P1

$$\hat\beta \in \arg\min_\beta \frac{1}{n}\sum_{i=1}^n \rho(y_i - \langle X_i, \beta\rangle)$$

$$\begin{cases} \sigma^2 = \delta\, \mathbb{E}\, \Psi^2(\varepsilon + \sigma_x Z; b_x) \\ \frac{1}{\delta} = \mathbb{E}\, \Psi'(\varepsilon + \sigma_x Z; b_x) \end{cases}$$

Effective score $\Psi(z; b) = \frac{d}{dz} e_{bf}(z)$

AMP told us $\hat\beta - \beta_0 \simeq N(0, \sqrt{\delta}\sigma_x I_p)$

(Noureddine El Karoui 2012) 2018 print

leave one sample out

leave-one-out method WLOG $\beta_0 = 0$

$$r_i \triangleq y_i - \langle X_i, \hat\beta\rangle = \varepsilon_i - \langle X_i, \hat\beta\rangle$$

score func. $\psi \triangleq \rho'$

F.O.C. $\sum_{i=1}^n X_i \psi(r_i) = 0$ —— ①

$$\hat\beta_{(-k)} = \arg\min_\beta \sum_{i\neq k} \rho(y_i - \langle X_i, \beta\rangle)$$

leave the k-th sample

F.O.C. $\sum_{i\neq k} X_i \psi(r_{i;(-k)}) = 0$ —— ②

$r_{i;(-k)} \triangleq \varepsilon_i - \langle X_i, \hat\beta_{(-k)}\rangle$, $\hat\beta_{(-k)} \perp\!\!\!\perp X_k$

①-②

$$0 = X_k \psi(r_k) + \sum_{i\neq k} X_i [\psi(r_i) - \psi(r_{i;(-k)})]$$

$$0 = X_k \psi(r_k) + \sum_{i\neq k} X_i \psi'(r_{i;(-k)})(r_i - r_{i;(-k)}) + \cdots$$ by Taylor expansion

$$0 = X_k \psi(r_k) + \sum_{i\neq k} X_i \underbrace{\psi'(r_{i;(-k)})}_{\text{scalar}} X_i^T(\hat\beta_{(-k)} - \hat\beta)$$

$$0 = X_k \psi(r_k) + \underbrace{\left[\sum_{i\neq k} \psi'(r_{i;(-k)}) X_i X_i^T\right]}_{(p\times p)\ \triangleq\ S_k \in \mathbb{R}^{p\times p}\ \perp\!\!\!\perp X_k} (\hat\beta_{(-k)} - \hat\beta)$$

不包括(k)

$$\hat\beta_{(-k)} - \hat\beta = -S_k^{-1} X_k \psi(r_k)$$

$$r_k - r_{k;(-k)} = \langle X_k, \hat\beta_{(-k)} - \hat\beta\rangle$$

$$= -\langle X_k, S_k^{-1} X_k \psi(r_k)\rangle$$

$$= -\langle X_k, S_k^{-1} X_k\rangle \psi(r_k)$$

$$= -\frac{1}{n}\mathrm{tr}(S_k^{-1})\psi(r_k)$$

if $S_k \triangleq$ sample covariance

$\mathrm{tr}(S_k^{-1})$ use stilties transformation

claim: $\max_{k=1\ldots p} \left|\frac{1}{n}\mathrm{tr}(S_k^{-1}) - C_x\right| = 0$ w.h.p.

$$r_k - r_{k;(-k)} = -C_x \psi(r_k)$$

③

p2 leave one predictor out

$$x_i = \begin{pmatrix} X_{i,(p-1)} \\ X_{i,p} \end{pmatrix} \triangleq \begin{pmatrix} V_i \\ X_{i,p} \end{pmatrix} \begin{matrix} - \mathbb{R}^{(p-1)\times 1} \\ - \text{scalar} \end{matrix}$$

$$\hat\beta = \begin{pmatrix} \hat\beta_{(p-1)} \\ \hat\beta_p \end{pmatrix} \begin{matrix} \in \mathbb{R}^{p-1} \\ \in \mathbb{R} \end{matrix}$$

$$\beta^{(-p)} = \underset{\beta \in \mathbb{R}^{p-1}}{\arg\min} \sum_{i=1} \rho(Y_i - \langle V_i, \beta \rangle)$$

F.O.C. $\sum_{i=1}^{n} V_i \psi(r_{i,(-p)}) = 0_{p-1} \in \mathbb{R}^{p-1}$ ④(1)

note: $r_{i,(-p)} \triangleq \varepsilon_i - \langle V_i, \beta^{(-p)} \rangle$ --- ④

