CPT: An Assumption-Free Exact Test For Linear Model With Exchangeable Errors

Lihua Lei and Peter J. Bickel

Department of Statistics
UC Berkeley

Stanford-Berkeley Joint Colloquium, 2017

Table of Contents

Table of Contents

Setup

Consider a linear model:

$$y = X\beta + \epsilon$$

where

- $y \in \mathbb{R}^n$: response vector;
- $X \in \mathbb{R}^{n\times p}$: **fixed** design matrix;
- $\beta \in \mathbb{R}^p$: coefficient vector;
- $\epsilon \in \mathbb{R}^n$: random error terms.

Goal: Test $H_0 : \beta_1 = 0$ (v.s. $H_1 : \beta_1 > 0$ or $H_1 : \beta_1 \neq 0$)

Existing Methods: Regression methods

Examples:

- marginal t-test on OLS estimator;
- marginal z-test on general M-estimator, etc.

Pros:

- easily computed;
- intuitive and interpretable;
- implemented in standard software;

Cons:

- requires distributional conditions on $\mathcal{L}(\epsilon)$;
- requires geometric conditions on X;
- only controls Type-I error asymptotically.

Existing Methods: Permutation Tests

Examples:

- permute X_1 and recomputes $\hat{\beta}_1$ as the null population;
- permute the regression residuals and recomputes $\hat{\beta}_1$ as the null population (Freedman and Lane, 1983);

Pros:

- easily computed;
- only requires exchangeability of ϵ_i's;

Cons:

- only works for random designs;
- only controls Type-I error asymptotically when $p > 1$.

Existing Methods: Fiducial Methods

Examples:

- Group-bound (Meinshausen, 2015).

Pros:

- requires no assumption on X;
- works for $p > n$ in the sparse case;
- valid in finite-samples;

Cons:

- requires extremely strong assumption on $\mathcal{L}(\epsilon)$;
- potentially low power due to the artificial factors.

Our Contribution

We derive a test, referred to as **Cyclic Permutation Test (CPT)**,

Our Contribution

We derive a test, referred to as **Cyclic Permutation Test (CPT)**,

- which is valid in finite-sample;

Our Contribution

We derive a test, referred to as **Cyclic Permutation Test (CPT)**,

- which is valid in finite-sample;
- with exact coverage α;

Our Contribution

We derive a test, referred to as **Cyclic Permutation Test (CPT)**,

- which is valid in finite-sample;
- with exact coverage α;
- for arbitrary *fixed* design matrix X;

Our Contribution

We derive a test, referred to as **Cyclic Permutation Test (CPT)**,

- which is valid in finite-sample;
- with exact coverage α;
- for arbitrary *fixed* design matrix X;
- and for arbitrary exchangeable errors ϵ;

Our Contribution

We derive a test, referred to as **Cyclic Permutation Test (CPT)**,

- which is valid in finite-sample;
- with exact coverage α;
- for arbitrary *fixed* design matrix X;
- and for arbitrary exchangeable errors ϵ;
- with reasonable power in various practical situations.

Table of Contents

Run linear regression with X_1:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{pmatrix} \sim \begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ X_{31} & X_{32} & \cdots & X_{3p} \\ X_{41} & X_{42} & \cdots & X_{4p} \\ X_{51} & X_{52} & \cdots & X_{5p} \\ X_{61} & X_{62} & \cdots & X_{6p} \end{pmatrix}$$

$$\left(\hat{\beta}_1^{(0)}, \ \hat{\beta}_2^{(0)}, \ \ldots, \ \hat{\beta}_p^{(0)} \right)$$

A pool of estimates: $\hat{\beta}_1^{(0)}$

Motivation

Run linear regression with permuted X_1:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{pmatrix} \sim \begin{pmatrix} X_{31} & X_{12} & \cdots & X_{1p} \\ X_{61} & X_{22} & \cdots & X_{2p} \\ X_{11} & X_{32} & \cdots & X_{3p} \\ X_{21} & X_{42} & \cdots & X_{4p} \\ X_{41} & X_{52} & \cdots & X_{5p} \\ X_{51} & X_{62} & \cdots & X_{6p} \end{pmatrix}$$

$$\left(\hat{\beta}_1^{(1)}, \ \hat{\beta}_2^{(1)}, \ \ldots, \ \hat{\beta}_p^{(1)} \right)$$

A pool of estimates: $\hat{\beta}_1^{(0)}$, $\hat{\beta}_1^{(1)}$

Motivation

Run linear regression with permuted X_1:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{pmatrix} \sim \begin{pmatrix} X_{41} & X_{12} & \cdots & X_{1p} \\ X_{51} & X_{22} & \cdots & X_{2p} \\ X_{31} & X_{32} & \cdots & X_{3p} \\ X_{61} & X_{42} & \cdots & X_{4p} \\ X_{11} & X_{52} & \cdots & X_{5p} \\ X_{21} & X_{62} & \cdots & X_{6p} \end{pmatrix}$$

$$\left(\hat{\beta}_1^{(2)}, \hat{\beta}_2^{(2)}, \ldots, \hat{\beta}_p^{(2)} \right)$$

