Method-of-Moments Inference for GLMs and Doubly Robust Functionals under Proportional Asymptotics

陈星宇

上海交通大学

Draft: https://arxiv.org/pdf/2408.06103

Slides: https://cxy0714.github.io/

May 8, 2025

Coauthors

刘林

上海交通大学

Rajarshi Mukherjee

Harvard T.H. Chan School of Public Health

benefited from discussion with Subhabrata Sen (Harvard Stats) and Pragya Sur (Harvard Stats)

Outline

Motivations: Estimating Functionals of High-Dimensional GLMs

Our proposal – Moments-based estimators

Numerical experiments

Conclusion and open ends

What is a Generalized Linear Model (GLM)?

- GLM is a classical statistical model, generalizing linear regression:

$$\mathsf{E}(\mathrm{y}|\mathbf{x}) = \phi(\mathbf{x}^\top \boldsymbol{\beta})$$

  where ϕ is a known, smooth, monotonic link function.

- Common choices include:

  logistic regression: $\phi(t) = \frac{1}{1+e^{-t}}$
  Poisson regression: $\phi(t) = e^t$
  ...

- In statistics, we care not just about prediction, but also:

  Is a feature x_j truly associated with the outcome?
  Is the effect positive or negative, and how strong?
  What is the confidence interval for β_j?

- **High-dimensional example: gene expression analysis**

  y = whether a patient responds to a drug (yes/no)
  $\mathbf{x}$ = expression levels of 10,000 genes

  Goal: Which genes affect response? How strong is the signal? Can we quantify uncertainty?

High-dimensional Generalized Linear Models

- Data: $(\mathrm{y}_i \in \mathbb{R}, \mathbf{x}_i \in \mathbb{R}^p)_{i=1}^n \overset{\text{i.i.d.}}{\sim} \mathbb{P}$
- Model:

$$\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \mathsf{E}(\mathrm{y}|\mathbf{x}) = \phi(\boldsymbol{\beta}^\top \mathbf{x}), \quad \mathsf{var}(\mathrm{y}|\mathbf{x}) = \boldsymbol{\Sigma}^2(\mathbf{x})$$

- High-dimensional regime:

$$\frac{p}{n} \to \delta \in (0, +\infty)$$

- Goal: estimate and conduct inference on

$$\text{(i) } \boldsymbol{\beta} \quad \text{(ii) } \boldsymbol{\beta}^\top \boldsymbol{\Sigma} \boldsymbol{\beta}$$

What Makes an Estimator "Good"? (from a Statistics Viewpoint)

- $\hat{\boldsymbol{\beta}}$ is called **consistent** if it converges to the true $\boldsymbol{\beta}$ as $n \to \infty$.

- **Root-n consistency**:

$$\|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\| = \mathcal{O}_p(1/\sqrt{n})$$

- **Root-n Consistent & Asymptotic Normality ($\sqrt{n}$ CAN)**:

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(0, \mathrm{Cov})$$

  That is, for each coordinate $j = 1, \ldots, p$:

$$\sqrt{n}(\hat{\beta}_j - \beta_j) \xrightarrow{d} \mathcal{N}(0, \boldsymbol{\Sigma}_j^2)$$

- $\sqrt{n}$ CAN enables:

  Confidence intervals –for each coordinate j:

$$\hat{\beta}_j \pm z_{1-\alpha/2} \cdot \frac{\hat{\boldsymbol{\Sigma}}_j}{\sqrt{n}}$$

  where $z_{1-\alpha/2}$ is the standard normal quantile.
  Hypothesis testing –e.g., test $H_0 : \beta_j = 0$

A General Theme in the Literature: Debiasing

- In high dimensions, many works build on a biased initial estimator $\hat{\boldsymbol{\beta}}_{\text{init}}$, often obtained via *MLE or Penalized MLE*:

$$\hat{\boldsymbol{\beta}}_{\text{init}} \in_{\boldsymbol{\beta}\in\mathbb{R}^p} \left\{ \frac{1}{n}\sum_{i=1}^{n} \ell(\mathrm{y}_i, \mathbf{x}_i; \boldsymbol{\beta}) + g_\lambda(\boldsymbol{\beta}) \right\}$$

  This leads to a debiased estimator of the form:

$$\hat{\boldsymbol{\beta}}_{\text{db}} = \frac{1}{\widehat{\text{scale}}} \left(\hat{\boldsymbol{\beta}}_{\text{init}} - \widehat{\text{bias}} \right)$$

- There is a long list of works for logistic regression and other types of GLMs when $p/n \to \delta$
  Sur & Candés, PNAS 19; Zhao, Sur & Candés, Bernoulli 23; Massa et al. 22 and many others

- In the general form in Bellec 23, and (to our knowledge) first appeared for **Ridge Penalized MLE** in Pragya Sur's thesis:

$$\hat{\boldsymbol{\beta}}_{\text{db}} = \hat{\boldsymbol{\beta}}_{\text{init}} + \frac{1}{\hat{n}} \boldsymbol{\Sigma}^{-1} \sum_{i=1}^{n} \mathbf{x}_i \nabla \ell(\mathrm{y}_i, \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}_{\text{init}})$$

  where $\hat{n}$ is derived as part of a fixed-point system using Approximate Message Passing(AMP) machinery
  (Sur & Candés, PNAS 19; Zhao, Sur & Candés, Bernoulli 23)

Theoretical guarantees

Theorem 1 (Bellec 23)

Under regularity conditions, $\sqrt{n}$-consistency & CAN in aggregated sense: $\xi_j \sim \mathrm{N}(0,1)$, $j = 1, \cdots, p$

$$\frac{1}{p}\sum_{j=1}^{p} \mathsf{E}\left[\left(\sqrt{n}(\boldsymbol{\Sigma}^{-1})_{j,j}\frac{\hat{n}}{n}(\hat{r}\hat{\beta}_{\mathrm{db},j} - \hat{t}\beta_j) - \xi_j\right)^2\right] \to 0$$

where $\hat{r}, \hat{t}, \hat{n}$ are part of the solution to the AMP fixed-point equations

- Entry-wise CAN is still an open question for $\hat{\boldsymbol{\beta}}_{\mathrm{db}}$ for general GLM and $p > n$.
- Our estimator can achieve Entry-wise CAN!

Our Contribution: Entrywise Inference without Sparsity

- We propose a new estimator based on **method of moments**:

  Classical, easy-to-understand and easy-to-extend framework;
  No sparsity;
  No difficult AMP to learn;
  No tuning parameters
  Consistent estimator of the variance.

- Under **known** $\mathbf{\Sigma}$, our method achieves:

  Entrywise $\sqrt{n}$ **CAN**;
  Extensible to non-Gaussian $\mathbf{x}$

- Under **unknown** $\mathbf{\Sigma}$, our method achieves:

  when $p < n$, **Entrywise** $\sqrt{n}$ **consistent**;
  Extensible to non-Gaussian $\mathbf{x}$
  when $p > n$, **Entrywise consistent**;
  Extensible to non-Gaussian $\mathbf{x}$

Outline

Motivations: Estimating Functionals of High-Dimensional GLMs

Our proposal – Moments-based estimators

Numerical experiments

Conclusion and open ends

Roadmap: From Easy to Hard

Case I **Gaussian, known $\boldsymbol{\mu} = 0$ & known $\boldsymbol{\Sigma}$**
$\mathbf{x} \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$

$\Downarrow$

Case II **Gaussian, unknown $\boldsymbol{\mu}$ & known $\boldsymbol{\Sigma}$**
$\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{\mu}$ unknown

$\Downarrow$

Case III **Gaussian, unknown $\boldsymbol{\mu}$ & unknown $\boldsymbol{\Sigma}$**
$\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ unknown

$\Downarrow$

Case IV **Non-Gaussian, unknown $\boldsymbol{\mu}$ & unknown $\boldsymbol{\Sigma}$**
$\mathbf{x}$ non-Gaussian, both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ unknown

$\Downarrow$

Inference **Bootstrap variance estimators**
Bootstrap and Delta method

GLM model with Gaussian Covariates

Assume $\boldsymbol{\Sigma}$ is known. Consider the GLM model with Gaussian covariates:

$$\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \mathbb{E}[\mathrm{y} \mid \mathbf{x}] = \phi(\boldsymbol{\beta}^\top \mathbf{x})$$

Let's check

$$\mathbb{E}[\mathrm{xy}] = \mathbb{E}[\mathrm{x}\mathbb{E}[\mathrm{y} \mid \mathrm{x}]] = \mathbb{E}[\mathrm{x}\phi(\boldsymbol{\beta}^\top \mathrm{x})]$$

Stein's Lemma to rescue!

$$\mathbb{E}[\mathrm{xy}] = \mathbb{E}[\mathrm{x}\,\phi(\boldsymbol{\beta}^\top \mathrm{x})] = \mathbb{E}[\phi'(\boldsymbol{\beta}^\top \mathrm{x})]\boldsymbol{\Sigma}\,\boldsymbol{\beta} + \mathbb{E}[\phi(\boldsymbol{\beta}^\top \mathrm{x})]\boldsymbol{\mu}$$

Here,

$$\boldsymbol{\beta}^\top \mathrm{x} \sim \mathcal{N}(\boldsymbol{\beta}^\top \boldsymbol{\mu}, \boldsymbol{\beta}^\top \boldsymbol{\Sigma}\boldsymbol{\beta})$$

Denote:

$$\lambda_{\boldsymbol{\beta}} := \boldsymbol{\beta}^\top \boldsymbol{\mu}, \quad \gamma_{\boldsymbol{\beta}}^2 := \boldsymbol{\beta}^\top \boldsymbol{\Sigma}\boldsymbol{\beta}, \quad \mathrm{f}_i(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2) := \mathbb{E}[\phi^{(i)}(\boldsymbol{\beta}^\top \mathrm{x})]$$

