Binary response model $\quad Prob(y_i=1|x_i)=F(x_i^T\theta)$

$\quad Prob(y_i=0|x_i)=1-F(x_i^T\theta)$

$\begin{cases} \text{Probit: normal } F \\ \text{Logit: } F(v)=\frac{\exp\{v\}}{1+\exp\{v\}} \end{cases}$

$\Leftrightarrow$ Latent variable model $\quad \begin{cases} y_i^* = x_i^T\theta+\varepsilon_i, \quad \varepsilon_i|x_i \sim F \\ y_i = 1\{y_i^*>0\} \end{cases}$

Estimation: MLE

$$\hat\theta_{MLE} = \underset{\theta\in\Theta}{\operatorname{argmax}} \frac{1}{n}\sum_{i=1}^n \ln f(y_i|x_i;\theta)$$
$$= \underset{\theta\in\Theta}{\operatorname{argmax}} \frac{1}{n}\sum_{i=1}^n \{1\{y_i=1\}\ln F(x_i^T\theta)+1\{y_i=0\}\ln(1-F(x_i^T\theta))\}$$

FOC: $\frac{1}{n}\sum_{i=1}^n \frac{\partial \ln f(y_i|x_i;\theta)}{\partial\theta}=0$

$$\frac{1}{n}\sum_{i=1}^n S(y_i|x_i;\theta_0)+H(y_i|x_i;\theta_0)(\hat\theta-\theta_0)=0$$
$$\sqrt{n}(\hat\theta-\theta_0) = -H(y_i|x_i;\theta_0)^{-1}\frac{\sqrt{n}}{n}\sum_{i=1}^n S(y_i|x_i;\theta_0)$$
$$\xrightarrow{d} EH(y_i|x_i;\theta_0)^{-1}N(0,E(S(y_i|x_i;\theta_0)S^T(y_i|x_i;\theta_0)))$$
$$= N(0, E(S(y_i|x_i;\theta_0)S^T(y_i|x_i;\theta_0))^{-1})$$

$$S(y_i|x_i;\theta_0)=\frac{\partial \ln f(y_i|x_i;\theta_0)}{\partial\theta}$$
$$=1\{y_i=1\}\frac{f(x_i^T\theta_0)}{F(x_i^T\theta_0)}x_i+1\{y_i=0\}\frac{-f(x_i^T\theta_0)}{1-F(x_i^T\theta_0)}x_i$$
$$E(SS^T|x_i)=x_ix_i^Tf^2(x_i^T\theta_0)[F(x_i^T\theta_0)(1-F(x_i^T\theta_0))]^{-1}$$
$$\hat V_\theta^0=[\frac{1}{n}\sum_{i=1}^n x_ix_i^Tf^2(x_i^T\hat\theta)[F(x_i^T\hat\theta)(1-F(x_i^T\hat\theta))]^{-1}]^{-1}$$

Interpretation:

Partial effect: $PE_j=\frac{\partial Pr(y_i=1|x_i)}{\partial x_{ij}}=f(x_i^T\theta)\,\theta_j$

Relative partial effect: $PE_j/PE_k=\theta_j/\theta_k$

Partial effect at average: $PEA_j=f(\frac{1}{n}\sum_{i=1}^n x_i^T\hat\theta)\,\hat\theta_j$

Average partial effect: $APE_j=\frac{1}{n}\sum_{i=1}^n f(x_i^T\hat\theta)\,\hat\theta_j$

Binary x_{ij}: $PE_j=F(x_{i1}\theta_1+\cdots+\theta_j+\cdots)-F(x_{i1}\theta_1+\cdots+0+\cdots)$

H_0: x_i is exogenous $\quad H_A$: x_i is endogenous with IV z_i

H_A: $y_i^*=x_i^T\theta+u_i \qquad \begin{bmatrix}u_i\\v_i\end{bmatrix}\sim N(0,\begin{bmatrix}\sigma_u^2 & \rho\sigma_u\sigma_v\\ \rho\sigma_u\sigma_v & \sigma_v^2\end{bmatrix})$

$\quad x_i=\pi z_i+v_i$

$u_i|v_i\sim N(\rho\frac{\sigma_u}{\sigma_v}v_i,\ \sigma_u^2(1-\rho^2))$

Control function: $e_i=u_i-\rho\frac{\sigma_u}{\sigma_v}v_i \qquad \begin{cases}E(e_i|v_i)=0\\ E(e_i^2)=\sigma_u^2(1-\rho^2)\end{cases}$

$y_i^*=x_i^T\theta+\rho\frac{\sigma_u}{\sigma_v}v_i+e_i$. $E(x_ie_i)=E(v_ie_i)=0$

