Achievable distributional robustness

when the robust risk is only partially identified

Julia Kostin, Nicola Gnecco, Fanny Yang

ETH Zurich, University College London

Team

Julia Kostin

ETH zürich

Nicola Gnecco

UCL

Fanny Yang

ETH zürich

Out-of-domain generalization

Training distribution $\mathbb{P}_{\text{train}}$

Test distribution $\mathbb{P}_{\text{test}}$

Out-of-domain generalization


$$\beta_{\text{train}} = \arg\min_{\beta} \mathscr{R}(\beta; \mathbb{P}_{\text{train}})$$

Low training risk

$$\mathscr{R}(\beta_{\text{train}}; \mathbb{P}_{\text{test}}) \gg \min_{\beta} \mathscr{R}(\beta; \mathbb{P}_{\text{test}})$$

High test risk


Training distribution $\mathbb{P}_{\text{train}}$

Test distribution $\mathbb{P}_{\text{test}}$

Distributional robustness

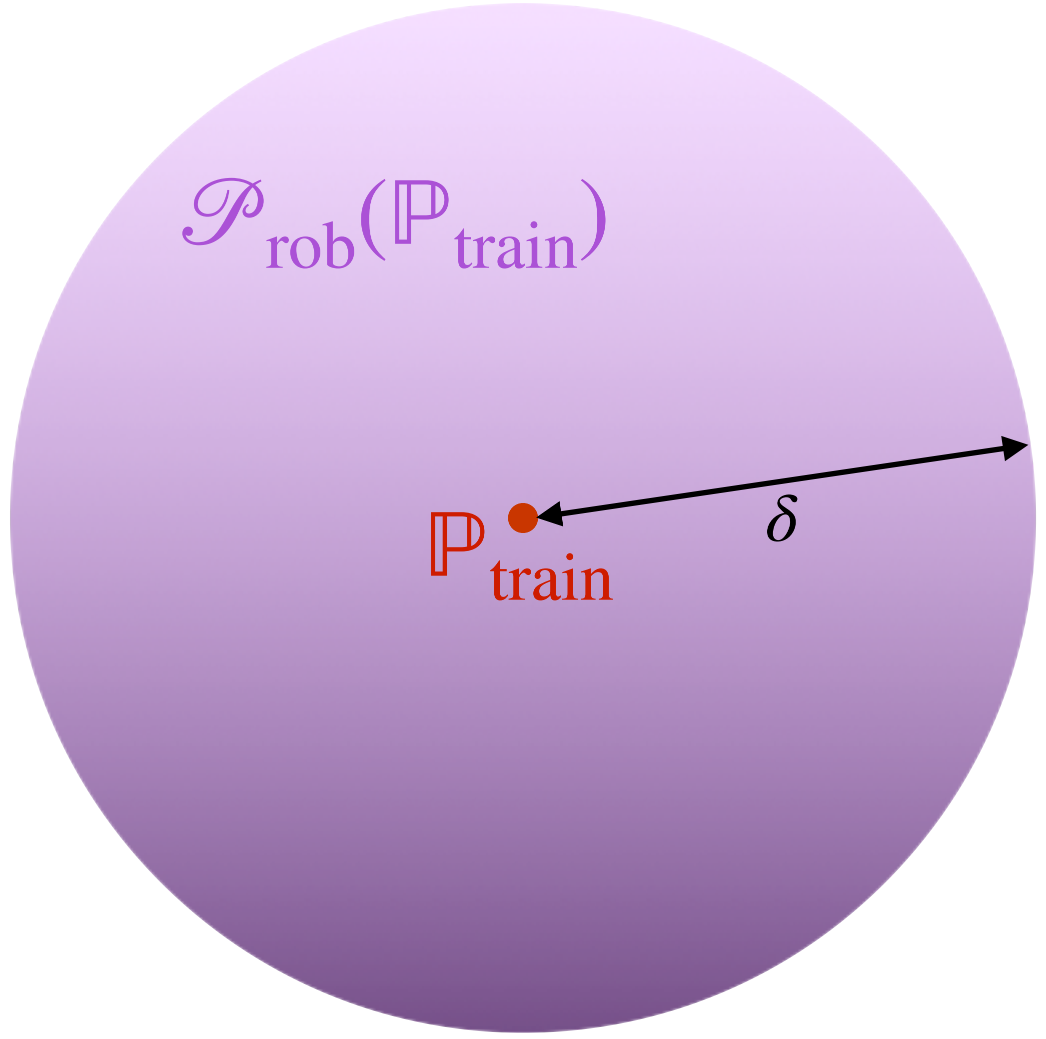
Goal: given training data, generalize to a set of feasible test distributions, called robustness set, by computing a minimiser of the robust risk

