

Regularization / Penalization : reduce variance at the cost of introduce bias

## Ridge Regression / $L_2$ -Shrinkage

Idea: Restrict  $\beta$  to be in  $L_2$ -ball of radius  $c$  around zero.

$$\{b \in \mathbb{R}^k : \|b\|^2 \leq c^2\}$$

$$\rightarrow \min SSR_n(b) \text{ s.t. } \|b\|^2 \leq c^2$$

$$\text{Lagrangian } \mathcal{L} = (y - Xb)'(y - Xb) + \lambda(b'b - c^2).$$

$$\text{FOC}[\beta] \rightarrow 0 = -2X'(y - X\hat{\beta}_R) + 2\hat{\lambda}_R I_k \hat{\beta}_R$$

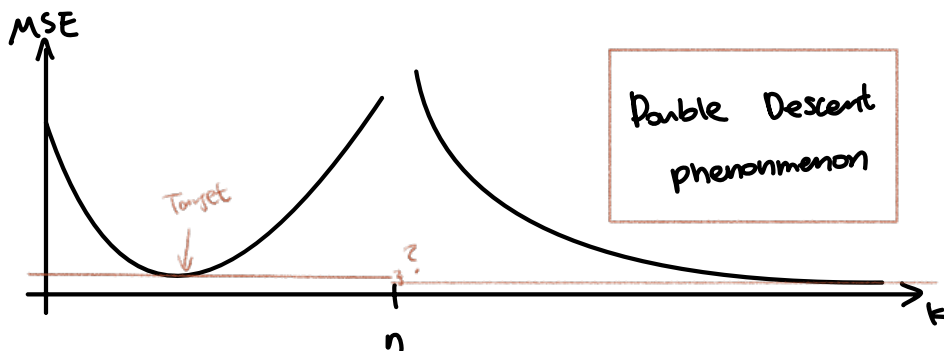
$$\rightarrow \hat{\beta}_R = (X'X + \hat{\lambda}_R I_k)^{-1} X'y$$

Bayesian interpretation:

1) Gaussian regression model :  $y_i = X_i'\beta + e_i$ ,  $e_i | X_i \sim N(0, \sigma^2)$ .

2) Gaussian prior for  $\beta$ :  $\beta \sim N(0, \frac{\sigma^2}{\lambda} I_k)$

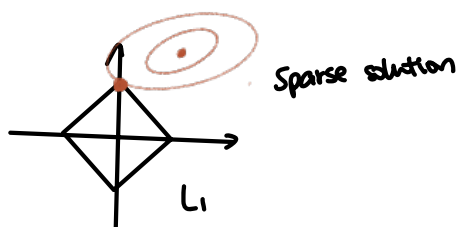
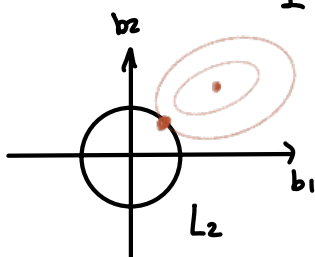
$\rightarrow$  can show : posterior mean for  $\beta$ ,  $E_{\pi}[\beta | (X_1, y_1), \dots, (X_n, y_n)] = (X'X + \lambda I_k)^{-1} X'y$ .



## LASSO / $L_1$ -Shrinkage

$$\hat{\beta}_{\text{lasso}} \text{ solves } \min_b SSR_n(b) \text{ s.t. } \|b\|_1 = \sum_{i=1}^n |b_i| \leq c$$

$$\mathcal{L} = (y - Xb)'(y - Xb) + \lambda(\|b\|_1 - c).$$



Oracle Property

## Restricted Estimation

LPM  $y_i = x_i' \beta + e_i, \quad E[x_i e_i] = 0$

Restriction  $R\beta = r$   
 $(q \times k) \quad (q \times 1)$

Ex. Exclusion restriction

$$\beta = \begin{bmatrix} \beta_1' & \beta_2' \end{bmatrix}' \rightarrow \beta_2 = 0$$

$k_1 \times 1 \quad k_2 \times 2$

$$\rightarrow R = \begin{bmatrix} 0 & I_{k_2} \end{bmatrix} \quad r = 0_{k_2 \times 1}$$

Ex. Cobb-Douglas prod. fn

$$y_i = A_i K_i^{\beta_1} L_i^{\beta_2}, \quad A_i \perp K_i, L_i$$

$$\rightarrow \log(y_i) = \beta_0 + \beta_1 \log K_i + \beta_2 \log L_i + \underbrace{(\log(A_i) - \beta_0)}_{e_i}$$

Constant return to scale,  $\beta_1 + \beta_2 = 1$ .