F.O.C. (origin) $\sum_{i=1}^{n} X_i \psi(r_i) = 0$ ④(2)

~~Find~~ ④(1) ~~β~~ expansion

$$0 = \sum_{i=1}^{n} X_{i,p} \psi(r_i) = \sum_{i=1}^{n} X_{i,p} \left[\psi(r_{i,(-p)}) + \psi'(r_{i,(-p)})(r_i - r_{i,(-p)}) + (\cdots) \right\}$$

$$\boxed{r_i - r_{i,(-p)}} = \varepsilon_i - \langle x_i, \hat\beta \rangle - (\varepsilon_i - \langle V_i, \hat\beta_{(-p)} \rangle)$$

$$r_i - r_{i,(-p)} = \cancel{\langle X_i, \beta \rangle} \ \langle V_i, \hat\beta_{(-p)} \rangle - \langle X_i, \hat\beta \rangle$$

$$\boxed{r_i - r_{i,(-p)} = \langle V_i, \hat\beta^{(-p)} - \hat\beta_{[1:p-1]} \rangle - X_{i,p} \hat\beta_p \quad (*)}$$

$$0 = \sum_{i=1}^{n} X_{i,p} \left\{ \psi(r_{i,(-p)}) + \psi'(r_{i,(-p)}) \left(\langle V_i, \underset{\text{leave-one-out est.}}{\hat\beta^{(-p)}} - \underset{\text{origin est. 去 p}}{\hat\beta_{[1:p-1]}} \rangle - X_{i,p} \hat\beta_p \right) \right\}$$

get ⑤ $$\hat\beta_p = \frac{\sum_{i=1}^{n} X_{i,p} \left\{ \psi(r_{i,(-p)}) + \psi'(r_{i,(-p)}) \left(\langle V_i, \hat\beta^{(-p)} - \hat\beta_{[1:p-1]} \rangle \right) \right\}}{\sum_{i=1}^{n} X_{i,p} \psi'(r_{i,(-p)}) X_{i,p}}$$

④(2) expansion

$$0 = \sum X_i \psi(r_i) = \begin{pmatrix} \sum_i V_i \psi(r_i) \\ \sum_i X_{i,p} \psi(r_i) \end{pmatrix}$$

$$0_{p-1} = \sum_i V_i \psi(r_i).$$

$$= \sum_i V_i \left\{ \underbrace{\psi(r_{i,(-p)})}_{\overset{\text{by ④(1)}}{=0}} + \psi'(r_{i,(-p)})(r_i - r_{i,(-p)}) \right\}$$

$$= \sum_i V_i \left\{ \psi'(r_{i,(-p)})(r_i - r_{i,(-p)}) \right\}$$

$$= \sum_{i=1}^{n} V_i \psi'(r_{i,(-p)}) \left(V_i^T (\hat\beta^{(-p)} - \hat\beta_{[1:p-1]}) - X_{i,p} \hat\beta_p \right) \quad \text{by } (*)$$

$$0_{p-1} = \underbrace{\left[\sum_{i=1}^{n} \psi'(r_{i,(-p)}) V_i V_i^T \right]}_{\triangleq S_{[-p]} \in \mathbb{R}^{(p-1)\times(p-1)}} (\hat\beta^{(-p)} - \hat\beta_{[1:p-1]}) - \hat\beta_p \underbrace{\sum_{i=1}^{n} \psi'(r_{i,(-p)}) V_i X_{i,p}}_{\triangleq U_{(-p)}}$$

⑥ $\Rightarrow$ $$\hat\beta^{[-p]} - \hat\beta_{[1:p-1]} = S_{[-p]}^{-1} U_{[-p]} \hat\beta_p$$

P3 from ⑤ and ⑥ (plug in ⑥ into ⑤)

$$\hat{\beta}_p = \frac{\sum_{i=1}^{n} X_{i,p} \{ \psi(r_{i,(-p)}) + \psi'(r_{i,(-p)}) V_i^T S_{[-p]}^{-1} U_{[-p]} \hat{\beta}_p \}}{\cdots}$$

$$\text{分子} = \Big\{ \sum_{i=1}^{n} X_{i,p} \psi(r_{i,(-p)}) + \cancel{\sum X_{i,p} V_i^T S} \; \underbrace{\sum X_{i,p} \psi'(r_{i,[-p]}) V_i^T}_{= U_{[-p]}^T} S_{[-p]}^{-1} U_{[-p]} \hat{\beta}_p \Big\}$$