A pool of estimates: $\hat{\beta}_1^{(0)}$, $\hat{\beta}_1^{(1)}$, $\hat{\beta}_1^{(2)}$

Run linear regression with permuted X_1:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{pmatrix} \sim \begin{pmatrix} X_{21} & X_{12} & \cdots & X_{1p} \\ X_{41} & X_{22} & \cdots & X_{2p} \\ X_{61} & X_{32} & \cdots & X_{3p} \\ X_{11} & X_{42} & \cdots & X_{4p} \\ X_{31} & X_{52} & \cdots & X_{5p} \\ X_{51} & X_{62} & \cdots & X_{6p} \end{pmatrix}$$

$$\left(\hat{\beta}_1^{(3)}, \; \hat{\beta}_2^{(3)}, \; \ldots, \; \hat{\beta}_p^{(3)} \right)$$

A pool of estimates: $\hat{\beta}_1^{(0)}, \hat{\beta}_1^{(1)}, \hat{\beta}_1^{(2)}, \hat{\beta}_1^{(3)}$

Motivation

Run linear regression with permuted X_1:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{pmatrix} \sim \begin{pmatrix} X_{61} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ X_{41} & X_{32} & \cdots & X_{3p} \\ X_{51} & X_{42} & \cdots & X_{4p} \\ X_{31} & X_{52} & \cdots & X_{5p} \\ X_{11} & X_{62} & \cdots & X_{6p} \end{pmatrix}$$

$$(\hat{\beta}_1^{(m)}, \hat{\beta}_2^{(m)}, \ldots, \hat{\beta}_p^{(m)})$$

A pool of estimates: $\hat{\beta}_1^{(0)}, \hat{\beta}_1^{(1)}, \hat{\beta}_1^{(2)}, \hat{\beta}_1^{(3)}, \ldots, \hat{\beta}_1^{(m)}$.

Motivation

A pool of estimates: $\hat{\beta}_1^{(0)}$, $\hat{\beta}_1^{(1)}$, $\hat{\beta}_1^{(2)}$, $\hat{\beta}_1^{(3)}$, ..., $\hat{\beta}_1^{(m)}$.

Motivation

A pool of estimates: $\hat{\beta}_1^{(0)}, \hat{\beta}_1^{(1)}, \hat{\beta}_1^{(2)}, \hat{\beta}_1^{(3)}, \ldots, \hat{\beta}_1^{(m)}$.

- Calculate the rank of $\hat{\beta}_1^{(0)}$ in the pool;
- Reject H_0 if the rank is extreme;
- Widely used in boosting, random forests, etc., to construct measure of variable importance;
- Unfortunately fails to control type-I error even with i.i.d. ϵ_i's.

Motivation

Generally, we consider a pool of statistics: $\eta_0^T y$, $\eta_1^T y$, ..., $\eta_m^T y$.

Motivation

Generally, we consider a pool of statistics: $\eta_0^T y$, $\eta_1^T y$, $\ldots$, $\eta_m^T y$.

- Calculate the rank R of $\eta_0^T y$ in the pool;
- Hope that R is "uniform" under H_0 but "extreme" under H_1;
- Reject H_0 if R is "extreme".

A Slight Detour: Marginal Rank Test

Call $(S_0, S_1, \ldots, S_m)$ invariant under *cyclic permutation group* (CPG) iff for any $j \in [m]$,

$$(S_0, S_1, \ldots, S_m) \stackrel{d}{=} (S_j, S_{j+1}, \ldots, S_{j+m}).$$

A Slight Detour: Marginal Rank Test

Call $(S_0, S_1, \ldots, S_m)$ invariant under *cyclic permutation group* (CPG) iff for any $j \in [m]$,

$$(S_0, S_1, \ldots, S_m) \overset{d}{=} (S_j, S_{j+1}, \ldots, S_{j+m}).$$

Proposition (L. and Bickel, 2017)

Assume that $(S_0, S_1, \ldots, S_m)$ is invariant under CPG. For any function $f : \mathbb{R} \to \mathbb{R}$, let R_f denote the rank of $f(S_0)$ (in a decreasing order with ties broken randomly). Then

1. $R \sim \mathrm{Unif}([m+1])$;
2. *Let $U \sim \mathrm{Unif}([0,1])$. Then*

$$p \triangleq \frac{R - U}{m+1} \sim U([0,1]).$$

Goal: construct $\eta_0, \ldots, \eta_m$ such that

Goal: construct $\eta_0, \ldots, \eta_m$ such that

- (Validity)

$$(\eta_0^T y, \ldots, \eta_m^T y)$$

is invariant under CPG under H_0;

Goal: construct $\eta_0, \ldots, \eta_m$ such that

- (Validity)
$$(\eta_0^T y, \ldots, \eta_m^T y)$$
is invariant under CPG under H_0;
- (High Power) there exists f such that
$$f(\eta_0^T y) \gg \max_{j \in [m]} f(\eta_j^T y)$$
under H_1.

