ch 9. Multicollinearity

Def. $x_1 . x_2$ have multicollinearity $\Leftrightarrow$ $\exists\ ax_1 + by_2 + c$. a, b, c: const.

왜 문제인가?

$\boxed{\text{Model 1}}$ $\quad y_i - \bar{y} = \beta_1 (x_{i1} - \bar{x}_1) + \beta_2 (x_{i2} - \bar{x}_2) + \varepsilon_i$

$$\Leftrightarrow \quad y = X\beta + \varepsilon. \qquad y = \begin{pmatrix} y_1 - \bar{y} \\ \vdots \\ y_n - \bar{y} \end{pmatrix} \qquad X = \begin{bmatrix} X_{11} - \bar{X}_1 & X_{12} - \bar{X}_2 \\ \vdots & \vdots \\ X_{n1} - \bar{X}_1 & X_{n2} - \bar{X}_2 \end{bmatrix} \qquad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}.$$

$$X^T X = \begin{bmatrix} \Sigma (X_{i1} - \bar{X}_1)^2 & \Sigma (X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2) \\ \Sigma (X_{i1} - \bar{X}_1)(X_{i2} - \bar{X}_2) & \Sigma (X_{i2} - \bar{X}_2)^2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{12} & S_{22} \end{bmatrix}$$

$$(X^T X)^{-1} = \frac{1}{S_{11} \cdot S_{12} - S_{12}^2} \begin{bmatrix} S_{22} & -S_{12} \\ -S_{12} & S_{11} \end{bmatrix}$$

$$X^T y = \begin{bmatrix} \Sigma (X_{i1} - \bar{X})(Y_i - \bar{Y}) \\ \Sigma (X_{i2} - \bar{X})(Y_i - \bar{Y}) \end{bmatrix} = \begin{pmatrix} S_{1y} \\ S_{2y} \end{pmatrix}$$

$$Var(\hat{\beta}) = \sigma^2 (X^T X)^{-1}.$$

$$Var(\hat{\beta}_1) = \sigma^2 \cdot \frac{1}{S_{11} S_{12} - S_{12}^2} \cdot S_{22}$$

$$= \frac{\sigma^2 \cdot 1/S_{11}}{1 - \frac{S_{12}^2}{S_{11}/S_{12}}} = \frac{\sigma^2}{\underline{1 - r_{12}^2}} \cdot \frac{1}{S_1}$$

$\hookrightarrow$ $|r_{12}| \simeq 1$ 이면 $Var\uparrow$.

$\rightarrow$ $\hat{y} = X\underline{\hat{\beta}}$ 의 의미 X.

var 너무 클 경우.

$\boxed{\text{Model 2}}$ $y_i^* = \beta_1 x_{i1}^* + \beta_2 x_{i2}^* + \varepsilon.$

$$y_i^* = \frac{y_i - \bar{y}}{\sqrt{\Sigma(y_i-\bar{y})^2}} = \frac{y_i - \bar{y}}{S_y\sqrt{n-1}}$$

$$x_{i1}^* = \frac{X_{i1} - \bar{X}_1}{\sqrt{\Sigma(X_{i1}-\bar{X})^2}}, \quad X_{i2}^* = \frac{X_{i2} - \bar{X}_2}{\sqrt{\Sigma(X_{i2}-\bar{X})^2}}$$

) ⇒ 합 = 0, 제곱합 = 1.

$\Leftrightarrow y = X\beta + \varepsilon.$

$$X^TX = \begin{bmatrix} 1 & r_{12} \\ \frac{\Sigma(X_{i1}-\bar{X}_1)(X_{i2}-\bar{X}_2)}{\sqrt{\Sigma(\ )^2\Sigma(\ )^2}} = r_{12} & 1 \end{bmatrix}$$

$$(X^TX)^{-1} = \frac{1}{1-r_{12}^2} \cdot \begin{bmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{bmatrix}$$

$Var(\hat{\beta}_j)$는 $|r_{12}| \to 1$에 따라 $\longrightarrow \infty$

In general, $VIF_j = C_{jj} = \frac{1}{1-R_j^2}$ R_j^2: $x_j \leftrightarrow X_{(j)}$ reg. 한 R^2.

* Multicollinearity : ① Var ↑.

② X^TX not invertible. (extremely many solutions).

Back to poly-reg.

$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon$ ⇒ centering 하면 $Corr((x-\bar{x}), (x-\bar{x})^2) = 0.$

다중공선성 체크.

① Corr. matrix. = X^TX의 off-diagonal term.

⇒ pair 에만 useful. (corr 크지 않아도 다중공선성 발생 가능).

② VIF : 5 or 10 이상인 경우.

③ eigensystem analysis of X^TX. → condition number $:= \frac{\lambda_{max}}{\lambda_{min}} > 100.$

↳ 크면 $(X^TX)^{-1}$ 불안정.

⇒ sol : Adding additional data.
Model selection
Ridge Reg.

Model Selection

Full: $y_i = \beta_1 X_{i1} + \beta_2 X_{i2} + \varepsilon_i \rightarrow Var(\hat\beta_1) = \sigma^2 \frac{1}{1-r_{12}^2} = MSE(\hat\beta_1)$

$(\because \ bias = 0)$.

Subset: $y_i = \beta_1 X_{i1} + \varepsilon_i$

$\hookrightarrow \hat\beta_{1S} = \frac{\sum X_{i1} y_i}{\sum X_{i1}^2} = \sum X_{i1} y_i$

$$E(\hat\beta_{1S}) = E(\sum X_{i1} y_i)$$

$$= \sum X_{i1} (\beta_1 X_{i1} + \beta_2 X_{i2}).$$

$$= \beta_1 \sum X_{i1}^2 + \beta_2 \sum X_{i1} X_{i2} = \beta_1 + \underbrace{\beta_2 r_{12}}_{\text{bias}}$$

$$MSE(\hat\beta_{1S}) = Var(\hat\beta_{1S}) + bias^2$$

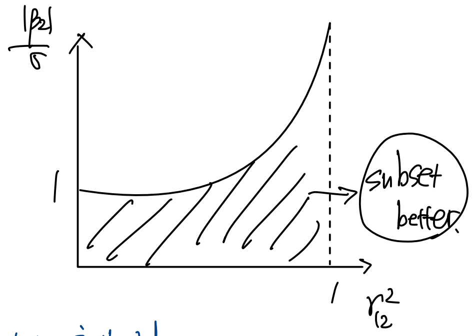
$$= \frac{\sigma^2}{\sum X_{i1}^2} + (\beta_2 r_{12})^2$$

$$= \sigma^2 + \beta_2^2 r_{12}^2$$

$\star \Rightarrow MSE(\hat\beta_{1S}) < MSE(\hat\beta_1)$

pf) $\sigma^2 + \beta_2^2 r_{12}^2 < \sigma^2 \left(\frac{1}{1-r_{12}^2} \right).$

$\frac{|\beta_2|}{\sigma} < \frac{1}{\sqrt{1-r_{12}^2}} \rightarrow |r_{12}| \rightarrow 1$ 이면 효과 크다.