① 2S: OLS: $x_i=\pi z_i+v_i \Rightarrow \hat\pi, \hat v_i$

Probit: $P(Y_i=1)=P(e_i>-x_i^T\theta-\rho\frac{\sigma_u}{\sigma_v}\hat v_i)$

$$=\Phi(\frac{x_i^T\theta+\rho\cancel{\frac{\sigma_u}{\sigma_v}}\hat v_i}{\sigma_u\sqrt{1-\rho^2}})$$

$\hat\theta_p=\frac{\theta}{\sigma_u\sqrt{1-\rho^2}}, \quad \hat\beta_p=\frac{\rho}{\sigma_v\sqrt{1-\rho^2}} \Rightarrow \sqrt{1-\rho^2}=[1+\hat\beta_p^2\sigma_v^2]^{-1}$

$\Rightarrow \hat\theta=\sigma_u\hat\theta_p[1+\hat\beta_p^2\sigma_v^2]^{-1}$

② MLE: $(\hat\theta,\hat\pi)=\underset{\theta,\pi}{\operatorname{argmax}}\frac{1}{n}\sum_{i=1}^n \ln f(y_i,x_i|z_i)$

$f(y_i,x_i|z_i)=f(y_i|x_i,z_i)f(x_i|z_i)$

$=f(y_i|x_i,v_i)f(x_i|z_i)$

$f(y_i|x_i,v_i)=Prob(x_i^T\theta+\rho\frac{\sigma_u}{\sigma_v}v_i+e_i>0)^{1\{y_i=1\}}P(\leq 0)^{1\{\ \}}$ 3

$f(x_i|z_i)=\frac{1}{\sigma_v}\phi(\frac{x_i-\pi z_i}{\sigma_v})$

H_0: $E(x_iu_i)=0 \Leftrightarrow E(u_iv_i)=0 \Leftrightarrow \rho=0 \Leftrightarrow \beta_p=0$

2nd stage: Probit $\quad Y_i\sim x_i^T\theta+\beta_p\hat v_i+e_i$

Multiple response model $\quad Prob(y_i=n)=\frac{F(x_i^T\theta_n)}{\sum_n F(x_i^T\theta_n)}$

Logit: $Pr(y_i=n|x_i)=\frac{\exp\{x_i^T\theta_n\}}{\sum_n\exp\{x_i^T\theta_n\}}$

Normalization: $\theta_0=0$. $Pr(y_i=n|x_i)=\frac{\exp\{x_i^T\theta_n\}}{1+\sum_n\exp\{x_i^T\theta_n\}}$

Estimation MLE

$$\ln L=\sum_{i=1}^n\sum_{m=1}^M 1\{y_i=m\}\ln Pr(y_i=m|x_i)$$

Interpretation

Log-odd ratio: $\ln(\frac{Pr(y_i=m|x_i)}{Pr(y_i=h|x_i)})=x_i^T(\theta_m-\theta_h)$

Conditional choice: $Pr(y_i=m|y_i=m \text{ or } h, x_i)$

$$=\frac{\exp\{x_i^T(\theta_m-\theta_h)\}}{1+\exp\{x_i^T(\theta_m-\theta_h)\}}=F(x_i^T(\theta_m-\theta_h))$$

Partial effect: $\frac{\partial Pr(y_i=m|x_i)}{\partial x_{ij}}=Pr(m)\{\theta_{mj}-\sum_h\theta_{hj}Pr(h)\}$

Latent utility model: $u_{im}^*=x_i^T\theta_m+\varepsilon_{im}$: $F(\varepsilon)=\exp\{-e^{-\varepsilon}\}$

$y_i=m$ if $u_{im}^*\geq u_{ih}^*\ \forall h$.

$$Pr(y_i=m|x_i)=Pr(x_i^T\theta_m+\varepsilon_{im}\geq x_i^T\theta_h+\varepsilon_{ih},\forall h\neq m|x_i)$$
$$=\int\prod_{h\neq m}\exp\{-e^{-\varepsilon_{im}}e^{x_i^T(\theta_h-\theta_m)}\}d\exp\{-e^{-\varepsilon_{im}}\}$$
$$=\int_0^1 x^\wedge\{\sum_{h\neq m}e^{x_i^T(\theta_h-\theta_m)}\}dx=\frac{\exp(x_i^T\theta_m)}{\sum_h\exp(x_i^T\theta_h)}$$

Conditional Logit model: $u_{im}^*=x_{im}^T\gamma+\varepsilon_{im}$

$\Downarrow$ Combination $\qquad y_i=m$ if $u_{im}^*\geq u_{ih}^*\ \forall h$

Latent utility: $u_{im}^*=w_i^T\theta_m+x_{im}^T\gamma+\varepsilon_{im}$

$y_i=m$ if $u_{im}^*\geq u_{ih}^*\ \forall h$

IIA: independence of irrelevant alternatives

$\frac{Pr(y_i=m|\cdot\cdot)}{Pr(y_i=h|\cdot\cdot)}=\exp\{w_i^T(\theta_m-\theta_h)+(x_{im}-x_{ih})'\gamma\}$

Other choice k does not impact the ratio.