Write X as $(X_1\ X_{[-1]})$ and β as $(\beta_1, \beta_{[-1]})$, then

$$\eta^T y = (\eta^T X_1)\beta_1 + (\eta^T X_{[-1]})\beta_{[-1]} + \eta^T \epsilon.$$

Cyclic Permutation Test: Step I (Validity)

Write X as $(X_1\ X_{[-1]})$ and β as $(\beta_1, \beta_{[-1]})$, then

$$\eta^T y = (\eta^T X_1)\beta_1 + (\eta^T X_{[-1]})\beta_{[-1]} + \eta^T \epsilon.$$

Under $H_0 : \beta_1 = 0$,

$$\eta^T y = \underbrace{(\eta^T X_{[-1]})\beta_{[-1]}}_{\text{Deterministic Part}} + \underbrace{\eta^T \epsilon}_{\text{Stochastic Part}}$$

Cyclic Permutation Test: Step I (Validity)

Write X as $(X_1 \ X_{[-1]})$ and β as $(\beta_1, \beta_{[-1]})$, then

$$\eta^T y = (\eta^T X_1)\beta_1 + (\eta^T X_{[-1]})\beta_{[-1]} + \eta^T \epsilon.$$

Under $H_0 : \beta_1 = 0$,

$$\eta^T y = \underbrace{(\eta^T X_{[-1]})\beta_{[-1]}}_{\text{Deterministic Part}} + \underbrace{\eta^T \epsilon}_{\text{Stochastic Part}}$$

Idea: matching the deterministic part and making the stochastic part invariant under CPG.

Cyclic Permutation Test: Step I (Validity)

Under $H_0 : \beta_1 = 0$,

$$\eta^T y = \underbrace{(\eta^T X_{[-1]})\beta_{[-1]}}_{\text{Deterministic Part}} + \underbrace{\eta^T \epsilon}_{\text{Stochastic Part}}$$

Construct $(\eta_0, \eta_1, \ldots, \eta_m)$ such that

- $X_{[-1]}^T \eta_j \equiv \gamma_{[-1]}$ for some $\gamma_{[-1]} \in \mathbb{R}^{p-1}$;
- $(\eta_0^T \epsilon, \eta_1^T \epsilon, \ldots, \eta_m^T \epsilon)$ is invariant under CPG.

Cyclic Permutation Test: Step I (Validity)

Assume n is divisible by $m+1$ with $n = (m+1)r$. Then $(\eta_0^T\epsilon, \eta_1^T\epsilon, \ldots, \eta_m^T\epsilon)$ is invariant under CPG if

$$\eta_0 = \eta^* = (\eta_1^*, \ldots, \eta_n^*),$$

and

$$\eta_j = \pi_L^r(\eta^*) \triangleq (\eta_{jr+1}^*, \ldots, \eta_{jr+n}^*)$$

where π_L is the *left-shift operator* on S_n, the permutation group with n elements.

Cyclic Permutation Test: Step I (Validity)

Assume n is divisible by $m + 1$ with $n = (m + 1)r$. Then $(\eta_0^T \epsilon, \eta_1^T \epsilon, \ldots, \eta_m^T \epsilon)$ is invariant under CPG if

$$\eta_0 = \eta^* = (\eta_1^*, \ldots, \eta_n^*),$$

and

$$\eta_j = \pi_L^r(\eta^*) \triangleq (\eta_{jr+1}^*, \ldots, \eta_{jr+n}^*)$$

where π_L is the *left-shift operator* on S_n, the permutation group with n elements.

Example: $n = 6, m = 2$,

$$\eta_0 = (\eta_1^*, \eta_2^*, \eta_3^*, \eta_4^*, \eta_5^*, \eta_6^*),$$

$$\eta_1 = (\eta_3^*, \eta_4^*, \eta_5^*, \eta_6^*, \eta_1^*, \eta_2^*),$$

$$\eta_2 = (\eta_5^*, \eta_6^*, \eta_1^*, \eta_2^*, \eta_3^*, \eta_4^*).$$

Cyclic Permutation Test: Step I (Validity)

Proof of the special case ($n = 6, m = 2$): By definition,

$$(\eta_0^T\epsilon, {\color{red}\eta_1^T\epsilon}, {\color{blue}\eta_2^T\epsilon})$$

$$= \begin{pmatrix} \eta_1^* & \eta_2^* & \eta_3^* & \eta_4^* & \eta_5^* & \eta_6^* \\ {\color{red}\eta_3^*} & {\color{red}\eta_4^*} & {\color{red}\eta_5^*} & {\color{red}\eta_6^*} & {\color{red}\eta_1^*} & {\color{red}\eta_2^*} \\ {\color{blue}\eta_5^*} & {\color{blue}\eta_6^*} & {\color{blue}\eta_1^*} & {\color{blue}\eta_2^*} & {\color{blue}\eta_3^*} & {\color{blue}\eta_4^*} \end{pmatrix} \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{pmatrix}$$

$$= \begin{pmatrix} \epsilon_1 & \epsilon_2 & \epsilon_3 & \epsilon_4 & \epsilon_5 & \epsilon_6 \\ {\color{red}\epsilon_5} & {\color{red}\epsilon_6} & {\color{red}\epsilon_1} & {\color{red}\epsilon_2} & {\color{red}\epsilon_3} & {\color{red}\epsilon_4} \\ {\color{blue}\epsilon_3} & {\color{blue}\epsilon_4} & {\color{blue}\epsilon_5} & {\color{blue}\epsilon_6} & {\color{blue}\epsilon_1} & {\color{blue}\epsilon_2} \end{pmatrix} \begin{pmatrix} \eta_1^* \\ \eta_2^* \\ \eta_3^* \\ \eta_4^* \\ \eta_5^* \\ \eta_6^* \end{pmatrix}$$