$$\frac{(\quad) + U_{[-p]}^T S_{[-p]}^{-1} U_{[-p]} \hat{\beta}_p}{(\cdots)}$$

i.e.
$$\hat{\beta}_p = \frac{\sum_{i=1}^{n} X_{i,p} \psi(r_{i(-p)}) + U_{[-p]}^T S_{(-p)}^{-1} U_{(-p)} \hat{\beta}_p}{\sum_{i=1}^{n} X_{i,p}^2 \psi'(r_{i,(-p)})}$$

solve $\hat{\beta}_p \Rightarrow$
$$\boxed{\hat{\beta}_p = \frac{\sum_{i=1}^{n} X_{i,p} \psi(r_{i(-p)})}{\underbrace{\sum_{i=1}^{n} X_{i,p}^2 \psi'(r_{i(-p)}) - U_{(-p)}^T S_{(-p)}^{-1} U_{(-p)}}_{(*1)}}}$$

recall the def:
$$\begin{cases} S_{(-p)} = \sum_{i=1}^{n} \psi'(r_{i(-p)}) V_i V_i^T \in \mathbb{R}^{(p-1)\times(p-1)} \\ U_{(-p)} = \sum^{n} \psi'(r_{i(-p)}) V_i X_{i,p} \in \mathbb{R}^{p-1} \end{cases}$$

$$D \triangleq \mathrm{diag}\left(\{\psi'(r_{i[-p]})\}_{i\in[n]}\right) \in \mathbb{R}^{n\times n}$$

$$X_{(p)} = (X_{1,p} \cdots X_{k,p}) \in \mathbb{R}^n \qquad V = \begin{bmatrix} V_1^T \\ \vdots \\ V_n^T \end{bmatrix} \in \mathbb{R}^{n\times(p-1)}$$

rewrite
$$S_{(-p)} = V^T D V \in \mathbb{R}^{(p-1)\times(p-1)}$$
$$U_{(-p)} = V^T D X_{(p)} \in \mathbb{R}^{p-1}$$

rewrite 二次型
$$U_{(-p)}^T S_{(-p)} U_{(-p)} = \langle X_{(p)}, D V (V^T D V)^{-1} V^T D X_{(p)} \rangle$$

projection matrix
$$P_V = D^{\frac{1}{2}} V (V^T D^{\frac{1}{2}} D^{\frac{1}{2}} V)^{-1} V^T D^{\frac{1}{2}}$$

$$\text{分母}(*1) = \langle X_{[p]}, (D - D^{\frac{1}{2}} P_V D^{\frac{1}{2}}) X_{[p]} \rangle \qquad X_{[p]} \perp\!\!\!\perp (D, V)$$

i.e. $X_{[p]} \perp D - D^{\frac{1}{2}} P_V D^{\frac{1}{2}}$

$$\approx \frac{1}{n} \mathrm{Tr}(D - D^{\frac{1}{2}} P_V D^{\frac{1}{2}})$$
$$= \frac{1}{n} \mathrm{Tr}(D - D P_V)$$
$$= \frac{1}{n} \mathrm{Tr}(D(I - P_V))$$
$$= \frac{1}{n} \sum_i \psi'(r_{i[-p]})(1 - P_{ii})$$

$$P_{ii} = \psi'(r_{i(-p)}) V_i^T S_{(-p)}^{-1} V_i$$
$$= \psi'(r_{i(-p)}) V_i^T \Big(\underbrace{\sum_{k\neq i} \psi'(r_{k(-p)}) V_k V_k^T}_{S_{[-p],-i}} + \psi'(r_{i(-p)}) V_i V_i^T \Big)^{-1} V_i$$

by woodbury identity
$$(S + uv^T)^{-1} = S^{-1} - \frac{S^{-1} u v^T S^{-1}}{1 + v^T S^{-1} u}$$

$$= \psi'(r_{i(-p)}) V_i^T \Big\{ S_{[-p],-i}^{-1} - \frac{S_{[-p],-i}^{-1} V_i V_i^T S_{[-p],-i}^{-1} \psi'(r_{i(-p)})}{1 + \psi'(r_{i(-p)}) V_i^T S_{[-p],-i}^{-1} V_i} \Big\} V_i$$

