Residual permutation test for high-dimensional regression coefficient testing

Yuhao Wang

Institute for Interdisciplinary Information Sciences (IIIS), Tsinghua University

yuhaow@tsinghua.edu.cn

Joint work with Kaiyue Wen & Tengyao Wang

January 3, 2024

Problem set up

We revisit one of the most basic problems in statistics: regression coefficient test

Problem set up

We revisit one of the most basic problems in statistics: regression coefficient test

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{Z}b + \varepsilon.$$

- $(\boldsymbol{X} \in \mathbb{R}^{n \times p}, \boldsymbol{Z} \in \mathbb{R}^{n})$: fixed-design;
- $\varepsilon \in \mathbb{R}^{n}$: random noise vector.

Problem set up

We revisit one of the most basic problems in statistics: regression coefficient test

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{Z}b + \varepsilon.$$

- $(\boldsymbol{X} \in \mathbb{R}^{n \times p}, \boldsymbol{Z} \in \mathbb{R}^{n})$: fixed-design;
- $\varepsilon \in \mathbb{R}^{n}$: random noise vector.

Our goal: Test whether

$$H_0 : b = 0 \quad \text{v.s.} \quad H_1 : b \neq 0$$

when $p \asymp n$.

$\hookrightarrow$ Sometimes also referred to as moderately high-dimensional regime.

Types of validity

- Asymptotic validity: *asymptotically* correct size control as $n \to \infty$.
  - Ex: OLS fit based, Freedman and Lane (1983), residual bootstrap, DiCiccio and Romano (2017), Toulis (2019), debiased lasso.

- Finite-population validity: valid size control with arbitrary n.

$\Rightarrow$ Our **target of interest**

Finite-population validity: history

- ANOVA (Fisher 1921): requires ε to be i.i.d. Gaussian;

Finite-population validity: history

- ANOVA (Fisher 1921): requires ε to be i.i.d. Gaussian;

- Hartigan (1970), Meinshuasen (2015): symmetric around zero or rotationally invariant;

Finite-population validity: history

- ANOVA (Fisher 1921): requires ε to be i.i.d. Gaussian;
- Hartigan (1970), Meinshuasen (2015): symmetric around zero or rotationally invariant;
- **Distribution-free valid test** (Lei and Bickle, 2021): just requires ε to be exchangeable for correct size control;

Finite-population validity: history

- ANOVA (Fisher 1921): requires ε to be i.i.d. Gaussian;
- Hartigan (1970), Meinshuasen (2015): symmetric around zero or rotationally invariant;
- **Distribution-free valid test** (Lei and Bickle, 2021): just requires ε to be exchangeable for correct size control;
  - Limitation: strong assumptions on dimension of $\boldsymbol{X}$:

$$n/p > 1/\alpha + 1$$

⇑ prespecified Type-I error

$\alpha = 0.01, n = 300 : p < 3.$

Our contributions

- Finite-population & distribution-free validity: whenever $p < n/2$;

Our contributions

- Finite-population & distribution-free validity: whenever $p < n/2$;
- Heavy-tail friendly: non-trivial power even when $\mathbb{E}[\varepsilon_i^2] = \infty$:

Our contributions

- Finite-population & distribution-free validity: whenever $p < n/2$;

- Heavy-tail friendly: non-trivial power even when $\mathbb{E}[\varepsilon_i^2] = \infty$:

when $\varepsilon_1, \varepsilon_2, \ldots$ are independent with uniformly bounded $(1+t)$-th order moment for $t \in [0,1]$, our test can still have power even when b is as small as $n^{-t/(1+t)}$.

Our contributions

- Finite-population & distribution-free validity: whenever $p < n/2$;

- Heavy-tail friendly: non-trivial power even when $\mathbb{E}[\varepsilon_i^2] = \infty$:

when $\varepsilon_1, \varepsilon_2, \ldots$ are independent with uniformly bounded $(1+t)$-th order moment for $t \in [0,1]$, our test can still have power even when b is as small as $n^{-t/(1+t)}$.

- Minimax rate optimality: we prove that $n^{-t/(1+t)}$ matches the minimax lower bound rate for coefficient test with heavy-tailed noises.