Trancated Regression $\quad y_i^*=x_i^T\beta+\varepsilon_i, \quad \varepsilon_i|x_i\sim N(0,\sigma^2)$

$y_i=y_i^*1\{y_i^*>c\}$

$y\sim N(\mu,\sigma^2) \quad E(y|y>c)=\mu+\sigma\frac{\phi(v)}{1-\Phi(v)} \quad v=\frac{c-\mu}{\sigma}$

$Var(y|y>c)=\sigma^2\{1+\frac{v\phi(v)}{1-\Phi(v)}-[\frac{\phi(v)}{1-\Phi(v)}]^2\}$

OLS X: $E(y_i|y_i>c)=x_i^T\beta+\sigma\frac{\phi(v)}{1-\Phi(v)}, \quad v=\frac{c-x_i\beta}{\sigma}$

$y_i=x_i^T\beta+\underline{\sigma\frac{\phi(v)}{1-\Phi(v)}+\eta_i}.\ \varepsilon_i. \quad E(x_i\varepsilon_i)\neq 0$

MLE V: $f(y_i|y_i>c,x_i)=\frac{f(y_i|x_i)}{1-F(c|x_i)}=\frac{\phi(\frac{y_i-x_i^T\beta}{\sigma})}{1-\Phi(\frac{c-x_i^T\beta}{\sigma})}$

Censored Regression $\quad y_i^*=x_i^T\beta+\varepsilon_i. \quad \varepsilon_i|x_i\sim N(0,\sigma^2)$

$y_i=(y_i^*-c)1\{y_i^*>c\}+c$

MLE: $f(y_i|x_i)=\{\phi(\frac{y_i-x_i^T\beta}{\sigma})\}^{1(y_i>c)}\{\Phi(\frac{c-x_i^T\beta}{\sigma})\}^{1(y_i=c)}$

Sample Selection

$y_i = x_i^T\beta + u_i$ $\quad \begin{bmatrix} u_i \\ v_i \end{bmatrix} \Big| x_i \sim N\left(0, \begin{bmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{bmatrix}\right)$

$T_i = 1\{z_i^T\gamma + v_i > 0\}$

y_i is observed iff. $T_i = 1$

$\mathrm{Corr}(u_i, v_i) = \rho \neq 0 \Rightarrow E(u_i x_i \mid T_i = 1) \neq 0$ Endogeneity

OLS: $E(y_i \mid x_i, T_i = 1) = x_i^T\beta + E(u_i \mid x_i, v_i > -z_i^T\gamma)$

$= x_i^T\beta + \rho\sigma \dfrac{\phi(z_i^T\gamma)}{\Phi(z_i^T\gamma)}$

Heckit 2 stage:

① Probit: $T_i = 1\{z_i^T\gamma + v_i > 0\} \Rightarrow \hat{\gamma} \Rightarrow \dfrac{\phi(z_i^T\hat{\gamma})}{\Phi(z_i^T\hat{\gamma})}$

② OLS: $y_i = x_i^T\beta + \rho\sigma \dfrac{\phi(z_i^T\hat{\gamma})}{\Phi(z_i^T\hat{\gamma})} + \varepsilon_i \Rightarrow \hat{\beta}$

GMM: $y_i = x_i^T\beta + \rho\sigma \dfrac{\phi(z_i^T\gamma)}{\Phi(z_i^T\gamma)} + \varepsilon_i$, (β, γ)

$E(\varepsilon_i \mid x_i, z_i, T_i = 1) = 0 \Rightarrow$ moments

MLE: $(\hat{\beta}, \hat{\gamma}) = \underset{(\beta,\gamma)}{\arg\max} \sum_{i=1}^{n_0} \ln P(T_i = 0 \mid x_i, z_i)$

$+ \sum_{i=n_0+1}^{n} \ln f(y_i, T_i = 1 \mid x_i, z_i)$

$P(T_i = 0 \mid x_i, z_i) = P(z_i^T\gamma + v_i \leq 0 \mid z_i) = \Phi(-z_i^T\gamma)$