Note $\mathrm{f}_i(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2)$ is known since ϕ is known. Then:

The Redemption of Stein's Lemma

$$\mathbb{E}[\mathrm{xy}] = \mathrm{f}_1(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2)\boldsymbol{\Sigma}\,\boldsymbol{\beta} + \mathrm{f}_0(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2)\boldsymbol{\mu} \tag{1}$$

GLM model with Gaussian Covariates

Assume $\mathbf{\Sigma}$ is known. Consider the GLM model with Gaussian covariates:

$$\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \mathbf{\Sigma}), \quad \mathbb{E}[\mathrm{y} \mid \mathbf{x}] = \phi(\boldsymbol{\beta}^\top \mathbf{x})$$

Let's check

$$\mathbb{E}[\mathbf{x}\mathrm{y}] = \mathbb{E}[\mathbf{x}\mathbb{E}[\mathrm{y} \mid \mathbf{x}]] = \mathbb{E}[\mathbf{x}\phi(\boldsymbol{\beta}^\top \mathbf{x})]$$

Stein's Lemma to rescue!

$$\mathbb{E}[\mathrm{xy}] = \mathbb{E}[\mathrm{x}\,\phi(\boldsymbol{\beta}^\top \mathrm{x})] = \mathbb{E}[\phi'(\boldsymbol{\beta}^\top \mathrm{x})]\mathbf{\Sigma}\,\boldsymbol{\beta} + \mathbb{E}[\phi(\boldsymbol{\beta}^\top \mathrm{x})]\boldsymbol{\mu}$$

Here,

$$\boldsymbol{\beta}^\top \mathrm{x} \sim \mathcal{N}(\boldsymbol{\beta}^\top \boldsymbol{\mu}, \boldsymbol{\beta}^\top \mathbf{\Sigma}\boldsymbol{\beta})$$

Denote:

$$\lambda_{\boldsymbol{\beta}} := \boldsymbol{\beta}^\top \boldsymbol{\mu}, \quad \gamma_{\boldsymbol{\beta}}^2 := \boldsymbol{\beta}^\top \mathbf{\Sigma}\boldsymbol{\beta}, \quad \mathrm{f}_i(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2) := \mathbb{E}[\phi^{(i)}(\boldsymbol{\beta}^\top \mathrm{x})]$$

Note $\mathrm{f}_i(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2)$ is known since ϕ is known. Then:

The Redemption of Stein's Lemma

$$\mathbb{E}[\mathrm{xy}] = \mathrm{f}_1(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2)\mathbf{\Sigma}\,\boldsymbol{\beta} + \mathrm{f}_0(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2)\boldsymbol{\mu} \tag{1}$$

GLM model with Gaussian Covariates

Assume $\mathbf{\Sigma}$ is known. Consider the GLM model with Gaussian covariates:

$$\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \mathbf{\Sigma}), \quad \mathbb{E}[\mathrm{y} \mid \mathbf{x}] = \phi(\boldsymbol{\beta}^\top \mathbf{x})$$

Let's check

$$\mathbb{E}[\mathbf{x}\mathrm{y}] = \mathbb{E}[\mathbf{x}\mathbb{E}[\mathrm{y} \mid \mathbf{x}]] = \mathbb{E}[\mathbf{x}\phi(\boldsymbol{\beta}^\top \mathbf{x})]$$

Stein's Lemma to rescue!

$$\mathbb{E}[\mathbf{x}\mathrm{y}] = \mathbb{E}[\mathbf{x}\,\phi(\boldsymbol{\beta}^\top \mathbf{x})] = \mathbb{E}[\phi'(\boldsymbol{\beta}^\top \mathbf{x})]\mathbf{\Sigma}\,\boldsymbol{\beta} + \mathbb{E}[\phi(\boldsymbol{\beta}^\top \mathbf{x})]\boldsymbol{\mu}$$

Here,

$$\boldsymbol{\beta}^\top \mathrm{x} \sim \mathcal{N}(\boldsymbol{\beta}^\top \boldsymbol{\mu}, \boldsymbol{\beta}^\top \mathbf{\Sigma}\boldsymbol{\beta})$$

Denote:

$$\lambda_{\boldsymbol{\beta}} := \boldsymbol{\beta}^\top \boldsymbol{\mu}, \quad \gamma_{\boldsymbol{\beta}}^2 := \boldsymbol{\beta}^\top \mathbf{\Sigma}\boldsymbol{\beta}, \quad \mathrm{f}_i(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2) := \mathbb{E}[\phi^{(i)}(\boldsymbol{\beta}^\top \mathbf{x})]$$

Note $\mathrm{f}_i(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2)$ is known since ϕ is known. Then:

The Redemption of Stein's Lemma

$$\mathbb{E}[\mathbf{x}\mathrm{y}] = \mathrm{f}_1(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2)\mathbf{\Sigma}\,\boldsymbol{\beta} + \mathrm{f}_0(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2)\boldsymbol{\mu} \tag{1}$$

GLM model with Gaussian Covariates

Assume $\mathbf{\Sigma}$ is known. Consider the GLM model with Gaussian covariates:

$$\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \mathbf{\Sigma}), \quad \mathbb{E}[\mathrm{y} \mid \mathbf{x}] = \phi(\boldsymbol{\beta}^\top \mathbf{x})$$

Let's check

$$\mathbb{E}[\mathbf{x}\mathrm{y}] = \mathbb{E}[\mathbf{x}\mathbb{E}[\mathrm{y} \mid \mathbf{x}]] = \mathbb{E}[\mathbf{x}\phi(\boldsymbol{\beta}^\top \mathbf{x})]$$

Stein's Lemma to rescue!

$$\mathbb{E}[\mathbf{x}\mathrm{y}] = \mathbb{E}[\mathbf{x}\,\phi(\boldsymbol{\beta}^\top \mathbf{x})] = \mathbb{E}[\phi'(\boldsymbol{\beta}^\top \mathbf{x})]\mathbf{\Sigma}\,\boldsymbol{\beta} + \mathbb{E}[\phi(\boldsymbol{\beta}^\top \mathbf{x})]\boldsymbol{\mu}$$

Here,

$$\boldsymbol{\beta}^\top \mathbf{x} \sim \mathcal{N}(\boldsymbol{\beta}^\top \boldsymbol{\mu}, \boldsymbol{\beta}^\top \mathbf{\Sigma}\boldsymbol{\beta})$$

Denote:

$$\lambda_{\beta} := \boldsymbol{\beta}^\top \boldsymbol{\mu}, \quad \gamma_{\beta}^2 := \boldsymbol{\beta}^\top \mathbf{\Sigma}\boldsymbol{\beta}, \quad \mathrm{f}_i(\lambda_{\beta}, \gamma_{\beta}^2) := \mathbb{E}[\phi^{(i)}(\boldsymbol{\beta}^\top \mathbf{x})]$$

Note $\mathrm{f}_i(\lambda_{\beta}, \gamma_{\beta}^2)$ is known since ϕ is known. Then:

The Redemption of Stein's Lemma

$$\mathbb{E}[\mathbf{x}\mathrm{y}] = \mathrm{f}_1(\lambda_{\beta}, \gamma_{\beta}^2)\mathbf{\Sigma}\,\boldsymbol{\beta} + \mathrm{f}_0(\lambda_{\beta}, \gamma_{\beta}^2)\boldsymbol{\mu} \tag{1}$$

GLM model with Gaussian Covariates

Assume $\boldsymbol{\Sigma}$ is known. Consider the GLM model with Gaussian covariates:

$$\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \mathbb{E}[\mathrm{y} \mid \mathbf{x}] = \phi(\boldsymbol{\beta}^\top \mathbf{x})$$

Let's check

$$\mathbb{E}[\mathbf{x}\mathrm{y}] = \mathbb{E}[\mathbf{x}\mathbb{E}[\mathrm{y} \mid \mathbf{x}]] = \mathbb{E}[\mathbf{x}\phi(\boldsymbol{\beta}^\top \mathbf{x})]$$

Stein's Lemma to rescue!

$$\mathbb{E}[\mathbf{x}\mathrm{y}] = \mathbb{E}[\mathbf{x}\,\phi(\boldsymbol{\beta}^\top \mathbf{x})] = \mathbb{E}[\phi'(\boldsymbol{\beta}^\top \mathbf{x})]\boldsymbol{\Sigma}\,\boldsymbol{\beta} + \mathbb{E}[\phi(\boldsymbol{\beta}^\top \mathbf{x})]\boldsymbol{\mu}$$

Here,

$$\boldsymbol{\beta}^\top \mathbf{x} \sim \mathcal{N}(\boldsymbol{\beta}^\top \boldsymbol{\mu}, \boldsymbol{\beta}^\top \boldsymbol{\Sigma}\boldsymbol{\beta})$$

Denote:

$$\lambda_{\boldsymbol{\beta}} := \boldsymbol{\beta}^\top \boldsymbol{\mu}, \quad \gamma_{\boldsymbol{\beta}}^2 := \boldsymbol{\beta}^\top \boldsymbol{\Sigma}\boldsymbol{\beta}, \quad \mathrm{f}_i(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2) := \mathbb{E}[\phi^{(i)}(\boldsymbol{\beta}^\top \mathbf{x})]$$

Note $\mathrm{f}_i(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2)$ is known since ϕ is known. Then:

The Redemption of Stein's Lemma

$$\mathbb{E}[\mathbf{x}\mathrm{y}] = \mathrm{f}_1(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2)\boldsymbol{\Sigma}\,\boldsymbol{\beta} + \mathrm{f}_0(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2)\boldsymbol{\mu} \tag{1}$$