Numerical analysis of ANOVA's validity

Simulations for general noise:

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon$$
$$\boldsymbol{Z} = \boldsymbol{X}\beta^{Z} + \boldsymbol{e}$$

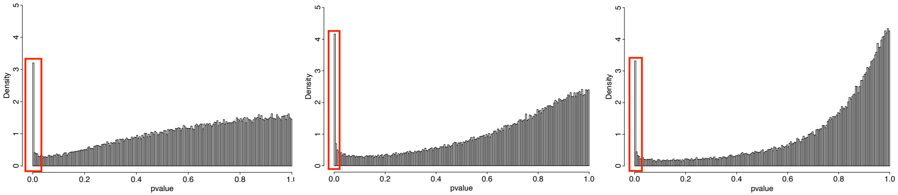
- $(n, p) = (300, 100), (600, 100), (600, 200)$;
- $\boldsymbol{X}$: Gaussian design, t_1 design;
- $\boldsymbol{e}, \varepsilon$: t_1 noise, t_2 noise, Gaussian noise.

Validity of ANOVA

n	p	X type	noise type	0.01	0.005
300	100	Gaussian	Gaussian	0.0101	0.0050
300	100	Gaussian	t_1	0.0181	0.0160
300	100	Gaussian	t_2	0.0153	0.0107
300	100	t_1	Gaussian	0.0101	0.0050
300	100	t_1	t_1	0.0243	0.0208
300	100	t_1	t_2	0.0180	0.0130
600	200	Gaussian	Gaussian	0.0101	0.0049
600	200	Gaussian	t_1	0.0141	0.0122
600	200	Gaussian	t_2	0.0150	0.0104
600	200	t_1	Gaussian	0.0101	0.0049
600	200	t_1	t_1	0.0202	0.0173
600	200	t_1	t_2	0.0170	0.0120

Table: empirical size with nominal levels $\alpha = 0.01$ and 0.005

Histogram of ANOVA's p-values



(a) (b) (c)

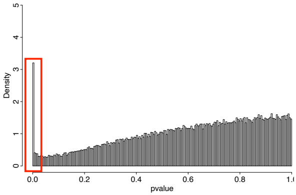
(a) $n = 300, p = 100$, Gaussian design, t_1 noises;

(b) $n = 300, p = 100$, t_1 design, t_1 noises;

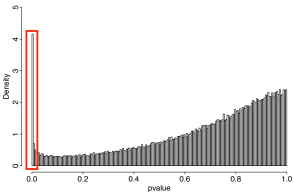
(c) $n = 600, p = 100$, Gaussian design, t_1 noises;

$\Rightarrow$ highest spike in heavy-tail design + heavy-tail noise.

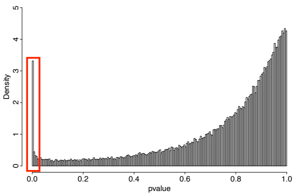
Histogram of ANOVA's p-values



(d)



(e)



(f)

(a) $n = 300, p = 100$, Gaussian design, t_1 noises;

(b) $n = 300, p = 100$, t_1 design, t_1 noises;

(c) $n = 600, p = 100$, Gaussian design, t_1 noises;

$\Rightarrow$ highest spike in heavy-tail design + heavy-tail noise.

This shows the importance of developing a distribution-free & finite-population valid test!!

Residual permutation test (RPT)

Residual permutation test (RPT)

1. Given permutation matrices $\boldsymbol{P}_1, \cdots, \boldsymbol{P}_K$:

Residual permutation test (RPT)

1. Given permutation matrices $\boldsymbol{P}_1, \cdots, \boldsymbol{P}_K$:
   - $\tilde{\boldsymbol{V}}_k \in \mathbb{R}^{n\times(n-2p)}$: orthonormal matrix perpendicular to $\mathrm{span}(\boldsymbol{X}) \cup \mathrm{span}(\boldsymbol{P}_k\boldsymbol{X})$.

Residual permutation test (RPT)

1. Given permutation matrices $\boldsymbol{P}_1, \cdots, \boldsymbol{P}_K$:
   - $\tilde{\boldsymbol{V}}_k \in \mathbb{R}^{n \times (n-2p)}$: orthonormal matrix perpendicular to $\mathrm{span}(\boldsymbol{X}) \cup \mathrm{span}(\boldsymbol{P}_k \boldsymbol{X})$.