$f(y_i, T_i = 1 \mid x_i, z_i) = f(y_i \mid x_i, z_i)\, P(T_i = 1 \mid y_i, x_i, z_i)$

$= f(x_i^T\beta + u_i)\, P(z_i^T\gamma + v_i > 0 \mid u_i = y_i - x_i^T\beta)$

$= \dfrac{1}{\sigma}\phi\left(\dfrac{y_i - x_i^T\beta}{\sigma}\right) \Phi\left(\dfrac{z_i^T\gamma\sigma + \rho(y_i - x_i^T\beta)}{\sqrt{1-\rho^2}}\right)$

Resampling Method

1. Jackknife: distribution from n leave-one-out estimators

$$\hat V_{\hat\theta}^{jack} = \frac{n-1}{n}\sum_{i=1}^n (\hat\theta_{-i}-\bar\theta)(\hat\theta_{-i}-\bar\theta)^T \qquad \bar\theta = \frac1n\sum_{i=1}^n \hat\theta_{-i}$$

2. Bootstrap: distribution from iid sampling with replacement

Prob(Observation i in Bootstrap) $= 1-(1-\frac1n)^n \to 1-e^{-1}=0.6$

Bootstrap estimation: $\hat\theta^* = \{\hat\theta^{*(b)}: b=1,\dots,B\}$

$$\hat V_{\hat\theta}^{boot} = \frac{1}{B-1}\sum_{b=1}^B(\hat\theta^*(b)-\bar\theta^*)(\hat\theta^*(b)-\bar\theta^*), \quad \bar\theta^* = \frac1B\sum_{b=1}^B\hat\theta^*(b)$$

Percentile interval: $C^{PI} = [q^*_{\frac\alpha2}, q^*_{1-\frac\alpha2}]$

Bootstrap quantile: $\frac1B\sum_{b=1}^B 1\{\hat\theta^*(b)\le q^*_{\frac\alpha2}\} = \frac\alpha2$

Transformation-respecting: $C^{PI}(m(\theta)) = m(C^{PI}(\theta))$

Bootstrap distribution.

Real: $Y\sim F$, $G_n(u,F) = Prob(\hat\theta\le u|F)$

Boot: $Y^*\sim F_n$, $G_n^*(u,F_n) = Prob(\hat\theta^*\le u|F_n)$

$$E^*(Y^*) = \sum_{i=1}^n Y_i P^*(Y^*=Y_i) = \frac1n\sum_{i=1}^n Y_i = \bar Y$$

$$Var^*(Y^*) = \frac1n\sum_{i=1}^n(Y_i-\bar Y)(Y_i-\bar Y)^T = \frac1n\sum_{i=1}^n Y_iY_i^T - \bar Y\bar Y^T = \hat\Sigma$$

$\bar Y^* = \frac1n\sum_{i=1}^n Y_i^*$ $\quad E^*(\bar Y^*) = \frac1n\sum_{i=1}^n E^*(Y_i^*) = \bar Y$

$$Var^*(\bar Y^*) = \frac{1}{n^2}\sum_{i=1}^n Var^*(Y_i^*) = \frac1n\hat\Sigma$$

iid $\{Y_i\}_{i=1}^n$, $\mu = E[h(Y)]$, $\theta=g(\mu)$ $\quad E\|h(Y)\|^2<+\infty$

$\hat\mu = \frac1n\sum_{i=1}^n h(Y_i)$, $\hat\theta = g(\hat\mu)$

$\hat\mu^* = \frac1n\sum_{i=1}^n h(Y_i^*)$ $\quad \hat\theta^* = g(\hat\mu^*)$

CLT: $\sqrt n(\hat\mu^*-\hat\mu)\xrightarrow{d^*} N(0,V)$, $V = E(h(Y)-\mu)(h(Y)-\mu)^T$

$$E^*\hat\mu^* = \frac1n\sum_{i=1}^n E^*h(Y_i^*) = E^*h(Y_i^*) = \frac1n\sum_{i=1}^n h(Y_i) = \hat\mu$$

$$Var^*\hat\mu^* = \frac1{n^2}\sum_{i=1}^n Var^*(h(Y_i^*)) = \frac1n E^*(h(Y_i^*)-\hat\mu)^2 = \frac1n\frac1n\sum_{i=1}^n(h(Y_i)-\hat\mu)^2$$

Delta: $\sqrt n(\hat\theta^*-\hat\theta)\xrightarrow{d^*} N(0, V_\theta = GVG^T)$, $G = \frac{\partial g}{\partial\mu^T}(\mu)$