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} \quad R = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix} \quad r = 1.$$

## Constrained Least Squares:

$\hat{\beta}_{CLS}$  minimizes  $SSR_n(b)$  s.t.  $Rb = r$ .

$$\rightarrow \mathcal{L} = \frac{1}{2} SSR_n(b) + \underbrace{\lambda' (Rb - r)}_{\substack{\uparrow \\ q \quad q \times 1}}$$

F.O.C (1) w.r.t.  $\lambda$  :  $0 = \frac{\partial \mathcal{L}}{\partial \lambda} = R\hat{\beta}_{OLS} - r$

$$(2) \text{ w.r.t. } \beta : 0 = \frac{\partial \mathcal{L}}{\partial \beta} = \frac{1}{2} \frac{\partial}{\partial \beta} SSR_n(\beta) \Big|_{\beta = \hat{\beta}_{CLS}} + R'\lambda \quad (*)$$
$$= -X'(y - X\hat{\beta}_{CLS}) + R'\hat{\lambda}_{CLS}$$

$\rightarrow$  Solve for  $\hat{\beta}_{CLS}, \hat{\lambda}_{CLS}$ .

$$\rightarrow \hat{\beta}_{CLS} : 0 = -X'y + X'X\hat{\beta}_{CLS} + R'\hat{\lambda}_{CLS} \Rightarrow \hat{\beta}_{CLS} = \underbrace{(X'X)^{-1}X'y}_{\hat{\beta}_{OLS}} - (X'X)^{-1}R'\hat{\lambda}_{OLS}$$

$\rightarrow$  Solve for  $\hat{\lambda}_{OLS}$  :

Pre-multiply (\*) with  $R(X'X)^{-1}$  :  $0 = - \underbrace{R(X'X)^{-1}X'Y}_{\hat{\beta}_{LS}} + \underbrace{R(X'X)^{-1}X'X}_{=I} \hat{\beta}_{CLS} + R(X'X)^{-1}R'\hat{\gamma}_{CLS}$

$$\Rightarrow R\hat{\beta}_{LS} - r = R(X'X)^{-1}R'\hat{\gamma}_{CLS}$$

$$\Rightarrow \hat{\gamma}_{CLS} = [R(X'X)^{-1}R']^{-1} [R\hat{\beta}_{LS} - r]$$

$$\hookrightarrow \hat{\beta}_{CLS} = \hat{\beta}_{LS} - \underbrace{(X'X)^{-1}R'}_{C'C} [R(X'X)^{-1}R']^{-1} [R\hat{\beta}_{LS} - r].$$

$$= C'CX'Y - C' \underbrace{[CR'[RC'C R']^{-1}RC']}_{P_{CR'}} [CX'Y - r].$$

$$= C'(I - P_{CR'})CX'Y$$

Homoskedastic LRM  $\rightarrow \text{var}(X'Y) = \sigma^2(X'X).$

$$\Rightarrow \text{var}(\hat{\beta}_{CLS}) = \sigma^2 C'(I - P_{CR'}) \underbrace{CX'X}_{I} C'(I - P_{CR'})C$$

$$= \sigma^2 C'(I - P_{CR'})C$$

$$= \underbrace{\sigma^2 C'C}_{\substack{(X'X)^{-1} \\ \text{Var}(\hat{\beta}_{LS})}} - \underbrace{\sigma C'P_{CR'}C}_{\text{P.S.D.}}$$

Also,  $\text{var}(\hat{\gamma}_{CLS}) = \sigma^2 [R(X'X)^{-1}R']^{-1}$  (under homoskedasticity).