Proof of the special case ($n = 6, m = 2$): By definition,

$$(\eta_1^T \epsilon, \eta_2^T \epsilon, \eta_0^T \epsilon)$$

$$= \begin{pmatrix} \eta_3^* & \eta_4^* & \eta_5^* & \eta_6^* & \eta_1^* & \eta_2^* \\ \eta_5^* & \eta_6^* & \eta_1^* & \eta_2^* & \eta_3^* & \eta_4^* \\ \eta_1^* & \eta_2^* & \eta_3^* & \eta_4^* & \eta_5^* & \eta_6^* \end{pmatrix} \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{pmatrix}$$

$$= \begin{pmatrix} \epsilon_5 & \epsilon_6 & \epsilon_1 & \epsilon_2 & \epsilon_3 & \epsilon_4 \\ \epsilon_3 & \epsilon_4 & \epsilon_5 & \epsilon_6 & \epsilon_1 & \epsilon_2 \\ \epsilon_1 & \epsilon_2 & \epsilon_3 & \epsilon_4 & \epsilon_5 & \epsilon_6 \end{pmatrix} \begin{pmatrix} \eta_1^* \\ \eta_2^* \\ \eta_3^* \\ \eta_4^* \\ \eta_5^* \\ \eta_6^* \end{pmatrix}$$

Proof of the special case ($n = 6, m = 2$): ϵ is exchangeable,

$$\begin{pmatrix} \epsilon_1 & \epsilon_2 & \epsilon_3 & \epsilon_4 & \epsilon_5 & \epsilon_6 \\ \epsilon_5 & \epsilon_6 & \epsilon_1 & \epsilon_2 & \epsilon_3 & \epsilon_4 \\ \epsilon_3 & \epsilon_4 & \epsilon_5 & \epsilon_6 & \epsilon_1 & \epsilon_2 \end{pmatrix} \stackrel{d}{=} \begin{pmatrix} \epsilon_5 & \epsilon_6 & \epsilon_1 & \epsilon_2 & \epsilon_3 & \epsilon_4 \\ \epsilon_3 & \epsilon_4 & \epsilon_5 & \epsilon_6 & \epsilon_1 & \epsilon_2 \\ \epsilon_1 & \epsilon_2 & \epsilon_3 & \epsilon_4 & \epsilon_5 & \epsilon_6 \end{pmatrix}$$

As a consequence,

$$(\eta_0^T \epsilon, \eta_1^T \epsilon, \eta_2^T \epsilon) \stackrel{d}{=} (\eta_1^T \epsilon, \eta_2^T \epsilon, \eta_0^T \epsilon)$$

Cyclic Permutation Test: Step I (Validity)

Want to construct $(\eta_0, \eta_1, \ldots, \eta_m)$ such that

- $X_{[-1]}^T \eta_j \equiv \gamma_{[-1]}$ for some $\gamma_{[-1]} \in \mathbb{R}^{p-1}$;
- ~~$(\eta_0^T \epsilon, \eta_1^T \epsilon, \ldots, \eta_m^T \epsilon)$ is invariant under CPG.~~

Cyclic Permutation Test: Step I (Validity)

Want to construct $(\eta_0, \eta_1, \ldots, \eta_m)$ such that

- $X_{[-1]}^T \eta_j \equiv \gamma_{[-1]}$ for some $\gamma_{[-1]} \in \mathbb{R}^{p-1}$;
- ~~$(\eta_0^T \epsilon, \eta_1^T \epsilon, \ldots, \eta_m^T \epsilon)$ is invariant under CPG.~~
- $\eta_j = \pi_L^{rj}(\eta^*)$.

Cyclic Permutation Test: Step I (Validity)

Want to construct $(\eta_0, \eta_1, \ldots, \eta_m)$ such that

- $X_{[-1]}^T \eta_j \equiv \gamma_{[-1]}$ for some $\gamma_{[-1]} \in \mathbb{R}^{p-1}$;
- ~~$(\eta_0^T \epsilon, \eta_1^T \epsilon, \ldots, \eta_m^T \epsilon)$ is invariant under CPG.~~
- $\eta_j = \pi_L^{rj}(\eta^*)$.

