[M_{2r}^{2r}+M_{2r+\epsilon}^{2r}\sum_{i=1}^{n}i^{r-1}\alpha(i)^{f(r,\epsilon)}\right],$$

where C does not depend on the distribution of $\{\xi_i, i \geq 1\}$ or on n.

S1.2 Proof of Theorem 1

First, for $1 \leq j \leq p$, define $L_j = \omega_j t_j^2$, $\sigma^2 = var(\sum_{j=1}^{p} L_j)$, and $\mu_j = E(L_j)$. By applying the big- block-little-block method by Rosenblatt (1956), we Partition the sequence

$$\sigma^{-1}\{L_j - \mu_j\}, 1 \leq j \leq p$$

into N blocks, where each block contains t variables such that $Nt \leq p < (N+1)t$. Further, for each $1 \leq i \leq N$, we partition the ith block into two sub-blocks with a larger one Q_{i1} and a smaller one Q_{i2}. Suppose each Q_{i1} has t_1 variables and each Q_{i2} has $t_2 = t - t_1$ variables. We require $N \to \infty, t_1 \to \infty, t_2 \to \infty, Nt_1/p \to 1$ and $Nt_2/p \to 0$as $p \to \infty$.We write

$$Q_{i1} = \textstyle\sum_{j=1}^{t_1}\left[L_{\{(i-1)t+j\}} - \mu_{\{(i-1)t+j\}}\right]$$
$$Q_{i2} = \textstyle\sum_{j=1}^{t_2}\left[L_{\{(i-1)t+t_1+j\}} - \mu_{\{(i-1)t+t_1+j\}}\right].$$

And then define

$$\mathcal{L}_1 = \sigma^{-1}\textstyle\sum_{i=1}^{N} Q_{i1}$$
$$\mathcal{L}_2 = \sigma^{-1}\textstyle\sum_{i=1}^{N} Q_{i2}$$
$$\mathcal{L}_3 = \sigma^{-1}\textstyle\sum_{j=Nt+1}^{p}\{L_j - \mu_j\}.$$

Now we have

$$\sigma^{-1}\sum_{j=1}^{p}\{L_j-\mu_j\}=\mathcal{L}_1+\mathcal{L}_2+\mathcal{L}_3.$$

The big- block-little-block method makes Q_{i1}'s "almost" independent, thus the study of $\mathcal{L}_1$ may be related to the well-studied cases of sums of independent random variables. Also, since t_2 is small compared with t_1, the sum $\mathcal{L}_2$ and $\mathcal{L}_3$ will be small compared with the total sum of variables in the sequence, i.e., $\sigma^{-1}\sum_{j=1}^{p}\{L_j-\mu_j\}$. We next show

$$\sigma^{-1}\sum_{j=1}^{p}\{L_j-\mu_j\}=\mathcal{L}_1+o_p(1).$$

As $E(\mathcal{L}_2)=E(\mathcal{L}_3)=0$, it is sufficient to prove that $var(\mathcal{L}_2)=var(\mathcal{L}_3)=o(1)$. Consider $var(\mathcal{L}_2)$ and we have

$$\begin{aligned}\operatorname{var}(\mathcal{L}_2)&=\sigma^{-2}\operatorname{var}\left\{\sum_{i=1}^{N}Q_{i2}\right\}\\&\leq\sigma^{-2}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{j_1=1}^{t_2}\sum_{j_2=1}^{t_2}\left|\operatorname{cov}\left\{L_{(i_1t+t_1+j_1)},L_{(i_2t+t_1+j_2)}\right\}\right|.\end{aligned}$$

By Lemma A.1 & A.2, we have the following α-mixing inequality that for any $\epsilon>0$,

$$\operatorname{cov}\{L_i,L_j\}\leq C\alpha_X(|i-j|)^{\epsilon/(2+\epsilon)}\max_i\left[\mathrm{E}\{L_i\}^{2+\epsilon}\right]^{(2+\epsilon)/2},$$

where C is a constant. Then, take $\epsilon = 1$,

$$
\begin{aligned}
&\left|\operatorname{cov}\left\{L_{(i_1t+t_1+j_1)}, L_{(i_2t+t_1+j_2)}\right\}\right| \\
&\leq B\alpha_X\left\{\left|(i_1t+t_1+j_1)-(i_2t+t_1+j_2)\right|\right\}^{1/3} \\
&\leq BM\delta^{|i_1t+j_1-i_2t-j_2|/3},
\end{aligned}
$$