Ridge

$\rightarrow$ biased est. (MSE$\downarrow$).

$$\hat\beta_\lambda = (X^TX + \lambda I)^{-1} X^T y.$$

properties

(a) $\hat\beta_\lambda$ 는 LSE 의 선형변환.

(b) biased. $= Z_\lambda \beta - \beta = -\lambda (X^TX + \lambda I)^{-1} \beta.$

(c) $Var(\hat\beta_\lambda) = \sigma^2 (X^TX + \lambda I)^{-1} (X^TX) (X^TX + \lambda I)^{-1}.$

(a) $\hat\beta_\lambda = (X^TX + \lambda I)^{-1} X^T y$

$$= (X^TX + \lambda I)^{-1} X^TX \hat\beta = \star\, Z_\lambda \hat\beta$$

(b) $bias(\hat\beta_\lambda) = E(\hat\beta_\lambda) - \beta = Z_\lambda \beta - \beta.$

$$= [(X^TX + \lambda I)^{-1} X^TX - I_p] \beta.$$

$$= \cancel{\ } [(X^TX+\lambda I)^{-1}(X^TX+\lambda I) - \lambda(X^TX+\lambda I)^{-1} - I]\beta .$$

$$= -\lambda(X^TX+\lambda I)^{-1}\beta .$$

(c). $Var(\hat\beta_\lambda) = Var(Z_k\hat\beta)$

$$= Z_k\, Var(\hat\beta)\, Z_k^T$$

$$= \sigma^2 (X^TX+\lambda I)^{-1}(X^TX)(X^TX+\lambda I)^{-1}$$

MSE of ridge est. . .

$$MSE(\hat\beta_\lambda) = E[(\hat\beta_\lambda-\beta)^T(\hat\beta_\lambda-\beta)]$$

$$= E\left(\sum_{j=1}^{p}(\hat\beta_{\lambda_j}-\beta_j)^2\right) = \sum_{j=1}^{p}E(\hat\beta_{\lambda,j}-\beta_j)^2.$$

$$= \sum_{j=1}^{p}(Var(\hat\beta_{\lambda,j}) + \text{bias}^2 \text{ in } \hat\beta_{\lambda_j}).$$

$$\sum var(\hat\beta_{\lambda,j}) = tr(Var(\hat\beta_j))$$

$$\sum \text{bias}^2 \text{ in } \hat\beta_{\lambda_j} = \text{bias}(\hat\beta_\lambda)^T \text{bias}(\hat\beta_\lambda).$$

$$= [E(\hat\beta_\lambda)-\beta]^T[E(\hat\beta_\lambda)-\beta].$$

If $\lambda\uparrow$: $\text{bias}^2\uparrow$, var $\downarrow$.

i) if $X^TX = I$, $\text{bias}(\hat\beta_\lambda) = -\lambda(I+\lambda I)^{-1}\beta$

$$= -\lambda[(1+\lambda)I]^{-1}\beta.$$

$$* \ \text{bias}(\hat\beta_\lambda)^T\text{bias}(\hat\beta_\lambda) = \lambda^2\beta^T((1+\lambda)I)^{-1}((1+\lambda)I)^{-1}\beta$$

$$= \sum_{j=1}^{p}\left(\frac{\lambda}{1+\lambda}\right)^2\beta_j^2.$$

↳ if $\lambda=0$, $\text{bias}^2=0$.
$\lambda\to 1$, $\text{bias}^2\uparrow$.

ii) If $X^TX = I$, $Var(\hat\beta_\lambda) = \sigma^2((1+\lambda)I)^{-1}((1+\lambda)I)^{-1}$

$$tr(Var(\hat\beta_\lambda)) = \sigma^2\frac{1}{(1+\lambda)^2} \rightarrow \lambda = 0 \text{이면 } \sigma^2$$

$\lambda\uparrow$: $\downarrow$.

How to compute $\hat{\beta}_\lambda$?

$$X_A = \begin{pmatrix} X \\ \sqrt{\lambda} I_p \end{pmatrix} \qquad y_A = \begin{pmatrix} y \\ 0_p \end{pmatrix} \quad \Rightarrow \quad (X_A^T X_A)^{-1} X_A^T y_A$$

$(n+p) \times p$. $\qquad (n+p) \times 1$.

$$= (X^T X + \lambda I)^{-1} X^T y$$

⊕ 추가설명.

1. $\hat{\beta}_\lambda$는 penalized least-squares 의 solution.

$$\min \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik})^2 \quad \text{subject to} \sum_{j=1}^{p} \beta_j^2 \le s.$$

$$\Leftrightarrow \min_{\beta} (y - X\beta)^T (y - X\beta) + \lambda \beta^T \beta.$$

$$\begin{pmatrix} * \ (y_A - X_A\beta)^T (y_A - X_A\beta) \\ = y_A^T y_A - \beta^T X_A^T y_A - y_A^T X_A \beta + \beta^T X_A^T X_A \beta \\ = y^T y - \beta^T X^T y - y^T X \beta + \beta^T X^T X \beta + \lambda \beta^T \beta. \end{pmatrix}$$

$$\left.\frac{\partial S(\beta)}{\partial \beta}\right|_{\hat{\beta} = \beta} = -2X^T y + 2X^T X \hat{\beta} + 2\lambda \hat{\beta}. = 0$$

$$(X^T X + \lambda I) \hat{\beta}_\lambda = X^T y.$$

2. Ridge est. $\hat{\beta}_\lambda$: shrinkage est. (OLSE $\to 0$).

* If $X^T X = I$, $\quad \hat{\beta}_\lambda = (X^T X + \lambda I)^{-1} X^T y = (X^T X + \lambda I)^{-1} X^T X \hat{\beta}$

$$\hat{\beta}_{\lambda j} = \frac{\hat{\beta}_j}{1 + \lambda}.$$

* Generally, use SVD : $\quad X = U D V^T \qquad D = \text{diag}(d_1 \ldots d_{p^*})$.