2. p-value:

$$\phi = \frac{1}{1+K}\left(1 + \sum_{k=1}^{K} \mathbb{1}\left\{\min_{1 \le k' \le K} T\left(\tilde{\boldsymbol{V}}_{k'}^\top \boldsymbol{Z}, \tilde{\boldsymbol{V}}_{k'}^\top \boldsymbol{Y}\right) \le T\left(\tilde{\boldsymbol{V}}_{k}^\top \boldsymbol{Z}, \tilde{\boldsymbol{V}}_{k}^\top \boldsymbol{P}_k \boldsymbol{Y}\right)\right\}\right)$$

$\Rightarrow$ Projecting $(\boldsymbol{Y}, \boldsymbol{P}_k \boldsymbol{Y})$ onto $\mathrm{span}(\tilde{\boldsymbol{V}}_k)$ & compare.

Why residual permutation test?

- Classical regression residual:

$$\hat{\boldsymbol{R}}_Y = (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top)\boldsymbol{Y}$$

$\rightsquigarrow$ Projecting $\boldsymbol{Y}$ onto the space orthogonal to $\boldsymbol{X}$;

Why residual permutation test?

- Classical regression residual:

$$\hat{\boldsymbol{R}}_Y = (\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\top \boldsymbol{X})^{-1}\boldsymbol{X}^\top)\boldsymbol{Y}$$

$\rightsquigarrow$ Projecting $\boldsymbol{Y}$ onto the space orthogonal to $\boldsymbol{X}$;

- $\tilde{\boldsymbol{V}}_k^\top \boldsymbol{Y}$:

$\rightsquigarrow$ a residual by regressing $\boldsymbol{Y}$ onto both $\boldsymbol{X}$ & $\boldsymbol{P}_k\boldsymbol{X}$ ...

Analysis of RPT's validity

- $\tilde{\boldsymbol{V}}_k$ is orthogonal to the space by $\boldsymbol{X}$ & $\boldsymbol{P}_k\boldsymbol{X}$ and we are under H_0:

Analysis of RPT's validity

- $\tilde{\boldsymbol{V}}_k$ is orthogonal to the space by $\boldsymbol{X}$ & $\boldsymbol{P}_k\boldsymbol{X}$ and we are under H_0:

$$\tilde{\boldsymbol{V}}_k^\top \boldsymbol{P}_k \boldsymbol{Y} \overset{\text{under } H_0}{=} \tilde{\boldsymbol{V}}_k^\top \boldsymbol{P}_k \boldsymbol{X}\beta + \tilde{\boldsymbol{V}}_k^\top \boldsymbol{P}_k \boldsymbol{\varepsilon}$$

Analysis of RPT's validity

- $\tilde{\boldsymbol{V}}_k$ is orthogonal to the space by $\boldsymbol{X}$ & $\boldsymbol{P}_k\boldsymbol{X}$ and we are under H_0:

$$\tilde{\boldsymbol{V}}_k^\top \boldsymbol{P}_k \boldsymbol{Y} \overset{\text{under } H_0}{=} \tilde{\boldsymbol{V}}_k^\top \boldsymbol{P}_k \boldsymbol{X}\beta + \tilde{\boldsymbol{V}}_k^\top \boldsymbol{P}_k \boldsymbol{\varepsilon} = \tilde{\boldsymbol{V}}_k^\top \boldsymbol{P}_k \boldsymbol{\varepsilon}$$

Analysis of RPT's validity

- $\tilde{\boldsymbol{V}}_k$ is orthogonal to the space by $\boldsymbol{X}$ & $\boldsymbol{P}_k\boldsymbol{X}$ and we are under H_0:

$$\tilde{\boldsymbol{V}}_k^\top \boldsymbol{P}_k \boldsymbol{Y} \overset{\text{under } H_0}{=} \tilde{\boldsymbol{V}}_k^\top \boldsymbol{P}_k \boldsymbol{X}\beta + \tilde{\boldsymbol{V}}_k^\top \boldsymbol{P}_k \boldsymbol{\varepsilon} = \tilde{\boldsymbol{V}}_k^\top \boldsymbol{P}_k \boldsymbol{\varepsilon}$$

$$\tilde{\boldsymbol{V}}_k^\top \boldsymbol{Y} = \tilde{\boldsymbol{V}}_k^\top \boldsymbol{\varepsilon}$$