$\hat V_\theta \xrightarrow{P^*} V_\theta$: $\hat V_\theta = G(\hat\mu^*)\hat V^* G(\hat\mu^*)^T$, $\hat V^* = \frac1n\sum_{i=1}^n[h(Y_i^*)-\hat\mu^*]^2$

Consistency of Bootstrap estimate of Variance

$\sqrt n(\hat\theta^*-\hat\theta)\xrightarrow{d^*}N(0,V_\theta=GVG^T)$

$\sqrt n(\hat\theta-\theta)\xrightarrow{d}N(0,V_\theta=GVG^T)$

$$\hat V_\theta^{boot,B} := \frac1{B-1}\sum_{b=1}^B(\hat\theta^{*b}-\bar{\hat\theta}^*)(\hat\theta^{*b}-\bar\theta^{*b})^T, \quad \bar\theta^* = \frac1B\sum_{b=1}^B\hat\theta^{*b}$$

$\xrightarrow{P^*} V_\theta$

Trimmed Estimator of Bootstrap Variance

$\hat\theta^{**(b)} = \hat\theta^{*(b)}1\{\|\hat\theta^{*b}\|\le\tau_n\}$ $\quad \tau_n$: threshold

$$\hat V_{boot}(\hat\theta^*) = \frac1{B-1}\sum_{b=1}^B(\hat\theta^{**b}-\bar\theta^{**})(\hat\theta^{**b}-\bar\theta^{**})^T$$

Bias Correction: $\tilde\theta = \hat\theta - \widehat{Bias}$, $\hat\theta = g(\bar Y_n)$, $\theta=g(\mu)$

$Bias = E\hat\theta-\theta$

$= Eg'(\mu)(\bar Y_n-\mu) + \frac12 g''(\mu)(\bar Y_n-\mu)^2 + o(\frac1n)$

$= \frac12 g''(\mu)\frac1n\sigma^2 + o(\frac1n)$

Consistency of the Percentile Interval $C^{PI} = [q^*_{\frac\alpha2}, q^*_{1-\frac\alpha2}]$

If $a_n(\hat\theta-\theta)\xrightarrow{d}\zeta$ then $P(\theta\in C^{PI})\to1-\alpha$ $(n\to+\infty)$

$a_n(\hat\theta^*-\hat\theta)\xrightarrow{d^*}\zeta$ (ζ: pivotal & symmetric)

ζ cont. & sym 0

$$\begin{cases} q_\alpha = T_\alpha S(\hat\theta)+\theta \\ q^*_\alpha = T^*_\alpha S(\hat\theta)+\hat\theta \end{cases} \Rightarrow \frac{q^*_\alpha-\hat\theta}{S(\hat\theta)} - z_\alpha = O(n^{-\frac12})$$

$$P(\theta\in PI) = P\left(\frac{q^*_{\frac\alpha2}-\hat\theta}{S(\hat\theta)} \le \frac{\hat\theta-\theta}{S(\hat\theta)} \le \frac{q^*_{1-\frac\alpha2}-\hat\theta}{S(\hat\theta)}\right)$$

$$= \Phi\left(\frac{q^*_{1-\frac\alpha2}-\hat\theta}{S(\hat\theta)}\right) - \Phi\left(\frac{q^*_{\frac\alpha2}-\hat\theta}{S(\hat\theta)}\right) + O(n^{-\frac12})$$

$$= \Phi(z_{1-\frac\alpha2}) - \Phi(z_{\frac\alpha2}) + \phi(z_{1-\frac\alpha2})\left[\frac{q^*_{1-\frac\alpha2}-\hat\theta}{S(\hat\theta)} - z_{1-\frac\alpha2}\right] + \cdots$$

$$= 1-\alpha + O(n^{-\frac12})$$

Bias-Corrected Percentile Interval $\quad$ unknown: $\begin{cases}\varphi: \varphi'>0 \\ z_0: \text{Const.}\end{cases}$

Suppose $Z = \varphi(\hat\theta)-\varphi(\theta)+z_0 \sim N(0,1)$. $\Phi(\cdot)$

Let $\hat p^* = \frac1B\sum_{b=1}^B 1\{\hat\theta^*_b\le\hat\theta\}$, $z_0 = \Phi^{-1}(\hat p^*)$