$\quad n \times p \quad n \times p \quad p \times p \quad p \times p$.

$$U^T U = I_p = V^T V = V V^T$$

$$(V^T = V^{-1})$$

Note that $X^T X = V D^T U^T U D V^T = V D^2 V^T$.

$$\hat{\beta}_\lambda = (X^T X + \lambda I)^{-1} X^T y$$

$$= (V D^2 V^T + \lambda I)^{-1} V D U^T y.$$

$$= (V D^2 V^T + \lambda V V^T)^{-1} V D U^T y.$$

$$= [V (D^2 + \lambda I) V^T]^{-1} V D U^T y.$$

↓ ②

$$= (V^T)^{-1}(D^2+\lambda I)^{-1}\, V^{-1}\, V D U^T y$$

$$= V (D^2+\lambda I)^{-1} V^T V D U^T y$$

$$= V (D^2+\lambda I)^{-1} D U^T y.$$

Define $\Phi = XV = UDV^TV. = UD.$

$\hat{\beta}_\lambda = V(D^2+\lambda I)^{-1}\Phi^T y.$ $\rightsquigarrow$ OLS로 표현 함을 보이자.

pf) Consider $y = X\beta + \varepsilon$

$$= XVV^T\beta + \varepsilon$$

$$= \Phi\alpha + \varepsilon \qquad (XV = \Phi,\ V^T\beta = \alpha)$$

Then, $\hat{\alpha}_{OLS} = (\Phi^T\Phi)^{-1}\Phi^T y$

$$= (DU^TUD)^{-1}\Phi^T y$$

$$= D^{-2} V^T X^T y.$$

$$\rightarrow \hat{\alpha}_\lambda = V^T\hat{\beta}_\lambda$$

$$= V^T V (D^2+\lambda I)^{-1} V^T X^T y$$

$$= (D^2+\lambda I)^{-1} D^2\, \hat{\alpha}_{OLS}$$

$$\therefore\ \hat{\alpha}_{\lambda,j} = \frac{d_j^2}{d_j^2+\lambda}\cdot \hat{\alpha}_{OLS,j}.$$

$\rightarrow$ if $\lambda = 0$, $\hat{\alpha}_{\lambda,j} = \hat{\alpha}_{OLS,j}$

$\lambda\uparrow$. $\hat{\alpha}_{\lambda,j}\downarrow < \hat{\alpha}_{OLS,j}$

shrink.

3. MSE $(\hat{\beta}_\lambda)$의 일반적인 계산. $(X^TX \neq I)$.

$$MSE(\hat{\beta}_\lambda) = E[(\hat{\beta}_\lambda-\beta)^T(\hat{\beta}_\lambda-\beta)]$$

$$= tr(Var(\hat{\beta}_\lambda)) + (E(\hat{\beta})-\beta)^T(E(\hat{\beta})-\beta).$$

i) $tr(Var(\hat{\beta}_\lambda)) = tr(\sigma^2(X^TX+\lambda I)^{-1}(X^TX)(X^TX+\lambda I)^{-1})$

$$= \sigma^2\, tr((VD^2V^T+\lambda VV^T)^{-1} VD^2V^T(VD^2V^T+\lambda VV^T)^{-1})$$

$$= \sigma^2\, tr(\cancel{V}(D^2+\lambda I)^{-1}\cancel{V^TV}D^2\cancel{V^TV}(D^2+\lambda I)^{-1}\cancel{V^T})$$

$$= \sigma^2\, tr((D^2+\lambda I)^{-1}D^2(D^2+\lambda I)^{-1}). = \sigma^2\sum_{j=1}^{p}\frac{d_j^2}{(d_j^2+\lambda)^2}$$

ii) $E(\hat{\beta}_\lambda)-\beta = E((X^TX+\lambda I)^{-1}X^Ty) - \beta$

$$= E((X^TX+\lambda I)^{-1}X^TX\hat{\beta}) - \beta$$

$$= (X^TX+\lambda I)^{-1}X^TX\beta - \beta$$

$$= [(VD^2V^T+\lambda VV^T)^{-1}VD^2V^T - VV^T]\beta$$

$$= [V(D^2+\lambda I)^{-1}\cancel{V^TV}D^2V^T - VV^T]\beta$$

$$= V[(D^2+\lambda I)^{-1}D^2 - I]V^T\beta$$

$$= V[(D^2+\lambda I)^{-1}D^2 - I]\cancel{V^TV}\alpha.$$

$$[E(\hat{\beta}_\lambda)-\beta]^T[E(\hat{\beta}_\lambda)-\beta] = \alpha^T[(D^2+\lambda I)^{-1}D^2 - I]^T\cancel{V^TV}[(D^2+\lambda I)^{-1}D^2 - I]\alpha.$$

$$= \sum_{j=1}^{p}\left(\frac{(d_j^2)^2}{(d_j^2+\lambda)^2}-1\right)\alpha_j^2 \qquad * \ \frac{d_j^2}{d_j^2+\lambda}-1 = -\frac{\lambda}{d_j^2+\lambda}$$

$$= \sum_{j=1}^{p}\frac{\lambda^2\alpha_j^2}{(d_j^2+\lambda)^2}$$

$$\Rightarrow MSE(\hat{\beta}_\lambda) = \sum_{j=1}^{p}\frac{\sigma^2 d_j^2}{(d_j^2+\lambda)^2} + \frac{\lambda^2\alpha_j^2}{(d_j^2+\lambda)^2}$$

↳ If $\lambda=0$, $MSE = p\sigma^2$.

$\lambda\uparrow$: $tr(Var(\hat{\beta}_\lambda))\downarrow$, $bias^2\uparrow$. → 적절히 선택.

⊕ Principal Component Reg.

$$y = X\beta+\varepsilon \iff y = XVV^T\beta+\varepsilon$$
$$= \Phi\alpha+\varepsilon.$$

↱ eigenvectors

$* \ V = [\ |\ |\ |\ ]$

$\Phi = [\phi_1 \dots \phi_p]$

↳ principal components.