where B is some big constant. The above result implies that

$$
\begin{aligned}
\operatorname{var}(\mathcal{L}_2) &\leq \sigma^{-2}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{j_1=1}^{t_2}\sum_{j_2=1}^{t_2}\left|\operatorname{cov}\left\{L_{(i_1t+t_1+j_1)}, L_{(i_2t+t_1+j_2)}\right\}\right| \\
&\leq \sigma^{-2}\sum_{i_1=1}^{N}\sum_{i_2=1}^{N}\sum_{j_1=1}^{t_2}\sum_{j_2=1}^{t_2} BM\delta^{|i_1t+j_1-i_2t-j_2|/3} \\
&= O(1)Nt_2/p,
\end{aligned}
$$

which goes to 0 as $p \to \infty$. This implies $\mathcal{L}_2 = o_p(1)$. Similarly, we can show that $\mathcal{L}_3 = o_p(1)$ under the strong mixing assumption. Therefore, we only need to focus on $\mathcal{L}_1$. Then for properly chosen N and t_2, we have

$$
\left|\mathrm{E}\left\{\exp\left(ir\mathcal{L}_1\right)\right\} - \mathrm{E}^N\left[\exp\left\{ir\sigma^{-1}Q_{1,1}\right\}\right]\right| \leq 16N\alpha_X(t_2) \to 0, \quad r \in \mathbb{R}
$$

by Lemma A.3, which implies that there exist independent random variables $\{\xi_i; i = 1, 2, \ldots, N\}$ such that ξ_i and Q_{i1} are identically distributed and $\mathcal{L}_1$ has the same asymptotic distribution as $\sigma^{-1}\sum_{i=1}^{N}\xi_i$.

Then, we only need to show that the central limit theorem holds for $\sigma^{-1}\sum_{i=1}^{N}\xi_i$. . This can be done by checking the Lyapunov condition. In particular,by Lemma A.4, the strong mixing assumption implies

$$\begin{aligned}
\mathrm{E}\left\{\sigma^{-1}Q_{1,1}\right\}^4 &= \sigma^{-4}\mathrm{E}\left[\sum_{j=1}^{t_1}\{L_j-\mu_j\}\right]^4 \\
&= O(1)\sigma^{-4}t_1^2\left\{B_1+B_2\sum_{j=1}^{t_1} j\alpha(j)^{\epsilon/(4+\epsilon)}\right\} \\
&= O(1)t_1^2/p^2 = O\left(t_1^2/p^2\right),
\end{aligned}$$

where B_1 and B_2 are constants. Thus we have $\sum_{i=1}^N \sigma^{-4}\mathrm{E}\xi_i^4 = O\left(Nt_1^2p^{-2}\right) = o(1)$ and the Lyapunov condition holds. Now we have proved the asymptotic normal distribution of T.

S1.3 Proof of Theorem 2

For $\kappa \in \Gamma = \{0 \le \kappa_s \le 1, s = 1, 2, \ldots, m\}$, we have proved the asymptotic normal distribution of $T(\kappa)$ in Theorem 1. For any linear combination of $T(\kappa)'$s with respect to different κ, a similar argument as proof of Theorem 1 gives the asymptotic normal distribution. Then the Cramér-Wold Theorem implies the asymptotic joint distribution of $\{T(\kappa), \kappa \in \Gamma\}$ when m is finite, and the covariance matrix is derived in Proposition 3.

S2 Proof of Propositions

Before presenting the proof of Propositions 1-3, we need the following three lemmas.

Lemma S.1. (Theorem in Hall et al. (1987)) Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables with mean μ and variance $\sigma^2 < \infty$ and $Z_n = \sqrt{n}(\bar{X} - \mu)/s$, where $\bar{X} = \sum_{i=1}^{n} X_i/n$ and $s^2 = n^{-1}\sum_{i=1}^{n} X_i^2 - \bar{X}^2$. Denote $F_n(x) = P(Z_n \leq x)$ and P_i as the polynomial of degree $3i - 1$ appearing in the formal Edgeworth expansion of the distribution of Z_n, if $E|X|^{k+2} < \infty$ for $k \geq 1$, then

$$F_n(x) = \Phi(x) + \sum_{i=1}^{k} n^{-i/2} p_i(x)\phi(x) + o\left(n^{-k/2}\right)$$

uniformly in x.

The coefficients in p_i are functions of $E(X), \ldots, E(X^{i+2})$, for example,

$$p_1(x) = 1/6\tau\left(2x^2 + 1\right),$$

$$p_2(x) = -x\left\{1/18\tau^2\left(x^4 + 2x^2 - 3\right) - 1/12\kappa\left(x^2 - 3\right) + 1/4\left(x^2 + 3\right)\right\},$$

where $\tau = E(X - \mu)^3/\sigma^3$ and $\kappa = E(X - \mu)^4/\sigma^4 - 3$.