Analysis of RPT's validity

- $\tilde{\boldsymbol{V}}_k$ is orthogonal to the space by $\boldsymbol{X}$ & $\boldsymbol{P}_k\boldsymbol{X}$ and we are under H_0:

$$\tilde{\boldsymbol{V}}_k^\top \boldsymbol{P}_k \boldsymbol{Y} \overset{\text{under } H_0}{=} \tilde{\boldsymbol{V}}_k^\top \boldsymbol{P}_k \boldsymbol{X}\beta + \tilde{\boldsymbol{V}}_k^\top \boldsymbol{P}_k \boldsymbol{\varepsilon} = \tilde{\boldsymbol{V}}_k^\top \boldsymbol{P}_k \boldsymbol{\varepsilon}$$
$$\tilde{\boldsymbol{V}}_k^\top \boldsymbol{Y} = \tilde{\boldsymbol{V}}_k^\top \boldsymbol{\varepsilon}$$

- We can rewrite the p-value as:

$$\phi = \frac{1}{1+K}\left(1 + \sum_{k=1}^{K} \mathbb{1}\left\{\min_{1\le k'\le K} T\left(\tilde{\boldsymbol{V}}_{k'}^\top \boldsymbol{Z}, \tilde{\boldsymbol{V}}_{k'}^\top \boldsymbol{\varepsilon}\right) \le T\left(\tilde{\boldsymbol{V}}_k^\top \boldsymbol{Z}, \tilde{\boldsymbol{V}}_k^\top \boldsymbol{P}_k \boldsymbol{\varepsilon}\right)\right\}\right)$$

Analysis of RPT's validity

- $\tilde{\boldsymbol{V}}_k$ is orthogonal to the space by $\boldsymbol{X}$ & $\boldsymbol{P}_k\boldsymbol{X}$ and we are under H_0:

$$\tilde{\boldsymbol{V}}_k^\top \boldsymbol{P}_k \boldsymbol{Y} \overset{\text{under } H_0}{=} \tilde{\boldsymbol{V}}_k^\top \boldsymbol{P}_k \boldsymbol{X}\beta + \tilde{\boldsymbol{V}}_k^\top \boldsymbol{P}_k \boldsymbol{\varepsilon} = \tilde{\boldsymbol{V}}_k^\top \boldsymbol{P}_k \boldsymbol{\varepsilon}$$
$$\tilde{\boldsymbol{V}}_k^\top \boldsymbol{Y} = \tilde{\boldsymbol{V}}_k^\top \boldsymbol{\varepsilon}$$

- We can rewrite the p-value as:

$$\phi = \frac{1}{1+K}\left(1 + \sum_{k=1}^{K} \mathbb{1}\left\{\min_{1\le k'\le K} T\left(\tilde{\boldsymbol{V}}_{k'}^\top \boldsymbol{Z}, \tilde{\boldsymbol{V}}_{k'}^\top \boldsymbol{\varepsilon}\right) \le T\left(\tilde{\boldsymbol{V}}_k^\top \boldsymbol{Z}, \tilde{\boldsymbol{V}}_k^\top \boldsymbol{P}_k \boldsymbol{\varepsilon}\right)\right\}\right)$$
$$\ge \frac{1}{1+K}\left(1 + \sum_{k=1}^{K} \mathbb{1}\left\{\min_{1\le k'\le K} T\left(\tilde{\boldsymbol{V}}_{k'}^\top \boldsymbol{Z}, \tilde{\boldsymbol{V}}_{k'}^\top \boldsymbol{\varepsilon}\right) \le \min_{1\le k'\le K} T\left(\tilde{\boldsymbol{V}}_{k'}^\top \boldsymbol{Z}, \tilde{\boldsymbol{V}}_{k'}^\top \boldsymbol{P}_k \boldsymbol{\varepsilon}\right)\right\}\right)$$