(P4)

$$= \psi'(r_{i(-p)}) V_i^T S_{[-p],-i}^{+} V_i - \frac{\left(\psi'(r_{i(-p)}) V_i^T S_{[-p]-i}^{+} V_i\right)^2}{1+\psi'(r_{i(p)}) V_i^T S_{[p]-i}^{+} V_i}$$

$$= \square - \frac{\square^2}{1+\square}$$

$$= \frac{\square}{1+\square}$$

$$= 1 - \frac{1}{1+\square}$$

by $V_i \perp\!\!\!\perp S^{+}_{[-p]-i}$

$$= 1 - \frac{1}{1+\psi'(r_{i(-p)}) \frac{1}{n}\mathrm{Tr}(S^{+}_{[-p]-i})}$$

$$1-P_{v,ii} = 1 - \frac{1}{1+\psi'(r_{i(-p)})\, C_*}$$

note : $tr(P_{v,ii}) = rank(P_{v,ii}) = p-1$

$\sum_{i=1}^{n}$ (i.e. take trace of both side)

$$\Rightarrow \boxed{\frac{1}{n}\sum_i \frac{1}{1+\psi'(\cdot)C_*} = 1 - \frac{p}{n}} \quad (7)$$

$$\text{解}(*1) = \frac{1}{n}\frac{1}{C_*}\sum_i \frac{C_*\psi'(r_{i(-p)})}{1+\psi'(r_{i(-p)})C_*}$$

$$= \frac{1}{C_* n}\sum_i \left(1 - \frac{1}{1+\psi'(r_{i(-p)})C_*}\right)$$

$$= \frac{1}{C_* n} \frac{p}{n} \times n = \frac{1}{C_*\delta} \quad \left(\delta \triangleq \frac{n}{p}\right)$$

thus $\beta_p = C_* \delta \sum_{i=1}^{n} X_{i,p}\, \psi(r_{i(-p)})$ ⫫ 独立

$$\mathbb{E}\beta_p^2 = C_*^2 \delta^2 \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\psi^2(r_{i(-p)})$$

$$\approx C_*^2 \delta^2 \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\psi^2(r_i)$$

$\times \frac{p}{n}$

$$\boxed{\frac{1}{n}\mathbb{E}\|\beta\|^2 = C_*^2 \delta \cdot \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\psi^2(r_i)}$$

recall $r_{k(-k)} = r_k + C_*\psi(r_k)$

$\Leftrightarrow r_k = prox_{C_*\rho}(r_{k;(-k)})$ by $prox_\rho(z) = x \Leftrightarrow x + \rho'(x) = z$

$\Rightarrow C_*\psi(r_k) = r_{k;(-k)} - prox_{C_*\rho}(r_{k(-k)})$ $\quad \psi = \rho'$

recall def. $r_{k(-k)} = \varepsilon_k - \langle X_k, \beta_{(-k)} \rangle$ ⫫

$\overset{d}{=} \varepsilon_k + \frac{\|\beta_{(-k)}\|}{\sqrt{n}} Z_k \quad Z_k \sim N(0,1)$

$$\sigma_*^2 \approx \frac{1}{n}\mathbb{E}\|\beta\|^2 = \cancel{C_*^2}\delta \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\cancel{C_*^2}} \mathbb{E}\Psi^2(\varepsilon_k + \sigma_* Z_k; C_*)$$

$$\Psi(\cdot) = \left[r_{k;(-k)} - prox_{C_*\rho}(r_{k(-k)}) \right]$$

remind $\hat\beta - \beta \sim N(0, \sqrt{\delta}\,\sigma_* I_p)$

$$\boxed{\sigma_*^2 = \delta\, \mathbb{E}\Psi^2(\varepsilon + \sigma_* Z; C_*)}$$

equ 1 fixed point

~~recall~~ rewrite (7)

$$\frac{1}{1+\psi'(r_{i(-p)})C_*} \text{ this term}$$

p5

note: $\text{prox}_\rho(z) + \rho'(\text{prox}_\rho(z)) = z$

$$\text{prox}_\rho'(z) + \rho''(\text{prox}_\rho(z))\, z + \rho'(\text{prox}_\rho(z))\, \text{prox}_\rho'(x) = 0$$

$$\Leftrightarrow \quad (1 + \rho''(\text{prox}_\rho(z))\, \text{prox}_\rho'(z) = 1$$

$$\Rightarrow \quad \text{prox}_f'(z) = \frac{1}{1+\rho''(\text{prox}_\rho(z))} = \frac{1}{1+\psi'(r_{i(-i)})C_x}$$

rewrite 7

$$1 - \frac{1}{\delta} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\, \frac{1}{1 + C_x \psi'(r_{k(-k)})}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\, \text{prox}_{C_x\rho}'(r_{k(-k)})$$

$$\overset{d}{\approx} \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\, \text{prox}_{C_x\rho}'(\varepsilon_k + \sigma_x Z_k)$$

$$\boxed{1 - \frac{1}{\delta} = \mathbb{E}\, \Phi'(\varepsilon + r_x Z;\, C_x)}$$

eq42 fixed point

<u>remark</u> rigorous version is worked out with Ridge penalty

for other ~~~~ without strong convexity e.g. robust regression, this method can be ~~probb~~ problematic!