Lemma S.2. For $k = 2$, denote $W = Z_n^2$ and $H(x) = P(W \leq x)$, if $E(X^4) < \infty$,

$$H(x) = \Phi(\sqrt{x}) - \Phi(-\sqrt{x}) + O\left(n^{-1}\right)$$

uniformly in x.

Proof. From Lemma S.1.

Lemma S.3. For $k \geq 1$, $E|X|^{k+2} < \infty$, then

$$\left| E(Z_n^4) - \int_{-\infty}^{\infty} x^4 d\Phi(x) \right| = O(n^{-1}).$$

Proof. From Lemma S.1 and S.2,

$$\begin{aligned}
&\left| E(Z_n^4) - \int_{-\infty}^{\infty} x^4 d\Phi(x) \right| \\
=&\left| E(W^2) - \int_{-\infty}^{\infty} x^4 d\Phi(x) \right| \\
=&\left| \int_{0}^{\infty} x^2 dH(x) - \int_{-\infty}^{\infty} x^4 d\Phi(x) \right| \\
=&\left| \int_{0}^{\infty} x^2 d\left\{ \Phi(\sqrt{x}) - \Phi(-\sqrt{x}) + \sum_{i=2}^{\infty} n^{-i/2} p_i(\sqrt{x})\phi(\sqrt{x}) \right\} - \int_{-\infty}^{\infty} x^4 d\Phi(x) \right| \\
=&\left| \int_{0}^{\infty} x^2 d\left\{ \sum_{i=2}^{\infty} n^{-i/2} p_i(\sqrt{x})\phi(\sqrt{x}) \right\} \right| \\
=&\left| \sum_{i=2}^{\infty} n^{-i/2} \int_{0}^{\infty} 2x p_i(\sqrt{x})\phi(\sqrt{x}) dx \right| = O(n^{-1}).
\end{aligned}$$

Now we are ready to prove Propositions 1-3.

S2.1 Proof of Proposition 1

Define $G_{j,0}(x)$ and $g_{j,0}(x)$ as the cumulative distribution function and probability density function of t_j^2 under the H_0, respectively. Now under H_0 :

$\boldsymbol{\mu_1} = \boldsymbol{\mu_2}$, by Lemma S.2 and S.3,

$$\begin{aligned}
\mu_j &= E\left\{I_j\kappa t_j^2 + (1-I_j)(1-(1-\kappa)Rt_j^{-2})t_j^2\right\} \\
&= E\left\{(\kappa-1)I_jt_j^2 + t_j^2 + (1-\kappa)RI_j - (1-\kappa)R\right\} \\
&= (\kappa-1)E(I_jt_j^2) + E(t_j^2) + (1-\kappa)RE(I_j) - R(1-\kappa) \\
&= (1-\kappa)\left\{RE(I_j) - E(I_jt_j^2)\right\} + E(t_j^2) - (1-\kappa)R \\
&= (1-\kappa)\left\{RG_{j,0}(R) - (RG_{j,0}(R) - \int_0^R G_{j,0}(x)dx)\right\} \\
&\quad + \int_0^\infty xg_{j,0}(x)dx - (1-\kappa)R \\
&= (1-\kappa)\left\{\int_0^R F(x)dx - R\right\} + \int_0^\infty xf(x)dx + O(1/n),
\end{aligned}$$

where $F(x)$ and $f(x)$ denote the cumulative distribution function and probability density function of a χ_1^2 distribution, respectively.

S2.2 Proof of Proposition 2 & 3

Under $H_0 : \boldsymbol{\mu_1} = \boldsymbol{\mu_2}$, by Lemma S.2 and S.3,

$$\begin{aligned}
&\varsigma^2 \\
=&Var\left\{K_j\right\} \\
=&E\left\{(\kappa-1)I_jt_j^2 + t_j^2 + (1-\kappa)RI_j\right\}^2 - E^2\left\{(\kappa-1)I_jt_j^2 + t_j^2 + (1-\kappa)RI_j\right\} \\
=&E\left\{I_j(1-\kappa)(R-t_j^2)\left[(1-\kappa)(R-t_j^2)+2t_j^2\right]\right\} + E\left\{t_j^4\right\} - (\kappa-1)^2\left\{E(I_jt_j^2) - RE(I_j)\right\}^2 \\
&- E^2(t_j^2) - 2(\kappa-1)E(t_j^2)\left\{E(I_jt_j^2) - RE(I_j)\right\} \\
=&\int_0^R (1-\kappa)(R-x)\left[(1-\kappa)(R-x)+2x\right]g_{j,0}(x)dx + \int_0^\infty x^2g_{j,0}(x)dx - (\kappa-1)^2\left\{\int_0^R G_{j,0}(x)dx\right\}^2 \\
&- \left[\int_0^\infty xg_{j,0}(x)dx\right]^2 - 2(\kappa-1)\int_0^R G_{j,0}(x)dx\int_0^\infty xg_{j,0}(x)dx \\
=&\int_0^R (1-\kappa)(R-x)\left[(1-\kappa)(R-x)+2x\right]f(x)dx + \int_0^\infty x^2f(x)dx - (\kappa-1)^2\left\{\int_0^R F(x)dx\right\}^2
\end{aligned}$$

$$-\left[\int_0^\infty xf(x)dx\right]^2 - 2(\kappa-1)\int_0^R F(x)dx\int_0^\infty xf(x)dx + O(1/n).$$