GLM model with Gaussian Covariates

Assume $\boldsymbol{\Sigma}$ is known. Consider the GLM model with Gaussian covariates:

$$\mathbf{x} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \quad \mathbb{E}[\mathrm{y} \mid \mathbf{x}] = \phi(\boldsymbol{\beta}^\top \mathbf{x})$$

Let's check

$$\mathbb{E}[\mathbf{x}\mathrm{y}] = \mathbb{E}[\mathbf{x}\mathbb{E}[\mathrm{y} \mid \mathbf{x}]] = \mathbb{E}[\mathbf{x}\phi(\boldsymbol{\beta}^\top \mathbf{x})]$$

Stein's Lemma to rescue!

$$\mathbb{E}[\mathbf{x}\mathrm{y}] = \mathbb{E}[\mathbf{x}\,\phi(\boldsymbol{\beta}^\top \mathbf{x})] = \mathbb{E}[\phi'(\boldsymbol{\beta}^\top \mathbf{x})]\boldsymbol{\Sigma}\,\boldsymbol{\beta} + \mathbb{E}[\phi(\boldsymbol{\beta}^\top \mathbf{x})]\boldsymbol{\mu}$$

Here,

$$\boldsymbol{\beta}^\top \mathbf{x} \sim \mathcal{N}(\boldsymbol{\beta}^\top \boldsymbol{\mu}, \boldsymbol{\beta}^\top \boldsymbol{\Sigma}\boldsymbol{\beta})$$

Denote:

$$\lambda_{\boldsymbol{\beta}} := \boldsymbol{\beta}^\top \boldsymbol{\mu}, \quad \gamma_{\boldsymbol{\beta}}^2 := \boldsymbol{\beta}^\top \boldsymbol{\Sigma}\boldsymbol{\beta}, \quad \mathrm{f}_i(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2) := \mathbb{E}[\phi^{(i)}(\boldsymbol{\beta}^\top \mathbf{x})]$$

Note $\mathrm{f}_i(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2)$ is known since ϕ is known. Then:

The Redemption of Stein's Lemma

$$\mathbb{E}[\mathbf{x}\mathrm{y}] = \mathrm{f}_1(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2)\boldsymbol{\Sigma}\,\boldsymbol{\beta} + \mathrm{f}_0(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2)\boldsymbol{\mu} \tag{1}$$

Identification under **Case I**: Gaussian, known $\boldsymbol{\mu} = 0$ & known $\boldsymbol{\Sigma}$

$$\mathbb{E}[\mathbf{x}\mathrm{y}] = \mathrm{f}_1(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2)\boldsymbol{\Sigma}\,\boldsymbol{\beta} + \mathrm{f}_0(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2)\boldsymbol{\mu}, \quad \lambda_{\boldsymbol{\beta}} = \boldsymbol{\beta}^\top \boldsymbol{\mu}, \quad \gamma_{\boldsymbol{\beta}}^2 = \boldsymbol{\beta}^\top \boldsymbol{\Sigma}\boldsymbol{\beta}$$

When $\boldsymbol{\mu} = \mathbf{0}$, the expression simplifies:

$$\mathbb{E}[\mathbf{x}\mathrm{y}] = \mathrm{f}_1(0, \gamma_{\boldsymbol{\beta}}^2)\boldsymbol{\Sigma}\,\boldsymbol{\beta}$$

Identification Equations for $\boldsymbol{\mu} = \mathbf{0}$

$$\begin{aligned} \mathrm{m}_{\mathrm{xy},2} &:= \mathbb{E}[\mathrm{y}_1\mathrm{x}_1^\top \boldsymbol{\Sigma}^{-1}\mathrm{x}_2\mathrm{y}_2] = \mathrm{f}_1^2(0, \gamma_{\boldsymbol{\beta}}^2)\cdot \gamma_{\boldsymbol{\beta}}^2 =: \Psi(\gamma_{\boldsymbol{\beta}}^2) \\ \mathrm{m}_{\beta_j} &:= \mathbb{E}[\mathrm{y}\mathrm{x}^\top]\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_j = \mathrm{f}_1(0, \gamma_{\boldsymbol{\beta}}^2)\cdot \beta_j \end{aligned} \tag{2}$$

Here, $\Psi(\gamma_{\boldsymbol{\beta}}^2)$ is strictly monotonic if ϕ is strictly monotonic —so $\gamma_{\boldsymbol{\beta}}^2$ can be recovered via inversion.
Let $\mathbb{U}_{n,1}[h] := \frac{1}{n}\sum_{i=1}^n h(Z_i)$ and $\mathbb{U}_{n,2}[h] := \frac{1}{n(n-1)}\sum_{i\neq j} h(Z_i, Z_j)$ where $Z_i = (\mathrm{y}_i, \mathrm{x}_i)$.

Estimators for $\boldsymbol{\mu} = \mathbf{0}$

$$\hat{\mathrm{m}}_{\mathrm{xy},2} := \mathbb{U}_{n,2}[\mathrm{y}_1\mathrm{x}_1^\top \boldsymbol{\Sigma}^{-1}\mathrm{x}_2\mathrm{y}_2], \quad \hat{\mathrm{m}}_{\beta_j} := \mathbb{U}_{n,1}[\mathrm{y}_1\mathrm{x}_1^\top \boldsymbol{\Sigma}^{-1}\boldsymbol{e}_j]$$

$$\hat{\gamma}_{\boldsymbol{\beta}}^2 = \Psi^{-1}(\hat{\mathrm{m}}_{\mathrm{xy},2}), \quad \hat{\beta}_j = \frac{\hat{\mathrm{m}}_{\beta_j}}{f_1(0, \hat{\gamma}_{\boldsymbol{\beta}}^2)} \tag{3}$$

Identification under **Case I**: Gaussian, known $\mu = 0$ & known $\boldsymbol{\Sigma}$

$$\mathbb{E}[\mathbf{x}\mathrm{y}] = \mathrm{f}_1(\lambda_{\boldsymbol{\beta}}, \gamma^2_{\boldsymbol{\beta}})\boldsymbol{\Sigma}\,\boldsymbol{\beta} + \mathrm{f}_0(\lambda_{\boldsymbol{\beta}}, \gamma^2_{\boldsymbol{\beta}})\boldsymbol{\mu}, \quad \lambda_{\boldsymbol{\beta}} = \boldsymbol{\beta}^\top\boldsymbol{\mu}, \quad \gamma^2_{\boldsymbol{\beta}} = \boldsymbol{\beta}^\top\boldsymbol{\Sigma}\boldsymbol{\beta}$$

When $\boldsymbol{\mu} = \mathbf{0}$, the expression simplifies:

$$\mathbb{E}[\mathbf{x}\mathrm{y}] = \mathrm{f}_1(0, \gamma^2_{\boldsymbol{\beta}})\boldsymbol{\Sigma}\,\boldsymbol{\beta}$$

Identification Equations for $\boldsymbol{\mu} = \mathbf{0}$

$$\begin{aligned} \mathrm{m}_{\mathrm{xy},2} &:= \mathbb{E}[\mathrm{y}_1\mathbf{x}_1^\top\boldsymbol{\Sigma}^{-1}\mathbf{x}_2\mathrm{y}_2] = \mathrm{f}_1^2(0,\gamma^2_{\boldsymbol{\beta}}) \cdot \gamma^2_{\boldsymbol{\beta}} =: \Psi(\gamma^2_{\boldsymbol{\beta}}) \\ \mathrm{m}_{\boldsymbol{\beta}_j} &:= \mathbb{E}[\mathrm{y}\mathbf{x}^\top]\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_j = \mathrm{f}_1(0,\gamma^2_{\boldsymbol{\beta}}) \cdot \boldsymbol{\beta}_j \end{aligned} \tag{2}$$

Here, $\Psi(\gamma^2_{\boldsymbol{\beta}})$ is strictly monotonic if ϕ is strictly monotonic —so $\gamma^2_{\boldsymbol{\beta}}$ can be recovered via inversion.

Let $\mathbb{U}_{n,1}[h] := \frac{1}{n}\sum_{i=1}^n h(Z_i)$ and $\mathbb{U}_{n,2}[h] := \frac{1}{n(n-1)}\sum_{i\neq j} h(Z_i, Z_j)$ where $Z_i = (\mathrm{y}_i, \mathrm{x}_i)$.