### Misspecification

Suppose that projection coefficient  $\beta$  satisfies  $R\beta = r^* \neq r$ .

$$\Rightarrow \hat{\beta}_{CLS} = \hat{\beta}_{LS} - \underbrace{(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}[R\hat{\beta}_{LS} - r^* + (r^* - r)]}_{\text{biased}}.$$

Pseudo-true value :  $\beta^* := \beta - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}[r^* - r].$

Nonlinear constraints ?

$$R(\beta) = r \quad \rightarrow \mathcal{L} = \text{SSR}_n(b) + \lambda'(R(\beta) - r).$$

$$R: \mathbb{R}^k \rightarrow \mathbb{R}^q$$

## Testing:

$$H_0: R(\beta) = \theta_0$$

$$H_1: R(\beta) \neq \theta_0$$

$$R: \mathbb{R}^k \rightarrow \mathbb{R}^q$$

Approach 1: Look at  $(R(\hat{\beta}_{LS}) - \theta_0)$

$$\text{under } H_0, \sqrt{n}(R(\hat{\beta}_{LS}) - \theta_0) \xrightarrow{d} N(0, V_\theta)$$

← if  $R$  is ctly diff'ble  
+ LS asymptotically normal

$$V_\theta = R_\beta' V_\beta R_\beta \quad \text{where} \quad R_\beta := \frac{\partial}{\partial \beta} R(\beta)$$

$$\text{Wald statistic: } W_n := n(R(\hat{\beta}_{LS}) - \theta_0)' \hat{V}_\theta^{-1} (R(\hat{\beta}_{LS}) - \theta_0)$$

$$Z_n := \sqrt{n} \hat{V}_\theta^{-1/2} (R(\hat{\beta}_{LS}) - \theta_0) \xrightarrow{d} N(0, I_q)$$

$$W_n = Z_n' Z_n = \sum_{r=1}^q Z_r^2 \xrightarrow{d} \chi^2(q)$$

$$H_0: \beta_1 = 0$$

vs.

$$H_0': \exp(\beta_1) = 1$$

← Equivalent testing

$$H_1: \beta_1 \neq 0$$

$$H_1': \exp(\beta_1) \neq 1$$

$$R(\beta) = \beta_1 - 0$$

$$\tilde{R}(\beta) = \exp(\beta_1) - 1$$

↑  
More accurate

Approach 2: Impose constraint

$$\rightarrow \hat{\beta}_{CLS}, \hat{\Sigma}_{CLS}$$

↗ Lagrange multiplier  $\in \mathbb{R}^q$

$$\text{showed } \hat{\Sigma}_{CLS} = [R(X'X)^{-1}R']^{-1} \overbrace{(R(X'X)^{-1}X'Y - \theta_0)}^{\hat{\beta}_{LS}}$$

$$\sqrt{n} \hat{\Sigma}_{CLS} |_{H_0} \xrightarrow{d} N(0, V_\Sigma)$$

Lagrangian multiplier statistic:

$$LM_n := n \hat{\Sigma}_n' \hat{V}_\Sigma^{-1} \hat{\Sigma}_n \xrightarrow{d} \chi^2(q)$$

$$\tilde{Z}_n := \sqrt{n} \hat{V}_\Sigma^{-1/2} \hat{\Sigma}_{CLS} \xrightarrow{d} N(0, I_q)$$

Compare  $W_n$ ,  $LM_n$  for :

• Linear constraint  $R\beta = \theta_0$

• homoskedastic error  $\hat{V}_{\hat{\beta}} := \hat{S}_n^2 \left( \frac{1}{n} X'X \right)^{-1}$ ,  $\hat{S}_n^2 = \frac{1}{n-k} \hat{e}'_u \hat{e}_u$

$$\rightarrow W_n = \frac{n}{\hat{S}_n^2} (R \hat{\beta}_{LS} - \theta_0)' [R (X'X)^{-1} R']^{-1} (R \hat{\beta}_{LS} - \theta_0)$$

$$\rightarrow \hat{V}_{\hat{\beta}} = \frac{\hat{S}_n^2}{n} (R (X'X)^{-1} R')^{-1}$$

$$\begin{aligned} LM_n &= \frac{n}{\hat{S}_n^2} (R \hat{\beta}_{LS} - \theta_0)' [R (X'X)^{-1} R']^{-1} (R (X'X)^{-1} R') [R (X'X)^{-1} R']^{-1} (R \hat{\beta}_{LS} - \theta_0) \\ &= \frac{n}{\hat{S}_n^2} (R \hat{\beta}_{LS} - \theta_0)' [R (X'X)^{-1} R']^{-1} (R \hat{\beta}_{LS} - \theta_0) \end{aligned}$$

