

# Penalised Regression

NGSchool 2022

Nuno Sepúlveda, 16.09.2022  
[nuno.Sepulveda@mini.pw.edu.pl](mailto:nuno.Sepulveda@mini.pw.edu.pl)  
<http://www.immune-stats.net>

# Objectives

**Basics on penalised regression**

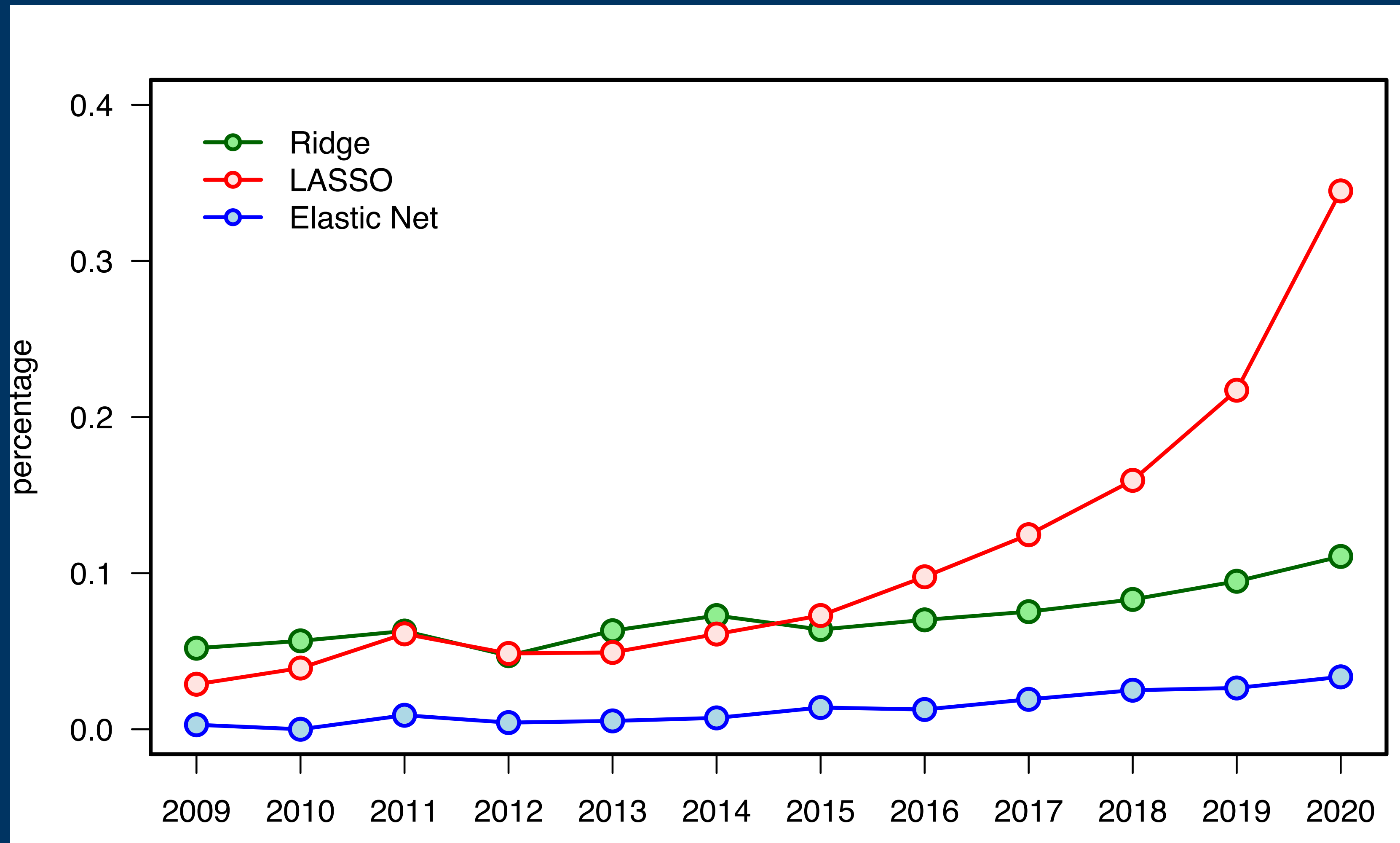
Ridge Regression

LASSO Regression

Elastic Net Regression

**Use R to conduct data analysis**

# Why learning penalised regression?



**Estimation**



**Model selection**

**Accuracy**



**Bias**

# Penalised regression

$$\hat{\mathbf{b}} = \underset{\mathbf{b}}{\operatorname{argmin}} \left\{ \sum_{i=1}^n \left( y_i - b_0 - \sum_{j=1}^p b_j x_i \right)^2 \right\} .$$

subject to a constraint

$$pen \leq \lambda$$

*pen* = penalty function

$\lambda$  = tuning parameter

# Ridge Regression

$$\hat{\mathbf{b}} = \operatorname{argmin}_{\mathbf{b}} \left\{ \sum_{i=1}^n \left( y_i - b_0 - \sum_{j=1}^p b_j x_i \right)^2 \right\},$$