= X들의 linear comb.

$$\Rightarrow \hat{\alpha}_{LSE} = (\Phi^T\Phi)^{-1}\Phi^Ty$$
$$= D^{-1}\Phi y.$$

$$(\because \Phi^T\Phi = V^TX^TXV = V^TVD^2V^TV = D^2).$$
$$\triangleq D.$$

↳ diagonal ⇒ 항상 invertible.

$$Var(\hat{\alpha}) = \sigma^2 (\Phi^T \Phi)^{-1} = \sigma^2 D^{-1}$$

$$Var(\hat{\beta}) = Var(V\hat{\alpha}) = V\, Var(\hat{\alpha})\, V^T$$

$$= \sigma^2 V D^{-1} V^T \;\rightarrow\; \sigma^2 a^T V D^{-1} V^T a.$$

$$Var(\hat{\beta}_j) = \sigma^2 \left(\frac{\sum_{j=1}^{p} V_{ij}^2}{d_i} \right).$$

Ch10. Variable Selection.

↳ motivation: to improve $\underline{\text{precision}}$ of param. estimates.
= MSE.

Why useful?

Consider model 2. : $y = \beta_1 x_1 + \beta_2 x_2 + \varepsilon$. (full).

$$\rightarrow \begin{cases} Var(\hat{\beta}_1) = \sigma^2 \dfrac{1}{1-r_{12}^2} \\ E(\hat{\beta}_1) = \beta_1 \quad \text{(unbiased)}. \end{cases}$$

Consider subset model ~ $y = \beta_1 x_1 + \varepsilon$.

$$\rightarrow \begin{cases} Var(\hat{\beta}_1) = \sigma^2. \\ E(\hat{\beta}_1) = \beta_1 + r_{12}\beta_2. \;\star \end{cases}$$

hw4 참고.

pf) $\hat{\beta}_{1,subset} = \dfrac{S_{1y}}{S_{11}} = S_{1y} = \sum (x_{i1} - \bar{x}) y_i$

$Var(\hat{\beta}_1) = \sum (x_{i1} - \bar{x}_1)^2 \sigma^2 = S_{11}\sigma^2 = \sigma^2.$

$E(\hat{\beta}_1) = \sum (x_{i1} - \bar{x})(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}) = \beta_1 + \beta_2 r_{12}.$

↳ $\beta_0 \sum (x_{i1} - \bar{x}) + \beta_1 \sum (x_{i1} - \bar{x}_1)^2 + \beta_2 \sum (x_{i1} - \bar{x}_1) \sum (x_{i2} - \bar{x}_2).$

Compare MSE : $\dfrac{|\beta_2|}{\sigma} < \dfrac{1}{\sqrt{1-r_{12}^2}}$

$\uparrow$ MSE($\hat{\beta}_1$, S) $\qquad \uparrow$ MSE($\hat{\beta}_1$, full).

if $r_{12}^2 \simeq 1$: subset

$|\beta_2| < \sigma$: subset.

(noise보다 영향↓).

Criteria for Subset Selection.

① $R_p^2 = \frac{SS_R(p)}{SS_T} = 1 - \frac{SS_E(p)}{SS_T}$. $\quad p = 1, \dots, k+1$. variable의 개수.

- $\binom{k}{p-1}$ values for each p.
- $p \uparrow$, $R_p^2 \uparrow$. $\quad$ (max : $p = k+1$).

② Adjusted R_p^2. $= 1 - \frac{n-1}{n-p}(1 - R_p^2)$

$$= 1 - \frac{n-1}{n-p} \cdot \frac{SS_E(p)}{SS_T} = 1 - \frac{n-1}{SS_T} \cdot MSE.$$

(min MSE $\Leftrightarrow$ max $R^2_{adj,p}$).

③ $MS_E = \frac{SS_E(p)}{n-p}$.

④ Mallow's C_p. $= \frac{SS_E(p)}{\hat{\sigma}^2} - n + 2p$.

Let $SS_E = y^T(I - P_p)y$.

$$E(SS_E(p)) = \mu^T(I-P_p)\mu + (n-p)\sigma^2. \quad (\because \text{quadratic form}).$$

$$= \underbrace{\mu^T(I-P_p)\mu + p\sigma^2}_{E(ME)} + (n-2p)\sigma^2$$

$$= E(ME) + (n-2p)\sigma^2.$$

$$\Leftrightarrow \frac{E(ME)}{\sigma^2} = \frac{E(SS_E(p))}{\sigma^2} - n + 2p. \quad \leftarrow \quad \underline{C_p \text{ is est. of }} \ \frac{E(ME)}{\sigma^2}$$

$$E(C_p) = \frac{E(ME)}{\sigma^2}.$$

* model error의 이해.

$$ME = \sum(\mu_i - \hat{\mu}_i)^2 = (\mu - \hat{\mu})^T(\mu - \hat{\mu}).$$

$$= \|\mu - \hat{\mu}\|^2$$

$$= \|\mu - Py\|^2$$

$$= \|\mu - P(\mu + \varepsilon)\|^2$$

$$= \|\mu - P\mu\|^2 + \|P\varepsilon\|^2.$$

$$= \mu^T(I-P)\mu + \varepsilon^T P \varepsilon.$$

$$E(ME) = \mu^T(I-P)\mu + \sigma^2 tr(P) = \mu^T(I-P)\mu + \sigma^2 p.$$

$$E(C_p) \simeq \frac{E(SSE(p))}{\sigma^2} - n + 2p.$$

$$= \frac{1}{\sigma^2}\left(\mu^T (I-P)\mu + (n-p)\sigma^2 \right) - n + 2p.$$

$$= \frac{\mu^T (I-P)\mu}{\sigma^2} + p. = \frac{E(ME)}{\sigma^2}.$$

$$\rightarrow E(C_p) = p. \quad \text{if } \mu = X_p\beta_p$$ (μ가 p로 완벽히 설명).

bias = 0.

* Prediction error.