Similarly,

$$\begin{aligned}
&\varsigma'^2\\
=&Cov(K_i(\kappa_s), K_i(\kappa_t))\\
=&\int_0^R \left[(1-\kappa_s)(1-\kappa_t)(R-x)^2 + (2-\kappa_s-\kappa_t)(R-x)x\right] g_{j,0}(x)dx + \int_0^\infty x^2 g_{j,0}(x)dx\\
&-(1-\kappa_s)(1-\kappa_t)\left\{\int_0^R G_{j,0}(x)dx\right\}^2 - \left[\int_0^\infty xg_{j,0}(x)dx\right]^2\\
&-(2-\kappa_s-\kappa_t)\int_0^R G_{j,0}(x)dx\int_0^\infty xg_{j,0}(x)dx\\
=&\int_0^R \left[(1-\kappa_s)(1-\kappa_t)(R-x)^2 + (2-\kappa_s-\kappa_t)(R-x)x\right] f(x)dx + \int_0^\infty x^2 f(x)dx\\
&-(1-\kappa_s)(1-\kappa_t)\left\{\int_0^R F(x)dx\right\}^2 - \left[\int_0^\infty xf(x)dx\right]^2\\
&-(2-\kappa_s-\kappa_t)\int_0^R F(x)dx\int_0^\infty xf(x)dx + O(1/n).
\end{aligned}$$

S3 Proof of Consistency of Estimators

S3.1 Consistency of $\hat{\nu}$

According to Proposition 1, we estimate ν by $\hat{\nu} = (1-\kappa)\left\{\int_0^R F(x)dx - R\right\} + 1$ and $\nu - \hat{\nu} = O(1/n)$. With $p/n^2 = o_p(1)$ in C.3, $\hat{\nu} = \nu + o_P(1/p^{1/2})$.

S3.2 Consistency of $\hat{\zeta}^2$

Based on the estimation of ρ_{ij}, we have $\rho_{ij} = \hat{\rho}_{ij}(1+O_p(\frac{1}{\sqrt{p}}))$. Also, based on the $\alpha-$mixing assumption, we have $\rho_{ij} \leq M\delta^{|i-j|}$. Here we take $L = c\log p$

for some constant $c > 0$.

$$\begin{aligned}
\zeta^2 - \hat{\zeta}^2 &= \frac{\varsigma^2}{p} \sum_{i,j \le p} \rho_{ij} - \mathfrak{p}(|i-j|/L)\hat{\rho}_{ij} \\
&\le \frac{\varsigma^2}{p}\Big(\sum_{k=0}^{\frac{L}{2}} (1 - \mathfrak{p}(k/L) + O_p(\frac{1}{\sqrt{p}}))pM\delta^k + \sum_{k=\frac{L}{2}+1}^{p} (1 + O_p(\frac{1}{\sqrt{p}}))pM\delta^k \Big) \\
&\le \frac{\varsigma^2}{p}\Big(\sum_{k=0}^{\frac{L}{2}} (1 - \mathfrak{p}(k/L) + O_p(\frac{1}{\sqrt{p}}))pM\delta^k + \sum_{k=\frac{L}{2}+1}^{p} (1 + O_p(\frac{1}{\sqrt{p}}))pM\delta^k \Big) \\
&\le \frac{\varsigma^2}{p}\Big(A + B + C \Big),
\end{aligned}$$

where,

$$\begin{aligned}
A &= \sum_{k=0}^{\frac{L}{2}} (1 - \mathfrak{p}(k/L))pM\delta^k \\
&= \sum_{k=0}^{\frac{L}{2}} 6((\frac{k}{L})^2 - (\frac{k}{L})^3)pM\delta^k \\
&\le \frac{6}{L^3} \sum_{k=0}^{\infty} (k^2 L + k^3)pM\delta^k \\
&\le \frac{6}{L^3}(C_1 L + C_2)p \\
&= \frac{6(cC_1 \log p + C_2)p}{(c \log p)^3} \\
&= o(p),
\end{aligned}$$

for some constants $C_1 > 0$ and $C_2 > 0$.