$x(\alpha) = \Phi(z_\alpha+2z_0)$, $\quad z_\alpha = \Phi^{-1}(\alpha)$

BC: $C^{bc} = [q^*_{x(\frac\alpha2)}, q^*_{x(1-\frac\alpha2)}]$

$P(\theta\in C^{bc}) = 1-\alpha$

Pivotal: $\begin{cases} P(\varphi(\hat\theta)-\varphi(\theta)+z_0\le x) = \Phi(x) \\ P^*(\varphi(\hat\theta^*)-\varphi(\hat\theta)+z_0\le x) = \Phi(x) \Rightarrow z_0 = \Phi^{-1}(P^*(\hat\theta^*\le\hat\theta)) \end{cases}$

$x(\alpha) = P^*(\varphi(\hat\theta^*)-\varphi(\hat\theta)+z_0\le z_\alpha+2z_0)$

$= P^*(\hat\theta^*\le\varphi^{-1}[z_\alpha+z_0+\varphi(\hat\theta)])$

$\Rightarrow q^*_{x(\alpha)} = \varphi^{-1}[z_\alpha+z_0+\varphi(\hat\theta)]$

$P(\theta\in C^{bc}) = P(z_{\frac\alpha2}+z_0+\varphi(\hat\theta)\le\varphi(\theta)\le z_{1-\frac\alpha2}+z_0+\varphi(\hat\theta))$

Percentile-t Interval

$T = \frac{\hat\theta-\theta}{s(\hat\theta)}\xrightarrow{d}\zeta$

$T^* = \frac{\hat\theta^*-\hat\theta}{s(\hat\theta^*)}\xrightarrow{d^*}\zeta \Rightarrow \{T^{*b}\}_{b=1}^B \Rightarrow q^*_{1-\frac\alpha2}, q^*_{\frac\alpha2}$ $\quad (q^*_\alpha\xrightarrow{P}q_\alpha)$

$$C^{pt} = [\hat\theta - s(\hat\theta)q^*_{1-\frac\alpha2}, \hat\theta - s(\hat\theta)q^*_{\frac\alpha2}] = [\hat\theta+\hat s q^*_{\frac\alpha2}, \hat\theta+\hat s q^*_{1-\frac\alpha2}]$$

$$P(\theta\in C^{pt}) = P\left(q^*_{\frac\alpha2}\le\frac{\hat\theta-\theta}{s(\hat\theta)}\le q^*_{1-\frac\alpha2}\right)$$

$$= 1-\alpha+\phi(\cdot)(q^*_{\frac\alpha2}-q_{\frac\alpha2}) + \cdots + O(\tfrac1n)$$

$$= 1-\alpha+O(\tfrac1n)$$

Edgeworth expansion on $T(\theta) = \frac{\hat\theta-\theta}{s(\hat\theta)}\xrightarrow{d}N(0,1)$

$$P(T(\theta)\le x) = \Phi(x)+n^{-\frac12}\underbrace{p_1(x)}_{\text{even polynomial}}\phi(x)+n^{-1}\underbrace{p_2(x)}_{\text{odd polynomial}}\phi(x)+o(\tfrac1n)$$

Cornish-Fisher expansion on quantile q_α of t-ratio T:

$q_\alpha = z_\alpha + n^{-\frac12}p_{11}(z_\alpha)+O(\frac1n)$ $\quad p_{11}(\cdot)$ even polynomial

$q^*_\alpha = z_\alpha+n^{-\frac12}p^*_{11}(z_\alpha)+O_p(\frac1n)$ $\quad p^*_{11}(z_\alpha) = p_{11}(z_\alpha)+O_p(\frac1n)$

$\Rightarrow q_\alpha - q^*_\alpha = n^{-\frac12}[p_{11}(z_\alpha)-p^*_{11}(z_\alpha)]+O_p(\frac1n) = O_p(\frac1n)$

T and $T+q_\alpha-q^*_\alpha = T+O_p(\frac1n)$ have same Edgew- expansion:

$P(T\le q^*_\alpha) = P(T+q_\alpha-q^*_\alpha\le q_\alpha) = P(T\le q_\alpha)+O(\frac1n)$

$= \alpha+O(\frac1n)$

$\text{Bias}^* = \frac{1}{B}\sum_{b=1}^{B} \hat{\theta}^{*b} - \hat{\theta}$

$= \frac{1}{B}\sum_{b=1}^{B} [g'(\bar{y}_n)(\bar{y}_n^{*b} - \bar{y}_n) + \frac{1}{2}g''(\bar{y}_n)(\bar{y}_n^{*b} - \bar{y}_n)^2 + o_p(\frac{1}{n})]$

$\xrightarrow{P^*} \frac{1}{2}g''(\bar{y}_n)\frac{1}{n}S_n^2 + o(\frac{1}{n}) \xrightarrow{P} \text{Bias}$