- $\hat\beta_p = (X_p^T X_p)^{-1} X_p^T y.$ $\quad \hat y_p = X_p \hat\beta_p.$
- new obs $y_i^* = \mu_i + \varepsilon_i^*.$

$$E(PE) = \Sigma E(y_i^* - \hat y_{ip})^2.$$

$$= \Sigma E\left((y_i^* - \mu_i) - (\hat y_{ip} - \mu_i) \right)^2$$

$$= \Sigma \left[E(y_i^* - \mu_i)^2 - 2\,\cancel{E(\varepsilon_i^* (\hat y_{ip} - \mu_i))}^{\,0} + E(\hat y_{ip} - \mu_i)^2 \right].$$

$$= \Sigma \underbrace{E(\varepsilon_i^*)^2}_{\sigma^2} + \Sigma E(\hat y_{ip} - \mu_i)^2$$

$$\Sigma E(\hat y_{ip} - \mu_i)^2 = \Sigma E\left(\hat y_{ip} - E(\hat y_{ip}) + E(\hat y_{ip}) - \mu_i \right)^2$$

$$= \underline{\Sigma Var(\hat y_{ip})} + 2\Sigma E\left(\cancel{\hat y_{ip} - E(\hat y_{ip})}^{\,0}\right)\underbrace{\left(E(\hat y_{ip}) - \mu_i\right)}_{\text{상수}}$$

$$+ \Sigma \left(E(\hat y_{ip}) - \mu_i \right)^2$$

* $Var(\hat y_p)$

$= Var(X_p^T \hat\beta)$

$= X_p^T \sigma^2 (X_p^T X_p)^{-1} X_p$

$$= tr(Var(\hat y_{ip})) + \Sigma (E(\hat y_{ip}) - \mu_i)^2$$

$$= \sigma^2 \cdot p + \Sigma (\mu_{ip} - \mu_i)^2$$

$$\therefore E(PE) = n\sigma^2 + p\sigma^2 + \Sigma(\mu_{ip} - \mu_i)^2. = \mu^T\mu$$

$$\mu_p = E(Py) = P\,E(y) = P\mu. = \{\mu_{ip}\}_{i=1,\dots n}.$$

$$\Sigma(\mu_{ip} - \mu_i)^2 = \mu^T\mu - \mu^T P\mu = \mu^T(I-P)\mu.$$

?

$2\Sigma\, \underline{\mu_{ip} \times \mu_i}$.

$$E(PE) = (n+p)\sigma^2 + \underbrace{\mu^T(I-P)\mu}_{\|\mu - P\mu\|^2}.$$

* estimator of EPE.

$$SSE(p) = \Sigma(y_i - \hat{y}_{ip})^2 \qquad \text{cf. } EPE = E\Sigma(y_i^* - \hat{y}_{ip})^2$$
$$= \mu^T(I-P)\mu + (n+p)\sigma^2$$

$$E(SSE(p)) = \mu^T(I-P)\mu + (n-p)\sigma^2.$$
$$= EPE - 2p\sigma^2.$$
$$\widehat{EPE} = SSE(p) + 2p\hat{\sigma}^2$$

↳ Cp와의 관계: $C_p = \frac{SSE(p)}{\hat{\sigma}^2} - n + 2p.$
$$= \frac{\widehat{EPE}}{\hat{\sigma}^2} - n. \quad \rightarrow PE\downarrow \equiv C_p\downarrow.$$
$$\hat{p}^* = \min_p C(p).$$

* Cross-Validation.

- LOOCV := $\frac{1}{n}\sum_{i=1}^n (y_i - x_i^T\hat{\beta}^{-i})^2$ ($\hat{\beta}^{-i}$: estimate of β w/o i^{th} obs)

- PE와의 관계: $PE = \Sigma(y_i^* - \hat{y}_{ip})^2$

$y_1 \cdots y_i \cdots y_n.$
↳ y_i^* (new obs)로 생각.

→ CV는 PE mimic하는 것.

- $CV = \frac{1}{n}\Sigma\left(\frac{y_i - x_i^T\hat{\beta}}{1-h_{ii}}\right)^2 = \frac{1}{n}\Sigma(\boxed{y_i - x_i^T\hat{\beta}^{-i}})^2$ = PRESS residual. $e_{(i)}$

pf) $$y_i - x_i^T\hat{\beta}^{-i} = y_i - x_i^T\left(\hat{\beta} - \frac{(X^TX)^{-1}x_i(y_i - x_i^T\hat{\beta})}{1-h_{ii}}\right)$$
$$= \frac{y_i - x_i^T\hat{\beta}}{1-h_{ii}}$$
$$= \frac{e_i}{1-h_{ii}}.$$

- $GCV_{(p)} = \frac{1}{n}\Sigma\left(\frac{y_i - \hat{y}_{ip}}{1 - \frac{tr(H)}{n}}\right)^2 \xrightarrow{\text{ridge:}} \frac{1}{n}\Sigma\left(\frac{y_i - \hat{y}_{i,\lambda}}{1-\frac{1}{n}tr(P_\lambda)}\right)^2 \rightarrow \lambda^*$ 찾는데 이용.

($P_\lambda = X(X^TX + \underline{\lambda I})^{-1}X$ λ에 depend).

* CV(λ), MSE(λ), $\underset{=EPE}{PSE(\lambda)}$의 관계: (목표: CV → PSE → MSE).

$$\begin{cases} CV(\lambda) = \frac{1}{n}\Sigma(y_i - \hat{f}_\lambda^{-i}(x_i))^2 \\ MSE(\lambda) = \frac{1}{n}\Sigma E(\hat{f}_\lambda(x_i) - f(x_i))^2 \\ PSE(\lambda) = \frac{1}{n}\Sigma E(y_i^* - \hat{f}_\lambda(x_i))^2 \end{cases}$$

$$\underline{E(CV(\lambda))} \overset{?}{\simeq} E\left(y_i - \hat{f}_\lambda^{-i}(x_i)\right)^2$$

$$= E\left(y_i - f(x_i) + f(x_i) - \hat{f}_\lambda^{-i}(x_i)\right)^2$$

$$= E\left(y_i - f(x_i)\right)^2 + 2E\left(y_i - f(x_i)\right)\left(f(x_i) - \hat{f}_\lambda^{-i}(x_i)\right)$$

연관 x

$$+ E\left(f(x_i) - \hat{f}_\lambda^{-i}(x_i)\right)^2$$

$$= \sigma^2 + E\left(f(x_i) - \hat{f}_\lambda^{-i}(x_i)\right)^2$$

$$E\left(y_i^* - \hat{f}_\lambda(x_i)\right)^2 = E\left(y_i^* - f(x_i) + f(x_i) - \hat{f}_\lambda(x_i)\right)^2$$

$$= E\left(y_i^* - f(x_i)\right)^2 + 2E\left(y_i^* - f(x_i)\right)\left(f(x_i) - \hat{f}_\lambda(x_i)\right)$$

$$+ E\left(f(x_i) - \hat{f}_\lambda(x_i)\right)^2.$$

$$= \sigma^2 + E\left(f(x_i) - \hat{f}_\lambda(x_i)\right)^2.$$

$\Rightarrow$ assume that $\hat{f}_\lambda^{-i}(x_i) \simeq \hat{f}_\lambda(x_i)$, $E(CV(\lambda)) \simeq PSE(\lambda)$.

