## 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.

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

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

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

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

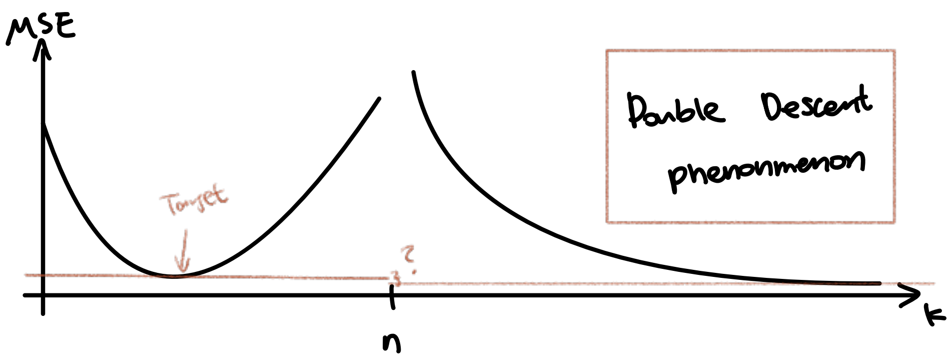
$$\rightarrow \hat{\beta}_R = (X'X + \hat{\lambda}_R I_k)^{-1} X'y$$

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 $\beta$, $E_\pi[\beta | (X_1, Y_1), \dots, (X_n, Y_n)] = (X'X + \lambda I_k)^{-1} X'y$.

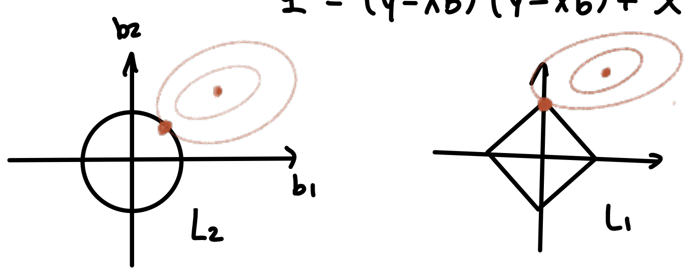



Double Descent phenomenon

### LASSO / $L_1$ - Shrinkage

$\beta_{Lasso}$ solves $\min_b SSR_n(b)$ s.t. $\|b\|_1 = \sum_{\ell=1}^{n} |b_\ell| \le c$

$$\mathcal{L} = (y - Xb)'(y - Xb) + \lambda(\|\beta_1\| - c).$$




Sparse solution

Oracle Property

Leeb & Pötscher (2007)

## Restricted Estimation

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

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

Ex. Exclusion restriction

$$\beta = [\underset{k_1\times 1}{\beta_1'} \quad \underset{k_2\times 2}{\beta_2'}]' \longrightarrow \beta_2 = 0$$

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

Ex. Cobb-Douglas prod. fxn

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

$$\rightarrow \log(Y_i) = \beta_0 + \beta_1 \log K_i + \beta_2 \log L_i + \overbrace{(\log(A_i) - \beta_0)}^{e_i}$$

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

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

## Constrained Least Squares :

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

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

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

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

$$= -X'(Y - X\hat\beta_{CLS}) + R'\hat\lambda_{CLS}$$

$\rightarrow$ Solve for $\hat\beta_{CLS}$, $\hat\lambda_{CLS}$.

$\rightarrow \hat\beta_{CLS}$ : $0 = -X'Y + X'X\hat\beta_{CLS} + R'\hat\lambda_{CLS} \Rightarrow \hat\beta_{CLS} = \underbrace{(X'X)^{-1}X'Y}_{\hat\beta_{LS}} - (X'X)^{-1}R'\hat\lambda_{OLS}$

$\rightarrow$ Solve for $\hat\lambda_{OLS}$ :

Pre-multiply (*) with $R(X'X)^{-1}$ : $0 = -R\underbrace{(X'X)^{-1}X'Y}_{\hat\beta_{LS}} + \underbrace{R\cancel{(X'X)^{-1}X'X}\hat\beta_{CLS}}_{=r} + R(X'X)^{-1}R'\hat\lambda_{CLS}$

$$\Rightarrow \quad R\hat\beta_{LS} - r = R(X'X)^{-1}R'\hat\lambda_{CLS}$$

$$\Rightarrow \quad \hat\lambda_{CLS} = [R(X'X)^{-1}R']^{-1}[R\hat\beta_{LS} - r]$$

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

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

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

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

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

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

$$= \underbrace{\sigma^2\underbrace{C'C}_{(X'X)^{-1}}}_{Var(\hat\beta_{LS})} - \underbrace{\sigma C'P_{RC'}C}_{p.s.d.}$$

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

<u>Misspecification</u>

S'pose that projection coefficient $\beta$ satisfies $R\beta = r^* \neq r$.

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

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

Nonlinear constraints ?