Estimators for $\boldsymbol{\mu} = \mathbf{0}$

$$\hat{\mathrm{m}}_{\mathrm{xy},2} := \mathbb{U}_{n,2}[\mathrm{y}_1\mathbf{x}_1^\top\boldsymbol{\Sigma}^{-1}\mathbf{x}_2\mathrm{y}_2], \quad \hat{\mathrm{m}}_{\boldsymbol{\beta}_j} := \mathbb{U}_{n,1}[\mathrm{y}_1\mathbf{x}_1^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_j]$$

$$\hat{\gamma}^2_{\boldsymbol{\beta}} = \Psi^{-1}(\hat{\mathrm{m}}_{\mathrm{xy},2}), \quad \hat{\boldsymbol{\beta}}_j = \frac{\hat{\mathrm{m}}_{\boldsymbol{\beta}_j}}{f_1(0,\hat{\gamma}^2_{\boldsymbol{\beta}})} \tag{3}$$

Identification under **Case I**: Gaussian, known $\boldsymbol{\mu} = 0$ & known $\boldsymbol{\Sigma}$

$$\mathbb{E}[\mathbf{x}\mathrm{y}] = \mathrm{f}_1(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2)\boldsymbol{\Sigma}\,\boldsymbol{\beta} + \mathrm{f}_0(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2)\boldsymbol{\mu}, \quad \lambda_{\boldsymbol{\beta}} = \boldsymbol{\beta}^\top \boldsymbol{\mu}, \quad \gamma_{\boldsymbol{\beta}}^2 = \boldsymbol{\beta}^\top \boldsymbol{\Sigma}\boldsymbol{\beta}$$

When $\boldsymbol{\mu} = \mathbf{0}$, the expression simplifies:

$$\mathbb{E}[\mathbf{x}\mathrm{y}] = \mathrm{f}_1(0, \gamma_{\boldsymbol{\beta}}^2)\boldsymbol{\Sigma}\,\boldsymbol{\beta}$$

Identification Equations for $\boldsymbol{\mu} = \mathbf{0}$

$$\begin{aligned} \mathrm{m}_{\mathrm{xy},2} &:= \mathbb{E}[\mathrm{y}_1 \mathbf{x}_1^\top \boldsymbol{\Sigma}^{-1} \mathbf{x}_2 \mathrm{y}_2] = \mathrm{f}_1^2(0, \gamma_{\boldsymbol{\beta}}^2) \cdot \gamma_{\boldsymbol{\beta}}^2 =: \Psi(\gamma_{\boldsymbol{\beta}}^2) \\ \mathrm{m}_{\boldsymbol{\beta}_j} &:= \mathbb{E}[\mathrm{y}\mathbf{x}^\top]\boldsymbol{\Sigma}^{-1}\boldsymbol{e}_j = \mathrm{f}_1(0, \gamma_{\boldsymbol{\beta}}^2) \cdot \boldsymbol{\beta}_j \end{aligned} \tag{2}$$

Here, $\Psi(\gamma_{\boldsymbol{\beta}}^2)$ is strictly monotonic if ϕ is strictly monotonic —so $\gamma_{\boldsymbol{\beta}}^2$ can be recovered via inversion.

Let $\mathbb{U}_{n,1}[h] := \frac{1}{n}\sum_{i=1}^n h(Z_i)$ and $\mathbb{U}_{n,2}[h] := \frac{1}{n(n-1)}\sum_{i\neq j} h(Z_i, Z_j)$ where $Z_i = (\mathrm{y}_i, \mathbf{x}_i)$.

Estimators for $\boldsymbol{\mu} = \mathbf{0}$

$$\begin{gathered} \hat{\mathrm{m}}_{\mathrm{xy},2} := \mathbb{U}_{n,2}[\mathrm{y}_1 \mathbf{x}_1^\top \boldsymbol{\Sigma}^{-1} \mathbf{x}_2 \mathrm{y}_2], \quad \hat{\mathrm{m}}_{\boldsymbol{\beta}_j} := \mathbb{U}_{n,1}[\mathrm{y}_1 \mathbf{x}_1^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{e}_j] \\ \hat{\gamma}_{\boldsymbol{\beta}}^2 = \Psi^{-1}(\hat{\mathrm{m}}_{\mathrm{xy},2}), \quad \hat{\boldsymbol{\beta}}_j = \frac{\hat{\mathrm{m}}_{\boldsymbol{\beta}_j}}{f_1(0, \hat{\gamma}_{\boldsymbol{\beta}}^2)} \end{gathered} \tag{3}$$

Identification under **Case II**: Gaussian, unknown $\boldsymbol{\mu}$ & known $\boldsymbol{\Sigma}$

Identification Equations for Unknown $\boldsymbol{\mu}$

$$
\begin{aligned}
&\mathrm{m}_1 := \mathrm{m}_{\mathrm{y}} = \mathrm{f}_0(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2),\\
&\mathrm{m}_2 := \mathrm{m}_{\mathrm{xy},2} + \mathrm{m}_{\mathrm{y}}^2 \cdot \mathrm{m}_{\mathrm{x},2} - 2 \cdot \mathrm{m}_{\mathrm{y}} \cdot \mathrm{m}_{\mathrm{xy},\mathrm{x}} = \mathrm{f}_1^2(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2) \cdot \gamma_{\boldsymbol{\beta}}^2,\\
&\Psi_{GLM} : (\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2) \to (m_1, \mathrm{m}_2).\\
&\mathrm{m}_{\nu_j} := \mathbb{E}[\mathbf{x}]^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{e}_j = \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{e}_j = \boldsymbol{\nu}^\top \boldsymbol{e}_j = \nu_j,\\
&\mathrm{m}_{\boldsymbol{\beta}_j} := \mathbb{E}[\mathrm{y}\mathbf{x}^\top]\boldsymbol{\Sigma}^{-1} \boldsymbol{e}_j = \mathrm{f}_0(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2) \cdot \nu_j + \mathrm{f}_1(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2) \cdot \boldsymbol{\beta}_j.
\end{aligned}
\tag{4}
$$

Estimators for Unknown $\boldsymbol{\mu}$

$$
\begin{aligned}
&\hat{m}_1 := \hat{m}_{\mathrm{y}} := \mathbb{U}_{n,1}[\mathrm{y}], \quad \hat{m}_2 := \hat{m}_{\mathrm{xy},2} + \hat{m}_{\mathrm{y}}^2 \cdot \hat{m}_{\mathrm{x},2} - 2 \cdot \hat{m}_{\mathrm{y}} \cdot \hat{m}_{\mathrm{xy},\mathrm{x}},\\
&\hat{m}_{\mathrm{x},2} := \mathbb{U}_{n,2}[\mathrm{x}_1^\top \boldsymbol{\Sigma}^{-1} \mathrm{x}_2], \quad \hat{m}_{\mathrm{xy},\mathrm{x}} := \mathbb{U}_{n,2}[\mathrm{y}_1 \mathrm{x}_1^\top \boldsymbol{\Sigma}^{-1} \mathrm{x}_2],\\
&\hat{m}_{\mathrm{xy},2} := \mathbb{U}_{n,2}[\mathrm{y}_1 \mathrm{x}_1^\top \boldsymbol{\Sigma}^{-1} \mathrm{x}_2 \mathrm{y}_2], \quad \hat{m}_{\nu_j} := \mathbb{U}_{n,1}[\mathrm{x}^\top]\boldsymbol{\Sigma}^{-1} \boldsymbol{e}_j,\\
&\hat{m}_{\beta_j} := \mathbb{U}_{n,1}[\mathrm{y}\mathrm{x}^\top]\boldsymbol{\Sigma}^{-1} \boldsymbol{e}_j.\\
&(\hat{\lambda_{\beta}}, \hat{\gamma_{\beta}^2}) := \Psi_{GLM}^{-1}(\hat{m}_1, \hat{m}_2), \quad \hat{\beta}_j := \frac{\hat{m}_{\beta_j} - \mathrm{f}_0(\hat{\lambda}_{\beta}, \hat{\gamma_{\beta}}^2) \cdot \hat{m}_{\nu_j}}{\mathrm{f}_1(\hat{\lambda}_{\beta}, \hat{\gamma_{\beta}}^2)}.
\end{aligned}
\tag{5}
$$

Identification under **Case II**: Gaussian, unknown $\boldsymbol{\mu}$ & known $\boldsymbol{\Sigma}$

Identification Equations for Unknown $\boldsymbol{\mu}$

$$
\begin{aligned}
&\mathrm{m}_1 := \mathrm{m}_{\mathrm{y}} = \mathrm{f}_0(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2),\\
&\mathrm{m}_2 := \mathrm{m}_{\mathrm{xy},2} + \mathrm{m}_{\mathrm{y}}^2 \cdot \mathrm{m}_{\mathrm{x},2} - 2 \cdot \mathrm{m}_{\mathrm{y}} \cdot \mathrm{m}_{\mathrm{xy},\mathrm{x}} = \mathrm{f}_1^2(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2) \cdot \gamma_{\boldsymbol{\beta}}^2,\\
&\Psi_{GLM} : (\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2) \to (m_1, \mathrm{m}_2).\\
&\mathrm{m}_{\nu_j} := \mathbb{E}[\mathbf{x}]^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{e}_j = \boldsymbol{\mu}^\top \boldsymbol{\Sigma}^{-1} \boldsymbol{e}_j = \boldsymbol{\nu}^\top \boldsymbol{e}_j = \nu_j,\\
&\mathrm{m}_{\boldsymbol{\beta}_j} := \mathbb{E}[\mathrm{y}\mathbf{x}^\top]\boldsymbol{\Sigma}^{-1} \boldsymbol{e}_j = \mathrm{f}_0(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2) \cdot \nu_j + \mathrm{f}_1(\lambda_{\boldsymbol{\beta}}, \gamma_{\boldsymbol{\beta}}^2) \cdot \boldsymbol{\beta}_j.
\end{aligned}
\tag{4}
$$