↳ est. of PSE.

또한 $$PSE(\lambda) = \frac{1}{n}\sum\left(\sigma^2 + E\left(f(x_i) - \hat{f}_\lambda(x_i)\right)^2\right)$$

$$= \sigma^2 + \frac{1}{n}\sum E\left(f(x_i) - \hat{f}_\lambda(x_i)\right)$$

$$= \sigma^2 + MSE(\lambda)$$

$\therefore \underset{\lambda}{\operatorname{argmin}}\, PSE(\lambda) = \underset{\lambda}{\operatorname{argmin}}\, MSE(\lambda).$

$\Rightarrow$ PE와 MSE는 equivalent.

④ $$AIC = -2\log(L) + 2p.$$

$$= n\log\left(\frac{SSE(p)}{n}\right) + 2p. \quad (\text{in OLS})$$

⑤ $$BIC = -2\log(L) + p\log(n)$$

$$= n\log\left(\frac{SSE(p)}{n}\right) + p\log(n).$$

* Variable Selection.

① All possible.

total : $2^p = \sum_{k=0}^{p} \binom{p}{k}$

② Forward Selection

total: $1 + \sum_{k=1}^{p} k = 1 + \frac{p(p+1)}{2}$.

1. M_0: null model. $y = \beta_0 + \varepsilon$
2. M_1 . R^2 largest.

   $\vdots$

   M_k.
3. stopping rule : $\left(\begin{array}{l} p^* \text{ 미리설정} \\ \text{partial } F < F^* \end{array} \right.$

   혹은 $M_1 \ldots M_k$ 中 C_p 가장 낮은 모델.

③ Backward Elimination

④ Stepwise.

Ch 12. Nonlinear Reg.

1. linearize.
2. $y = \frac{\theta_1 x}{x + \theta_2} + \varepsilon$. $\mapsto f(x, \theta)$

$$\frac{1}{y} = \frac{x + \theta_2}{\theta_1 x} + \frac{1}{\varepsilon} \rightarrow \text{안될 수도.}$$

< Nonlinear Reg. 에서의 추정 >.

$$y_i = f(x_i, \theta) + \varepsilon_i.$$

$$S(\theta) = \sum (y_i - f(x_i, \theta))^2 \rightarrow \text{difficult to solve.}$$

$\Rightarrow$ Gauss – Newton

$$f(x_i : \theta) \simeq f(x_i, \theta_0) + \sum_{j=1}^{p} \left(\frac{\partial f(x_i, \theta)}{\partial \theta_j} \right)_{\theta = \theta_0} (\theta_j - \theta_{j0}).$$

$$= f_i^0 + \sum_{j=1}^{p} z_{ij}^0 \beta_j$$

$$\Leftrightarrow \quad y_i - f_i^0 = \sum_{j=1}^{p} z_{ij}^0 \beta_j^0$$

$$y_0 = Z_0 \beta_0 + \varepsilon. \quad \underset{L-S.}{\Rightarrow} \quad \hat{\beta} = (Z_0^T Z_0)^{-1} Z_0^T y_0.$$

$$\hat{\theta} = \hat{\beta}_0 - \theta_0.$$

$\hookrightarrow$ $\hat{\theta}_1$ 이라 하자. ....

$$\therefore \hat{\theta}_{k+1} = \hat{\theta}_k + \hat{\beta}_k.$$

Ch 13. GLM.

$\theta = g(\mu) = \eta$. function of location param.

* Exponential family.

$$f(y:\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right)$$

θ: canonical param.

ϕ: dispersion param.

① Poisson. Dist'n

$$f(y,\mu) = \frac{\mu^y \cdot e^{-y}}{y!} = \exp(y\log\mu - \mu - \log(y!)).$$

$\log\mu = \theta$. $b(\theta) = \exp(\theta)$. $a(\phi) = 1$. $c(y,\phi) = -\log(y!)$.

② Binomial Dist'n.

$$f(y,p) = \binom{n}{y} p^y (1-p)^{n-y} = \exp\left(y\log\left(\frac{p}{1-p}\right) + n\log(1-p) + \log\binom{n}{y}\right).$$

$\log\frac{p}{1-p} = \theta$. , $b(\theta) = n\log(1+e^\theta)$, $a(\phi) = 1$. $c(y,\phi) = \log\binom{n}{y}$.

$\Leftrightarrow p = e^\theta(1-p)$.

$\Leftrightarrow p = \frac{e^\theta}{1+e^\theta}$.

③ Normal Dist'n

$$f(y,\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(y-\mu)^2} = \exp\left(\frac{y\mu - \frac{1}{2}\mu^2}{\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right)$$

$\mu = \theta$. $b(\theta) = \frac{1}{2}\theta^2$. $a(\phi) = \phi$. $c(y,\phi) = -\frac{1}{2}\left(\frac{y^2}{\sigma^2} + \log(2\pi\sigma^2)\right)$

$\sigma^2 = \phi$

< Exponential Family 의 성질 >.

① $E\left(\frac{\partial l}{\partial\theta}\right) = 0$. $(l(\theta,\phi,y) = \log f(y:\theta,\phi))$.

pf)

$$E\left(\frac{\partial \log f(y,\theta,\phi)}{\partial\theta}\right) = \int \frac{\partial \log f(y,\theta,\phi)}{\partial\theta} f(y;\theta,\phi)\, dy$$

$$= \int \frac{\partial f(y,\theta,\phi)}{\partial\theta} \cdot \frac{1}{\cancel{f(y,\theta,\phi)}} \cdot \cancel{f(y,\theta,\phi)}\, dy$$

$$= \frac{\partial}{\partial\theta}\int f(y,\theta,\phi)\, dy = \frac{\partial}{\partial\theta} 1 = 0.$$

② $E\left(\frac{\partial l^2}{\partial\theta^2}\right) + E\left(\left(\frac{\partial l}{\partial\theta}\right)^2\right) = 0$.

pf)

$$0 = \frac{\partial}{\partial \theta} \int \frac{\partial \log f(y,\theta,\phi)}{\partial \theta} \cdot f(y,\theta,\phi)\, dy$$