subject to  $\sum_{j=1}^p b_j^2 \leq \lambda_2$

$$\lambda_2 \in \left[ 0, \sum_{j=1}^p (\hat{b}_j^*)^2 \right]$$

↑  
OLS estimates

# Geometrical interpretation (2D)

$$\sum_{j=1}^2 b_j^2 \leq \lambda_2$$

$$b_1 = r \cos \theta$$

$$b_2 = r \sin \theta$$

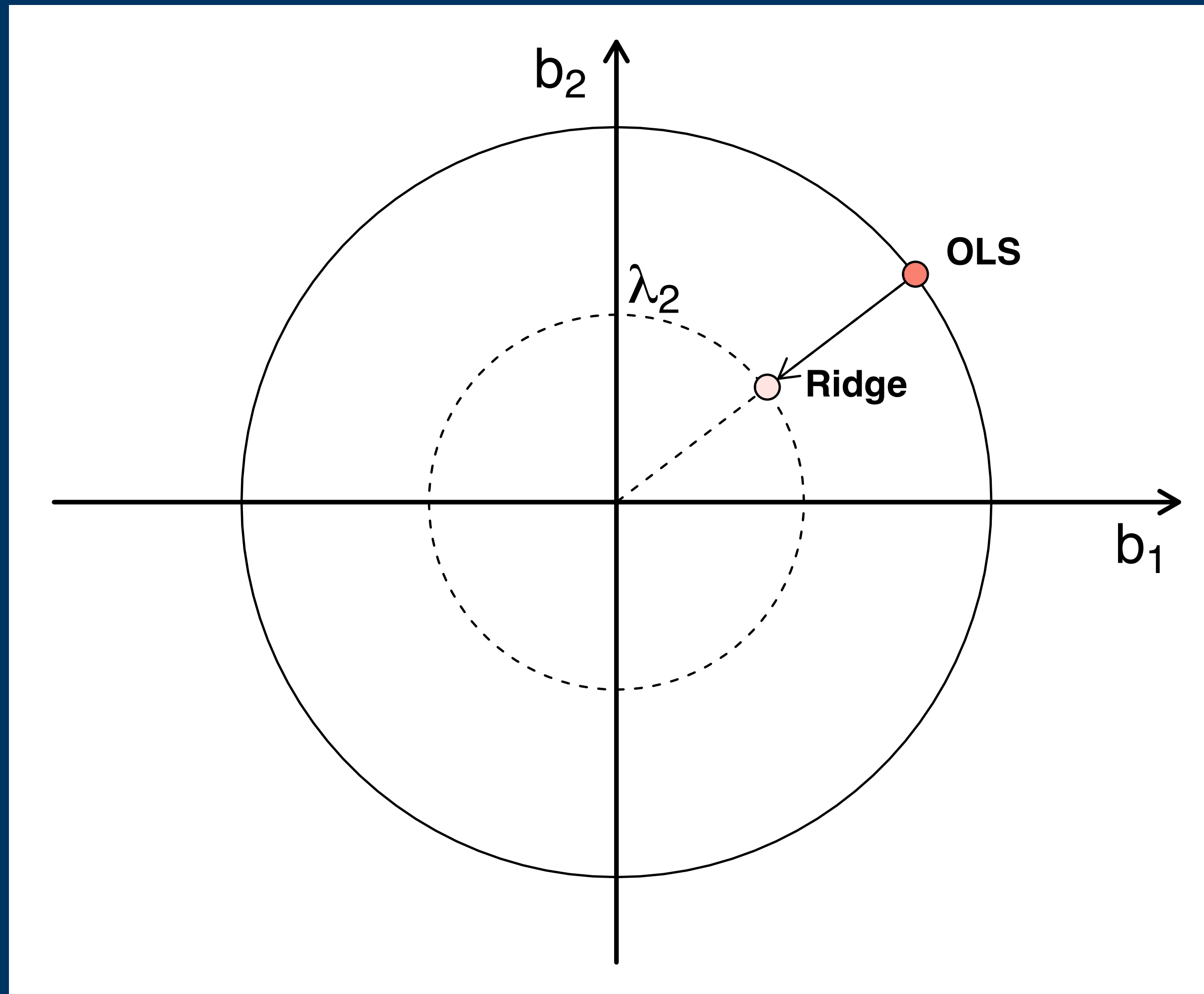
$$r^2(\cos^2 \theta + \sin^2 \theta) \leq \lambda_2$$

$$r^2 \leq \lambda_2$$

Ridge estimator is only dependent on the radius and not on the angle



# Geometrical interpretation (2D)



# Ordinary least squares estimator

$$\hat{\mathbf{b}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

## Ridge estimator

$$\hat{\mathbf{b}} = (\mathbf{X}^T \mathbf{X} + \lambda_2 \mathbf{I})^{-1} \mathbf{X}^T \mathbf{Y}$$