Estimators for Unknown $\boldsymbol{\mu}$

$$
\begin{aligned}
&\hat{m}_1 := \hat{m}_{\mathrm{y}} := \mathbb{U}_{n,1}[\mathrm{y}], \quad \hat{m}_2 := \hat{m}_{\mathrm{xy},2} + \hat{m}_{\mathrm{y}}^2 \cdot \hat{m}_{\mathrm{x},2} - 2 \cdot \hat{m}_{\mathrm{y}} \cdot \hat{m}_{\mathrm{xy},\mathrm{x}},\\
&\hat{m}_{\mathrm{x},2} := \mathbb{U}_{n,2}[\mathbf{x}_1^\top \boldsymbol{\Sigma}^{-1} \mathbf{x}_2], \quad \hat{m}_{\mathrm{xy},\mathrm{x}} := \mathbb{U}_{n,2}[\mathrm{y}_1 \mathbf{x}_1^\top \boldsymbol{\Sigma}^{-1} \mathbf{x}_2],\\
&\hat{m}_{\mathrm{xy},2} := \mathbb{U}_{n,2}[\mathrm{y}_1 \mathbf{x}_1^\top \boldsymbol{\Sigma}^{-1} \mathbf{x}_2 \mathrm{y}_2], \quad \hat{m}_{\nu_j} := \mathbb{U}_{n,1}[\mathbf{x}^\top]\boldsymbol{\Sigma}^{-1} \boldsymbol{e}_j,\\
&\hat{m}_{\boldsymbol{\beta}_j} := \mathbb{U}_{n,1}[\mathrm{y}\mathbf{x}^\top]\boldsymbol{\Sigma}^{-1} \boldsymbol{e}_j.\\
&(\hat{\lambda_{\boldsymbol{\beta}}}, \hat{\gamma_{\boldsymbol{\beta}}^2}) := \Psi_{GLM}^{-1}(\hat{m}_1, \hat{m}_2), \quad \hat{\boldsymbol{\beta}}_j := \frac{\hat{m}_{\boldsymbol{\beta}_j} - \mathrm{f}_0(\hat{\lambda}_{\boldsymbol{\beta}}, \hat{\gamma_{\boldsymbol{\beta}}}^2) \cdot \hat{m}_{\nu_j}}{\mathrm{f}_1(\hat{\lambda}_{\boldsymbol{\beta}}, \hat{\gamma_{\boldsymbol{\beta}}}^2)}.
\end{aligned}
\tag{5}
$$

$\sqrt{n}$-consistency and CAN

Theorem 2 (C., Liu, Mukherjee, 24)

Under some mild conditions, when $\mathbf{x} \sim \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the following is a $\sqrt{n}$-consistent and CAN estimator of $(\boldsymbol{\beta}_j, \lambda_{\boldsymbol{\beta}}, \gamma^2_{\boldsymbol{\beta}})$

$$(\hat{\lambda_{\boldsymbol{\beta}}}, \hat{\gamma^2_{\boldsymbol{\beta}}}) := \Psi^{-1}_{GLM}(\hat{m}_1, \hat{m}_2), \quad \hat{\boldsymbol{\beta}}_j := \frac{\hat{m}_{\boldsymbol{\beta}_j} - \mathrm{f}_0(\hat{\lambda}_{\boldsymbol{\beta}}, \hat{\gamma_{\boldsymbol{\beta}}}^2) \cdot \hat{m}_{\nu_j}}{\mathrm{f}_1(\hat{\lambda}_{\boldsymbol{\beta}}, \hat{\gamma_{\boldsymbol{\beta}}}^2)}.$$

Proof sketch(Delta method).

$\sqrt{n}$-consistency & CAN follow from (1) the $\sqrt{n}$-consistency & CAN of U-statistics; (2) Ψ_{GLM} is a diffeomorphism □

Identification under **Case III**: Gaussian, unknown μ & unknown Σ

- Identification Equations will be invariant.
- The knowledge of Σ influences the construction of moment estimators.
- One lazy method involves using a sample splitting strategy with weighted sample covariance under $I_1 \cup I_2 = [n]$, $|I_1| = |I_2| = n/2$

Moment Estimators with Unknown Σ (Sample Splitting)

$$\hat{m}_{\mathrm{xy},2} := \frac{1}{\frac{n}{2}\left(\frac{n}{2}-1\right)} \sum_{i_1 \neq i_2 \in I_1} \mathrm{y}_{i_1} \mathrm{x}_{i_1}^\top \tilde{\Sigma}^{-1} \mathrm{x}_{i_2} \mathrm{y}_{i_2},$$

$$\tilde{\Sigma} := \frac{1}{\frac{n}{2}-p-1} \sum_{j \in I_2} (\mathrm{x}_j - \bar{\mathrm{x}}_{I_2})(\mathrm{x}_j - \bar{\mathrm{x}}_{I_2})^\top, \quad \bar{\mathrm{x}}_{I_2} := \frac{2}{n} \sum_{j \in I_2} \mathrm{x}_j. \tag{6}$$

- When $p > n/2$, one can apply multiple rounds of sample splitting (e.g., leave-2-out) to construct unbiased estimators.

Identification under **Case III**: Gaussian, unknown $\boldsymbol{\mu}$ & unknown $\boldsymbol{\Sigma}$

- Identification Equations will be invariant.
- The knowledge of $\boldsymbol{\Sigma}$ influences the construction of moment estimators.
- One lazy method involves using a sample splitting strategy with weighted sample covariance under $I_1 \cup I_2 = [n]$, $|I_1| = |I_2| = n/2$

Moment Estimators with Unknown $\boldsymbol{\Sigma}$ (Sample Splitting)

$$
\begin{aligned}
\hat{m}_{\mathrm{xy},2} &:= \frac{1}{\frac{n}{2}\left(\frac{n}{2}-1\right)} \sum_{i_1 \neq i_2 \in I_1} \mathrm{y}_{i_1} \mathrm{x}_{i_1}^{\top} \tilde{\Sigma}^{-1} \mathrm{x}_{i_2} \mathrm{y}_{i_2}, \\
\tilde{\Sigma} &:= \frac{1}{\frac{n}{2}-p-1} \sum_{j \in I_2} (\mathrm{x}_j - \bar{\mathrm{x}}_{I_2})(\mathrm{x}_j - \bar{\mathrm{x}}_{I_2})^{\top}, \quad \bar{\mathrm{x}}_{I_2} := \frac{2}{n} \sum_{j \in I_2} \mathrm{x}_j.
\end{aligned}
\tag{6}
$$

- When $p > n/2$, one can apply multiple rounds of sample splitting (e.g., leave-2-out) to construct unbiased estimators.

Identification under **Case III**: Gaussian, unknown $\boldsymbol{\mu}$ & unknown $\boldsymbol{\Sigma}$

- Identification Equations will be invariant.
- The knowledge of $\boldsymbol{\Sigma}$ influences the construction of moment estimators.
- One lazy method involves using a sample splitting strategy with weighted sample covariance under $I_1 \cup I_2 = [n],\ |I_1| = |I_2| = n/2$

Moment Estimators with Unknown $\boldsymbol{\Sigma}$ (Sample Splitting)

$$
\begin{aligned}
\hat{m}_{\mathrm{xy},2} &:= \frac{1}{\frac{n}{2}\left(\frac{n}{2}-1\right)} \sum_{i_1 \neq i_2 \in I_1} \mathrm{y}_{i_1} \mathbf{x}_{i_1}^{\top} \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{x}_{i_2} \mathrm{y}_{i_2}, \\
\tilde{\boldsymbol{\Sigma}} &:= \frac{1}{\frac{n}{2}-p-1} \sum_{j \in I_2} (\mathbf{x}_j - \bar{\mathbf{x}}_{I_2})(\mathbf{x}_j - \bar{\mathbf{x}}_{I_2})^{\top}, \quad \bar{\mathbf{x}}_{I_2} := \frac{2}{n} \sum_{j \in I_2} \mathbf{x}_j.
\end{aligned}
\tag{6}
$$

- When $p > n/2$, one can apply multiple rounds of sample splitting (e.g., leave-2-out) to construct unbiased estimators.

Identification under **Case III**: Gaussian, unknown $\boldsymbol{\mu}$ & unknown $\boldsymbol{\Sigma}$

- Identification Equations will be invariant.
- The knowledge of $\boldsymbol{\Sigma}$ influences the construction of moment estimators.
- One lazy method involves using a sample splitting strategy with weighted sample covariance under $I_1 \cup I_2 = [n]$, $|I_1| = |I_2| = n/2$

Moment Estimators with Unknown $\boldsymbol{\Sigma}$ (Sample Splitting)

$$
\begin{aligned}
\hat{m}_{\mathrm{xy},2} &:= \frac{1}{\frac{n}{2}\left(\frac{n}{2}-1\right)} \sum_{i_1 \neq i_2 \in I_1} \mathrm{y}_{i_1} \mathbf{x}_{i_1}^{\top} \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{x}_{i_2} \mathrm{y}_{i_2}, \\
\tilde{\boldsymbol{\Sigma}} &:= \frac{1}{\frac{n}{2}-p-1} \sum_{j \in I_2} (\mathbf{x}_j - \bar{\mathbf{x}}_{I_2})(\mathbf{x}_j - \bar{\mathbf{x}}_{I_2})^{\top}, \quad \bar{\mathbf{x}}_{I_2} := \frac{2}{n} \sum_{j \in I_2} \mathbf{x}_j.
\end{aligned}
\tag{6}
$$

- When $p > n/2$, one can apply multiple rounds of sample splitting (e.g., leave-2-out) to construct unbiased estimators.

Identification under **Case III**: Gaussian, unknown $\boldsymbol{\mu}$ & unknown $\boldsymbol{\Sigma}$

- The sample splitting strategy will no longer useful when $p > \frac{n}{2}$.
- An alternative method involve the Chebyshev polynomial approximate first considered in Kong and Valiant 18

Moment Estimators Unknown $\boldsymbol{\Sigma}$ (Chebyshev) $\boldsymbol{\mu} = \mathbf{0}$

$$\begin{aligned} \boldsymbol{\Sigma}^{-1} &\approx \sum_{l=0}^{J} c_l \boldsymbol{\Sigma}^l, \\ \hat{m}_{\mathrm{xy},2} &:= \sum_{l=0}^{J} c_l \mathbb{U}_{n,l+2}\left[\mathrm{y}_1 \mathbf{x}_1^\top \left(\prod_{s=3}^{l+2} \mathbf{x}_s \mathbf{x}_s^\top\right) \mathbf{x}_2 \mathrm{y}_2\right]. \end{aligned} \tag{7}$$

- under same condition in Theorem 2, our estimators are consistent.

