IS STATISTICAL INFERENCE WITHOUT SPARSITY POSSIBLE IN HIGH-DIMENSIONS?

Jelena Bradic

www.jelenabradic.net
Department of Mathematics
University of California, San Diego
jbradic@ucsd.edu

Joint work with:

Gerda Claeskens

Jianqing Fan

Yinchu Zhu

Setup

Consider a **high dimensional linear regression setting,**

$$\mathsf{Y} = \mathsf{X}\boldsymbol{\gamma}^* + \mathsf{Z}\beta^* + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathsf{Z} \in \mathbb{R}^{n}$ and $\mathsf{X} \in \mathbb{R}^{n\times p}$ are the design matrices, $p \gg n$, $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ is the error term independent of the design with $\mathbb{E}(\boldsymbol{\varepsilon}) = 0$ and $\mathbb{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top) = \sigma_\varepsilon^2 \mathbb{I}_n$, and $\boldsymbol{\gamma}^*$ and β^* are unknown model parameters.

We focus on the problem of testing single entries of the model parameter, namely the following hypothesis:

$$\mathsf{H}_0 : \beta^* = \beta_0, \qquad \text{versus} \qquad \mathsf{H}_1 : \beta^* \neq \beta_0. \tag{2}$$

Sparsity assumption: $\|\boldsymbol{\gamma}^*\|_0 := s_\gamma \ll n$ and for inference procedures is such that $s_\gamma \log p/\sqrt{n} \to 0$ as $n \to \infty$.

What happens if we apply sparsity-based methods when the underlying model parameter is not sparse? Can we obtain misleading and spurious results ?

Example 1

- ⋆ Assume: $\mathsf{X} = \mathbb{I}_p$, ε_i are i.i.d. with $\mathcal{N}(0,1)$ and such that for $a \in [-10, 10]$

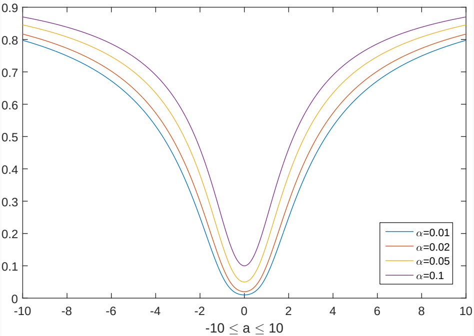
$$\beta^* = 0 \qquad \text{and} \qquad \boldsymbol{\gamma}^* = a p^{-1/2} \mathbf{1}_p,$$

- ⋆ We consider the "de-biasing" approach as formulated in Van de Geer et.al (2014) Let $\boldsymbol{\pi}^* = (\beta^*, {\boldsymbol{\gamma}^*}^\top)^\top \in \mathbb{R}^{p+1}$ and $\mathsf{W} = (\mathsf{Z}, \mathsf{X}) \in \mathbb{R}^{n \times (p+1)}$. The debiased estimator is then defined $\widetilde{\boldsymbol{\pi}} = \widehat{\boldsymbol{\pi}} + \mathbb{I}_{p+1} W^\top (\mathsf{Y} - \mathsf{W}\widehat{\boldsymbol{\pi}})/n$
- ⋆ Wald test rejects the hypothesis whenever $|\widetilde{\boldsymbol{\pi}}_1| > \Phi^{-1}(1 - \alpha/2)/\sqrt{n}$.

Theorem

In the above setup, we have $\lim_{n \to \infty} P\left(|\widetilde{\boldsymbol{\pi}}_1| > \Phi^{-1}(1 - \alpha/2)/\sqrt{n}\right) = F(\alpha, a)$, where $F(\alpha, a) = 2 - 2\Phi\left[\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)/\sqrt{1 + a^2}\right]$.

Figure: Plot of the asymptotic Type I error of Wald test



The horizontal axis denotes a and the vertical axis denotes $F(\alpha, a)$.