Analysis of RPT's validity

- $\tilde{\boldsymbol{V}}_k$ is orthogonal to the space by $\boldsymbol{X}$ & $\boldsymbol{P}_k\boldsymbol{X}$ and we are under H_0:

$$\tilde{\boldsymbol{V}}_k^\top \boldsymbol{P}_k \boldsymbol{Y} \overset{\text{under } H_0}{=} \tilde{\boldsymbol{V}}_k^\top \boldsymbol{P}_k \boldsymbol{X}\beta + \tilde{\boldsymbol{V}}_k^\top \boldsymbol{P}_k \boldsymbol{\varepsilon} = \tilde{\boldsymbol{V}}_k^\top \boldsymbol{P}_k \boldsymbol{\varepsilon}$$
$$\tilde{\boldsymbol{V}}_k^\top \boldsymbol{Y} = \tilde{\boldsymbol{V}}_k^\top \boldsymbol{\varepsilon}$$

- We can rewrite the p-value as:

$$\begin{aligned}
\phi &= \frac{1}{1+K}\left(1+\sum_{k=1}^{K}\mathbb{1}\left\{\min_{1\le k'\le K} T\left(\tilde{\boldsymbol{V}}_{k'}^\top \boldsymbol{Z}, \tilde{\boldsymbol{V}}_{k'}^\top \boldsymbol{\varepsilon}\right) \le T\left(\tilde{\boldsymbol{V}}_{k}^\top \boldsymbol{Z}, \tilde{\boldsymbol{V}}_{k}^\top \boldsymbol{P}_k\boldsymbol{\varepsilon}\right)\right\}\right) \\
&\ge \frac{1}{1+K}\left(1+\sum_{k=1}^{K}\mathbb{1}\left\{\min_{1\le k'\le K} T\left(\tilde{\boldsymbol{V}}_{k'}^\top \boldsymbol{Z}, \tilde{\boldsymbol{V}}_{k'}^\top \boldsymbol{\varepsilon}\right) \le \min_{1\le k'\le K} T\left(\tilde{\boldsymbol{V}}_{k'}^\top \boldsymbol{Z}, \tilde{\boldsymbol{V}}_{k'}^\top \boldsymbol{P}_k\boldsymbol{\varepsilon}\right)\right\}\right) \\
&= \frac{1}{1+K}\left(1+\sum_{k=1}^{K}\mathbb{1}\left\{g(\boldsymbol{\varepsilon}) \le g(\boldsymbol{P}_k\boldsymbol{\varepsilon})\right\}\right)
\end{aligned}$$

for some $g(\cdot)$ depending only on $\boldsymbol{X}, \boldsymbol{Z}, \mathcal{P} := \{\boldsymbol{P}_0 = \boldsymbol{I}, \boldsymbol{P}_1, \cdots, \boldsymbol{P}_K\}$.

Analysis of RPT's validity

Remaining challenge:

$$\text{Prove } \phi \geq \frac{1}{1+K}\left(1+\sum_{k=1}^{K}\mathbb{1}\left\{g(\varepsilon) \leq g(\boldsymbol{P}_k\varepsilon)\right\}\right) \text{ is a valid p-value} \quad (1)$$

However, here $g(\cdot)$ depends on $\mathcal{P}$.

Analysis of RPT's validity

Remaining challenge:

$$\text{Prove } \phi \geq \frac{1}{1+K}\left(1+\sum_{k=1}^{K} \mathbb{1}\left\{g(\boldsymbol{\varepsilon}) \leq g(\boldsymbol{P}_k\boldsymbol{\varepsilon})\right\}\right) \text{ is a valid p-value} \quad (1)$$

However, here $g(\cdot)$ depends on $\mathcal{P}$.

Lemma

Suppose we construct $\mathcal{P} := \{\boldsymbol{P}_0 := \boldsymbol{I}, \boldsymbol{P}_1, \ldots, \boldsymbol{P}_K\}$ *s.t. it formalizes a group:*

$$\forall \boldsymbol{P}_i, \boldsymbol{P}_j \in \mathcal{P}, \exists \boldsymbol{P}_\ell \ s.t. \ \boldsymbol{P}_\ell := \boldsymbol{P}_i \boldsymbol{P}_j.$$

Then (1) *is a valid p-value.*

Finite-population validity of RPT

Theorem

Suppose

- *the set of permutation matrices $\mathcal{P}$ formalizes a group;*

Finite-population validity of RPT

Theorem

Suppose

- *the set of permutation matrices $\mathcal{P}$ formalizes a group;*
- *ε is exchangeable;*

Finite-population validity of RPT

Theorem

Suppose

- *the set of permutation matrices* $\mathcal{P}$ *formalizes a group;*
- ε *is exchangeable;*
- $p < n/2$;

Finite-population validity of RPT

Theorem

Suppose

- *the set of permutation matrices* $\mathcal{P}$ *formalizes a group;*
- ε *is exchangeable;*
- $p < n/2$*;*

under H_0, ϕ *is a valid p-value:* $\mathbb{P}(\phi \leq \alpha) \leq \alpha \ \forall \alpha \in [0, 1]$.