?

$$= \int \left[\frac{\partial \log^2 f(y,\theta,\phi)}{\partial \theta^2} f(y,\theta,\phi) + \frac{\partial \log f(y,\theta,\phi)}{\partial \theta} \cdot \frac{\partial f(y,\theta,\phi)}{\partial \theta} \right] dy$$

$$= \int \left[\qquad '' \qquad + \left(\frac{\partial \log f(y,\theta,\phi)}{\partial \theta} \right)^2 f(y,\theta,\phi) \right] dy$$

$$\Rightarrow E\left(\frac{\partial l^2}{\partial \theta^2}\right) + E\left(\left(\frac{\partial l}{\partial \theta}\right)^2\right) = 0.$$

$$l = \frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi).$$

① $E\left(\frac{\partial l}{\partial \theta}\right) = \frac{E(y - b'(\theta))}{a(\phi)} = 0. \Leftrightarrow E(y) = b'(\theta).$

② $E\left(\frac{\partial^2 l}{\partial^2 \theta}\right) + E\left(\left(\frac{\partial l}{\partial \theta}\right)^2\right) = \frac{-b''(\theta)}{a(\phi)} + E\left(\frac{y-\mu}{a(\phi)}\right)^2$

$$= -\frac{b''(\theta)}{a(\phi)} + \frac{Var(y)}{a^2(\phi)} = 0 \Leftrightarrow Var(y) = a(\phi) b''(\theta).$$

Poi

$b''(\theta) = \mu.$

$a(\phi) = 1.$

$Var(y) = \mu.$

Binomial

$b''(\theta) = n\frac{e^\theta}{(1+e^\theta)^2} = np(1-p)$

$a(\phi) = 1$

$Var(y) = np(1-p).$

Normal

$b''(\theta) = 1$

$a(\phi) = \sigma^2$

$Var(y) = \sigma^2.$

* GLM의 concept.

일반적 $y = f(x,\beta) + \varepsilon.$

$E(y) = f(x,\beta). \Rightarrow y = E(y) + \varepsilon.$

$\mu = E(y)$ depends on $\eta = x^T\beta,$

$\eta = g(\mu).$

g: link function.

* Logistic Reg.

response var. 이 binary인 linear 모델 (의 문제).

$$y_i = x_i^T \beta + e_i$$

$\rightarrow$ Assume $y_i \sim Ber(\pi_i)$

이 때

$$
\left|\begin{array}{l}
E(y_i) = \pi_i \text{ 이므로 } \varepsilon_i = 1-\pi_i \text{ or } -\pi_i . \\
Var(y_i) = E(y_i^2) - (E(y_i))^2 \\
\qquad\qquad = \pi_i - \pi_i^2 \neq \text{const.} \\
0 \le E(y_i) \le 1 \text{ but } 0 \le \hat{y}_i \le 1 \text{ 보장 X.}
\end{array}\right.
$$

Consider logistic reg. $\quad y = E(y) + \varepsilon.$

$$E(y_i) = \frac{\exp(x_i\beta)}{1+\exp(x_i\beta)} = \pi_i, \qquad \eta_i = x_i^T\beta.$$

$\pi_i = h(\eta_i)$라 ~~두면~~ $\quad \eta_i = g(\pi_i). \quad g = h^{-1} =$ link fn.

① Logit model.

$$\pi = h(\eta) = \frac{\exp(\eta)}{1+\exp(\eta)} = E(y).$$

$$\eta = g(\pi) = \log\frac{\pi}{1-\pi}$$

(1) 계수의 의미 $\left(\text{odds} := \frac{\pi(x)}{1-\pi(x)}\right)$

regressor 하나인 경우 $\quad g(\pi(x)) = \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_0 + \beta_1 x$

$$g(\pi(x+1)) = \ln\left(\frac{\pi(x+1)}{1-\pi(x+1)}\right) = \beta_0 + \beta_1(x+1).$$

$$\ln\left(\frac{\pi(x+1)}{1-\pi(x)}\right) - \ln\left(\frac{\pi(x)}{1-\pi(x)}\right) = \beta_1 .$$

$$\text{odds ratio} = e^{\beta_1}.$$

x_1 1 증가시
odds of y가 e^{β}배.

(2) Estimation. $(\hat{y} = \hat{E(y)})$.

Note that $y_i \overset{iid}{\sim}$ Bernoulli $\rightarrow f(y_1)\cdots f(y_n)$ $\begin{cases} \pi_i & \text{if } y_i = 1 \\ 1-\pi_i & \text{if } y_i = 0 . \end{cases}$

$$L(\beta) = \prod_{i=1}^{n} (\pi(x_i))^{y_i} (1-\pi(x_i))^{1-y_i}$$

$$l = \log L(\beta) = \sum_{i=1}^{n} y_i \log \pi_i + \sum_{i=1}^{n} (1-y_i)\log(1-\pi_i).$$

$$\frac{\partial l}{\partial \beta} = \frac{\partial l}{\partial \pi_i} \cdot \frac{\partial \pi_i}{\partial \beta}.$$

$\hookrightarrow \dfrac{\partial l}{\partial \pi_i} = \sum y_i \dfrac{1}{\pi_i} - \sum \dfrac{1-y_i}{1-\pi_i} = \sum \dfrac{y_i - \pi_i}{\pi_i(1-\pi_i)}$

$\hookrightarrow \dfrac{\partial \pi_i}{\partial \beta} = \left[\left(\dfrac{\exp(x_i^T\beta)}{1+\exp(x_i^T\beta)} \right) - \left(\;''\; \right)^2 \right] x_i$

$= \pi_i (1-\pi_i) x_i$.

$\therefore \dfrac{\partial l}{\partial \beta} = \sum (y_i - \pi_i) x_i$.

$\hookrightarrow$ Newton – Raphson 이용해 해 찾기.

② probit model.

$y = E(y) + \varepsilon$
$= \pi + \varepsilon$
$= \Phi(\eta) + \varepsilon$, $\rightarrow$ Φ: std Normal 의 cdf. 즉 link ftn $= \Phi^{-1}(\pi)$.

$y_i = \Phi(x_i^T\beta) + \varepsilon_i$

③ Complementary log-log.

$y_i = \pi_i + \varepsilon_i$.
$= 1 - \exp(-\exp(x_i^T\beta)) + \varepsilon_i$. $\rightarrow$ $g(\pi) = \eta$
$= \log(-\log(1-\pi_i))$.

