Invariance-based Inference in High-Dimensional Regression with Finite-Sample Guarantees

Wenxuan Guo and Panos Toulis

Booth School of Business, University of Chicago

Setup

We focus on the perennial linear regression model:

$$y = X\beta + \varepsilon \tag{1}$$

- $y = (y_1, \dots, y_n)^\top$ is the outcome vector.
- $X \in \mathbb{R}^{n \times p}$ are covariates.
- $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)^\top$ are unobserved errors.

We want to test the global null hypothesis

$$H_0 : \beta = 0$$

in a high dimensional setup ($p < n$ but p grows with n or $p > n$).

- In our paper we also test for the partial nulls $H_0^S : \beta_S = 0$.
- This allows for inference.

Prior work

- F-test and its extensions for $p > n$ [Li et al., 2020, Zhong and Chen, 2011, Cui et al., 2018].
- Minimax optimal tests [Ingster et al., 2010].

However, these methods have limitations:

- Asymptotic methods that do not provide finite-sample guarantees for either type I or II error.
- Restrictive assumptions on ε, e.g., IIDness, homoskedasticity, bounded higher moments or sub-Gaussianity.
- These methods are not robust to heavy-tailed or heteroskedastic errors.

These limitations motivate us to study *invariance-based tests*.

Our work and contributions

- We study invariance-based inference, which relies on general invariance assumptions on the errors, e.g., sign symmetry $\varepsilon_i \overset{d}{=} -\varepsilon_i$.
- Invariance-based tests are also known as randomization tests [Lehmann and Romano, 2005].
- An alternative framework for testing and inference, different from the standard i.i.d. framework. [Chung and Romano, 2013, Toulis, 2019, Lei and Bickel, 2020, Dobriban, 2022, Wen et al., 2022].

We provide:

- Finite-sample valid tests.
- Nonasymptotic analysis on type II error.
- Minimax optimality against certain nonsparse alternatives.

Empirically, our invariance-based test has a more robust performance especially under multicollinearity and heavy-tailed data.

Component 1: Invariance assumption

- Assume a general form of invariance:

$$\varepsilon \overset{\mathrm{d}}{=} g\varepsilon \mid X \text{ for all } g \in \mathcal{G}_n \ .$$

- $\mathcal{G}_n$ is an algebraic group of $\mathbb{R}^n \to \mathbb{R}^n$ linear transformations under matrix multiplication as the group action.
- Sign symmetry: Consider $\mathcal{G}_n = \left\{ \begin{bmatrix} \pm 1 & & 0 \\ & \ddots & \\ 0 & & \pm 1 \end{bmatrix} \right\}$ then the invariance assumption boils down to

$$(\varepsilon_1, \ldots, \varepsilon_n) \overset{\mathrm{d}}{=} (\pm\varepsilon_1, \ldots, \pm\varepsilon_n) \mid X \ .$$

- Main difference from the i.i.d. framework: We require no further assumptions on X and ε beyond invariance.

Component 2: Test statistic

- Use ridge-based test statistic for $t(y, X)$

$$t(y, X) = \|X\widehat{\beta}_\lambda\|^2 , \quad \widehat{\beta}_\lambda = (X^\top X + \lambda I_p)^{-1} X^\top y , \qquad (2)$$

- We choose the ridge statistic for two main reasons:
    - Easy to compute and directly applicable for $p > n$.
    - Amenable to theoretical analysis. In particular, this choice leads to a minimax optimal test.
- Our method allows testing and inference with the ridge estimator, which is unexplored in the literature.

A concrete feasible test for the global null

1. Obtain the observed statistic, $T_n = t(y, X)$.
2. Compute $t(G_r y, X)$, where $G_r \overset{iid}{\sim} \text{Unif}(\mathcal{G}_n)$, $r = 1, \dots, R$.
3. Obtain the one-sided p-value:

$$\mathrm{pval} = \frac{1}{R+1}\left(1 + \sum_{r=1}^{R} \mathbb{I}\{t(G_r y, X) > T_n\}\right) . \tag{3}$$

4. Reject $H_0 : \beta = 0$ if $\psi_\alpha = \mathbb{I}\{\mathrm{pval} \leq \alpha\} = 1$.

Remarks.

- $\text{Unif}(\mathcal{G}_n)$ is the uniform distribution on $\mathcal{G}_n$.
- We draw R samples from $\text{Unif}(\mathcal{G}_n)$, as enumerating $\mathcal{G}_n$ can be computationally challenging.

Finite-sample validity

Theorem

Suppose that $H_0 : \beta = 0$ *is true. Then, for any* $n > 0$ *and any level* $\alpha \in (0, 1]$, *we have*

$$\mathbb{E}_0(\psi_\alpha \mid X) \leq \alpha \ .$$

Proof sketch:

- H_0 implies $y = \varepsilon$, so we have $y \overset{d}{=} gy \mid X$.
- $t(y, X) \overset{d}{=} t(gy, X)$ for any $g \in \mathcal{G}_n$.
- $\{T_n, t(G_1 y, X), \ldots, t(G_R y, X)\}$ is a finite-sample valid reference distribution for T_n.