Finite-population validity of RPT

Theorem

Suppose

- *the set of permutation matrices* $\mathcal{P}$ *formalizes a group;*
- ε *is exchangeable;*
- $p < n/2$;

under H_0, ϕ *is a valid p-value:* $\mathbb{P}(\phi \leq \alpha) \leq \alpha \; \forall \alpha \in [0, 1]$.

Remark

Finite-population validity of RPT

Theorem

Suppose

- *the set of permutation matrices $\mathcal{P}$ formalizes a group;*
- *ε is exchangeable;*
- *$p < n/2$;*

under H_0, ϕ is a valid p-value: $\mathbb{P}(\phi \leq \alpha) \leq \alpha \; \forall \alpha \in [0, 1]$.

Remark

1. Construction of $\tilde{\boldsymbol{V}}_k$ requires $p < n/2$;

Finite-population validity of RPT

Theorem

Suppose

- *the set of permutation matrices $\mathcal{P}$ formalizes a group;*
- *ε is exchangeable;*
- *$p < n/2$;*

under H_0, ϕ is a valid p-value: $\mathbb{P}(\phi \leq \alpha) \leq \alpha \ \forall \alpha \in [0, 1]$.

Remark

1. Construction of $\tilde{\boldsymbol{V}}_k$ requires $p < n/2$;
2. With prespecified α, one needs to choose $K > 1/\alpha$ to have power.

Power analysis

Model of Z:

$$\boldsymbol{Z} = \boldsymbol{X}\beta^Z + \boldsymbol{e}.$$

Power analysis

Model of Z:

$$\boldsymbol{Z} = \boldsymbol{X}\beta^Z + \boldsymbol{e}.$$

Theorem

Assume $\varepsilon_1, \cdots, \varepsilon_n \overset{i.i.d.}{\sim} \mathbb{P}_\varepsilon$ *&* $e_1, \cdots, e_n \overset{i.i.d.}{\sim} \mathbb{P}_e$ *and*

$$0 < \mathbb{E}[|e_1|^2] < \infty \qquad and \qquad 0 < \mathbb{E}[|\varepsilon_1|^{1+t}] < \infty$$

for $t \in [0, 1)$.

Power analysis

Model of Z:

$$\boldsymbol{Z} = \boldsymbol{X}\beta^Z + \boldsymbol{e}.$$

Theorem

Assume $\varepsilon_1, \cdots, \varepsilon_n \overset{i.i.d.}{\sim} \mathbb{P}_\varepsilon$ *&* $e_1, \cdots, e_n \overset{i.i.d.}{\sim} \mathbb{P}_e$ *and*

$$0 < \mathbb{E}[|e_1|^2] < \infty \qquad and \qquad 0 < \mathbb{E}[|\varepsilon_1|^{1+t}] < \infty$$

for $t \in [0, 1)$. *Then if* $n > (3+m)p$ *for const.* $m > 0$ *&* $b = \Omega(n^{-t/(1+t)})$,

Power analysis

Model of Z:

$$\boldsymbol{Z} = \boldsymbol{X}\beta^Z + \boldsymbol{e}.$$

Theorem

Assume $\varepsilon_1, \cdots, \varepsilon_n \overset{i.i.d.}{\sim} \mathbb{P}_\varepsilon$ *&* $e_1, \cdots, e_n \overset{i.i.d.}{\sim} \mathbb{P}_e$ *and*

$$0 < \mathbb{E}[|e_1|^2] < \infty \qquad and \qquad 0 < \mathbb{E}[|\varepsilon_1|^{1+t}] < \infty$$

for $t \in [0, 1)$. *Then if* $n > (3 + m)p$ *for const.* $m > 0$ *&* $b = \Omega(n^{-t/(1+t)})$,

$$\lim_{n\to\infty} \mathbb{P}\left(\phi > \frac{1}{K+1}\right) = 0.$$

Remarks about power analysis

- $\boldsymbol{Z}$ is a linear model w.r.t. $\boldsymbol{X}$ & all noises i.i.d. are just for simplicity of illustration, not **necessary conditions**;