# Ridge Regression

$$\hat{\mathbf{b}} = \operatorname{argmin}_{\mathbf{b}} \left\{ \sum_{i=1}^n \left( y_i - b_0 - \sum_{j=1}^p b_j x_i \right)^2 \right\},$$

0% shrinkage

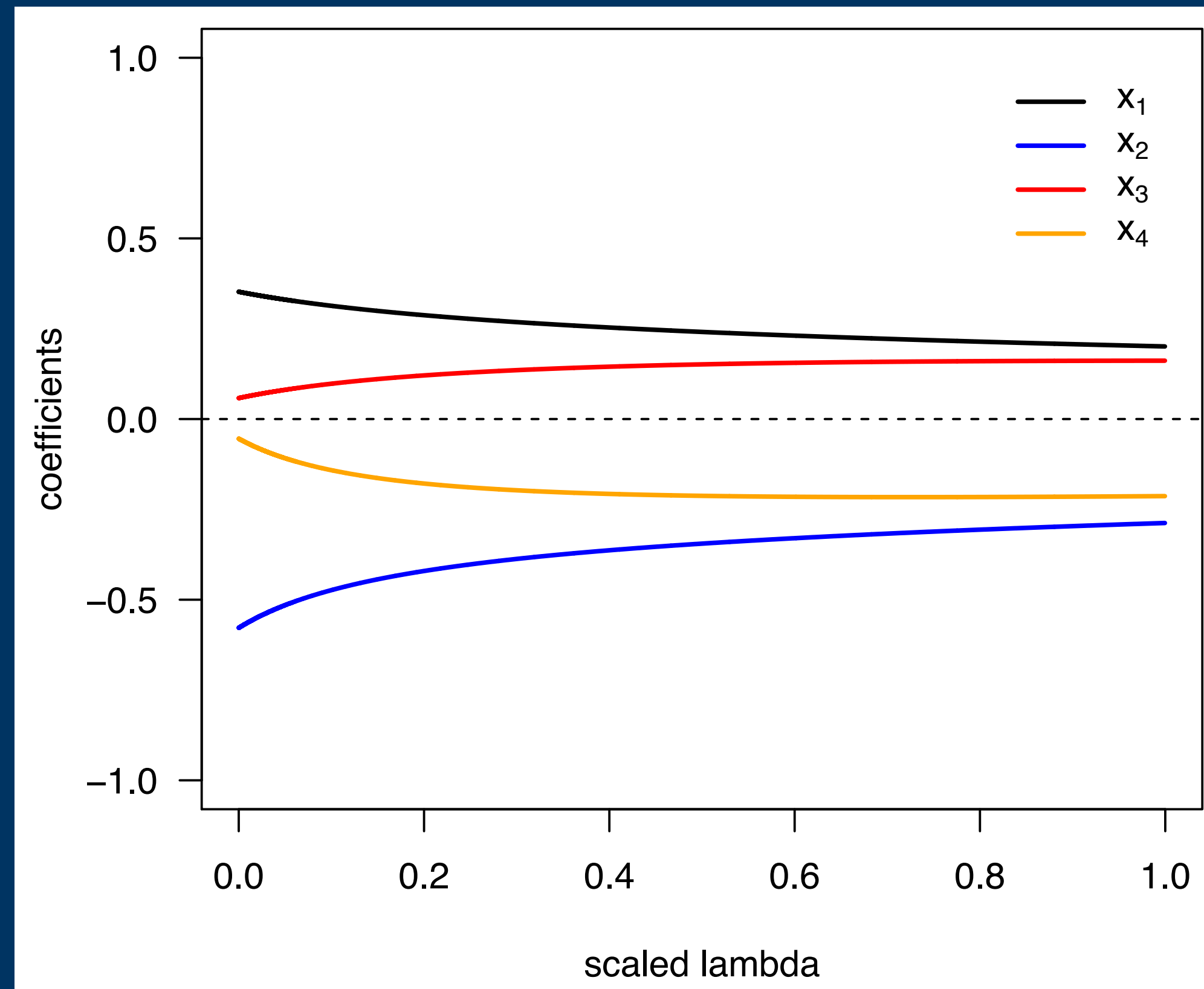
subject to

$$\frac{\sum_{j=1}^p b_j^2}{\sum_{j=1}^p (\hat{b}_j^*)^2} \leq 1 - \lambda^*$$

$$\lambda^* \in [0, 1]$$

“100%” shrinkage

# Ridge trace plot



# Ridge regression

## Advantages

Remove multicollinearity

Estimator with a closed form

Shrinkage

## Disadvantages

Biased estimators

No shrinkage to zero

(No model selection)