$$\beta_{\mathrm{rob}} = \arg\min_{\beta} \left[\mathscr{R}_{\mathrm{rob}}(\beta; \mathscr{P}_{\mathrm{rob}}(\theta_\star)) := \sup_{\mathbb{P} \in \mathscr{P}_{\mathrm{rob}}(\theta_\star)} \mathscr{R}(\beta; \mathbb{P}) \right]$$

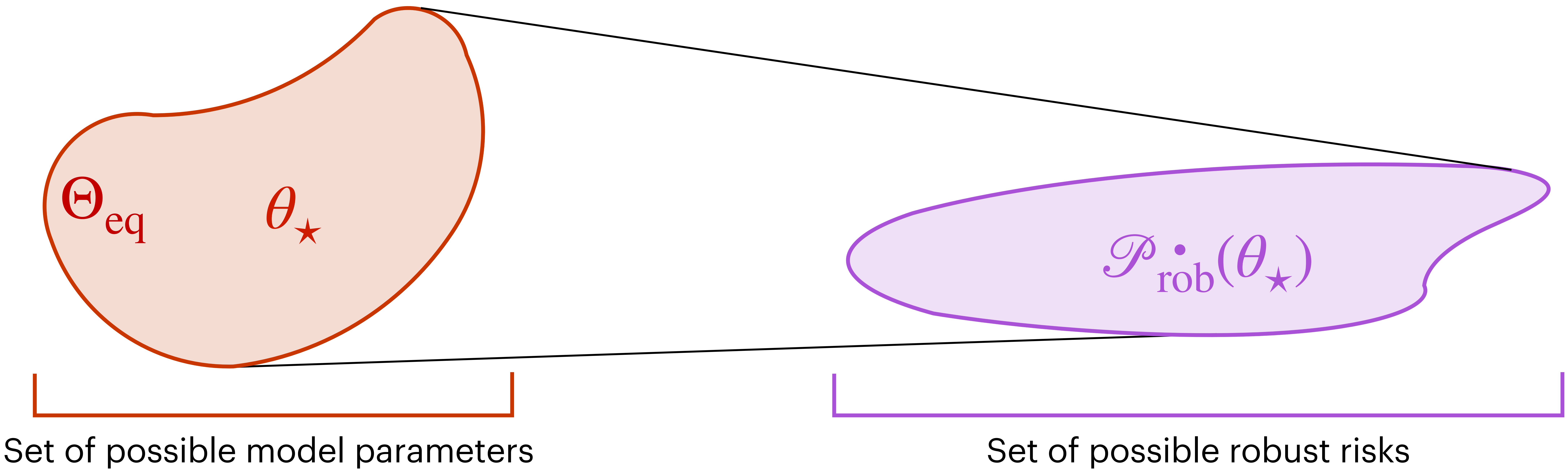
In previously considered robustness scenarios, the parameters $\theta_\star$ and/or the robustness set $\mathscr{P}_{\text{rob}}(\theta_\star)$ are considered to be **known**:

Distributionally robust optimization

$$\theta_\star = \mathbb{P}_{\text{train}};$$
$$\mathscr{P}_{\text{rob}}(\theta_\star) = \{\mathbb{P} : \; D(\mathbb{P}, \mathbb{P}_{\text{train}}) \leq \delta\}$$



Often, $\theta_\star$ and/or $\mathscr{P}_{\mathrm{rob}}(\theta_\star)$ are neither known nor computable from training data



Instead, they can be merely set identified.