* odds ratio vs. contingency table.

	$X=0$	$X=1$
$Y=0$	n_{00}	n_{01}
$Y=1$	n_{10}	n_{11}

$\Rightarrow$ odds ratio $= \dfrac{\;\frac{n_{11}}{n_{01}}\;}{\;\frac{n_{10}}{n_{00}}\;} = \dfrac{n_{11}\, n_{00}}{n_{01}\, n_{10}}$.

In logistic reg, $\pi = \dfrac{\exp(\beta_0 + \beta_1 X)}{1+\exp(\beta_0+\beta_1 X)}$

$\pi(X=0) = P(Y=1 \mid X=0) = \dfrac{e^{\beta_0}}{1+e^{\beta_0}}$

$\pi(X=1) = P(Y=1 \mid X=1) = \dfrac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}$

$\dfrac{\pi(X=0)}{1-\pi(X=0)} = e^{\beta_0} = \dfrac{P(Y=1 \mid X=0)}{P(Y=0 \mid X=0)} = \dfrac{n_{10}}{n_{00}}$.

$\dfrac{\pi(X=1)}{1-\pi(X=1)} = e^{\beta_0+\beta_1} = \dfrac{P(Y=1 \mid X=1)}{P(Y=0 \mid X=1)} = \dfrac{n_{11}}{n_{01}}$

$e^{\beta_1} = \dfrac{n_{11}\, n_{10}}{n_{10}\, n_{01}}$

$\Rightarrow$ odds ratio 일치.

* Poisson Reg. : canonical param. $\theta = \log \mu$. $\Leftrightarrow \mu = \exp\theta$.

$y = \mu + \varepsilon$. $= \exp(x^T \beta) + \varepsilon$.

β.

Ch 14. Time Series.

* First order Autoregressive Error

model: $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$

$\varepsilon_t = \rho \varepsilon_{t-1} + d_t$ = white noise.

($|\rho| < 1$. d_t are indep, $\sim N(0, \sigma^2)$)

$\varepsilon_{t-1} = \rho \varepsilon_{t-2} + d_{t-1}$.

$\varepsilon_{t-2} = \rho \varepsilon_{t-3} + d_{t-2}$

$\vdots$

$\Rightarrow \varepsilon_t = \rho(\rho \varepsilon_{t-2} + d_{t-1}) + d_t$.

$= \rho^2(\rho \varepsilon_{t-3} + d_{t-2})) + \rho d_{t-1} + d_t$.

$\vdots$

$= \sum_{u=0}^{\infty} \rho^u d_{t-u}$.

properties

① $E(\varepsilon_t) = 0$

② $Var(\varepsilon_t) = \sigma^2 \frac{1}{1-\rho^2}$

③ $Cov(\varepsilon_t, \varepsilon_{t-1}) = \sigma^2 \frac{\rho}{1-\rho^2}$ * $Cov(\varepsilon_t, \varepsilon_{t-n}) = \sigma^2 \frac{\rho^n}{1-\rho^2}$

④ $Corr(\varepsilon_t, \varepsilon_{t-1}) = \rho$ * $Corr(\varepsilon_t, \varepsilon_{t-n}) = \rho^n$.

pf) ① $E(\varepsilon_t) = E(\sum \rho^u d_{t-u}) = \sum \rho^u E(d_{t-u}) = 0$.

② $Var(\varepsilon_t) = Var(\sum \rho^u d_{t-u})$.

$= \sum \rho^{2u} Var(d_{t-u}) = \sigma^2 \sum \rho^{2u} = \sigma^2 \frac{1}{1-\rho^2}$, $|\rho| < 1$.

③ $Cov(\varepsilon_t, \varepsilon_{t-1}) = E(\varepsilon_t \varepsilon_{t-1}) - \cancel{E(\varepsilon_t)E(\varepsilon_{t-1})}^{\to 0}$.

$= E\left(\underbrace{(d_t + \rho d_{t-1} + \rho^2 d_{t-2} + \cdots)}_{d_t + \rho \varepsilon_{t-1}} \underbrace{(d_{t-1} + \rho d_{t-2} + \cdots)}_{\varepsilon_{t-1}}\right)$.

$= E\left(d_t (d_{t-1} + \cdots) + \rho (\underbrace{\quad}) ^2\right)$

$= E(\rho \varepsilon_{t-1}^2) = \rho \, Var(\varepsilon_{t-1}) = \sigma^2 \frac{\rho}{1-\rho^2}$.

④ $Corr(\varepsilon_t, \varepsilon_{t-1}) = \dfrac{Cov(\varepsilon_t, \varepsilon_{t-1})}{\sqrt{Var(\varepsilon_t)}\sqrt{Var(\varepsilon_{t-1})}}$

$= \dfrac{\cancel{\sigma^2} \frac{\rho}{\cancel{1-\rho^2}}}{\cancel{\sigma^2} \frac{1}{\cancel{1-\rho^2}}} = \rho$.

※ 추정: Iterative approach.

Let $\begin{cases} y_t' = y_t - \rho y_{t-1}. \\ x_t' = x_t - \rho x_{t-1}. \end{cases}$

$\begin{cases} y_t = \beta_0 + \beta_1 x_t + \underbrace{\varepsilon_t}_{\rho \varepsilon_{t-1} + d_t}. \\ \rho y_{t-1} = \rho\beta_0 + \rho\beta_1 x_{t-1} + \rho \varepsilon_{t-1} \end{cases}$

$\Rightarrow y_t' = \beta_0' + \beta_1' X_t' + d_t \overset{indep}{\sim} N(0, \sigma^2)$.

$= \varepsilon_t - \rho \varepsilon_{t-1}$.

but ρ를 모른다..

1. $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \to$ OLS로 e_t obtain. $\to \hat{\rho} = \dfrac{\sum e_{t-1} e_t}{\sum e_{t-1}^2}$

2. obtain $\begin{cases} y_t' = y_t - \hat{\rho} y_{t-1} \\ x_t' = x_t - \hat{\rho} x_{t-1}. \end{cases}$

3. OLS.

4. D-W.

5. 새로운 e_t로 update. (until H_0: $\rho = 0$ accepted).

First diff : $\rho = 1 \Rightarrow \begin{cases} \beta_0' = \beta_0(1-\rho). \\ y_t' = y_t - y_{t-1}. \\ x_t' = x_t - x_{t-1}. \end{cases}$