# LASSO Regression

$$\hat{\mathbf{b}} = \underset{\mathbf{b}}{\operatorname{argmin}} \left\{ \sum_{i=1}^n \left( y_i - b_0 - \sum_{j=1}^p b_j x_i \right)^2 \right\},$$

subject to  $\sum_{j=1}^p |b_j| \leq \lambda_1$

$$\lambda_1 \in \left[ 0, \sum_{j=1}^p |\hat{b}_j^*| \right]$$

OLS estimates

# Geometrical interpretation (2D)

$$\sum_{j=1}^2 |b_j| \leq \lambda_1$$

$$b_1 = r \cos \theta$$

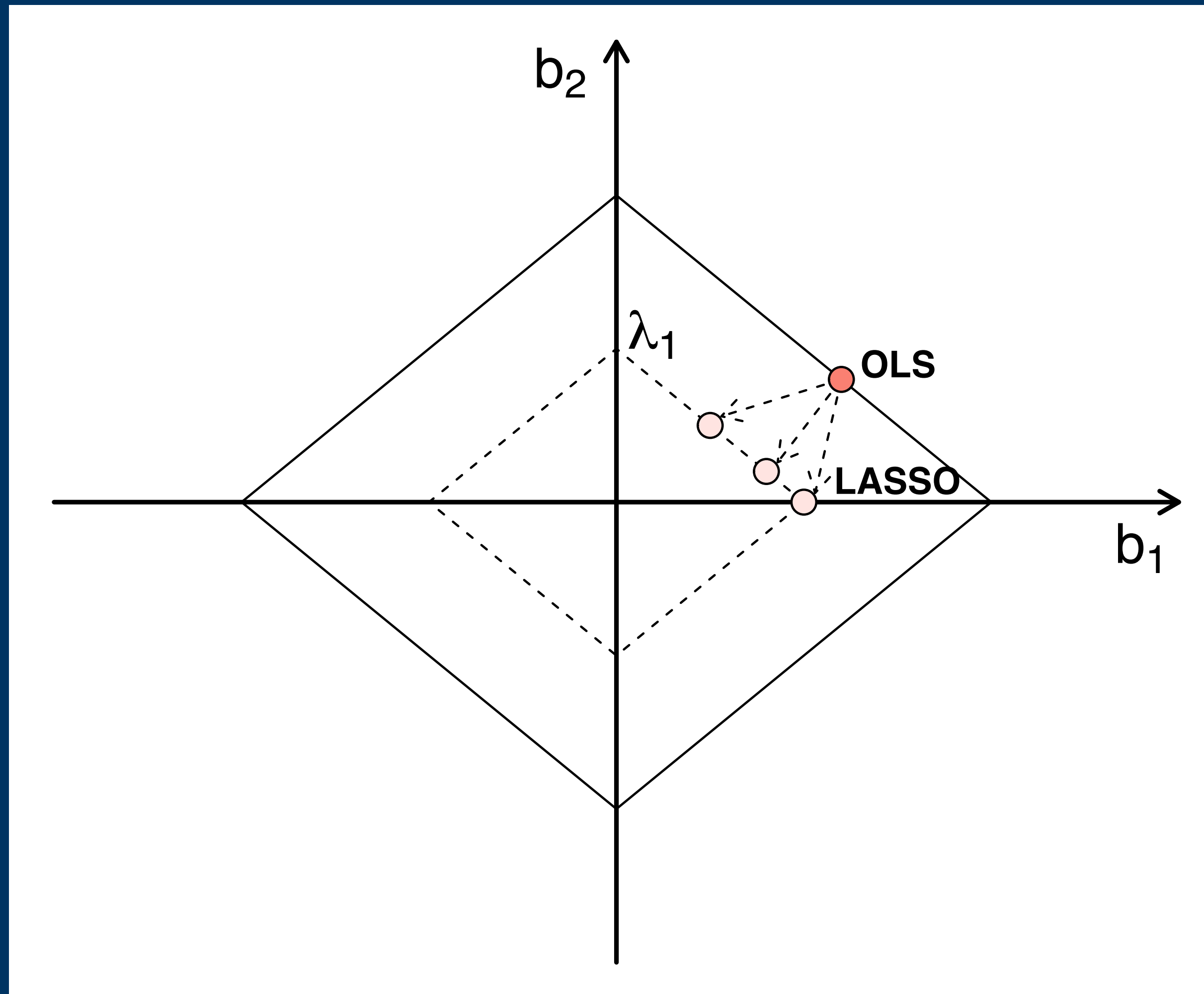
$$b_2 = r \sin \theta$$

$$r(\cos \theta + \sin \theta) \leq \lambda_2$$

$$r^2 \leq \lambda_2$$

LASSO estimator is dependent on  
both radius and angle

# Geometrical interpretation (2D)





# LASSO Regression

$$\hat{\mathbf{b}} = \underset{\mathbf{b}}{\operatorname{argmin}} \left\{ \sum_{i=1}^n \left( y_i - b_0 - \sum_{j=1}^p b_j x_i \right)^2 \right\},$$

subject to

$$\frac{\sum_{j=1}^p |b_j|}{\sum_{j=1}^p |b_j^*|} \leq 1 - \lambda^*$$

0% shrinkage (OLS)