We propose to minimise a new objective called the identifiable robust risk:

$$\mathscr{R}_{\mathrm{rob,ID}}(\beta; \Theta_{\mathrm{eq}}) := \sup_{\theta \in \Theta_{\mathrm{eq}}} \sup_{\mathbb{P} \in \mathscr{P}_{\mathrm{rob}}(\theta)} \mathscr{R}(\beta, \mathbb{P})$$

Best achievable distributional robustness:

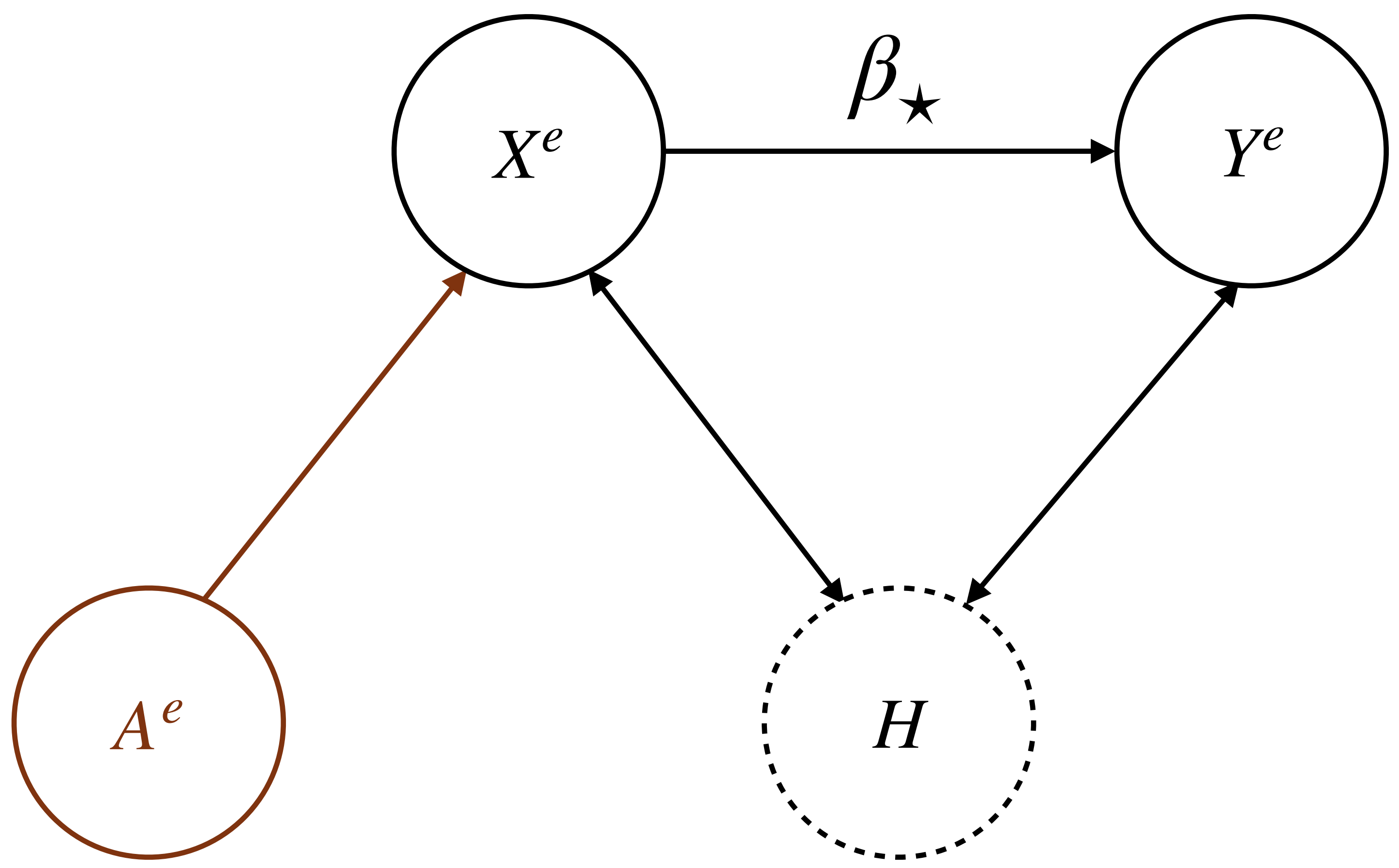
$$\mathfrak{M}(\Theta_{eq}) = \inf_{\beta \in \mathbb{R}^d} \mathscr{R}_{rob,ID}(\beta; \Theta_{eq})$$