The proof works for any test statistic.

Benefits of finite-sample validity

- No further assumptions on X and ε beyond the invariance.
- No asymptotics.
- Simple testing procedure.
- Robust to heavy tailed covariates and errors. See the following type I errors (%) evaluated on four simulation setups with different multicollinearity and heavy-tailed data.

Methods	small $\|\Sigma\|_F$, slow-decay	large $\|\Sigma\|_F$, slow-decay	small $\|\Sigma\|_F$, fast-decay	small $\|\Sigma\|_F$, fast-decay
Inv	5.24	4.70	4.73	5.00
SF	16.27	17.38	14.68	13.69
CGZ	7.32	7.26	5.99	6.06

SF: F test on randomly projected covariates, CGZ: Global test based on a U-statistic.

The test has a robust control of type I error, but is it powerful?

Power analysis

Though any choice of test statistic is valid in our procedure, the ridge statistic has the following properties in terms of the type II error:

- Simple and interpretable finite-sample bounds.
- Minimax optimal under certain conditions.

To develop the power theory with ridge, we make the following assumptions.

- Symmetric errors

$$(\varepsilon_1, \ldots, \varepsilon_n) \stackrel{\mathrm{d}}{=} (\pm\varepsilon_1, \ldots, \pm\varepsilon_n) \mid X \ . \tag{S1}$$

- $p < n$ but possibly $p \to \infty$.

Finite-sample type II error bounds

Theorem

Suppose $\sigma_{\min} > 0$*. If* $\lambda \leq \sigma_{\min}^2$*, then for any alternative hypothesis* $\beta \neq 0$*,*

$$\mathbb{E}_X(1-\psi_\alpha) = O\Big(\frac{p^2\kappa^4 x_*^2}{\sigma_{\min}^2}\Big) + O\Big(\frac{p\kappa^4\sigma_*^2}{\sigma_{\min}^2\|\beta\|^2}\Big) \ .$$

Remarks.

- Suppose $\sigma_{\min}^2 = O(n)$. The error bound reduces to

$$\mathbb{E}_X(1-\psi_\alpha) = O\Bigg(\underbrace{\frac{p}{n}}_{\text{problem dimension}} \cdot \underbrace{\kappa^4}_{\text{multicollinearity}} \cdot \Big(\underbrace{p x_*^2}_{\text{model leverage}} + \underbrace{\frac{\sigma_*^2}{\|\beta\|^2}}_{\text{SNR}}\Big)\Bigg).$$

- κ and $\sigma_{\min}$ denote the condition number and the minimum singular value of X.
- $x_* := \max_{i\in[n],j\in[p]} |X_{ij}|$.
- $\sigma_*^2 := \max_{i\in[n]} \mathbb{E}(\varepsilon_i^2)$.

From finite-sample results to consistency

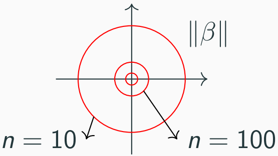
- Further suppose that κ, x_*, and σ_* are $O(1)$. Then

$$\mathbb{E}_X(1-\psi_\alpha) = O\left(\frac{p^2}{n}\right) + O\left(\frac{p}{n\|\beta\|^2}\right) .$$

- Suppose $p^2 = o(n)$. As $n \to \infty$, the test is consistent if

$$\|\beta\| = \Omega\left(\sqrt{\frac{p}{n}}\right) .$$

- Below, the red circles indicate regions with high type II errors, and shrink at a rate $\sqrt{p/n}$.



- This leads to the formal definition of detection radius.

Detection radius

- Consider the following alternative hypothesis space

$$H_1 : \beta \in \Theta(d),\ \Theta(d) = \{\ \beta \in \mathbb{R}^p : \|\beta\| \geq d\ \}\ .$$

- Define the worst-case type II error

$$\mathcal{B}(d, \psi) := \sup_{\beta \in \Theta(d)} \mathbb{E}(1 - \psi)\ .$$

Definition

We say a test ψ has a detection radius r_{np}, if for any sequence $d_{np} = \Omega(r_{np})$, it holds that $\lim_{n\to\infty} \mathcal{B}(d_{np}, \psi) = 0$.

- The detection radius provides a sufficient condition on how strong the signal should be to guarantee the consistency of a test.
- A smaller detection radius signifies a more powerful test.

Minimax optimality

Theorem

Suppose that

- X_i *are i.i.d. with* $\mathbb{E}X_i = 0$, $\mathbb{E}X_iX_i^\top = I_p$, *and sub-Gaussian tails.*
- ε_i *are i.i.d. with finite fourth moment, and* $\varepsilon \perp\!\!\!\perp X$.

Then, if

$$p = o(n^{0.5-\delta}) \text{ for some } \delta > 0 \text{ and } \lambda = o(n) \ ,$$

the invariance-based test, ψ_α, *for the global null hypothesis has a detection radius* $r_{np} = p^{1/4}/\sqrt{n}$ *and* ψ_α *is minimax optimal.*

- We provide the first result of minimax optimality of invariance-based tests for the global null under (S1).
- It is minimax optimal because $p^{1/4}/\sqrt{n}$ matches the least detectable signal strength, a known result established in [Ingster et al., 2010].