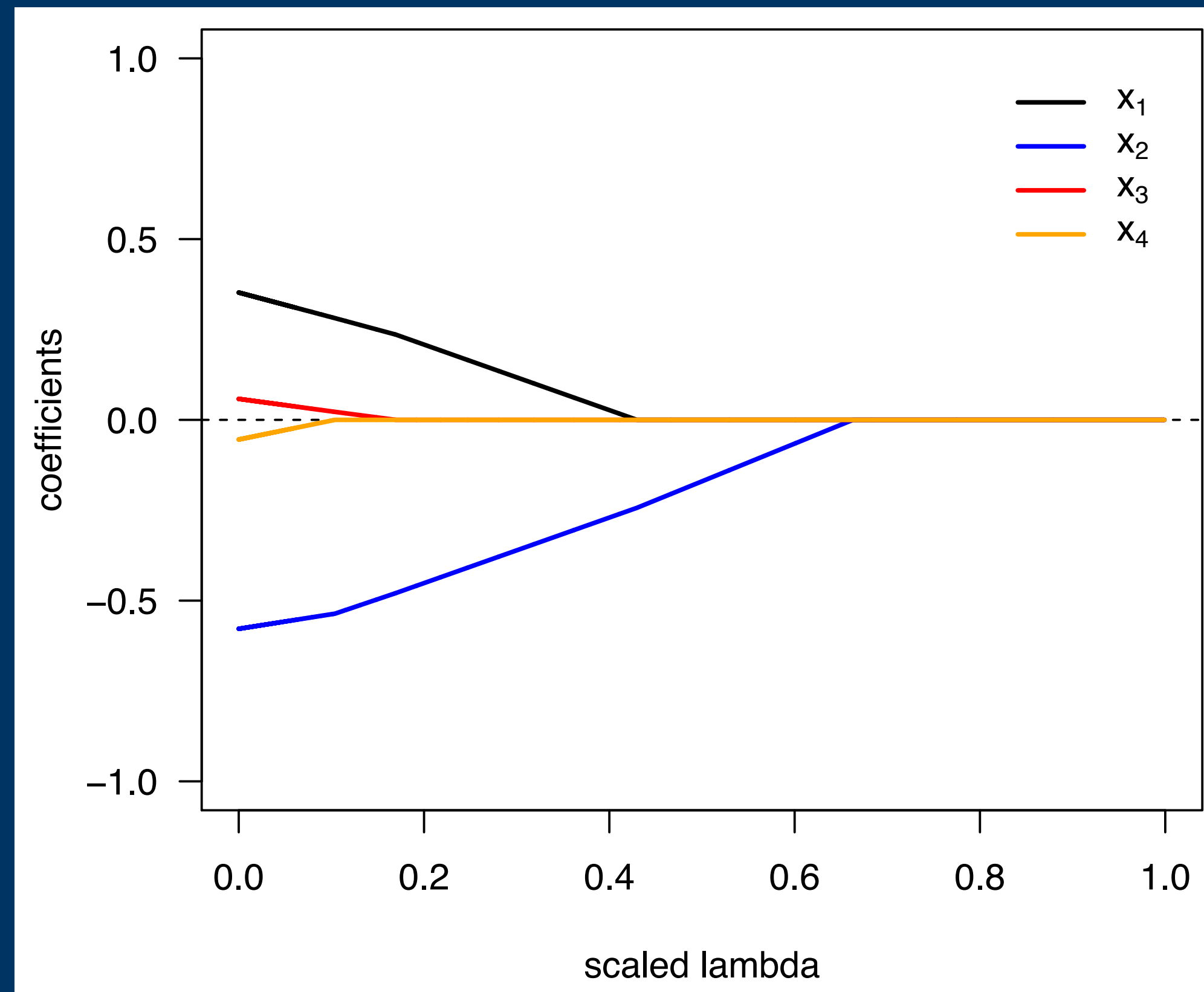


$$\lambda^* \in [0, 1]$$



100% shrinkage

# LASSO trace plot



# LASSO regression

## Advantages

Remove multicollinearity

Shrinkage to zero

(Model selection)

## Disadvantages

Random choice of highly correlated covariates

No closed-form expression

Problems with standard errors

# Elastic Net Regression

$$\hat{\mathbf{b}} = \operatorname{argmin}_{\mathbf{b}} \left\{ \sum_{i=1}^n \left( y_i - b_0 - \sum_{j=1}^p b_j x_i \right)^2 \right\},$$

subject to  $\alpha \|\mathbf{b}\|_1 + (1 - \alpha) \|\mathbf{b}\|^2 \leq \lambda$  for some  $\lambda$  and  $\alpha \in [0,1]$ .