Let $\Pi = \Pi_{L;r}$ be the permutation matrix that $\pi_L^r(z) = \Pi z$. The above conditions are equivalent to:

$$\begin{pmatrix} -I_{p-1} & X_{[-1]}^T \\ -I_{p-1} & X_{[-1]}^T \Pi \\ \vdots & \vdots \\ -I_{p-1} & X_{[-1]}^T \Pi^m \end{pmatrix} \begin{pmatrix} \gamma_{[-1]} \\ \eta^* \end{pmatrix} = 0.$$

Cyclic Permutation Test: Step I (Validity)

$$\begin{pmatrix} -I_{p-1} & X_{[-1]}^T \\ \vdots & \vdots \\ -I_{p-1} & X_{[-1]}^T \Pi^m \end{pmatrix} \begin{pmatrix} \gamma_{[-1]} \\ \eta^* \end{pmatrix} = 0. \qquad (1)$$

Theorem (L. and Bickel, 2017)

Assume n is divisible by $m+1$ with $n = (m+1)r$,

1. *(1) always has a non-zero solution if*

$$n + p - 1 > (m+1)(p-1) \leftrightarrow m < n/(p-1)$$

2. *for any solution $(\gamma_{[-1]}, \eta^*)$ of (1),*

$$({\eta^*}^T y, \pi_L^r(\eta^*)^T y, \cdots \pi_L^{rm}(\eta^*)^T y)$$

is invariant under CPG under H_0.

- We can construct (η_0, η_1) provided that $n \geq p$;
- In general, we want $m + 1 \geq 1/\alpha$ in order to avoid randomized p-values;
- For example, when $\alpha = 0.05$, a default choice is $m = 19$. In this case we need

$$n \geq 19(p-1) \Longleftarrow n/p \geq 19$$

which is reasonable in various applications.

Recall that

$$\eta_j^T y = \underbrace{(X_1^T \eta_j) \cdot \beta_1}_{\text{Deterministic Signal}} + \underbrace{\gamma_{[-1]}^T \beta_{[-1]}}_{\text{Deterministic Nuisance}} + \underbrace{\eta_j^T \epsilon}_{\text{Stochastic Part}} .$$

Cyclic Permutation Test: Step II (Power)

Recall that

$$\eta_j^T y = \underbrace{(X_1^T \eta_j) \cdot \beta_1}_{\text{Deterministic Signal}} + \underbrace{\gamma_{[-1]}^T \beta_{[-1]}}_{\text{Deterministic Nuisance}} + \underbrace{\eta_j^T \epsilon}_{\text{Stochastic Part}} .$$

To enhance power, we want

$$X_1^T \eta_0 \gg \max_j X_1^T \eta_j.$$

Cyclic Permutation Test: Step II (Power)

Recall that

$$\eta_j^T y = \underbrace{(X_1^T \eta_j) \cdot \beta_1}_{\text{Deterministic Signal}} + \underbrace{\gamma_{[-1]}^T \beta_{[-1]}}_{\text{Deterministic Nuisance}} + \underbrace{\eta_j^T \epsilon}_{\text{Stochastic Part}} .$$

To enhance power, we want

$$X_1^T \eta_0 \gg \max_j X_1^T \eta_j.$$

A simple setting:

$$X_1^T \eta_0 = \gamma_1 + \delta, \quad X_1^T \eta_j = \gamma_1, \ \forall j > 0.$$

When $\|\eta_j\|_2 = 1$, δ measures the *effective Signal-to-Noise Ratio*.

Cyclic Permutation Test: Step II (Power)

Now we have the following set of equations:

- $X_1^T \eta_0 = \gamma_1 + \delta$
- $X_1^T \eta_j = \gamma_1, \ \forall j > 0;$
- $X_{[-1]}^T \eta_j = \gamma_{[-1]} \ \forall j \geq 0;$
- $\eta_j = \Pi^j \eta^*.$

Cyclic Permutation Test: Step II (Power)

Now we have the following set of equations:

- $X_1^T \eta_0 = \gamma_1 + \delta$
- $X_1^T \eta_j = \gamma_1, \ \forall j > 0;$
- $X_{[-1]}^T \eta_j = \gamma_{[-1]} \ \forall j \geq 0;$
- $\eta_j = \Pi^j \eta^*.$

The goal is to maximize δ (after normalizing η_j's).

Cyclic Permutation Test: Step II (Power)

Now we have the following set of equations:

- $X_1^T \eta_0 = \gamma_1 + \delta$
- $X_1^T \eta_j = \gamma_1, \;\; \forall j > 0;$
- $X_{[-1]}^T \eta_j = \gamma_{[-1]} \;\; \forall j \geq 0;$
- $\eta_j = \Pi^j \eta^*.$

The goal is to maximize δ (after normalizing η_j's).

