

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

Ridge Regression / L₂-Shrinkage

Idea: Restrict β to be in L₂-ball of radius c around zero.

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

$$\rightarrow \min \text{SSR}_n(b) \text{ st. } \|b\|^2 \leq c^2$$

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

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

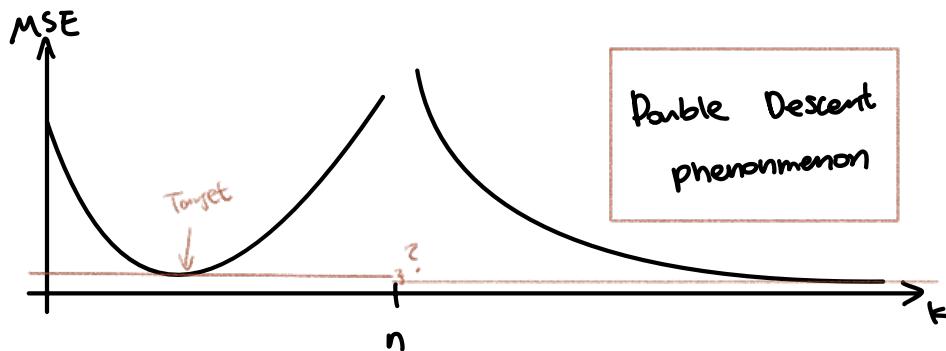
$$\rightarrow \hat{\beta}_R = (X'X + \hat{\lambda}_R I_k)^{-1} X'y$$

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

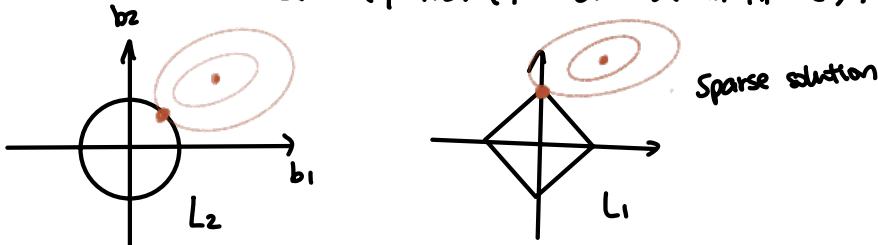
\rightarrow can show: posterior mean for β , $E_{\pi}[\beta | (x_1, y_1), \dots, (x_n, y_n)] = (X'X + \lambda I_k)^{-1} X'y$.



LASSO / L₁-Shrinkage

$\hat{\beta}_{\text{Lasso}}$ solves $\min_b \text{SSR}_n(b) \text{ st. } \|b\|_1 = \sum_{i=1}^n |b_i| \leq c$

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



Oracle Property

Restricted Estimation

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

$$\text{Restriction} \quad R\beta = r$$

(q x k) (q x 1)

Ex. Exclusion restriction

$$\beta = [\beta_1' \quad \beta_2']' \rightarrow \beta_2 = 0$$

k₁ x 1 k₂ x 2

$$\rightarrow R = [0 \quad I_{k_2}] \quad r = \underbrace{0}_{k_2 \times 1}$$

Ex. Cobb-Douglas prod. fxn

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

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

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

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

Constrained Least Squares :

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

$$\rightarrow J = \frac{1}{2} SSR_n(b) + \lambda' \underbrace{(Rb - r)}_{q \times 1}$$

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

$$\begin{aligned} \text{(2) w.r.t. } \beta : 0 &= \frac{\partial J}{\partial \beta} = \frac{1}{2} \frac{\partial}{\partial \beta} SSR_n(\beta) \Big|_{\beta = \hat{\beta}_{OLS}} + R' \lambda \quad (*) \\ &= -X'(Y - X \hat{\beta}_{OLS}) + R' \hat{\lambda}_{CLS} \end{aligned}$$

\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 $\hat{\lambda}_{OLS}$:

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

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

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

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

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

$$= C'(I - P_{CR'}) C X'y$$

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

$$\begin{aligned}\Rightarrow \text{var}(\hat{\beta}_{CLS}) &= \sigma^2 C'(I - P_{CR'}) \underbrace{C X'X}_{I} C' (I - P_{CR'}) C \\ &= \sigma^2 C'(I - P_{RC'}) C \\ &= \underbrace{\sigma^2 C'C}_{(X'X)^{-1}} - \underbrace{\sigma C' P_{RC'} C}_{\text{P.S.D.}} \\ &\quad \text{Var}(\hat{\beta}_{LS})\end{aligned}$$

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

Misspecification

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

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

bias

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

Nonlinear constraints ?