$\alpha = 0 \Rightarrow$  Ridge regression

$\alpha = 1 \Rightarrow$  LASSO regression

# Estimation of the tuning parameter(s)

Evaluate a grid of  
possible values



Highest  
accuracy

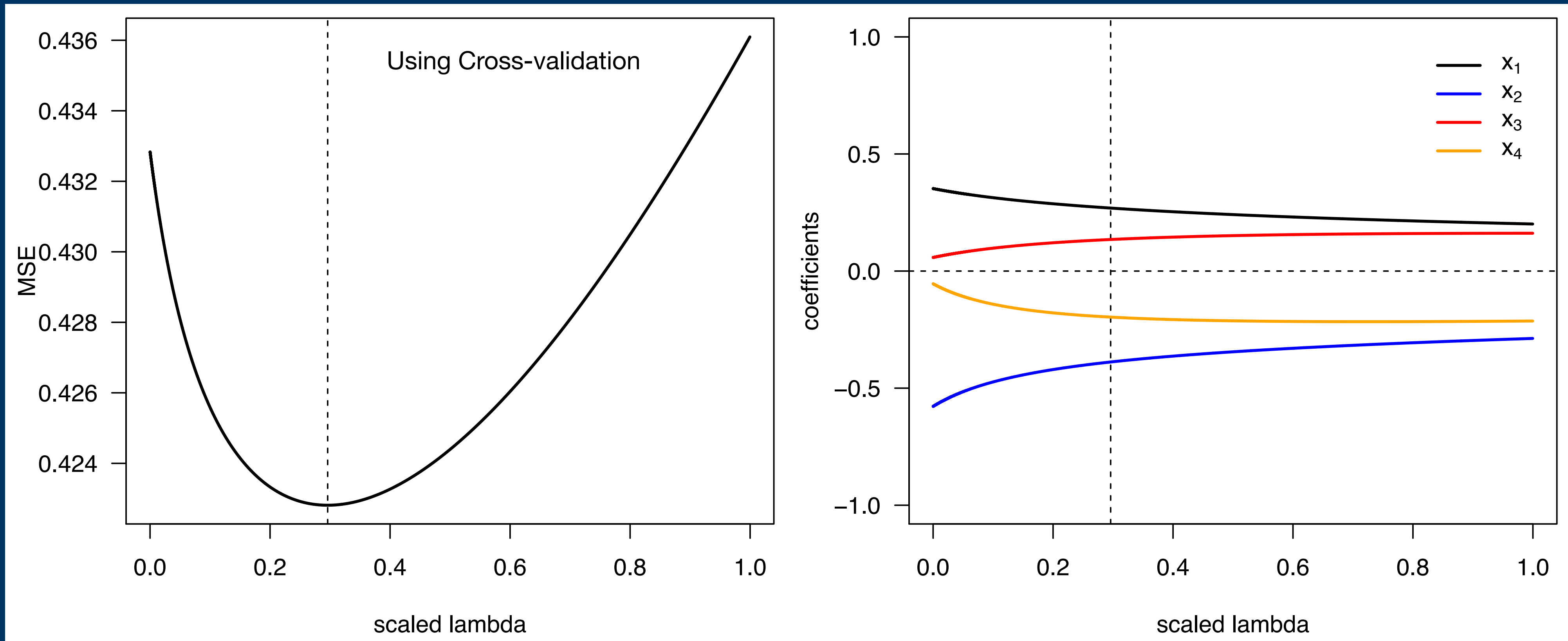


Cross-  
validation

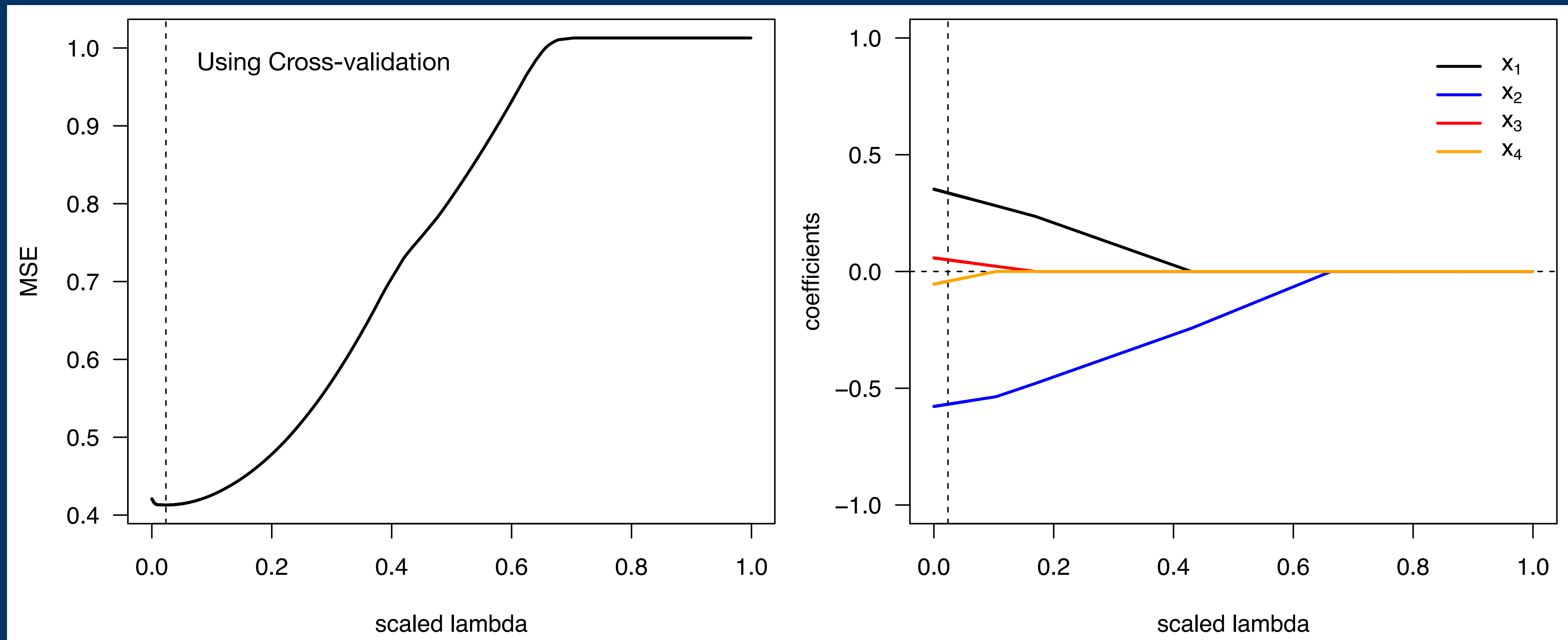