- ⋆ To develop sparsity-robust tests for the hypothesis (10)
  We say that a test is sparsity-robust if the Type I error is asymptotically bounded by the nominal level, regardless of whether or not $\boldsymbol{\gamma}^*$ is sparse.
- ⋆ Moreover, whenever the sparsity condition holds, our method is shown to be optimal and matches existing sparsity-based methods in terms of Type II errors.
- ⋆ We show minimax optimal power in certain dense models as well.
- ⋆ Our methodology is based on the idea of exploiting the implication of the null hypothesis.
  - ⋆ Instead of directly estimating the parameter under testing, we test a moment condition that is equivalent to the null hypothesis.

www.jelenabradic.net Department of Mathematics University of California, San Diego jbradic@ucsd.edu

We observe that a pseudo-response $V := Y - Z\beta_0$ satisfies a linear model

$$V = X\boldsymbol{\gamma}^* + e, \qquad e = Z(\beta^* - \beta_0) + \boldsymbol{\varepsilon}.$$

- ⋆ under H_0, X is uncorrelated with the error e
- ⋆ under H_1, e might have correlation with X through Z.

We formally introduce a model to account for the dependence between X and Z:

$$Z = X\boldsymbol{\theta}^* + u, \qquad i = 1, \ldots, n. \tag{3}$$

where $\boldsymbol{\theta}^* \in \mathbb{R}^p$ is sparse and $u \in \mathbb{R}^n$ is independent of X with mean zero and variance $\mathbb{E}(uu^\top) = \sigma_u^2 \mathbb{I}_n$.

We notice that

$$\mathbb{E}\left[(V - X\boldsymbol{\gamma}^*)^\top (Z - X\boldsymbol{\theta}^*)\right]/n = \sigma_u^2(\beta^* - \beta_0).$$

Hence, solving the inference problem (10) is equivalent to testing

$$H_0 : \mathbb{E}\left[(V - X\boldsymbol{\gamma}^*)^\top (Z - X\boldsymbol{\theta}^*)\right] = 0, \tag{4}$$

versus

$$H_1 : \mathbb{E}\left[(V - X\boldsymbol{\gamma}^*)^\top (Z - X\boldsymbol{\theta}^*)\right] \neq 0. \tag{5}$$

⋆ We define the following estimator

$$
\begin{aligned}
\widetilde{\gamma}(\sigma) \quad := & \quad \arg\min_{\gamma\in\mathbb{R}^p} \|\gamma\|_1 \\
\text{s.t.} \quad & \|n^{-1}X^\top(V - X\gamma)\|_\infty \le \eta_0\sigma \\
& \|V - X\gamma\|_\infty \le \|V\|_2/\log^2 n \\
& n^{-1}V^\top(V - X\gamma) \ge \rho_n n^{-1}\|V\|_2^2.
\end{aligned} \tag{6}
$$

for $\eta_0 = n^{-1/2}(1.1)\Phi^{-1}(1 - p^{-1}n^{-1})\sqrt{\max_{1\le j\le p} n^{-1}\sum_{i=1}^n x_{i,j}^2}$,
$\rho_n = 0.01/\sqrt{\log n}$.

⋆ $\widehat{\sigma}_\gamma = \arg\max\{\sigma : \sigma \in \mathcal{S}_\gamma\}$ and the set $\mathcal{S}_\gamma$ is defined as

$$
\mathcal{S}_\gamma = \left\{\sigma \ge \sqrt{\rho_n}\|V\|_2/\sqrt{n} : \quad 1.5\sigma \ge n^{-1/2}\|V - X\widetilde{\gamma}(\sigma)\|_2 \ge 0.5\sigma\right\}. \tag{7}
$$

→ When the estimation target fails to be sparse, the estimator is stable;
→ when the estimation target is sparse, the estimator automatically achieves consistency
→ does not require knowledge of the noise level.

We propose to consider the following correlation test (CorrT) statistic

$$T_n(\beta_0) = \frac{n^{-1/2}(V - X\hat{\gamma})^\top (Z - X\hat{\boldsymbol{\theta}})}{\hat{\sigma}_\varepsilon \hat{\sigma}_u}, \tag{8}$$

where

$\hat{\sigma}_\varepsilon = \|V - X\hat{\gamma}\|_2/\sqrt{n}$ and

$\hat{\sigma}_u = \|Z - X\hat{\boldsymbol{\theta}}\|_2/\sqrt{n}$.

Why does this work ?

We can show, without assuming sparsity of $\boldsymbol{\gamma}^*$, that

$$n^{-1/2}(V - X\hat{\gamma})^\top (Z - X\hat{\boldsymbol{\theta}}) = n^{-1/2}(V - X\hat{\gamma})^\top u + O_P(\sqrt{\log p}\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_1),$$

where under the null hypothesis, the first term on the right hand side has zero expectation and the second term vanishes fast enough.