$$B = \sum_{k=0}^{\frac{L}{2}} O_p(\frac{1}{\sqrt{p}})pM\delta^k$$

$$
\begin{aligned}
&\leq O_p(\frac{1}{\sqrt{p}})pLM \\
&= o_p(p).
\end{aligned}
$$

Then,

$$
\begin{aligned}
C &= \sum_{k=\frac{L}{2}+1}^{p} (1 + O_p(\frac{1}{\sqrt{p}}))pM\delta^k \\
&= M(1 + O_p(\frac{1}{\sqrt{p}})) \sum_{k=\frac{L}{2}+1}^{p} p\delta^k \\
&\leq M(1 + O_p(\frac{1}{\sqrt{p}}))C_3 p\delta^{\frac{L}{2}} \\
&\leq M(1 + O_p(\frac{1}{\sqrt{p}}))C_3 p \cdot p^{\frac{1}{2}c\log\delta} \\
&= o_p(p),
\end{aligned}
$$

for some constant $C_3 > 0$. Therefore, $A + B + C = o_p(p)$, implying $\hat{\zeta}/\zeta = 1 + o_p(1)$.

S4 Additional Simulation Results

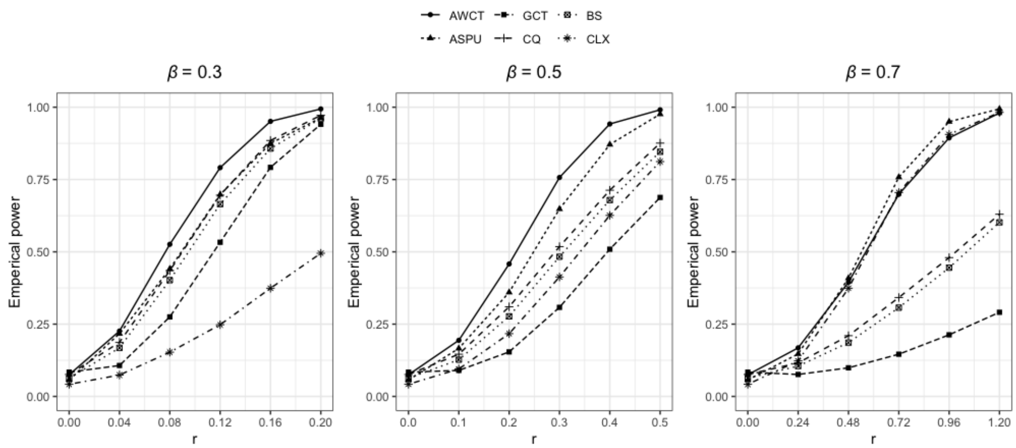

Figure 1: Power curves of the various tests against r under different sparsity levels of β based on Model (a) with normal innovations when $p = 400$, $n_1 = 200$, and $n_2 = 300$.

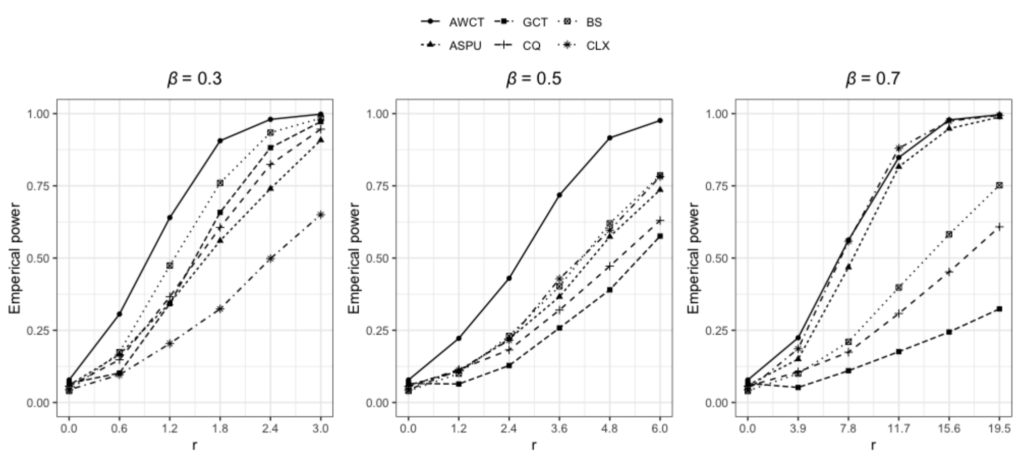

Figure 2: Power curves of the various tests against r under different sparsity levels of β based on Model (a) with heteroscedastic centered gamma(4,2) innovations when $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$.