Remarks about power analysis

- $\boldsymbol{Z}$ is a linear model w.r.t. $\boldsymbol{X}$ & all noises i.i.d. are just for simplicity of illustration, not **necessary conditions**;
- In our paper, we proved that the same conclusion still holds when $\boldsymbol{Z}$ is a nonlinear func. & noises are heteroschedastic:

Remarks about power analysis

- $\boldsymbol{Z}$ is a linear model w.r.t. $\boldsymbol{X}$ & all noises i.i.d. are just for simplicity of illustration, not **necessary conditions**;
- In our paper, we proved that the same conclusion still holds when $\boldsymbol{Z}$ is a nonlinear func. & noises are heteroschedastic:

Theorem

Consider a broader class of alternative where $\boldsymbol{Z}$ *is* **nonlinear** *w.r.t.* $\boldsymbol{X}$ *(i.e.,* $\boldsymbol{Z} = \boldsymbol{X}\beta^Z + \boldsymbol{h} + \boldsymbol{e}$*) and all noises are* **heteroschdastic**, *under*

Remarks about power analysis

- $\boldsymbol{Z}$ is a linear model w.r.t. $\boldsymbol{X}$ & all noises i.i.d. are just for simplicity of illustration, not **necessary conditions**;
- In our paper, we proved that the same conclusion still holds when $\boldsymbol{Z}$ is a nonlinear func. & noises are heteroschedastic:

Theorem

*Consider a broader class of alternative where $\boldsymbol{Z}$ is **nonlinear** w.r.t. $\boldsymbol{X}$ (i.e., $\boldsymbol{Z} = \boldsymbol{X}\beta^Z + \boldsymbol{h} + \boldsymbol{e}$) and all noises are **heteroschdastic**, under*

1. *Standard regularity conditions on degree of heteroschdasticity of ε & $\boldsymbol{e}$;*

Remarks about power analysis

- $\boldsymbol{Z}$ is a linear model w.r.t. $\boldsymbol{X}$ & all noises i.i.d. are just for simplicity of illustration, not **necessary conditions**;
- In our paper, we proved that the same conclusion still holds when $\boldsymbol{Z}$ is a nonlinear func. & noises are heteroschedastic:

Theorem

Consider a broader class of alternative where $\boldsymbol{Z}$ is **nonlinear** *w.r.t. $\boldsymbol{X}$ (i.e., $\boldsymbol{Z} = \boldsymbol{X}\beta^Z + \boldsymbol{h} + \boldsymbol{e}$) and all noises are* **heteroschdastic**, *under*

1. *Standard regularity conditions on degree of heteroschdasticity of ε & $\boldsymbol{e}$;*
2. $\limsup \|\boldsymbol{h}\|_2/\sqrt{n} \leq r \cdot \text{noise level}$
   $\Rightarrow r$: *const. depending only on* $\limsup(p/n)$ & $(\max_i \text{var}(e_i))/\bar{\sigma}_e^2$;

Remarks about power analysis

- $\boldsymbol{Z}$ is a linear model w.r.t. $\boldsymbol{X}$ & all noises i.i.d. are just for simplicity of illustration, not **necessary conditions**;
- In our paper, we proved that the same conclusion still holds when $\boldsymbol{Z}$ is a nonlinear func. & noises are heteroschedastic:

Theorem

*Consider a broader class of alternative where $\boldsymbol{Z}$ is **nonlinear** w.r.t. $\boldsymbol{X}$ (i.e., $\boldsymbol{Z} = \boldsymbol{X}\beta^Z + \boldsymbol{h} + \boldsymbol{e}$) and all noises are **heteroschdastic**, under*

1. *Standard regularity conditions on degree of heteroschdasticity of ε & $\boldsymbol{e}$;*
2. $\limsup \|\boldsymbol{h}\|_2/\sqrt{n} \leq r \cdot \text{noise level}$
   $\Rightarrow r$: *const. depending only on* $\limsup(p/n)$ *&* $(\max_i \mathrm{var}(e_i))/\bar{\sigma}_e^2$;

The signal detection rate $n^{-t/(1+t)}$ still holds.