Consider a **high dimensional linear regression setting,**

$$\mathsf{Y} = \mathsf{X}\boldsymbol{\beta}^* + \boldsymbol{\varepsilon}. \tag{9}$$

We focus on the problem of testing linear combinations of the model parameter, namely the following hypothesis:

$$\mathsf{H}_0 : \mathbf{a}^\top \boldsymbol{\beta}^* = \mathrm{g}_0, \qquad \text{versus} \qquad \mathsf{H}_1 : \mathbf{a}^\top \boldsymbol{\beta}^* \neq \mathrm{g}_0. \tag{10}$$

Sparsity assumption: $\|\mathbf{a}\|_0 :=??$ and $\|\boldsymbol{\beta}^*\|_0 :=??$

Prior knowledge of Σ_X

Let $\boldsymbol{\Omega}_X = \boldsymbol{\Sigma}_X^{-1}$.

For each of the features $\mathbf{x}_i \in \mathbb{R}^p$ consider the following decomposition:

$$\mathbf{x}_i = \mathbf{a} z_i + \mathbf{w}_i$$

with

$$z_i = \left(\frac{\boldsymbol{\Omega}_X \mathbf{a}}{\mathbf{a}^\top \boldsymbol{\Omega}_X \mathbf{a}} \right)^\top \mathbf{x}_i$$

and

$$\mathbf{w}_i = \left[\mathbb{I}_p - \frac{\mathbf{a}\mathbf{a}^\top \Omega_X}{\mathbf{a}^\top \boldsymbol{\Omega}_X \mathbf{a}} \right] \mathbf{x}_i$$

Notice that $\mathbf{a} z_i$ can be viewed as the projection of $\mathbf{x}_i$ onto the vector $\mathbf{a}$ – taking into account $\boldsymbol{\Omega}_X$, hence extracting information in $\mathbf{x}_i$ regarding the null hypothesis.

Now, we see that the original model can be reparametrized as

$$y_i = z_i(\mathbf{a}^\top \boldsymbol{\beta}^*) + \mathbf{w}_i^\top \boldsymbol{\beta}^* + \varepsilon_i,$$

which we refer to as **restructured regression**.

We observe that

$$\mathbb{E}[z_i(y_i - z_i g_0)] = \mathbb{E}[z_i^2(\mathbf{a}^\top \boldsymbol{\beta}^* - g_0)]$$

Hence, the original null is equivalent to the new null of the following kind

$$\mathbb{E}[z_i(y_i - z_i g_0)] = 0.$$

The test statistic then takes a simple form

$$\frac{n^{-1/2}\sum_{i=1}^{n} z_i(y_i - z_i g_0)}{\sqrt{n^{-1}\sum_{i=1}^{n} z_i^2(y_i - z_i g_0)^2}}$$

Remark

The novel methodology consists of two-stages. At the first stage, our procedure establishes a data-driven feature decomposition based on the structure of the null hypothesis directly. At the second stage, only “a moment condition” of the restructured regression is tested.

What if Σ_X is unknown

First, we pretend that $\boldsymbol{\Sigma}_X = \mathbb{I}_p$ and consider

$$z_i = \left(\frac{\mathbf{a}}{\mathbf{a}^\top \mathbf{a}}\right)^\top \mathbf{x}_i, \qquad \mathbf{w}_i = \left(\mathbb{I}_p - \frac{\mathbf{a}\mathbf{a}^\top}{\mathbf{a}^\top \mathbf{a}}\right) \mathbf{x}_i$$

Although the decomposition $\mathbf{x}_i = \mathbf{a}zi + \mathbf{w}_i$ still holds, features z_i and $\mathbf{w}_i$ might be highly correlated.