Write γ for $(\gamma_1, \gamma_{[-1]}^T)^T$ and $e_{1,p}$ for $(1, 0, \ldots, 0)^T \in \mathbb{R}^p$. Then

$$\max_{\delta \in \mathbb{R}, \gamma \in \mathbb{R}^p, \eta \in \mathbb{R}^n, \|\eta\|=1} \quad \delta \tag{2}$$

$$\text{s.t.} \quad \begin{pmatrix} -e_{1,p} & -I_p & X^T \\ 0 & -I_p & X^T \Pi \\ \vdots & \vdots & \vdots \\ 0 & -I_p & X^T \Pi^m \end{pmatrix} \begin{pmatrix} \delta \\ \gamma \\ \eta \end{pmatrix} = 0. \tag{3}$$

To simplify the notation, let

$$A(X) = \begin{pmatrix} -I_p & -I_p & \cdots & -I_p \\ X & \Pi^T X & \cdots & (\Pi^m)^T X \end{pmatrix} \in \mathbb{R}^{(n+p)\times(m+1)p}.$$

Then (2)-(3) is equivalent to

$$\max_{\delta\in\mathbb{R},\gamma\in\mathbb{R}^p,\eta\in\mathbb{R}^n,\|\eta\|=1} \quad \delta$$

$$\text{s.t.} \quad \left(-e_{1,p(m+1)} \vdots A(X)^T\right)\begin{pmatrix}\delta\\ \gamma\\ \eta\end{pmatrix} = 0.$$

Cyclic Permutation Test: Step II (Power)

Theorem (L. and Bickel, 2017)

1. (3) *always has a non-zero solution if*

$$n + p + 1 > (m + 1)p \iff m < (n + 1)/p.$$

2. *Let* $\begin{pmatrix} \tilde{\gamma} \\ \tilde{\eta} \end{pmatrix}$ *be the OLS estimator by regressing* $A(X)_1$ *on* $A(X)_{[-1]}$ *with* RSS *being the residual sum of squares. Then the optimal solution of* (2) *is given by*

$$\eta^*(X) = \frac{\tilde{\eta}}{\|\tilde{\eta}\|_2}, \quad \delta^*(X) = \frac{\mathrm{RSS}}{\|\tilde{\eta}\|_2}.$$

Cyclic Permutation Test: Step III (Best Ordering)

In general, for any permutation matrix $\Pi \in \mathbb{R}^{n\times n}$,

$$\delta^*(X) \neq \delta^*(\Pi X).$$

This leads to a secondary optimization problem:

$$\max_{\text{perm. } \Pi\in\mathbb{R}^{n\times n}} \delta^*(\Pi X).$$

Cyclic Permutation Test: Step III (Best Ordering)

In general, for any permutation matrix $\Pi \in \mathbb{R}^{n\times n}$,

$$\delta^*(X) \neq \delta^*(\Pi X).$$

This leads to a secondary optimization problem:

$$\max_{\text{perm. } \Pi \in \mathbb{R}^{n\times n}} \delta^*(\Pi X).$$

- This is a **Nonlinear Traveling Salesman Problem**;

In general, for any permutation matrix $\Pi \in \mathbb{R}^{n \times n}$,

$$\delta^*(X) \neq \delta^*(\Pi X).$$

This leads to a secondary optimization problem:

$$\max_{\text{perm. } \Pi \in \mathbb{R}^{n \times n}} \delta^*(\Pi X).$$

- This is a **Nonlinear Traveling Salesman Problem**;
- We do not need the exact maximizer;

Cyclic Permutation Test: Step III (Best Ordering)

In general, for any permutation matrix $\Pi \in \mathbb{R}^{n\times n}$,

$$\delta^*(X) \neq \delta^*(\Pi X).$$

This leads to a secondary optimization problem:

$$\max_{\text{perm. } \Pi \in \mathbb{R}^{n\times n}} \delta^*(\Pi X).$$

- This is a **Nonlinear Traveling Salesman Problem**;
- We do not need the exact maximizer;
- **Genetic Algorithm** can be efficient;

Cyclic Permutation Test: Step III (Best Ordering)

In general, for any permutation matrix $\Pi \in \mathbb{R}^{n\times n}$,

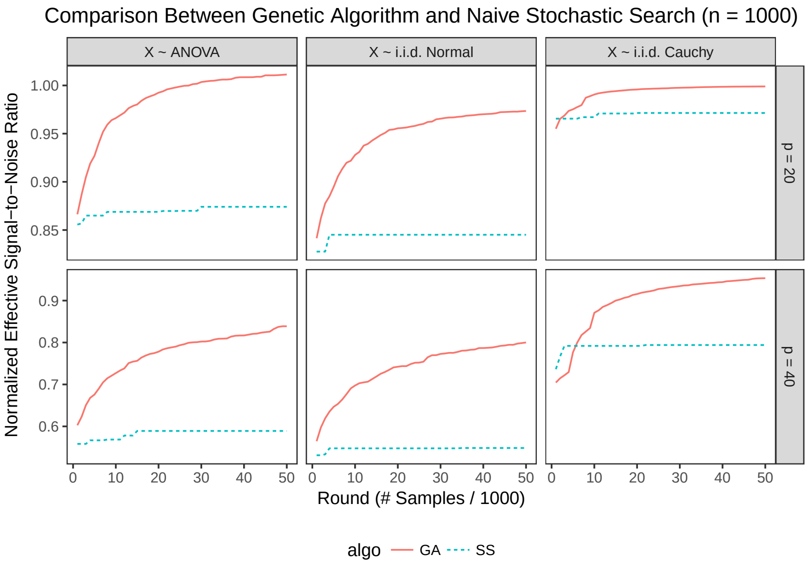
$$\delta^*(X) \neq \delta^*(\Pi X).$$

This leads to a secondary optimization problem:

$$\max_{\text{perm. } \Pi \in \mathbb{R}^{n\times n}} \delta^*(\Pi X).$$

- This is a **Nonlinear Traveling Salesman Problem**;
- We do not need the exact maximizer;
- **Genetic Algorithm** can be efficient;
- Implemented by gaoptim package in R.