Approach 3: Likelihood Ratio

Gaussian LRM  $y_i = x_i' \beta + e_i$ ,  $e_i | x_i \sim N(0, \sigma^2)$

Compare  $(\hat{\beta}_u, \hat{\sigma}_u^2) := \underset{\beta, \sigma^2}{\operatorname{argmax}} L_u(\beta, \sigma^2)$

$(\hat{\beta}_R, \hat{\sigma}_R^2) := \underset{\beta, \sigma^2}{\operatorname{argmax}} L(\beta, \sigma^2) \text{ s.t. } R(\beta) = \theta_0$

Showed  $\hat{\beta}_n = \hat{\beta}_{LS}$ ,  $\hat{\beta}_R = \hat{\beta}_{CLS}$ .

$$\begin{aligned} \Rightarrow L_n(\hat{\beta}_u, \hat{\sigma}_u^2) &= -\frac{n}{2} \log(2\pi \hat{\sigma}_u^2) - \frac{1}{2\hat{\sigma}_u^2} \underbrace{\hat{e}'_u \hat{e}_u}_{n \hat{\sigma}_u^2} \\ &= -\frac{n}{2} \log(\hat{\sigma}_u^2) - \frac{n}{2} (\log(2\pi) - 1) \end{aligned}$$

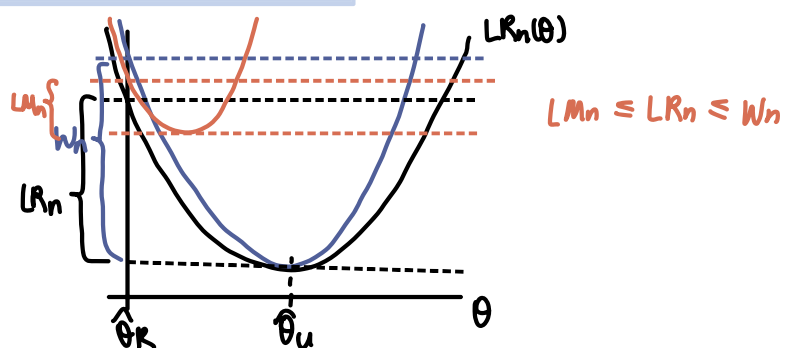
$$\text{Restricted model: } L_n(\hat{\beta}_R, \hat{\sigma}_R^2) = -\frac{n}{2} \log(\hat{\sigma}_R^2) - \frac{n}{2} (\log(2\pi) - 1)$$

$$\text{Likelihood ratio statistic: } LR_n := 2(L_n(\hat{\beta}_u, \hat{\sigma}_u^2) - L_n(\hat{\beta}_R, \hat{\sigma}_R^2))$$

$$= n(\log(\hat{\sigma}_R^2) - \log(\hat{\sigma}_u^2)) = n \log\left(\frac{\hat{\sigma}_R^2}{\hat{\sigma}_u^2}\right)$$

$$LR_n \xrightarrow{d} \chi^2(q)$$

→ "Trinity"  $\left\{ \begin{array}{l} \text{Wald} \\ LM \\ LR \end{array} \right.$



Special case: F-test for joint specification

$$y = X_1' \beta + X_2' \beta + e$$

$$X = [X_1 \ X_2]$$

$$\begin{aligned} \text{Hypothesis } H_0: \beta_2 &= 0 \\ H_1: \beta_2 &\neq 0 \end{aligned}$$

$$\rightarrow R\beta = [0, I_{k_2}] \beta = 0$$

$$\text{Wald stat. } W_n = n \hat{\beta}_2' (R \hat{V}_\beta R')^{-1} \hat{\beta}_2$$

$$\text{homoskedastic case: } \hat{V}_\beta = \hat{S}^2 \left( \frac{1}{n} X'X \right)^{-1}, \quad \hat{S}^2 = \frac{1}{n-k} \hat{e}_u' \hat{e}_u$$