However, by introducing a orthogonal matrix $\mathbf{U}_a$ such that

$$\mathbb{I}_p - \frac{\mathbf{a}\mathbf{a}^\top}{\mathbf{a}^\top \mathbf{a}} = \mathbf{U}_a \mathbf{U}_a^\top$$

we can construct

$$\widetilde{\mathbf{W}} = \mathbf{W}\mathbf{U}_a$$

and observe that

$$y_i = z_i(\mathbf{a}^\top \boldsymbol{\beta}^*) + \widetilde{\mathbf{w}}_i^\top \boldsymbol{\pi}_* + \epsilon_i,$$

for

$$\boldsymbol{\pi}_* = \mathbf{U}_z^\top \boldsymbol{\beta}^*.$$

Test

Introduce a feature model

$$z_i = \widetilde{\mathbf{w}}_i^\top \boldsymbol{\gamma}^* + u_i$$

where $\boldsymbol{\gamma}^*$ is the unknown parameter and u_i are independent of $\widetilde{w}_i$.

Then, consider the moment

$$H_0 : \mathbb{E}\left[(z_i - \widetilde{\mathbf{w}}_i^\top \boldsymbol{\gamma}^*)^\top \left(y_i - z_i g_0 - \widetilde{\mathbf{w}}_i^\top \boldsymbol{\pi}^*\right)\right] = 0.$$

and develop a test

$$T_n = \sqrt{n}\frac{(\mathbf{z} - \widetilde{\mathbf{W}}\widehat{\boldsymbol{\gamma}})\left(\mathbf{y} - \mathbf{z}g_0 - \widetilde{\mathbf{W}}\widehat{\boldsymbol{\pi}}\right)}{\|\mathbf{z} - \widetilde{\mathbf{W}}\widehat{\boldsymbol{\gamma}}\|_2 \|\mathbf{y} - \mathbf{z}g_0 - \widetilde{\mathbf{W}}\widehat{\boldsymbol{\pi}}\|_2}$$

Condition

Let $W = (Z, X)$ and $w_i = (z_i, x_i^\top)^\top$. The matrix $\boldsymbol{\Sigma}_W = \mathbb{E}[W^\top W]/n \in \mathbb{R}^{p\times p}$ satisfies that $\kappa_1 \leq \sigma_{\min}(\boldsymbol{\Sigma}_W) \leq \sigma_{\max}(\boldsymbol{\Sigma}_W) \leq \kappa_2$. The vectors $\boldsymbol{\Sigma}_W^{-1/2}\mathbf{w}_1$ are centered with sub-Gaussian norms upper bounded by κ_3 and $\mathbb{E}|\varepsilon_1|^{2+\delta} \leq \kappa_4$. Moreover, $\log p = o\left(n^{\delta/(2+\delta)} \wedge n\right)$.

→ For the designs, it is standard to impose well-behaved covariance matrices and sub-Gaussian properties.

Condition

$\|\boldsymbol{\gamma}^*\|_2 \leq \kappa_5$ and $s_\theta = o\left(\sqrt{n/\log n}/\log p\right)$, where $s_\theta = \|\boldsymbol{\theta}^*\|_0$.

→ The assumption on s_θ imposes sparsity in the first row of the precision matrix $\boldsymbol{\Sigma}_W$ and the rate for s_θ is stronger than the conditions in BCH and NL imposing $o(\sqrt{n}/\log p)$ and in VBRD imposing $o(n/\log p)$.

Theorem

Let Conditions 1 and 2 hold. Then under H_0

$$\forall\alpha \in (0,1),\ \lim_{n\to\infty} \mathbb{P}\left(|T_n(\beta_0)| > \Phi^{-1}(1-\alpha/2)\right) = \alpha.$$

→ Theorem 2 formally establishes that the new CorrT test is asymptotically exact in testing $\boldsymbol{\beta}^* = \boldsymbol{\beta}_0$. In particular, CorrT is robust to dense γ^* in the sense that even under dense γ^*, our procedure does not generate false positive results.

We say that a procedure for testing the hypothesis (10) is sparsity-adaptive if

(i) this procedure does not require knowledge of s_γ,

(ii) provides valid inference under any s_γ and

(iii) achieves efficiency with sparse $\boldsymbol{\gamma}^*$.

We now show the third property, efficiency under sparse $\boldsymbol{\gamma}^*$. To formally discuss our results, we consider testing $H_0 : \beta^* = \beta_0$ versus

$$H_{1,h} : \ \beta^* = \beta_0 + h/\sqrt{n}. \tag{11}$$

where $h \in \mathbb{R}$ is a fixed constant.