Identification under **Case III**: Gaussian, unknown $\boldsymbol{\mu}$ & unknown $\boldsymbol{\Sigma}$

- The sample splitting strategy will no longer useful when $p > \frac{n}{2}$.
- An alternative method involve the Chebyshev polynomial approximate first considered in Kong and Valiant 18

Moment Estimators Unknown $\boldsymbol{\Sigma}$ (Chebyshev) $\boldsymbol{\mu} = \mathbf{0}$

$$\begin{aligned}
\boldsymbol{\Sigma}^{-1} &\approx \sum_{l=0}^{J} c_l \boldsymbol{\Sigma}^l, \\
\hat{m}_{\mathrm{xy},2} &:= \sum_{l=0}^{J} c_l \mathbb{U}_{n,l+2} \left[\mathrm{y}_1 \mathbf{x}_1^\top \left(\prod_{s=3}^{l+2} \mathbf{x}_s \mathbf{x}_s^\top \right) \mathbf{x}_2 \mathrm{y}_2 \right].
\end{aligned} \tag{7}$$

- under same condition in Theorem 2, our estimators are consistent.

Identification under **Case III**: Gaussian, unknown $\boldsymbol{\mu}$ & unknown $\boldsymbol{\Sigma}$

- The sample splitting strategy will no longer useful when $p > \frac{n}{2}$.
- An alternative method involve the Chebyshev polynomial approximate first considered in Kong and Valiant 18

Moment Estimators Unknown $\boldsymbol{\Sigma}$ (Chebyshev) $\boldsymbol{\mu} = \mathbf{0}$

$$
\begin{aligned}
\boldsymbol{\Sigma}^{-1} &\approx \sum_{l=0}^{J} c_l \boldsymbol{\Sigma}^l, \\
\hat{m}_{\mathrm{xy},2} &:= \sum_{l=0}^{J} c_l \mathbb{U}_{n,l+2} \left[\mathrm{y}_1 \mathbf{x}_1^\top \left(\prod_{s=3}^{l+2} \mathbf{x}_s \mathbf{x}_s^\top \right) \mathbf{x}_2 \mathrm{y}_2 \right].
\end{aligned}
\tag{7}
$$

- under same condition in Theorem 2, our estimators are consistent.

Identification under **Case III**: Gaussian, unknown $\boldsymbol{\mu}$ & unknown $\boldsymbol{\Sigma}$

- The sample splitting strategy will no longer useful when $p > \frac{n}{2}$.
- An alternative method involve the Chebyshev polynomial approximate first considered in Kong and Valiant 18

Moment Estimators Unknown $\boldsymbol{\Sigma}$ (Chebyshev) $\boldsymbol{\mu} = \mathbf{0}$

$$\begin{aligned} \boldsymbol{\Sigma}^{-1} &\approx \sum_{l=0}^{J} c_l \boldsymbol{\Sigma}^l, \\ \hat{m}_{\mathrm{xy},2} &:= \sum_{l=0}^{J} c_l \mathbb{U}_{n,l+2}\left[\mathrm{y}_1 \mathbf{x}_1^\top \left(\prod_{s=3}^{l+2} \mathbf{x}_s \mathbf{x}_s^\top\right) \mathbf{x}_2 \mathrm{y}_2\right]. \end{aligned} \tag{7}$$

- under same condition in Theorem 2, our estimators are consistent.

Identification under **Case IV**: **Non-Gaussian**

- When $\mathbf{x}$ validate the Gaussian assumption, the Identification Equations above will no longer hold.

- But under some assumption, Identification Equations can hold approximately

Lemma 1 (C., Liu, Mukherjee, 24)

When $\boldsymbol{\Sigma}^{-1/2}(\mathbf{X}-\boldsymbol{\mu})$ has zero mean and unit variance, the above Identification Equations approximately with approximation error $\mathcal{O}\left(p^{-3/4}\right)=\mathcal{O}\left(n^{-3/4}\right)$ as $n \to \infty$.

- Thus, above results of $\sqrt{n}$-consistency or consistency still hold.

- As for CAN property we need some assumption on limiting distribution on $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$: $\sqrt{p}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\beta} \overset{\mathcal{W}_8}{\to} \mathrm{b}$ and $\sqrt{p}\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\mu} \overset{\mathcal{W}_8}{\to} \mathrm{u}$ where $\mathrm{b} \sim \rho$ and $\mathrm{u} \sim \varrho$ respectively for some probability measures ρ and ϱ supported on $\mathbb{R}$ and both ρ and ϱ have bounded first and second moments.

Identification under **Case IV**: **Non-Gaussian**

- When $\mathbf{x}$ validate the Gaussian assumption, the Identification Equations above will no longer hold.

- But under some assumption, Identification Equations can hold approximately

Lemma 1 (C., Liu, Mukherjee, 24)

When $\mathbf{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$ has zero mean and unit variance, the above Identification Equations approximately with approximation error $\mathcal{O}\left(p^{-3/4}\right) = \mathcal{O}\left(n^{-3/4}\right)$ as $n \to \infty$.

- Thus, above results of $\sqrt{n}$-consistency or consistency still hold.

- As for CAN property we need some assumption on limiting distribution on β and $\boldsymbol{\mu}$: $\sqrt{p}\Sigma^{1/2}\beta \xrightarrow{\mathcal{W}_8}$ b and $\sqrt{p}\Sigma^{-1/2}\boldsymbol{\mu} \xrightarrow{\mathcal{W}_8}$ u where b $\sim \rho$ and u $\sim \varrho$ respectively for some probability measures ρ and ϱ supported on $\mathbb{R}$ and both ρ and ϱ have bounded first and second moments.

Identification under **Case IV**: **Non-Gaussian**

- When $\mathbf{x}$ validate the Gaussian assumption, the Identification Equations above will no longer hold.

- But under some assumption, Identification Equations can hold approximately

Lemma 1 (C., Liu, Mukherjee, 24)

When $\mathbf{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$ has zero mean and unit variance, the above Identification Equations approximately with approximation error $\mathcal{O}\left(p^{-3/4}\right) = \mathcal{O}\left(n^{-3/4}\right)$ as $n \to \infty$.

- Thus, above results of $\sqrt{n}$-consistency or consistency still hold.

- As for CAN property we need some assumption on limiting distribution on β and $\boldsymbol{\mu}$: $\sqrt{p}\Sigma^{1/2}\beta \overset{\mathcal{W}_8}{\to}$ b and $\sqrt{p}\Sigma^{-1/2}\boldsymbol{\mu} \overset{\mathcal{W}_8}{\to}$ u where b $\sim \rho$ and u $\sim \varrho$ respectively for some probability measures ρ and ϱ supported on $\mathbb{R}$ and both ρ and ϱ have bounded first and second moments.

Identification under **Case IV**: **Non-Gaussian**

- When $\mathbf{x}$ validate the Gaussian assumption, the Identification Equations above will no longer hold.

- But under some assumption, Identification Equations can hold approximately

Lemma 1 (C., Liu, Mukherjee, 24)

When $\mathbf{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$ has zero mean and unit variance, the above Identification Equations approximately with approximation error $\mathcal{O}\left(p^{-3/4}\right) = \mathcal{O}\left(n^{-3/4}\right)$ as $n \to \infty$.

- Thus, above results of $\sqrt{n}$-consistency or consistency still hold.

- As for CAN property we need some assumption on limiting distribution on $\boldsymbol{\beta}$ and $\boldsymbol{\mu}$: $\sqrt{p}\mathbf{\Sigma}^{1/2}\boldsymbol{\beta} \overset{\mathcal{W}_8}{\to}$ b and $\sqrt{p}\mathbf{\Sigma}^{-1/2}\boldsymbol{\mu} \overset{\mathcal{W}_8}{\to}$ u where b $\sim \rho$ and u $\sim \varrho$ respectively for some probability measures ρ and ϱ supported on $\mathbb{R}$ and both ρ and ϱ have bounded first and second moments.

Bootstrap variance estimators

- Confidence intervals can also be built by using the following bootstrap procedure
- Taking the estimator $\hat{\mathrm{m}} := \mathbb{U}_{n,2}[\mathrm{y}_1\mathbf{x}_1^\top\boldsymbol{\Sigma}^{-1}\mathbf{x}_2\mathrm{y}_2]$ of $\mathrm{m} := \mathsf{E}[\mathrm{y}\mathbf{x}^\top]\boldsymbol{\Sigma}^{-1}\mathsf{E}[\mathbf{x}\mathrm{y}]$ as an example
- Drawing weights $\{\mathrm{w}_i^{(b)}\}_{i=1}^n \sim \mathsf{multinom}(n; 1/n, \cdots, 1/n)$ for $b = 1, \cdots, B$
- For $b = 1, \cdots, B$, compute
  $\hat{\mathrm{m}}^{(b)} := \mathbb{U}_{n,2}[\mathrm{w}_1^{(b)}\mathrm{y}_1\mathbf{x}_1^\top\boldsymbol{\Sigma}^{-1}\mathrm{w}_2^{(b)}\mathbf{x}_2\mathrm{y}_2]$,
  $\hat{\mathrm{m}}_{\mathrm{center}}^{(b)} := \mathbb{U}_{n,2}[(\mathrm{w}_1^{(b)} - 1)\mathrm{y}_1\mathbf{x}_1^\top\boldsymbol{\Sigma}^{-1}(\mathrm{w}_2^{(b)} - 1)\mathbf{x}_2\mathrm{y}_2]$
- Estimate the variance of $\hat{\mathrm{m}}$ by

$$\hat{V} := \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\mathrm{m}}^{(b)} - \frac{1}{B}\sum_{b'=1}^{B}\hat{\mathrm{m}}^{(b')}\right)^2 - \frac{2}{B}\sum_{b=1}^{B}\left(\hat{\mathrm{m}}_{\mathrm{center}}^{(b)} - \frac{1}{B}\sum_{b'=1}^{B}\hat{\mathrm{m}}_{\mathrm{center}}^{(b')}\right)^2$$