Cyclic Permutation Test: Step III (Best Ordering)



Cyclic Permutation Test: Summary

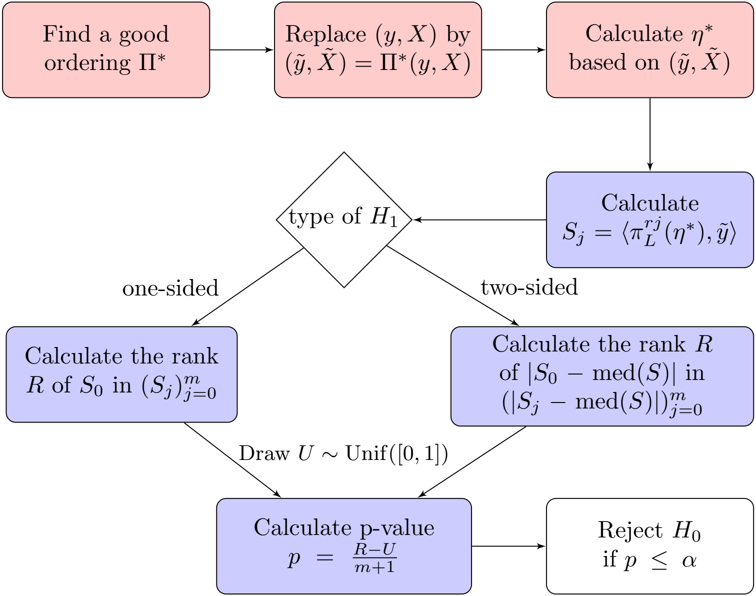
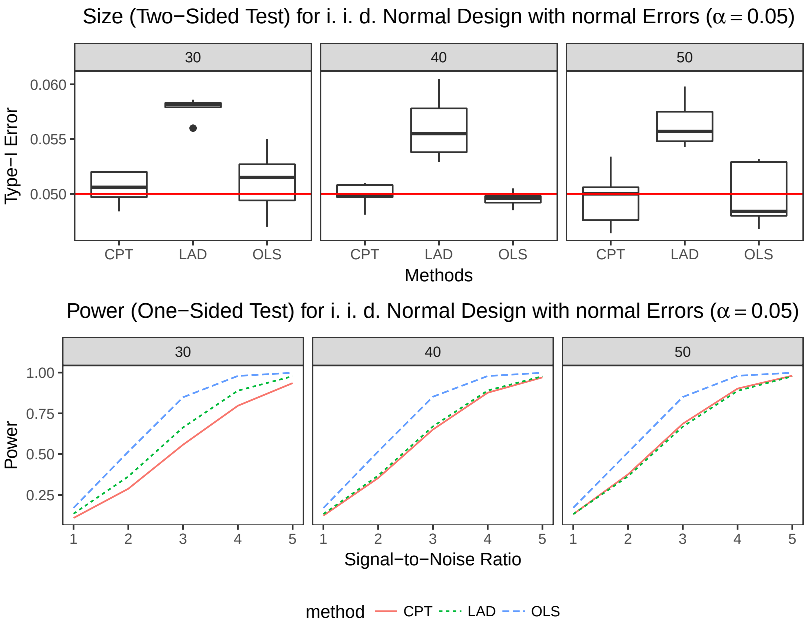
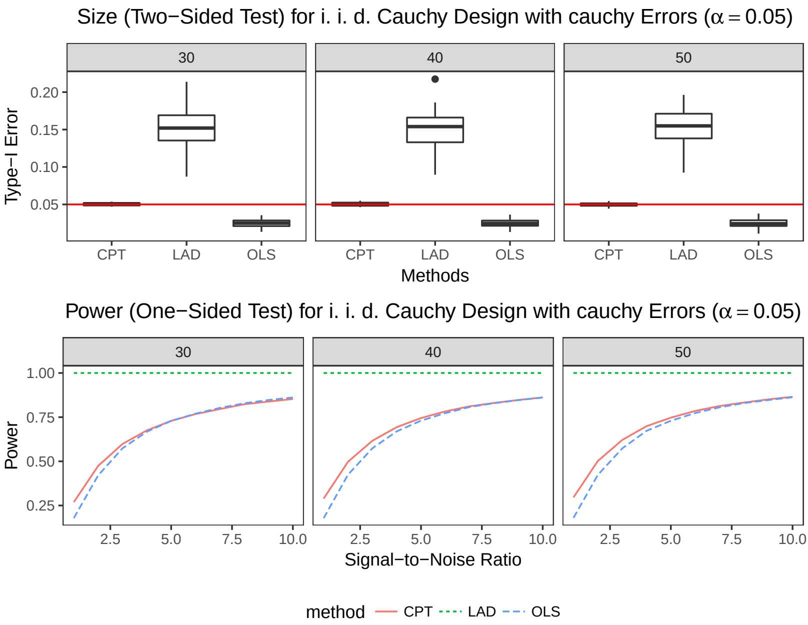