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

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

## Testing :

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

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

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

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

under $H_0$, $\sqrt{n}\,(R(\hat\beta_{LS}) - \theta_0) \xrightarrow{d} N(0, V_{\hat\theta})$ ← if $R$ is ctsly diff'ble + LS asymptotically normal

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

Wald statistic : $W_n := n\,(R(\hat\beta_{LS}) - \theta_0)' \hat V_\theta^{-1} (R(\hat\beta) - \theta_0)$

$\hat V_\theta^{-\frac12}\, \hat V_\theta\, \hat V_\theta^{-\frac12}$

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

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

chi

$H_0 : \beta_1 = 0$ vs. $H_0' : \exp(\beta_1) = 1$ ← Equivalent testing

$H_1 : \beta_1 \neq 0$ $\quad H_1' : \exp(\beta_1) \neq 1$

$R(\beta) = \beta_1 - 0$ $\qquad \tilde R(\beta) = \exp(\beta_1) - 1$

More accurate

**Approach 2 :** Impose constraint

$\rightarrow \hat\beta_{CLS}, \hat\lambda_{CLS}$ Lagrange multiplier $\in \mathbb{R}^q$

Showed $\hat\lambda_{CLS} = [R(X'X)^{-1}R']^{-1} (\overbrace{R(X'X)^{-1}X'Y}^{\hat\beta_{LS}} - \theta_0)$

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

Lagrangian multiplier statistic :

$LM_n := n \hat\lambda_n' \hat V_\lambda^{-1} \hat\lambda_n \xrightarrow{d} \chi^2(q)$

$\tilde z_n := \sqrt{n}\, \hat V_\lambda^{-\frac12} \hat\lambda_{CLS} \xrightarrow{d} N(0, I_q)$

## Compare $W_n$, $LM_n$ for :

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

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

$$\rightarrow \hat{V}_{\hat{\lambda}} = \frac{\hat{s}_{cls}^2}{n} (R(X'X)^{-1} R')^{-1}$$

$$LM_n = \frac{n}{\hat{s}_{cls}^2} (R\hat{\beta}_{ls} - \theta_0)' [R(X'X)^{-1} R']^{-1} (R(X'X)^{-1}R') [R(X'X)^{-1} R']^{-1} (R\hat{\beta}_{ls} - \theta_0)$$

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

## Approach 3 : Likelihood Ratio

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

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

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

Showed $\hat{\beta}_n = \hat{\beta}_{ls}$ , $\hat{\beta}_R = \hat{\beta}_{cls}$ .

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

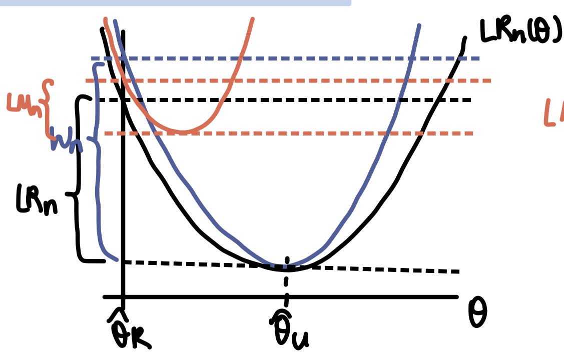
$$= -\frac{n}{2} \log(\hat{\sigma}_n^2) - \frac{n}{2} (\log(2\pi) - 1)$$

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

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

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




$LM_n \leq LR_n \leq W_n$

$\rightarrow$ "Trinity" $\begin{cases} \text{Wald} \\ LM \\ LR \end{cases}$

Special case : F-test for joint specification

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

$$X = [X_1 \ X_2]$$

Hypothesis $H_0: \beta_2 = 0$
$H_1: \beta_2 \neq 0$

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

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

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

Claim: $W_n = (n-k)\dfrac{\hat{e}_R'\hat{e}_R - \hat{e}_u'\hat{e}_u}{\hat{e}_u'\hat{e}_u} = F_n$ ... F-statistic

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

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

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

$$= (X_2'(I - P_{X_1})X_2)^{-1} \quad \leftarrow \text{Partitioned inverse formula}$$

$$\rightarrow R\hat{V}_\beta R' = \hat{S}_n^2(X_2'(I - P_{X_1})X_2)$$

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

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

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

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

## Confidence sets

$C_n$ is a set estimator.

$C_n$ has confidence size $1-\alpha$ if $1-\alpha = \inf_{\beta, F} \mathbb{P}_{\beta, F}(\beta \in C_n)$

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

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

$\rightarrow CI = \{ b \in \mathbb{R} : |t_n(b)| \leq z_{1-\frac{\alpha}{2}} \}$

$\hookrightarrow$ $1-\frac{\alpha}{2}$ quantile of $N(0,1)$.

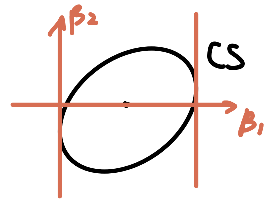
$= [\hat{\beta}_{nj} - s.e.\, z_{1-\frac{\alpha}{2}} , \ \hat{\beta}_{nj} + s.e.\, z_{1-\frac{\alpha}{2}}]$

Joint confidence set for $q$ coefficients :

$H_0 : \beta_2 = b \in \mathbb{R}^q$
$H_1 : \beta_2 \neq b$

$\Rightarrow \quad W_n(b) = n(\hat{\beta}_{2LS} - b)' \hat{V}_{\beta_2}^{-1} (\hat{\beta}_{2LS} - b) \xrightarrow{d} \chi^2_{(q)}$

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




<u>Bootstrap</u> :

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

Statistic $\quad T_n := T_n(z_1, \dots z_n)$

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

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

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