Bootstrap variance estimators

- Confidence intervals can also be built by using the following bootstrap procedure

- Taking the estimator $\hat{\mathrm{m}} := \mathbb{U}_{n,2}[\mathrm{y}_1 \mathbf{x}_1^\top \boldsymbol{\Sigma}^{-1} \mathbf{x}_2 \mathrm{y}_2]$ of $\mathrm{m} := \mathsf{E}[\mathrm{y}\mathbf{x}^\top] \boldsymbol{\Sigma}^{-1} \mathsf{E}[\mathbf{x}\mathrm{y}]$ as an example

- Drawing weights $\{\mathrm{w}_i^{(b)}\}_{i=1}^n \sim \mathrm{multinom}(n; 1/n, \cdots, 1/n)$ for $b = 1, \cdots, B$

- For $b = 1, \cdots, B$, compute
  $\hat{\mathrm{m}}^{(b)} := \mathbb{U}_{n,2}[\mathrm{w}_1^{(b)} \mathrm{y}_1 \mathbf{x}_1^\top \boldsymbol{\Sigma}^{-1} \mathrm{w}_2^{(b)} \mathbf{x}_2 \mathrm{y}_2]$,
  $\hat{\mathrm{m}}_{\mathrm{center}}^{(b)} := \mathbb{U}_{n,2}[(\mathrm{w}_1^{(b)} - 1) \mathrm{y}_1 \mathbf{x}_1^\top \boldsymbol{\Sigma}^{-1} (\mathrm{w}_2^{(b)} - 1) \mathbf{x}_2 \mathrm{y}_2]$

- Estimate the variance of $\hat{\mathrm{m}}$ by

$$\hat{V} := \frac{1}{B} \sum_{b=1}^{B} \left(\hat{\mathrm{m}}^{(b)} - \frac{1}{B} \sum_{b'=1}^{B} \hat{\mathrm{m}}^{(b')} \right)^2 - \frac{2}{B} \sum_{b=1}^{B} \left(\hat{\mathrm{m}}_{\mathrm{center}}^{(b)} - \frac{1}{B} \sum_{b'=1}^{B} \hat{\mathrm{m}}_{\mathrm{center}}^{(b')} \right)^2$$

Bootstrap variance estimators

- Confidence intervals can also be built by using the following bootstrap procedure

- Taking the estimator $\hat{\mathrm{m}} := \mathbb{U}_{n,2}[\mathrm{y}_1 \mathbf{x}_1^\top \boldsymbol{\Sigma}^{-1} \mathbf{x}_2 \mathrm{y}_2]$ of $\mathrm{m} := \mathsf{E}[\mathrm{y}\mathbf{x}^\top] \boldsymbol{\Sigma}^{-1} \mathsf{E}[\mathbf{x}\mathrm{y}]$ as an example

- Drawing weights $\{\mathrm{w}_i^{(b)}\}_{i=1}^n \sim \mathsf{multinom}(n; 1/n, \cdots, 1/n)$ for $b = 1, \cdots, B$

- For $b = 1, \cdots, B$, compute
  $\hat{\mathrm{m}}^{(b)} := \mathbb{U}_{n,2}[\mathrm{w}_1^{(b)} \mathrm{y}_1 \mathbf{x}_1^\top \boldsymbol{\Sigma}^{-1} \mathrm{w}_2^{(b)} \mathbf{x}_2 \mathrm{y}_2]$,
  $\hat{\mathrm{m}}_{\mathrm{center}}^{(b)} := \mathbb{U}_{n,2}[(\mathrm{w}_1^{(b)} - 1) \mathrm{y}_1 \mathbf{x}_1^\top \boldsymbol{\Sigma}^{-1} (\mathrm{w}_2^{(b)} - 1) \mathbf{x}_2 \mathrm{y}_2]$

- Estimate the variance of $\hat{\mathrm{m}}$ by

$$\hat{V} := \frac{1}{B} \sum_{b=1}^{B} \left(\hat{\mathrm{m}}^{(b)} - \frac{1}{B} \sum_{b'=1}^{B} \hat{\mathrm{m}}^{(b')} \right)^2 - \frac{2}{B} \sum_{b=1}^{B} \left(\hat{\mathrm{m}}_{\mathrm{center}}^{(b)} - \frac{1}{B} \sum_{b'=1}^{B} \hat{\mathrm{m}}_{\mathrm{center}}^{(b')} \right)^2$$

Bootstrap variance estimators

- Confidence intervals can also be built by using the following bootstrap procedure
- Taking the estimator $\hat{\mathrm{m}} := \mathbb{U}_{n,2}[\mathrm{y}_1 \mathbf{x}_1^\top \mathbf{\Sigma}^{-1} \mathbf{x}_2 \mathrm{y}_2]$ of $\mathrm{m} := \mathsf{E}[\mathrm{y}\mathbf{x}^\top]\mathbf{\Sigma}^{-1}\mathsf{E}[\mathbf{x}\mathrm{y}]$ as an example
- Drawing weights $\{\mathrm{w}_i^{(b)}\}_{i=1}^n \sim \mathsf{multinom}(n; 1/n, \cdots, 1/n)$ for $b = 1, \cdots, B$
- For $b = 1, \cdots, B$, compute
  $\hat{\mathrm{m}}^{(b)} := \mathbb{U}_{n,2}[\mathrm{w}_1^{(b)} \mathrm{y}_1 \mathbf{x}_1^\top \mathbf{\Sigma}^{-1} \mathrm{w}_2^{(b)} \mathbf{x}_2 \mathrm{y}_2]$,
  $\hat{\mathrm{m}}^{(b)}_{\mathrm{center}} := \mathbb{U}_{n,2}[(\mathrm{w}_1^{(b)} - 1) \mathrm{y}_1 \mathbf{x}_1^\top \mathbf{\Sigma}^{-1} (\mathrm{w}_2^{(b)} - 1) \mathbf{x}_2 \mathrm{y}_2]$
- Estimate the variance of $\hat{\mathrm{m}}$ by

$$\hat{V} := \frac{1}{B}\sum_{b=1}^{B}\left(\hat{\mathrm{m}}^{(b)} - \frac{1}{B}\sum_{b'=1}^{B}\hat{\mathrm{m}}^{(b')}\right)^2 - \frac{2}{B}\sum_{b=1}^{B}\left(\hat{\mathrm{m}}^{(b)}_{\mathrm{center}} - \frac{1}{B}\sum_{b'=1}^{B}\hat{\mathrm{m}}^{(b')}_{\mathrm{center}}\right)^2$$

Bootstrap variance estimators

- Confidence intervals can also be built by using the following bootstrap procedure
- Taking the estimator $\hat{\mathrm{m}} := \mathbb{U}_{n,2}[\mathrm{y}_1 \mathbf{x}_1^\top \boldsymbol{\Sigma}^{-1} \mathbf{x}_2 \mathrm{y}_2]$ of $\mathrm{m} := \mathsf{E}[\mathrm{y}\mathbf{x}^\top] \boldsymbol{\Sigma}^{-1} \mathsf{E}[\mathbf{x}\mathrm{y}]$ as an example
- Drawing weights $\{\mathrm{w}_i^{(b)}\}_{i=1}^n \sim \mathsf{multinom}(n; 1/n, \cdots, 1/n)$ for $b = 1, \cdots, B$
- For $b = 1, \cdots, B$, compute
  $\hat{\mathrm{m}}^{(b)} := \mathbb{U}_{n,2}[\mathrm{w}_1^{(b)} \mathrm{y}_1 \mathbf{x}_1^\top \boldsymbol{\Sigma}^{-1} \mathrm{w}_2^{(b)} \mathbf{x}_2 \mathrm{y}_2]$,
  $\hat{\mathrm{m}}_{\mathrm{center}}^{(b)} := \mathbb{U}_{n,2}[(\mathrm{w}_1^{(b)} - 1) \mathrm{y}_1 \mathbf{x}_1^\top \boldsymbol{\Sigma}^{-1} (\mathrm{w}_2^{(b)} - 1) \mathbf{x}_2 \mathrm{y}_2]$
- Estimate the variance of $\hat{\mathrm{m}}$ by

$$\hat{V} := \frac{1}{B} \sum_{b=1}^{B} \left(\hat{\mathrm{m}}^{(b)} - \frac{1}{B} \sum_{b'=1}^{B} \hat{\mathrm{m}}^{(b')} \right)^2 - \frac{2}{B} \sum_{b=1}^{B} \left(\hat{\mathrm{m}}_{\mathrm{center}}^{(b)} - \frac{1}{B} \sum_{b'=1}^{B} \hat{\mathrm{m}}_{\mathrm{center}}^{(b')} \right)^2$$