Without pivotal distribution: $\theta = \mu$. $\hat{\theta} = \bar{y}$

$T = \sqrt{n}(\bar{y} - \mu) \xrightarrow{d} N(0, \sigma^2) = F_\sigma$

$T^* = \sqrt{n}(\bar{y}^{*b} - \bar{y}) \xrightarrow{d^*} N(0, S^2) = F_S^* \rightarrow q^*$

$P(T \le x) = F_\sigma(x) + n^{-\frac{1}{2}} P_{even}(x, F_\sigma) + n^{-1} P_{odd}(x, F_\sigma) + o(\frac{1}{n})$

$P(T^* \le x) = F_S^*(x) + n^{-\frac{1}{2}} P_{even}(x, F_S^*) + n^{-1} P_{odd}(x, F_S^*) + o(\frac{1}{n})$

$\Rightarrow P(T^* \le x) - P(T \le x) = O(n^{-\frac{1}{2}})$

$P(\theta \in [\hat{\theta} + q^*_{\frac{\alpha}{2}}, \hat{\theta} + q^*_{1-\frac{\alpha}{2}}]) = P(q^*_{\frac{\alpha}{2}} \le \hat{\theta} - \theta \le q^*_{1-\frac{\alpha}{2}})$

$= 1 - \alpha + O(n^{-\frac{1}{2}})$

Consistency but no higher order accuracy

Bootstrap Hypothesis Testing

$H_0: \theta = \theta_0 \quad H_A: \theta \ne \theta_0$

Scalar: T statistic: $T = \frac{\hat{\theta} - \theta_0}{s(\hat{\theta})}$ $\quad P = 1 - G_n(|T|)$ null dist.

Bootstrap: $T^{*b} = \frac{\hat{\theta}^{*b} - \hat{\theta}}{s(\hat{\theta}^{*b})}$ $\quad P^* = \frac{1}{B}\sum_{b=1}^{B} 1\{|T^{*b}| > |T|\}$

Bootstrap $100\alpha\%$ critical value: $q^*_{1-\alpha}$: $(1-\alpha)$-th quantile $|T^{*b}|$

Reject $H_0 \Leftrightarrow |T| > q^*_{1-\alpha} \Leftrightarrow p^* < \alpha$

Smooth Function Model:

If $q^*_{1-\alpha} = \bar{z}_{1-\alpha} + o_p(\frac{1}{n}) \Rightarrow q^*_{1-\alpha} - q_{1-\alpha} = o_p(\frac{1}{n})$

then $P(|T| > \bar{z}_{1-\alpha} \mid H_0) = 1-\alpha + O(\frac{1}{n})$

$P(|T| > q^*_{1-\alpha} \mid H_0) = 1-\alpha + o(\frac{1}{n})$

Vector: $H_0: a(\theta) = 0 \quad H_A: a(\theta) \ne 0$

Wald statistic $W = (\hat{\theta} - \theta_0)^T \hat{V}_{\hat{\theta}}^{-1} (\hat{\theta} - \theta_0)$

Bootstrap: $W^{*b} = (\hat{\theta}^{*b} - \hat{\theta})^T \hat{V}_{\hat{\theta}}^{*b-1} (\hat{\theta}^{*b} - \hat{\theta})$

Criterion-based statistic $J = \min_{a(\theta)=0} J(\theta) - \min J(\theta)$

$J^{*b} = \min_{a(\theta) = a(\hat{\theta})} J^*(\theta) - \min J^*(\theta)$

Likelihood statistic:

$LR = 2[L(\hat{\theta}) - L(\tilde{\theta})] \quad \hat{\theta} = \underset{\theta}{\text{argmax}}\, L(\theta)$

$\tilde{\theta} = \underset{a(\theta)=0}{\text{argmax}}\, L(\theta)$

$LR^* = 2[L^*(\hat{\theta}^*) - L^*(\tilde{\theta}^*)] \quad \hat{\theta}^* = \underset{\theta}{\text{argmax}}\, L^*(\theta)$

$\tilde{\theta}^* = \underset{a(\theta)=a(\hat{\theta})}{\text{argmax}}\, L^*(\theta)$

$\alpha\%$ critical value: $q^*_{1-\alpha} \leftarrow \{LR^{*b}\}_{b=1}^{B}$

p-value: $p^* = \frac{1}{B}\sum_{b=1}^{B} 1\{LR^{*b} \ge LR\}$

Reject $H_0 \Leftrightarrow LR > q^*_{1-\alpha} \Leftrightarrow p^* < \alpha$

Pair Bootstrap. Sample: $\{y_i, x_i\}_{i=1}^n$

Bootstrap: $\{y_i^*, x_i^*\}_{i=1}^n$

$\forall b = 1, \dots, B$: $y_i^* = x_i^{*T}\hat{\beta} + \varepsilon_i^*$.