Theorem

Let Conditions 1 and 2 hold. Suppose that $s_\gamma = o\left(n/\log(p \vee n)\right)$ and $\sigma_u/\sigma_\varepsilon \to \kappa_0$ for some constant $\kappa_0 > 0$. Then, under $H_{1,h}$ in (11),

$$P\left(|T_n(\beta_0)| > \Phi^{-1}(1-\alpha/2)\right) \to \Psi(\alpha, \kappa_0, h),$$

where $\Psi(h, \kappa_0, \alpha) = 2 - \Phi\left(\Phi^{-1}(1-\alpha/2) + h\kappa_0\right) - \Phi\left(\Phi^{-1}(1-\alpha/2) - h\kappa_0\right)$.

→ Theorem 3 establishes the local power of CorrT. It turns out that this local power matches that of existing sparsity-based methods, such as VBRD, NL and BCH, that are shown to be efficient.

Theorem

Let Conditions 1 and 2 hold together with $\log p = o(n)$. Let $\boldsymbol{\Sigma}_X = E[\mathbf{x}_i \mathbf{x}_i^\top] \in \mathbb{R}^{(p-1)\times(p-1)}$. Suppose that

$$\|\boldsymbol{\Sigma}_X \boldsymbol{\gamma}^*\|_\infty \sqrt{n \log p} = o(1),$$

and with $n \to \infty$ and some $\kappa > 0$, $(\boldsymbol{\gamma}^{*\top} \boldsymbol{\Sigma}_X \boldsymbol{\gamma}^* + \sigma_\varepsilon^2)\sigma_u^{-2} \to \kappa$. Then, under $H_{1,h}$ in (11),

$$\lim_{n,p\to\infty} P_{\beta*}\left(|T_n| > \Phi^{-1}(1-\alpha)\right) = \Psi(h, \kappa, \alpha),$$

where $\Psi(h, \kappa, \alpha)$ is defined in Theorem 3.

- For $n, p \to \infty$, $\sqrt{\log p}/n = o(1)$ (i.e. $n/p \to 0$), the Type II error of the proposed CorrT test, against alternatives that are larger than $O(n^{-1/2})$, converges to zero.
- If $\boldsymbol{\Sigma}_X = \mathbb{I}_p$, the condition $\|\boldsymbol{\Sigma}_X \boldsymbol{\gamma}^*\|_\infty \sqrt{n \log p} = o(1)$ is satisfied for all $\boldsymbol{\gamma}^*$ for which

$$\|\boldsymbol{\gamma}^*\|_\infty = o(1/\sqrt{n \log p}), \|\boldsymbol{\gamma}^*\|_2 = O(\sqrt{n}/\log p);$$

- If $\max_{1 \le j \le p} \|\boldsymbol{\Sigma}_{X,j}\|_1 = o(\sqrt{p/(n \log p)})$, we can consider all

$$\boldsymbol{\gamma}^* = \mathbf{c}/\sqrt{p}$$

  with $\|\mathbf{c}\|_\infty = O(1)$.
- Minimax testing of one coordinate (not the whole parameter) in dense high-dimensional testing is possible!

Theorem

Let Conditions 1 and 2 hold together with $\log p = o(n)$. Let $\boldsymbol{\Sigma}_X = E[x_i x_i^\top] \in \mathbb{R}^{(p-1)\times(p-1)}$. Suppose that

$$\boldsymbol{\gamma}^* = \boldsymbol{\pi}^* + \boldsymbol{\mu}^*$$

for $\boldsymbol{\pi}^*$ and $\boldsymbol{\mu}^*$ satisfying $\|\boldsymbol{\pi}^*\|_0 = o(\sqrt{n}/\log p)$, $(\boldsymbol{\mu}^{*\top}\boldsymbol{\Sigma}_X\boldsymbol{\mu}^* + \sigma_\varepsilon^2)\sigma_u^{-2} \to \kappa$ and $\|\boldsymbol{\Sigma}_X\boldsymbol{\mu}^*\|_\infty\sqrt{n\log p} = o(1)$ for some $\kappa > 0$ as $n \to \infty$. Then, under $H_{1,h}$ in (11),

$$\lim_{n,p\to\infty} P_{\beta*}\left(|T_n| > \Phi^{-1}(1-\alpha)\right) = \Psi(h,\kappa,\alpha),$$

where $\Psi(h,\kappa,\alpha)$ is defined in Theorem 3.