Outline

Motivations: Estimating Functionals of High-Dimensional GLMs

Our proposal – Moments-based estimators

Numerical experiments

Conclusion and open ends

A peek at some numerical results: GLMs

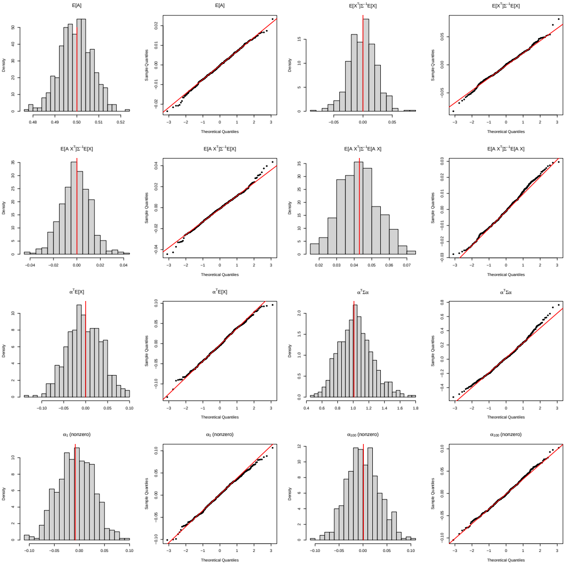

Figure: α_j's in logistic regression: known $\boldsymbol{\Sigma}$ Gaussian design and $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_p) \overset{\text{i.i.d.}}{\sim} \mathrm{Uniform}([-\sqrt{3/p}, \sqrt{3/p}])$, $p = 1.2n$, $n = 5000$

A peek at some numerical results: GLMs

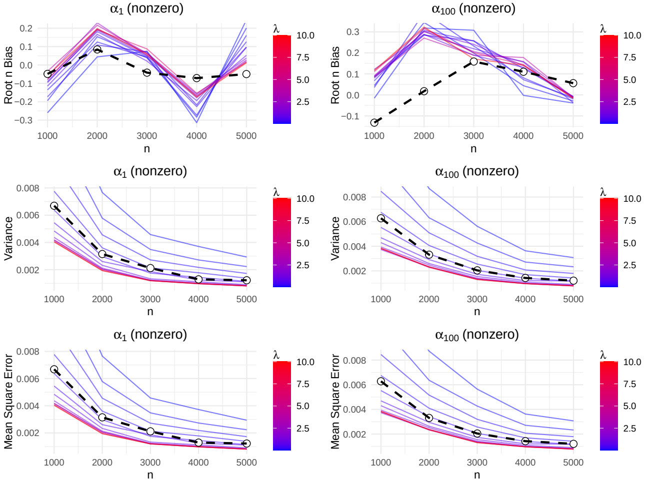

Figure: Comparision with Bellec 23: known $\boldsymbol{\Sigma}$ Gaussian design and $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_p) \overset{\text{i.i.d.}}{\sim} \text{Uniform}([-\sqrt{3/p}, \sqrt{3/p}])$, $p = 1.2n$, $n = 5000$

Boostrap variance estimator: GLM

Table: Bootstrap Variance Estimators vs. Monte Carlo Variances under (Gaussian design and dense regression coefficients), Based on 500 Monte Carlo Simulations with $n = 5000$, $p/n = 1.2$. Here $\boldsymbol{\mu}$ is unknown but $\boldsymbol{\Sigma}$ is known.

	MC Var	Mean Est. Var	$\frac{\text{Mean Est. Var}}{\text{MC Var}}$	Std Est. Var	MSE
$\mathbb{E}A$	4.81e-05	5.01e-05	1.041	2.24e-06	3.00e-06
$\mathbb{E}[A\mathbf{X}^\top]\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$	4.38e-04	4.77e-04	1.091	7.04e-05	8.08e-05
$\mathbb{E}[A\mathbf{X}^\top]\boldsymbol{\Sigma}^{-1}\mathbb{E}[A\mathbf{X}]$	1.85e-04	1.87e-04	1.009	2.83e-05	2.84e-05
$\mathbb{E}[A\mathbf{X}^\top]\boldsymbol{\Sigma}^{-1}\mathbb{E}[A\mathbf{X}]$	1.41e-04	1.36e-04	0.965	2.03e-05	2.09e-05
$\boldsymbol{\alpha}^\top\boldsymbol{\mu}$	2.96e-03	3.01e-03	1.017	1.62e-03	1.62e-03
$\boldsymbol{\alpha}^\top\boldsymbol{\Sigma}\boldsymbol{\alpha}$	6.42e-02	6.76e-02	1.053	3.69e-02	3.71e-02
α_1	1.15e-03	1.20e-03	1.043	1.02e-04	1.13e-04
α_{100}	1.13e-03	1.20e-03	1.060	1.08e-04	1.28e-04

A peek at some numerical results: Estimating $\mathsf{E}[\mathrm{y}]$ under MAR

Compared with Celentano & Wainwright, 23: based on debiased Lasso + AMP theory under Gaussian design

Our approach: a system of moment equations can be used to identify $\psi = \mathsf{E}[\mathrm{y}] = \boldsymbol{\beta}^\top \boldsymbol{\mu}$ in the following model:

$$\mathrm{y} = \boldsymbol{\beta}^\top \mathbf{x} + \varepsilon, \mathrm{t}|\mathbf{x} \sim \mathrm{Bern}(\phi(\boldsymbol{\alpha}^\top \mathbf{x}))$$

based on estimating the following moments

$$\mathsf{E}(\mathrm{t}), \mathsf{E}(\mathrm{ty}), \mathsf{E}(\mathbf{x}^\top)\boldsymbol{\Sigma}^{-1}\mathsf{E}(\mathbf{x}), \mathsf{E}(\mathrm{t}\mathbf{x}^\top)\boldsymbol{\Sigma}^{-1}\mathsf{E}(\mathbf{x}),$$
$$\mathsf{E}(\mathrm{t}\mathbf{x}^\top)\boldsymbol{\Sigma}^{-1}\mathsf{E}(\mathbf{x}\mathrm{t}), \mathsf{E}(\mathbf{x}^\top)\boldsymbol{\Sigma}^{-1}\mathsf{E}(\mathbf{x}\mathrm{ty}), \mathsf{E}(\mathrm{t}\mathbf{x}^\top)\boldsymbol{\Sigma}^{-1}\mathsf{E}(\mathbf{x}\mathrm{ty})$$

A peek at some numerical results: Estimating $\mathsf{E}[\mathrm{y}]$ under MAR

Compared with Celentano & Wainwright, 23: based on debiased Lasso + AMP theory under Gaussian design

Our approach: a system of moment equations can be used to identify $\psi = \mathsf{E}[\mathrm{y}] = \boldsymbol{\beta}^\top \boldsymbol{\mu}$ in the following model:

$$\mathrm{y} = \boldsymbol{\beta}^\top \mathbf{x} + \varepsilon, \mathrm{t}|\mathbf{x} \sim \mathrm{Bern}(\phi(\boldsymbol{\alpha}^\top \mathbf{x}))$$

based on estimating the following moments

$$\mathsf{E}(\mathrm{t}), \mathsf{E}(\mathrm{ty}), \mathsf{E}(\mathbf{x}^\top)\boldsymbol{\Sigma}^{-1}\mathsf{E}(\mathbf{x}), \mathsf{E}(\mathrm{t}\mathbf{x}^\top)\boldsymbol{\Sigma}^{-1}\mathsf{E}(\mathbf{x}),$$
$$\mathsf{E}(\mathrm{t}\mathbf{x}^\top)\boldsymbol{\Sigma}^{-1}\mathsf{E}(\mathbf{x}\mathrm{t}), \mathsf{E}(\mathbf{x}^\top)\boldsymbol{\Sigma}^{-1}\mathsf{E}(\mathbf{x}\mathrm{ty}), \mathsf{E}(\mathrm{t}\mathbf{x}^\top)\boldsymbol{\Sigma}^{-1}\mathsf{E}(\mathbf{x}\mathrm{ty})$$

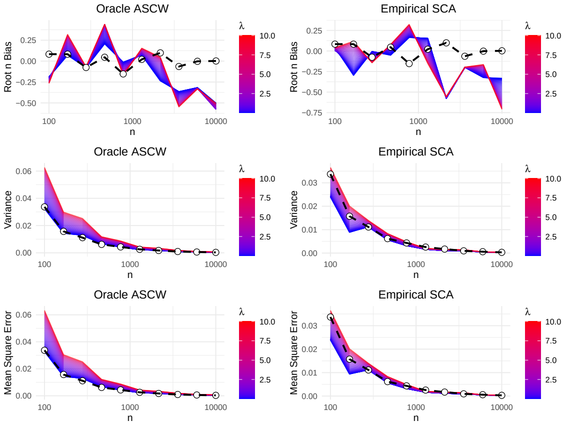

Figure: α_j's in logistic regression: known $\boldsymbol{\Sigma}$, Gaussian design and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_p) \overset{\text{i.i.d.}}{\sim} \mathrm{Uniform}([-\sqrt{3/p}, \sqrt{3/p}])$. Here $\phi(\mathbf{x}^\top \boldsymbol{\alpha}) = 0.1 + 0.9 \cdot \mathrm{expit}(\mathbf{x}^\top \boldsymbol{\alpha})$, , $p = 1.25n$, $n = 5000$.

Outline

Motivations: Estimating Functionals of High-Dimensional GLMs

Our proposal – Moments-based estimators

Numerical experiments

Conclusion and open ends

Conclusion and open ends

- We propose a new estimator based on **method of moments**:

  No sparsity;
  No difficult AMP to learn;
  No tuning parameters
  Consistent estimator of the variance;
  Classical, easy-to-understand and easy-to-extend framework.

- Open ends:

  - More general designs – semi-random, right orthogonally invariant, etc.
  - Better numerical algorithms for inverting the nonlinear maps?
  - Model misspecification
  - $\cdots$

Thank you!

Xingyu's Homepage: https://cxy0714.github.io/
arXiv Paper: https://arxiv.org/abs/2408.06103
GitHub Repo: https://github.com/cxy0714/Method-of-Moments-Inference-for-GLMs