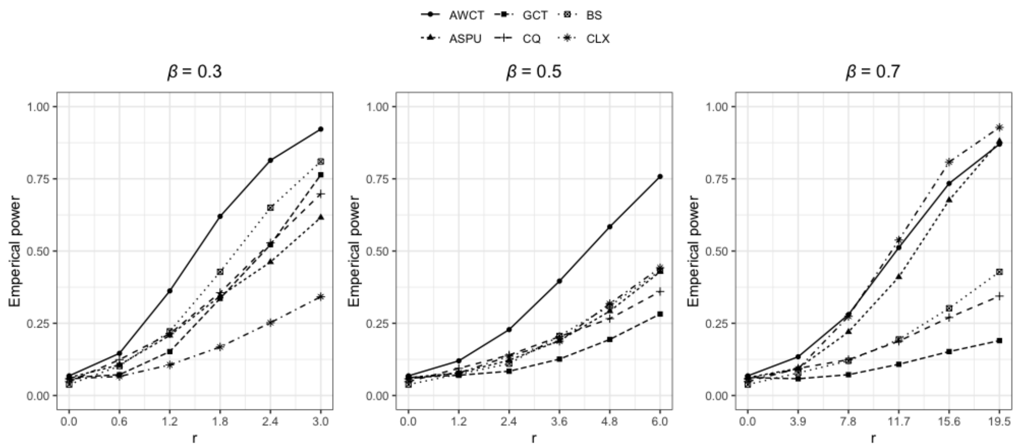

Figure 3: Power curves of the various tests against r under different sparsity levels of β based on Model (a) with heteroscedastic centered gamma(4,2) innovations when $\mathbf{\Sigma}_2 = 2\mathbf{\Sigma}_1$.

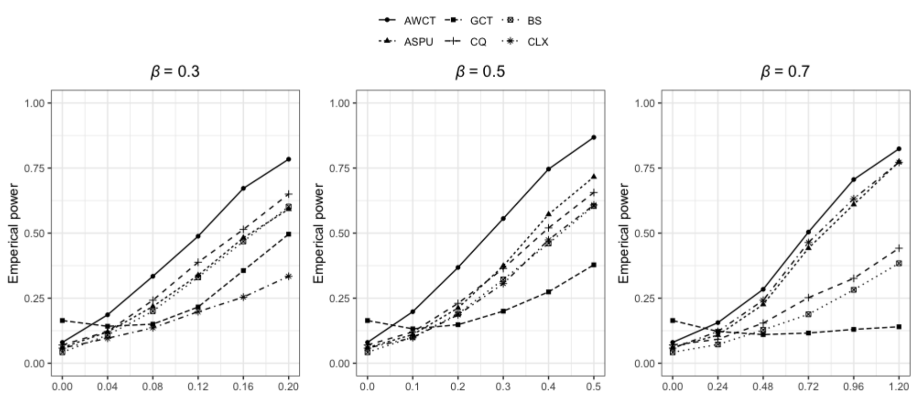

Figure 4: Power curves of the various tests against r under different sparsity levels of β based on Model (a) with normal innovations and $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$ when $p = 100$ and $n_1 = n_2 = 100$.

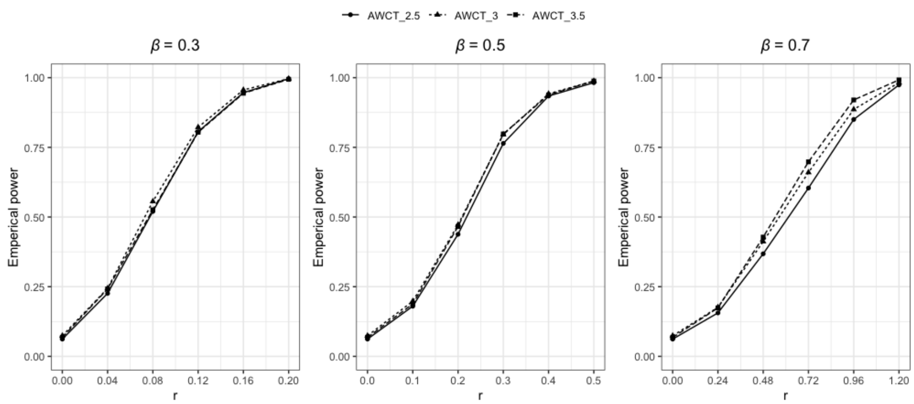

Figure 5: Power curves of the AWCT tests with different choices of R based on Model (a) with normal innovations and $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$ (AWCT_x: AWCT when R=x).