Setting of structural causal models

Data model: linear structural causal model (SCM) with unobserved confounding, environments differ via **additive shifts** A^e:

$$X^e = A^e + \eta;$$
$$Y^e = \beta_\star^\top X^e + \xi,$$

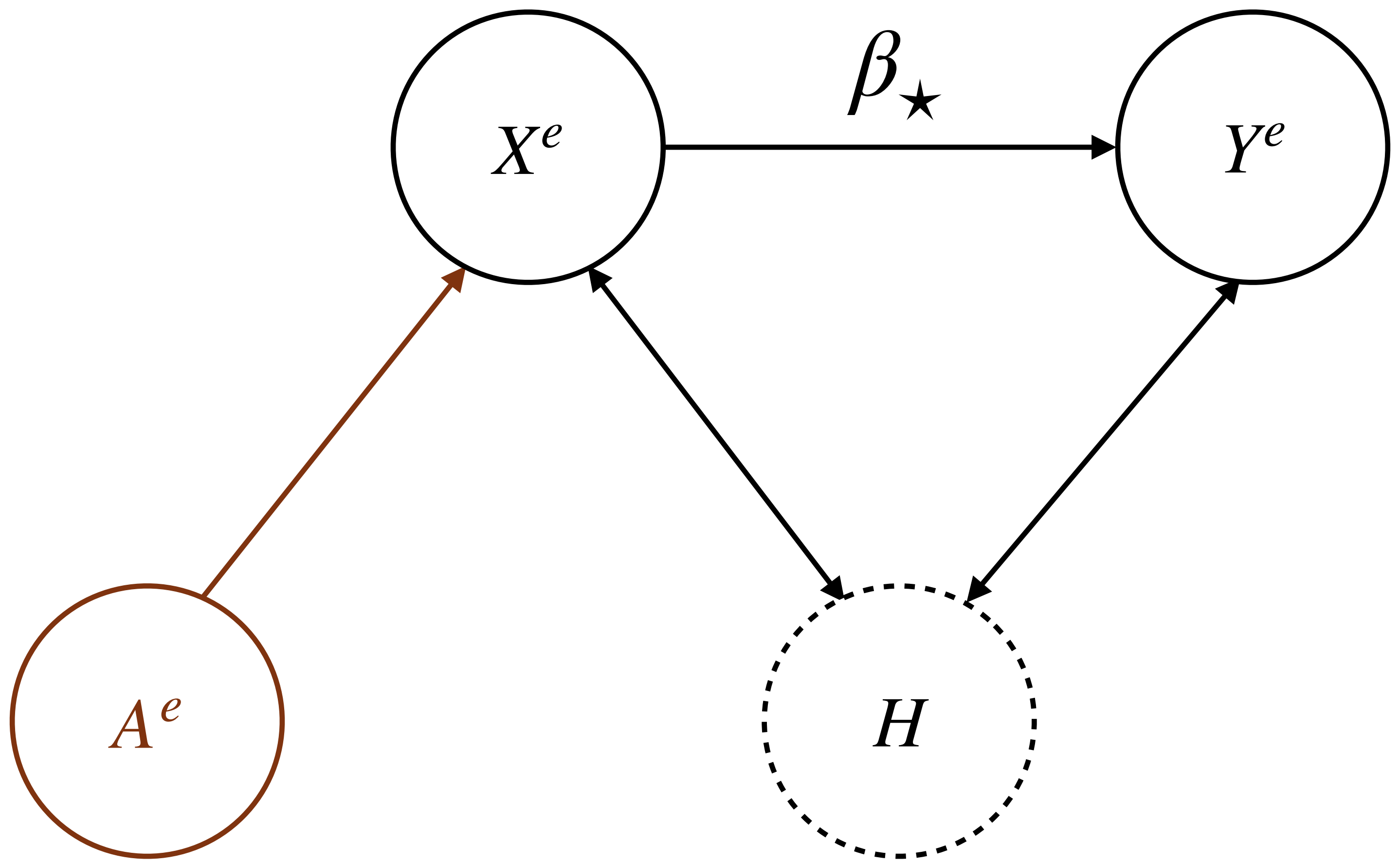
where $(\eta, \xi) \sim \mathcal{N}(0, \Sigma_\star)$ and $\theta_\star = (\Sigma_\star, \beta_\star)$ are the model parameters.