Lowest mean  
squared error

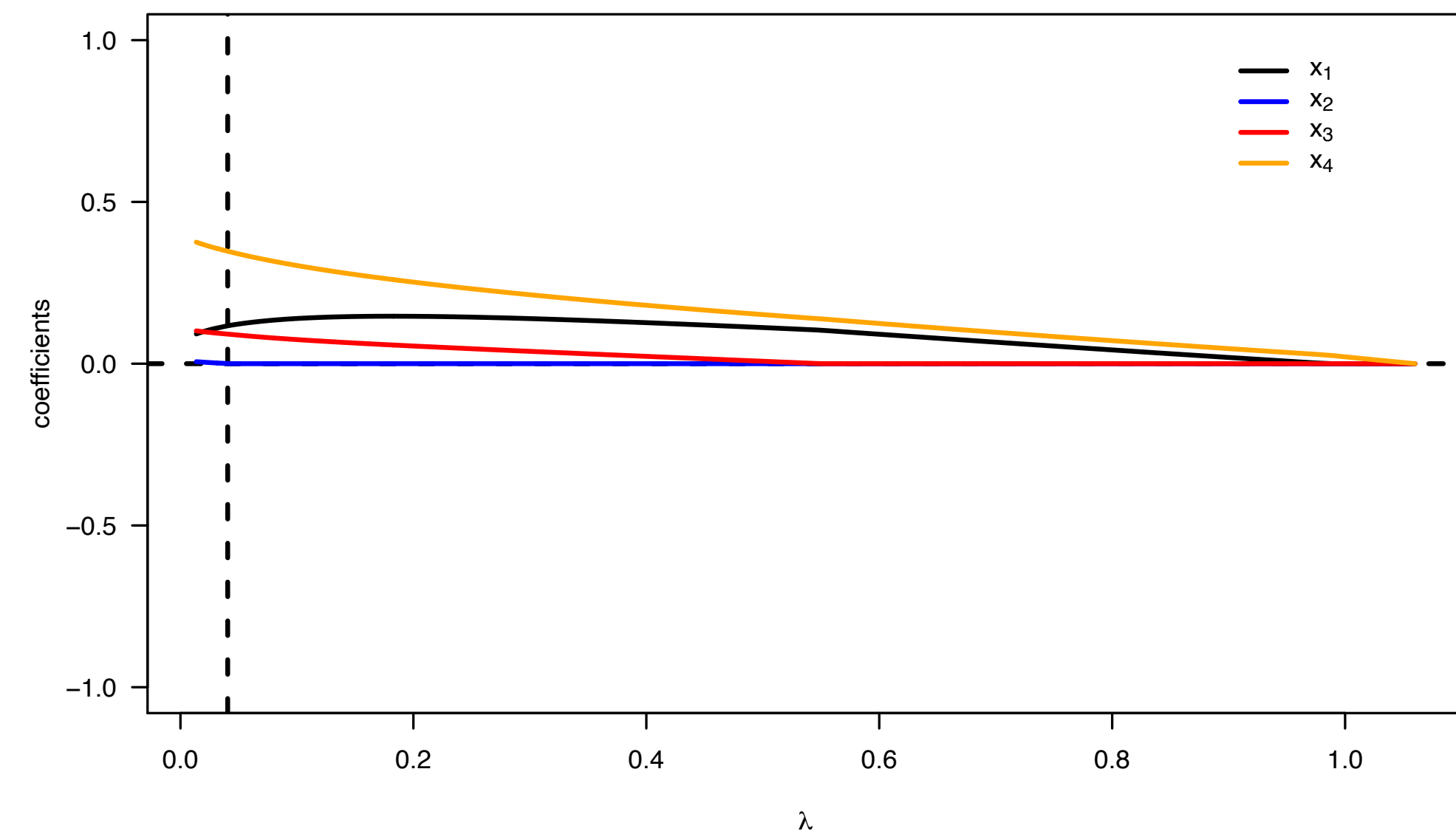
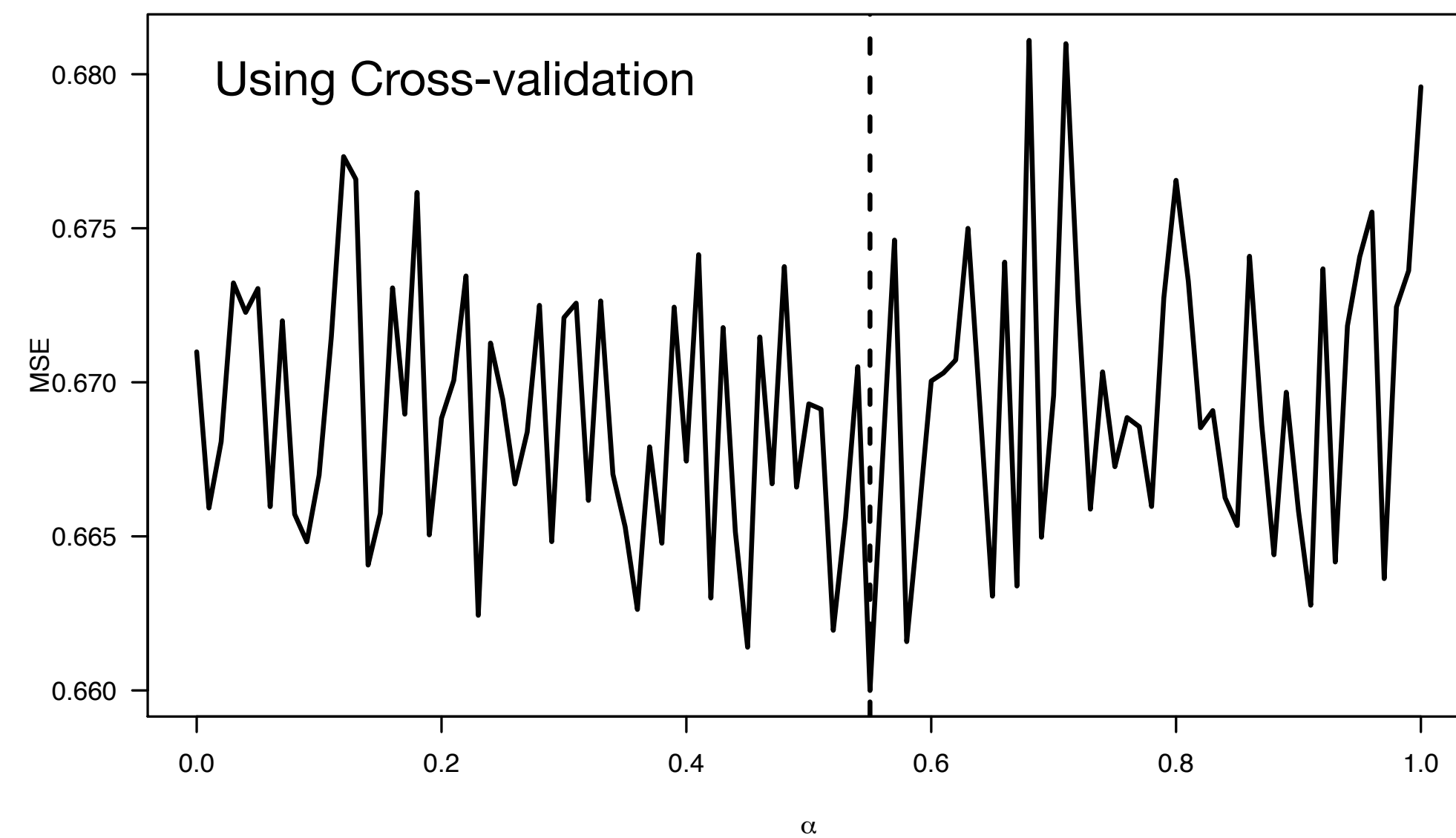
# Example: Ridge Regression



# Example: LASSO Regression



# Example: Elastic Net Regression





## Other methods

- Adaptive LASSO
- Smoothly clipped Absolute Deviations (SCAD)
- Minimax Concave Penalty (MCP)
- Bayesian LASSO

# Acknowledgements

## Immune-Stats Group

André Fonseca

## MiNI@WUT

Przemysław Biecek

## NGSchool 2022

Katarzyna Kedzierska

Urszula Smyczynska

Gladys Wojciechowska



ref. UIDB/00006/2020



ref. PPN/ULM/2020/1/00069/U/00001