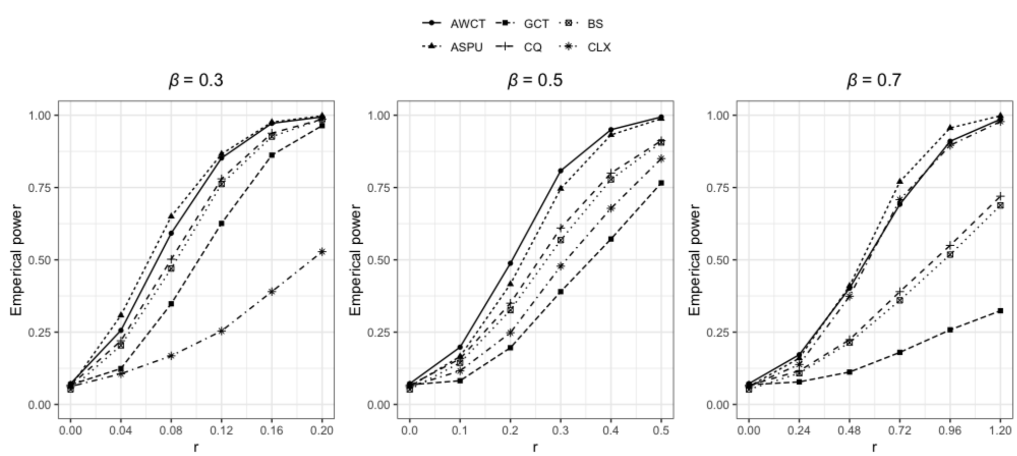

Figure 6: Power curves of the various tests against r under different sparsity levels of β based on Model (b) with normal innovations when $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$.

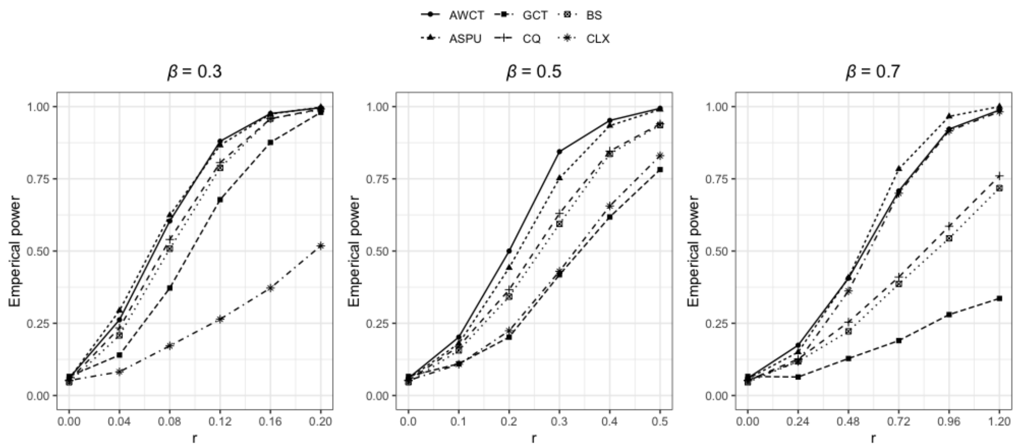

Figure 7: Power curves of the various tests against r under different sparsity levels of β based on Model (c) with normal innovations when $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$.

S5 Real data application: additional results for DNA methylation data

Table 1 summarizes the test results after the elimination of significant GcPs. In this case, it is not surprising to see that the CLX aimed at testing sparse signals fails to reject the null hypothesis. The BS also fails to reject the null hypothesis. One partial reason is that the assumption of equal covariance structures between the two samples is strongly violated by the data, which is required by the BS. In contrast, the other tests, including AWCT, ASPU, GCT and CQ can significantly reject the null hypothesis.

Table 1: The p-values of the various tests for testing equality of the DNA methylation levels measured by β-values on each chromosome after the elimination of significant GcPs.

Chr No.	1	2	3	4	5	6	7	8
AWCT	0	0	0	0	0	0	0	0
ASPU	0	0	0	0	0	0	0	0
GCT	0	0	0	0	0	0	0	0
CQ	3.34×10^{-5}	3.84×10^{-5}	3.35×10^{-5}	2.89×10^{-6}	5.37×10^{-5}	2.66×10^{-5}	1.10×10^{-4}	2.85×10^{-4}
BS	0.41	0.45	0.41	0.30	0.37	0.36	0.43	0.45
CLX	0.32	0.33	0.34	0.33	0.33	0.35	0.33	0.33
Chr No.	9	10	11	12	13	14	15	16
AWCT	0	0	0	0	0	0	0	0
ASPU	0	0	0	0	0	0	0	0
GCT	0	0	0	0	0	0	0	0
CQ	7.89×10^{-5}	3.90×10^{-4}	1.95×10^{-4}	1.85×10^{-5}	2.31×10^{-4}	5.07×10^{-4}	2.53×10^{-3}	6.60×10^{-4}
BS	0.46	0.47	0.45	0.39	0.36	0.47	0.48	0.53
CLX	0.33	0.34	0.33	0.33	0.32	0.35	0.35	0.34
Chr No.	17	18	19	20	21	22	X	
AWCT	0	0	0	0	0	0	0	
ASPU	0	0	0	0	0	0	0	
GCT	0	0	0	0	0	0	0	
CQ	4.52×10^{-4}	4.98×10^{-4}	5.37×10^{-4}	8.96×10^{-4}	2.52×10^{-3}	0.01	3.93×10^{-10}	
BS	0.51	0.40	0.48	0.44	0.39	0.59	9.48×10^{-5}	
CLX	0.33	0.34	0.36	0.35	0.33	0.35	0.28	