$$R(\beta) = r \rightarrow f = \text{SSR}_n(\beta) + \lambda'(R(\beta) - r).$$

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

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 $(R(\hat{\beta}_{LS}) - \theta_0)$

under H_0 , $\sqrt{n}(R(\hat{\beta}_{LS}) - \theta_0) \xrightarrow{d} N(0, V_{\beta})$ ← if R is cont. diff. + LS asymptotically normal

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

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

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

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

$$\begin{array}{lll} H_0: \beta_1 = 0 & \text{vs.} & H_0': \exp(\beta_1) = 1 \\ H_1: \beta_1 \neq 0 & & H_1': \exp(\beta_1) \neq 1 \end{array} \quad \leftarrow \text{Equivalent testing}$$

$$R(\beta) = \beta_1 - 0 \quad \tilde{R}(\beta) = \exp(\beta_1) - 1$$

↑
More accurate

Approach 2: Impose constraint

$$\rightarrow \hat{\beta}_{CLS}, S_{CLS} \quad \checkmark \text{Lagrange multiplier} \in \mathbb{R}^q$$

$$\text{Showed } \hat{S}_{CLS} = [R(X'X)^{-1} R']^{-1} (R(X'X)^{-1} X'y - \theta_0)$$

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

Lagrangian multiplier statistic:

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

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

Compare W_n, LM_n for :

- Linear constraint $R\beta = \theta_0$
- homoskedastic error $\hat{V}_2 := \hat{S}_n^2 (\frac{1}{n} X'X)^{-1}, \hat{S}_n^2 = \frac{1}{n-k} \hat{e}_{LS}' \hat{e}_{LS}$

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

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

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

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

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

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

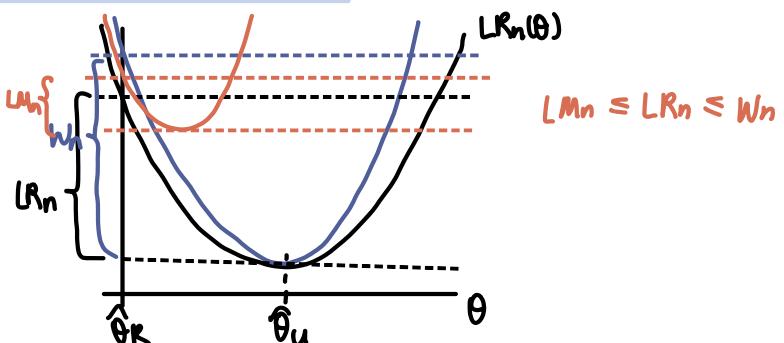
Restricted model : $L_n(\hat{\beta}_R, \hat{\sigma}_R^2) = -\frac{n}{2} \log(\hat{\sigma}_R^2) - \frac{n}{2} (\log(2\pi) - 1)$

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

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

$$X = [X_1 \ X_2]$$

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

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

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

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

$$\text{Claim: } W_n = (n-k) \frac{\hat{\beta}_2' \hat{e}_R - \hat{e}' \hat{e}}{\hat{e}' \hat{e}} = F_n \dots \text{F-statistic}$$

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

$$\hat{e}' = (I - P_X) Y$$

$$R(X'X)^{-1}R' = [0, I] \begin{bmatrix} X'_1 X_1 & X'_2 X_2 \\ X'_2 X_1 & X'_2 X_2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ I \end{bmatrix}$$

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

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

$$\hat{e}_R = (I - P_{X_1}) Y = (I - P_{X_1})(X_1 \hat{\beta}_1 + X_2 \hat{\beta}_2 + \hat{e}_u)$$

$$= (I - P_{X_1}) X_2 \hat{\beta}_2 + \hat{e}_u$$

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

Confidence sets

C_α is a set estimator.

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

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

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

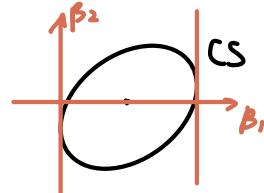
$$\begin{aligned} \rightarrow CI &= \{b \in \mathbb{R}: |t_n(b)| \leq z_{1-\frac{\alpha}{2}}\} \\ &\quad \hookrightarrow 1 - \frac{\alpha}{2} \text{ quantile of } N(0, 1). \\ &= [\hat{\beta}_{n,j} - \text{s.e. } z_{1-\frac{\alpha}{2}}, \hat{\beta}_{n,j} + \text{s.e. } z_{1-\frac{\alpha}{2}}] \end{aligned}$$

Joint confidence set for q coefficients:

$$H_0: \beta_2 = b \in \mathbb{R}^q \Rightarrow W_n(b) = n(\hat{\beta}_{2|S} - b)' V_{\beta_2}^{-1} (\hat{\beta}_{2|S} - b) \xrightarrow{d} \chi_q^2$$

$$H_1: \beta_2 \neq b$$

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



Bootstrap:

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

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

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

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

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