Minimax rate optimality

- We derive that the minimax lower bound rate of separation is of order $n^{-t/(1+t)}$ for heavy-tailed distribution;

$\Rightarrow$ matches the **pointwise** upper bound of RPT.

Minimax rate optimality

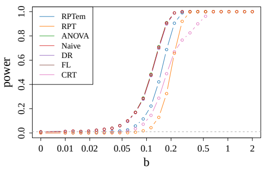
- We derive that the minimax lower bound rate of separation is of order $n^{-t/(1+t)}$ for heavy-tailed distribution;

$\Rightarrow$ matches the **pointwise** upper bound of RPT.

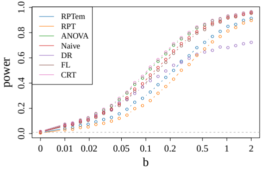
- We derive the uniform convergence rate of RPT is of $n^{-t/(1+t)+\delta}$ for any const. $\delta > 0$.

$\Rightarrow$ RPT nearly minimax rate optimal.

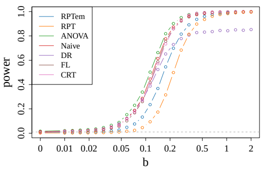
Power curves



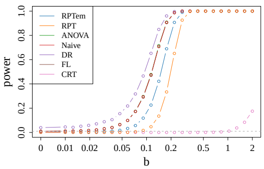
(g) Gaussian design, Gaussian noise



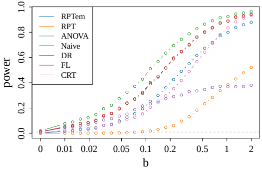
(h) Gaussian design, t_1 noise



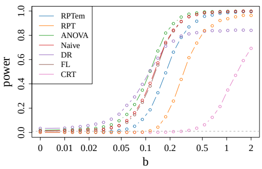
(i) Gaussian design, t_2 noise



(j) t_1 design, Gaussian noise



(k) t_1 design, t_1 noise



(l) t_1 design, t_2 noise

Figure: $n = 600, p = 100$

Power analysis

- **Theoretical power analysis:** RPT attains nearly *minimax optimal* rate as $n \to \infty$;

Power analysis

- **Theoretical power analysis:** RPT attains nearly *minimax optimal* rate as $n \to \infty$;

- **Finite population simulation:** when n is small, empirically RPT can still be more conservative than those invalid tests, especially for heavy-tailed ε;

Power analysis

- **Theoretical power analysis:** RPT attains nearly *minimax optimal* rate as $n \to \infty$;

- **Finite population simulation:** when n is small, empirically RPT can still be more conservative than those invalid tests, especially for heavy-tailed ε;

  $\Rightarrow$ A cost to pay for distribution-free & finite-population validity

Power analysis

- **Theoretical power analysis:** RPT attains nearly *minimax optimal* rate as $n \to \infty$;

- **Finite population simulation:** when n is small, empirically RPT can still be more conservative than those invalid tests, especially for heavy-tailed ε;

  $\Rightarrow$ A cost to pay for distribution-free & finite-population validity

- **Open question:** how to develop a distribution-free & finite-population valid test with better empirical power in small sample size.

Summary of contributions

- We have proposed RPT, which is distribution-free valid whenever $p < n/2$;

Summary of contributions

- We have proposed RPT, which is distribution-free valid whenever $p < n/2$;
- We analyze the signal detection rate of RPT and show that it nearly achieves the minimax lower bound rate;

Summary of contributions

- We have proposed RPT, which is distribution-free valid whenever $p < n/2$;
- We analyze the signal detection rate of RPT and show that it nearly achieves the minimax lower bound rate;
- We compared empirically with the other state of the art approaches;

Summary of contributions

- We have proposed RPT, which is distribution-free valid whenever $p < n/2$;
- We analyze the signal detection rate of RPT and show that it nearly achieves the minimax lower bound rate;
- We compared empirically with the other state of the art approaches;

For theoretical details and more simulation results, please see `https://arxiv.org/abs/2211.16182`