S6 Real data application: A semi-conductor manufacturing process

A complex modern semi-conductor manufacturing process is normally under consistent surveillance via the monitoring of signals or variables collected from sensors and or process measurement points. However, not all of these signals are equally valuable in a specific monitoring system. The measured signals contain a combination of useful information, irrelevant information as well as noise. It is often the case that useful information is buried in the latter two.

The dataset consists of 1567 observations each with 591 features and a label indicating a simple pass or failure for in-house line testing. The label of "-1" represents a pass and "1" represents a fail. Among these observations, 1463 are classified as pass and 104 are classified as fail. As the features have very different measurement units, we standardize the data by using their sample standard deviations. Also, some features are constant in the data set, which are removed in the analysis. The missing values are replaced with their sample means in the same group. Furthermore, before testing whether the mean vectors of the two groups (pass and fail) are equal, we also exclude those features with extremely large mean differences, which may dominate

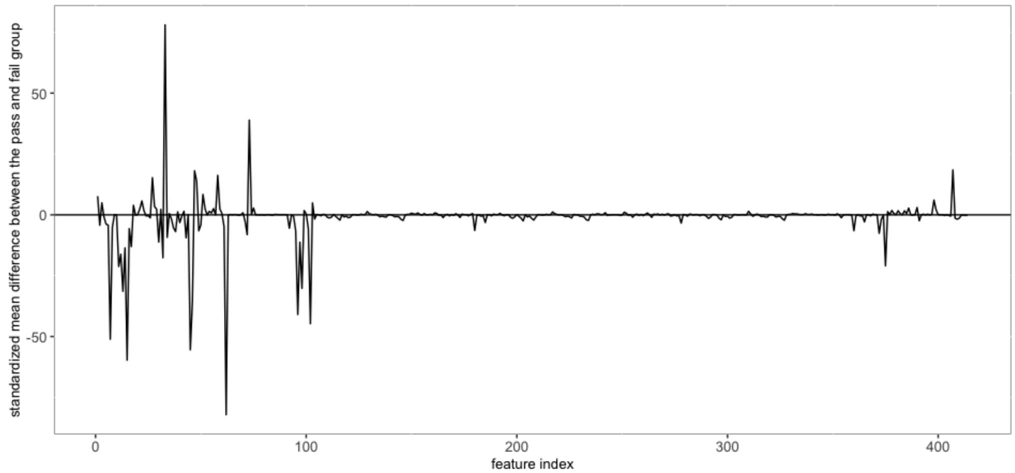

Figure 8: The standardized mean differences for the features between the pass and fail groups.

Table 2: The p-values of the various tests applied to the semi-conductor manufacturing process dataset.

	AWCT	ASPU	GCT	CQ	BS	CLX
p-value	0.000	0.978	0.000	0.680	0.000	0.961

the test results. In particular, we removed some features with p-values less than 0.01 based on the univariate two sample t test on each feature. After that, we only retain 414 features in the final analysis. Figure 8 plots the standardized mean differences for each of these 414 features. Clearly, there exists multiple spikes with relatively large magnitudes.

Table 2 presents the p-values of the above six methods for testing the equality of means between the pass and fail groups. The p-values of the

AWCT ($R = 3$), GCT and BS methods are zero, indicating that there exists a significant mean differences between the two groups. The ASPU, CQ and CLX methods fail to reject the equality of the two mean vectors after deleting those extremely large signals. Also, the κ adaptively chosen by the proposed AWCT method is 0.95, which indicates that the rejection is more likely to stem from accumulation of small differences on many components without those deleted extremely large signals.

References

Hall, P. et al. (1987). Edgeworth expansion for student's t statistic under minimal moment conditions. *The Annals of Probability 15*(3), 920–931.

Ibragimov, I. (1975). Independent and stationary sequences of random variables. *Wolters, Noordhoff Pub.*.

Kim, T. Y. (1994). Moment bounds for non-stationary dependent sequences. *Journal of applied probability 31*(3), 731–742.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America 42*(1), 43.

Yang, S. C. (2007). Maximal moment inequality for partial sums of strong mixing sequences and application. *Acta Mathematica Sinica, English Series 23*(6), 1013–1024.