Setting of structural causal models

Some structural knowledge about the strength and direction of the test shift:

$$\mathbb{E}[A^{\text{test}} A^{\text{test}\top}] \preceq M_{\text{test}} = \gamma \Pi_{\mathcal{M}}.$$



- Infinite **robustness** to arbitrary shifts **only possible if $\beta_\star$ known** (requires $\mathcal{O}(d)$ env's)

- However, $\beta_\star$ only identified on

$$\mathcal{S} = \text{range}\left(\sum_{e \in \mathcal{E}_{\text{train}}} \mathbb{E}[A^e A^{e\top}] \right)$$

Identifiable robustness for the SCM setting

We **compute** the identifiable robust risk explicitly:

$$\mathcal{R}_{rob,ID}(\beta;\Theta_{eq},\gamma\Pi_{\mathcal{M}}) = \underbrace{\mathcal{R}(\beta;\theta_\star)}_{\text{Reference risk}} + \underbrace{\gamma\|S^\top(\beta^{\mathcal{S}}-\beta)\|_2^2}_{\text{Invariance term}} + \underbrace{\gamma(C_{ker}+\|R^\top\beta\|_2)^2}_{\text{Non-identifiability term}},$$

where:

- S: test shift directions along which the causal model can be identified
- R: test shift directions along which the model is non-identifiable
- C_{ker}: max. norm of the model along non-identified directions

Identifiable robustness for the SCM setting

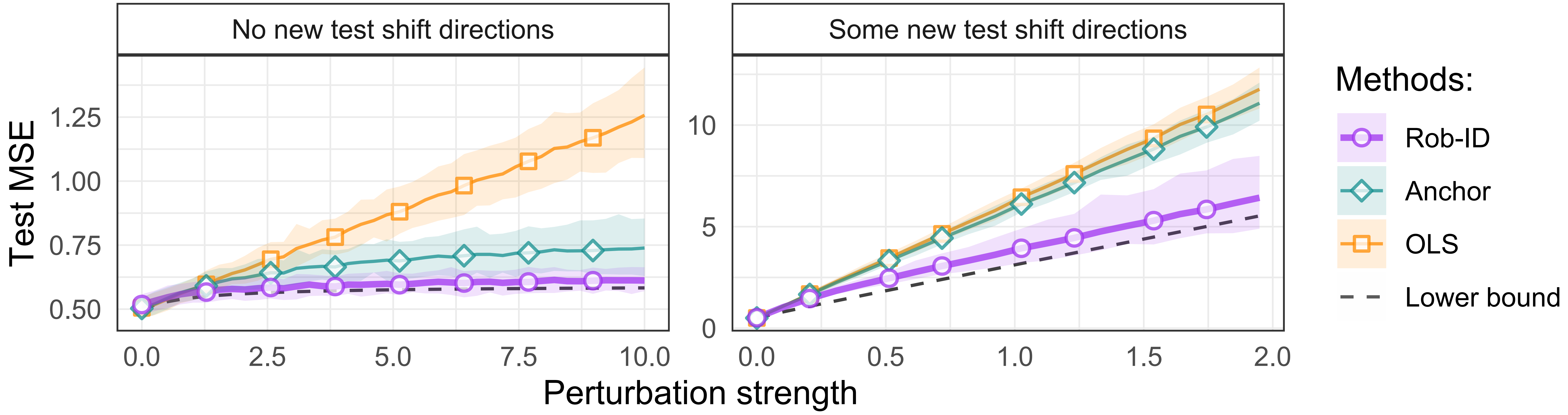
We **compute** the identifiable robust risk explicitly:

$$\mathcal{R}_{rob,ID}(\beta; \Theta_{eq}, \gamma\Pi_{\mathcal{M}}) = \mathcal{R}(\beta; \theta_\star) + \gamma\|S^\top(\beta^{\mathcal{S}} - \beta)\|_2^2 + \gamma(C_{ker} + \|R^\top\beta\|_2)^2,$$

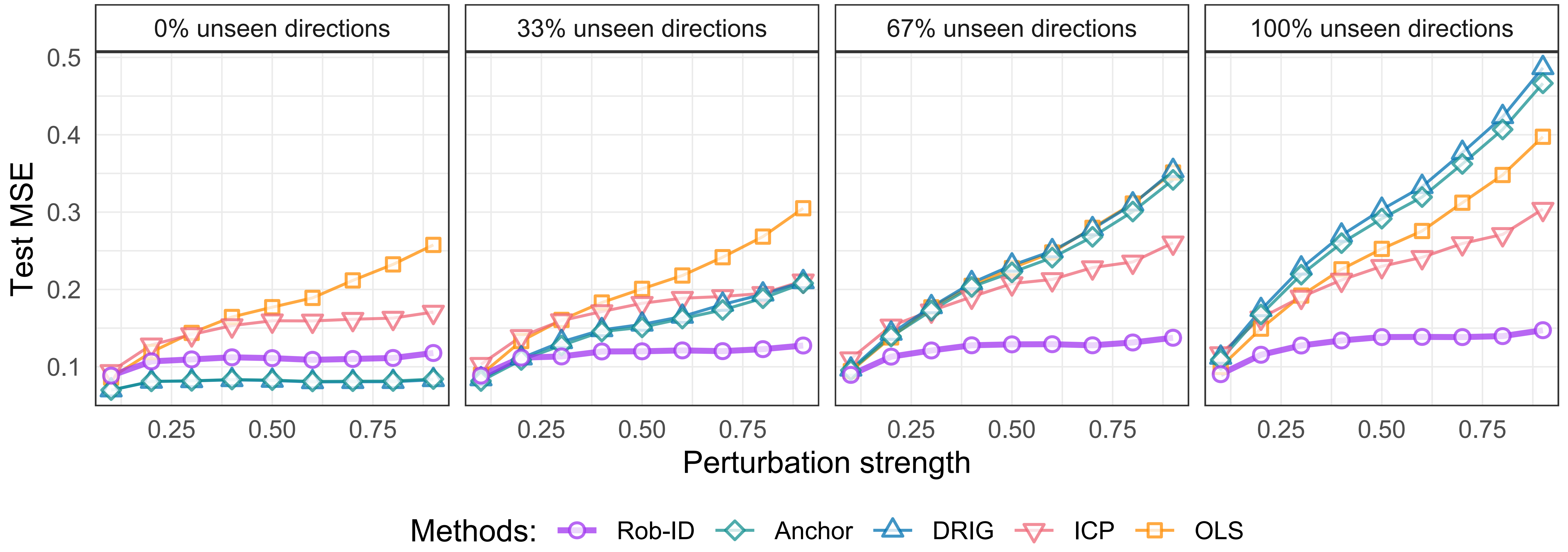
Reference risk Invariance term Non-identifiability term

- We prove a **lower bound** for the id. robust risk which is tight for large γ;
- For large γ, we prove **suboptimality of existing robustness methods** such as **anchor regression** [Rothenhäusler et al. 2021] and **DRIG** [Shen et al. 2023].

Simulations on **Gaussian SCM data**:



Experiments on **real-world gene expression dataset** [Replogle et al. 2022]:



Outlook

- Extension to classification
- Nonlinear models
- Use for active intervention selection
- Partially identifiable framework beyond causality

Thank you!