$$\text{Claim: } W_n = (n-k) \frac{\hat{e}_R' \hat{e}_R - \hat{e}_u' \hat{e}_u}{\hat{e}_u' \hat{e}_u} \equiv F_n \dots \text{F-statistic}$$

$$\hat{e}_R = (I - P_{X_1}) y$$

$$\hat{e}_u = (I - P_X) y$$

$$R(X'X)^{-1}R' = [0, I] \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ I \end{bmatrix}$$

$$\rightarrow R \hat{V}_\beta R' = \hat{S}_n^2 (X_2' (I - P_{X_1}) X_2)^{-1} \leftarrow \text{partitioned inverse formula}$$

$$W_n = n \hat{\beta}_2' (R \hat{V}_\beta R')^{-1} \hat{\beta}_2 = \frac{n \hat{\beta}_2' \left( \frac{1}{n} X_2' (I - P_{X_1}) X_2 \right) \hat{\beta}_2}{\hat{S}_n^2}$$

$$\begin{aligned} \hat{e}_R &= (I - P_{X_1}) y = (I - P_{X_1}) (X_1 \hat{\beta}_1 + X_2 \hat{\beta}_2 + \hat{e}_u) \\ &= (I - P_{X_1}) X_2 \hat{\beta}_2 + \hat{e}_u \end{aligned}$$

$$\rightarrow \hat{e}_R' \hat{e}_R = \hat{e}_u' \hat{e}_u + \hat{\beta}_2' X_2' (I - P_{X_1}) X_2 \hat{\beta}_2 + \underbrace{2 \hat{e}_u' (I - P_{X_1}) X_2 \hat{\beta}_2}_{=0}$$

### Confidence sets

$C_u$  is a set estimator.

$C_u$  has confidence size  $1 - \alpha$  if  $1 - \alpha = \inf_{\beta, F} P_{\beta, F}(\beta \in C_u)$

Ex. t-test for  $H_0: \beta = b$  vs.  $H_1: \beta \neq b$ .

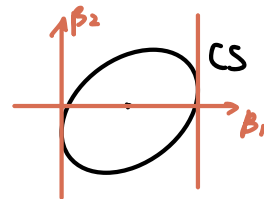
$$t_n(b) := \frac{\hat{\beta}_{nj} - b}{\text{se}(\hat{\beta}_j)} \Big|_{H_0} \xrightarrow{d} N(0, 1)$$

$$\begin{aligned} \rightarrow CI &= \{b \in \mathbb{R} : |t_n(b)| \leq z_{1-\frac{\alpha}{2}}\} \\ &\quad \hookrightarrow 1 - \frac{\alpha}{2} \text{ quantile of } N(0, 1). \\ &= [\hat{\beta}_{nj} - \text{s.e. } z_{1-\frac{\alpha}{2}}, \hat{\beta}_{nj} + \text{s.e. } z_{1-\frac{\alpha}{2}}] \end{aligned}$$

Joint confidence set for  $q$  coefficients:

$$\begin{aligned} H_0: \beta_2 = b \in \mathbb{R}^q &\Rightarrow W_n(b) = n(\hat{\beta}_{2LS} - b)' \hat{V}_{\beta_2}^{-1} (\hat{\beta}_{2LS} - b) \xrightarrow{d} \chi_{q,1}^2 \\ H_1: \beta_2 \neq b & \end{aligned}$$

$$CS := \{b \in \mathbb{R}^q : W_n(b) \leq \chi_{q,1}^2(1-\alpha)\}.$$



Bootstrap:

$$z_1, \dots, z_n \stackrel{iid}{\sim} F(z)$$

$$\text{Statistic } T_n := T_n(z_1, \dots, z_n)$$

$$\text{Exact distribution } G_n(t, F) := \mathbb{P}_F(T_n \leq t)$$

$$\text{Asymptotic principle: estimate } G_n(t, F) \text{ with } G_{\infty}(t, F) \equiv \lim_n G_n(t, F).$$

$$\text{Bootstrap principle: estimate } G_n(t, F) \text{ with } G_n(t, \hat{F}_n).$$