Table of Contents

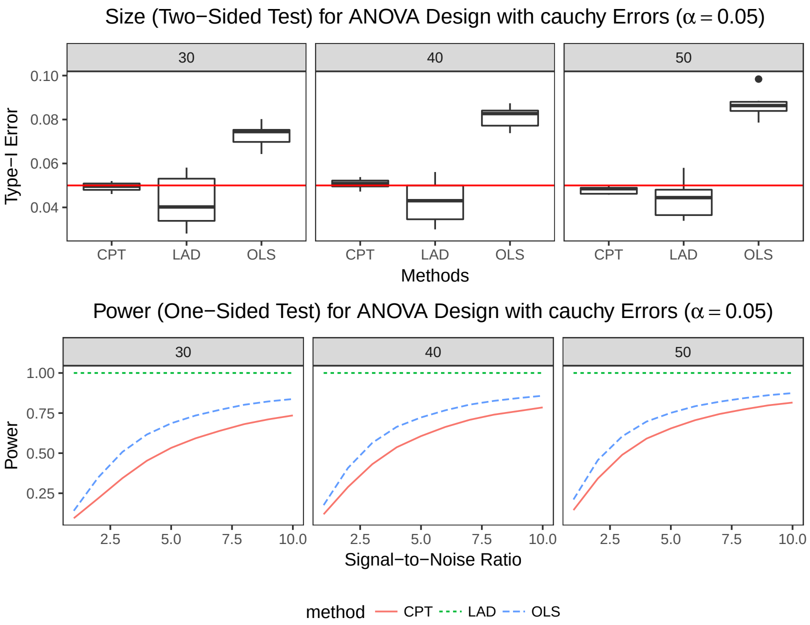
Example 1: i.i.d. normal design + normal error



Example 2: i.i.d. cauchy design + cauchy error



Example 3: ANOVA design + cauchy error



Example 4: ANOVA design + cauchy mixture error

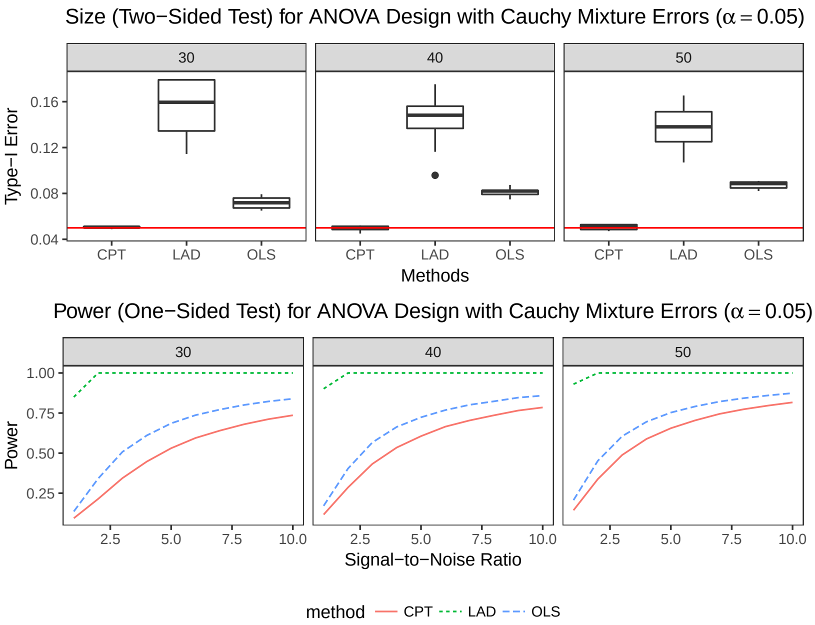


Table of Contents

Conclusion

We derive a test, referred to as **Cyclic Permutation Test (CPT)**,

- which is valid in finite-sample;
- with exact coverage α;
- for arbitrary *fixed* design matrix X;
- and for arbitrary exchangeable errors ϵ;
- with reasonable power in various practical situations.

Further comments:

- Easy to extend CPT to allow n not divisible by $m + 1$;
- Easy to extend CPT to test general linear hypothesis $H_0 : A\beta = 0$ where $A \in \mathbb{R}^{r \times p}$ with full rank r;
- The power can be analyzed asymptotically in various situations.

Open Problems

- Relation between n/p and confidence level α?
- Other construction that mimics general M-estimators, e.g. LAD estimator?
- Faster approach to find a good ordering?
- Why is OLS so robust? Self-normalizing property?
- Marginal Rank test in other applications?