Simulation (revisited)

Consider a $p > n$ setup from [Li et al., 2020].

We compare our method to "SF" and "CGZ" proposed in [Li et al., 2020, Cui et al., 2018].

- $X \in \mathbb{R}^{n \times p}$ with $(n, p) = (50, 500)$.
- $(X_i)_{i=1}^n \overset{iid}{\sim} \mathcal{N}(0, \Sigma)$ with the covariance Σ satisfying
  (1) "slow-decay" in eigenvalues: $\lambda_i = \log^{-2}(i+1)$.
  (2) "fast-decay": $\lambda_i = i^{-1}$.

  We fix $\|\Sigma\|_F = 100, 300$.
- $\beta_i \overset{iid}{\sim} \mathrm{Binom}(3, 0.3) + 0.3\mathcal{N}(0, 1)$. Rescale β to inspect $\|\beta\| = 0, 0.5, 1, 2$.
- $\varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, 1)$.

Simulation

		Slow-decay				Fast-decay			
		$\|\|\beta\|\|$				$\|\|\beta\|\|$			
		0	0.5	1	2	0	0.5	1	2
Panel A: Normal design, normal errors									
$\|\|\Sigma\|\|_F = 100$	Inv	4.76	22.94	**49.11**	**67.38**	5.14	22.34	**60.87**	**89.11**
	SF	4.93	8.13	12.42	15.10	5.09	11.60	25.99	41.52
	CGZ	5.22	**23.00**	41.86	51.88	5.02	**26.36**	55.89	74.99
$\|\|\Sigma\|\|_F = 300$	Inv	5.03	**43.23**	**65.10**	**73.34**	4.71	**51.13**	**85.33**	**95.23**
	SF	5.01	11.32	15.11	16.81	4.96	22.95	38.74	48.15
	CGZ	4.87	37.66	51.08	55.28	4.64	49.50	72.08	80.64
Panel B: t_1 design, t_1 errors									
$\|\|\Sigma\|\|_F = 100$	Inv	5.24	62.80	65.46	66.22	4.73	87.55	89.68	89.58
	SF	16.27	99.90	99.94	99.97	14.68	99.79	99.86	99.85
	CGZ	7.32	83.48	83.51	83.40	5.99	83.90	84.92	85.30
$\|\|\Sigma\|\|_F = 300$	Inv	4.70	53.12	53.04	53.98	5.00	79.12	80.13	80.01
	SF	17.38	99.96	99.92	99.93	13.69	99.80	99.81	99.82
	CGZ	7.26	82.99	83.06	83.03	6.06	85.25	85.02	84.84

- Panel A: All tests are valid (under $\|\beta\| = 0$) and "Inv" is **powerful** (under $\|\beta\| > 0$).
- Panel B: "Inv" is robust to heavy-tailed data, whereas other methods fail to control the type I error.

Concluding remarks

We develop invariance-based tests in high-dimensional linear models.

- For the global null, we propose a test with finite-sample guarantees on both type I-II errors. This procedure is also minimax optimal.
- We extend our method to test for partial nulls, based on the idea of residual randomization [Toulis, 2019]. Check out the paper!

Our work opens up interesting problems for future work:

- Explore the power theory for the global null with $p > n$.
- Extend invariance-based tests to nonlinear regression models, e.g., generalized linear models.

Thank you!

- Wenxuan Guo and Panos Toulis, "Invariance-based Inference in High-Dimensional Regression with Finite-Sample Guarantees," *preprint*, 2023

References i

Chung, E. and Romano, J. P. (2013).
Exact and asymptotically robust permutation tests.

Cui, H., Guo, W., and Zhong, W. (2018).
Test for high-dimensional regression coefficients using refitted cross-validation variance estimation.
Ann. Statist., 46(3):958–988.

Dobriban, E. (2022).
Consistency of invariance-based randomization tests.
The Annals of Statistics, 50(4):2443 – 2466.

Ingster, Y. I., Tsybakov, A. B., and Verzelen, N. (2010).
Detection boundary in sparse regression.
Electron. J. Stat., 4:1476–1526.

References ii

Lehmann, E. L. and Romano, J. P. (2005).
Testing statistical hypotheses.
Springer Texts in Statistics. Springer, New York, third edition.

Lei, L. and Bickel, P. J. (2020).
An assumption-free exact test for fixed-design linear models with exchangeable errors.
Biometrika.

Li, Y., Kim, I., and Wei, Y. (2020).
Randomized tests for high-dimensional regression: A more efficient and powerful solution.
In Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., and Lin, H., editors, *Advances in Neural Information Processing Systems*, volume 33, pages 4721–4732. Curran Associates, Inc.

References iii

Toulis, P. (2019).
Invariant inference via residual randomization.

Wen, K., Wang, T., and Wang, Y. (2022).
Residual permutation test for high-dimensional regression coefficient testing.

Zhong, P.-S. and Chen, S. X. (2011).
Tests for high-dimensional regression coefficients with factorial designs.
Journal of the American Statistical Association, 106(493):260–274.