Setting

- LTD Light-tailed design: $N(0, \Sigma_{(\rho)})$ with the (i, j) entry of $\Sigma_{(\rho)}$ being $\rho^{|i-j|}$.
- HTD Heavy-tailed design: each row of W is generated as $\Sigma_{(\rho)}^{1/2} U$, where $U \in \mathbb{R}^n$ contains i.i.d random variables of Student's t-distribution with 3 degrees of freedom normalized to have variance one. (the third moment does not exist.)

  The error term $\varepsilon \in \mathbb{R}^n$ contains i.i.d random variables from either $N(0, 1)$ (light-tailed error, or LTE) or Student's t-distribution with 6 degrees of freedom normalized to have variance one (heavy-tailed error, or HTE).

We set

$$\pi_j^* = \begin{cases} 2/\sqrt{n} & 2 \leq j \leq 4 \\ 0 & j > \max\{s, 4\} \\ U(0, 4)/\sqrt{n} & \text{otherwise.} \end{cases}$$

We test the hypothesis

$$H_0 : \pi_3^* = 2/\sqrt{n} + h.$$

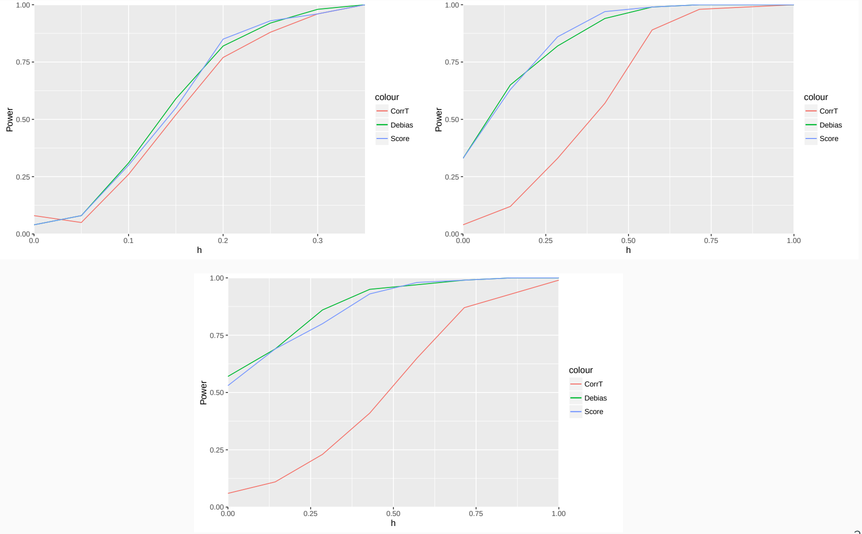
Table: Size properties (h = 0)

	LTD + LTE, $\rho = 0$			LTD + LTE, $\rho = -\frac{1}{2}$			HTD + HTE, $\rho = 0$		
	CorrT	Debias	Score	CorrT	Debias	Score	CorrT	Debias	Score
s = 1	0.03	0.05	0.04	0.05	0.04	0.05	0.06	0.04	0.02
s = 3	0.06	0.05	0.05	0.06	0.06	0.05	0.05	0.11	0.03
s = 5	0.09	0.09	0.09	0.07	0.11	0.10	0.07	0.04	0.04
s = 10	0.01	0.03	0.03	0.03	0.05	0.03	0.06	0.05	0.03
s = 20	0.08	0.12	0.11	0.03	0.06	0.06	0.03	0.12	0.04
s = 50	0.07	0.16	0.17	0.04	0.10	0.12	0.02	0.09	0.09
s = 100	0.05	0.29	0.28	0.01	0.15	0.14	0.05	0.20	0.21
s = n	0.04	0.35	0.33	0.04	0.27	0.27	0.04	0.38	0.38
s = p	0.07	0.54	0.52	0.04	0.39	0.40	0.05	0.57	0.53

	LTD + HTE, $\rho = 0$			LTD + HTE, $\rho = -\frac{1}{2}$			HTD + LTE, $\rho = 0$		
	CorrT	Debias	Score	CorrT	Debias	Score	CorrT	Debias	Score
s = 1	0.03	0.05	0.04	0.04	0.04	0.02	0.06	0.05	0.05
s = 3	0.06	0.05	0.05	0.11	0.06	0.06	0.03	0.07	0.04
s = 5	0.09	0.09	0.09	0.05	0.06	0.05	0.06	0.11	0.07
s = 10	0.01	0.03	0.03	0.03	0.04	0.03	0.09	0.11	0.10
s = 20	0.08	0.12	0.11	0.06	0.11	0.10	0.05	0.13	0.06
s = 50	0.07	0.16	0.17	0.07	0.16	0.15	0.06	0.19	0.14
s = 100	0.05	0.29	0.28	0.05	0.33	0.26	0.05	0.24	0.22
s = n	0.04	0.35	0.33	0.05	0.43	0.41	0.05	0.40	0.31
s = p	0.07	0.54	0.52	0.06	0.51	0.50	0.06	0.53	0.51

Power curves

Figure: Light-tailed errors



TRANSNOAH BREAST CANCER DATASET

- ⋆ Genome-wide gene expression profiling was performed using micro RNA from biopsies from 114 pre-treated patients with HER2+ breast cancer.
- ⋆ The complete data contains gene expression values of about 20000 genes located on different chromosomes.

- ⋆ BRCA1 is a human tumor suppressor gene that is normally expressed in the cells of breast and other tissue, where they help repair damaged DNA.
- ⋆ Research suggests that the BRCA1 proteins regulate the activity of other genes including tumor suppressors and regulators of the cell division cycle.
- ⋆ Moreover, it is believed that BRCA1 may regulate pathways that remove the damages in DNA introduced by the certain drugs.
- ⋆ Thus, understanding associations between BRCA1 and other genes provides a potentially important tool for tailoring chemotherapy in cancer treatment.

Gene	Biological association	Test Statistic		
		CorrT	Debias	Score
IGF2R[1]	breast cancer tumor suppressor	-4.692	-4.285	-4.445
Nmi[2]	endogenously associated with BRCA1	-4.239	-2.956	-2.669
RBBP4[3]	breast cancer	-4.186	-3.314	-2.806
NPM1[4]	breast cancer	-3.027	-2.112	-1.601
NARS2[5]	breast cancer	-4.163	-5.000	-4.983
B3GALNT1	lung cancer	1.151	2.082	2.065
C3orf62	lung cancer	-1.274	-2.143	-2.139
LTB	lung cancer	-0.131	-2.107	-2.143
TNFAIP1	lung cancer	1.231	2.181	2.118
CCPG1	prostate cancer	-1.597	-2.154	-2.251
LRRIQ3	colorectal cancer	-1.025	-2.480	-2.240
LOC100507537	bladder cancer	-0.137	-1.966	-1.135
ELOVL4	ataxia	-1.354	-2.152	-2.136

[1] sensitivity marker for radiation, chemotherapy, and endocrine therapy

[2] interactive binding protein

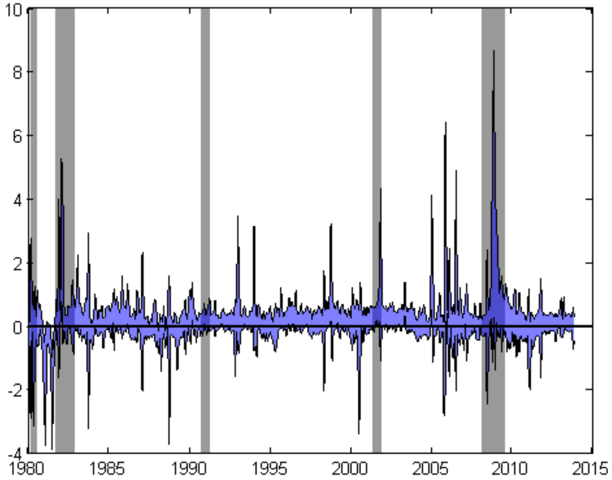
[3] retinoblastoma binding protein, a chromatin modeling factor

[4] blocks breast cancer cells

[5] partial or complete loss of

- Study the equity risk premia during different states from 1980-2014. of the economy
- The response is the excess return of the U.S stock market observed at time t, covariates are a large number of macroeconomic variables observed at time $t-1$ (McCracken, M. W. and Ng, S. (2015)) and s_t denotes the NBER recession indicator; $s_t = 1$ means that the economy is in recession at time t.
- Are risk premia in recessions higher than in expansions with the magnitude of difference that is economically meaningful?

Equity Risk Premia



Thank you for your attention!