$E^*(x_i^* \varepsilon_i^*) = E^*(x_i^*(y_i^* - x_i^{*T}\hat{\beta})) = \frac{1}{n}\sum_{i=1}^n (x_i(y_i - x_i^T\hat{\beta})) = 0$

$\sqrt{n}(\hat{\beta}^{*b} - \hat{\beta}) = \sqrt{n}(\frac{1}{n}\sum_{i=1}^n x_i^* x_i^{*T})^{-1}(\frac{1}{n}\sum_{i=1}^n x_i^* \varepsilon_i^*)$

$\xrightarrow{d^*} (\frac{1}{n}\sum_{i=1}^n x_i x_i^T)^{-1} N(0, \frac{1}{n}\sum_{i=1}^n x_i x_i^T \hat{\varepsilon}_i^2)$

Parametric Bootstrap

Residual-based: $y_i = x_i^T\beta + \varepsilon_i$. $E(\varepsilon_i \mid x_i) = 0$ $\quad \varepsilon_i \mid x_i \overset{iid}{\sim} (0, \sigma^2)$

① OLS $y_i \sim x_i$: $\hat{\beta}$. $\hat{\varepsilon}_i$

② Bootstrap $\{\hat{\varepsilon}_i^{*b}\}_{i=1}^n$: $y_i^{*b} = x_i^T\hat{\beta} + \hat{\varepsilon}_i^{*b}$

OLS $y_i^{*b} \sim x_i$: $\hat{\beta}^{*b}$

Wild Bootstrap. $\varepsilon_i^* = \hat{\varepsilon}_i V_i^*$ $\quad V_i^* \overset{iid}{\sim} (0, 1)$

Block Bootstrap: overlapping blocks $[1, b]\ [2, b+1] \dots$

$b \to +\infty$, $\frac{b}{n} \to 0$ randomly draw $\frac{n}{b}$ blocks. get $\{y_i^*, x_i^*\}_{i=1}^n$

Markov Bootstrap: $\hat{f}(Y_t \mid Y_{t-1}) = \hat{f}(Y_t, Y_{t-1}) / \hat{f}(Y_{t-1})$

Non-parametric Estimation

kernel density estimation : $g \in C^2$ Smoothness

$\hat{g}(x) = \frac{1}{nh}\sum_{i=1}^{n} k(\frac{x-x_i}{h})$ kernel $k(\cdot)$

Assumption on 2nd $k(\cdot)$: $\int k(u)du = 1$ $\int u k(u) du = 0$

$\int u^2 k(u) du = C_k$ $\int k^2(u) du = D_k$

$\hat{g}(x)$ is a valid density: $\hat{g}(x) \geq 0$, $\int \hat{g}(x) dx = 1$

$$\text{Bias}(x;h) = E\hat{g}(x) - g(x) \overset{iid}{=} \frac{1}{h}\int k(\frac{x-x_i}{h}) g(x_i) dx_i - g(x)$$
$$= \int k(u) g(x+uh) du - g(x)$$
$$= \int k(u)[g(x) + g'(x)uh + \frac{1}{2}g''(x)u^2h^2 + o(h^2)] du - g(x)$$
$$= \frac{1}{2} g''(x) C_k h^2 + o(h^2)$$

$$\text{Var}(x;h) = E(\hat{g}(x) - E\hat{g}(x))^2 = E\{\frac{1}{nh}\sum_{i=1}^{n}(k(\frac{x-x_i}{h}) - Ek(\frac{x-x_i}{h}))\}^2$$
$$\overset{iid}{=} \frac{1}{nh^2} E\{k(\frac{x-x_i}{h}) - Ek(\frac{x-x_i}{h})\}^2$$
$$= \frac{1}{nh^2} E k^2(\frac{x-x_i}{h}) - \frac{1}{nh^2}\{Ek(\frac{x-x_i}{h})\}^2$$
$$= \frac{1}{nh^2}\int k(\frac{x-x_i}{h})^2 g(x_i) dx - \frac{1}{nh^2}\{Ek(\frac{x-x_i}{h})\}^2$$
$$= \frac{1}{nh}\int k(u)^2 g(x+uh) du - \frac{h}{nh}\{\int k(u) g(x+uh) du\}^2$$
$$= \frac{1}{nh}\int k(u)^2 [g(x) + g'(x)uh + o(h)] du - o(\frac{1}{nh})$$
$$= D_k g(x) \frac{1}{nh} + o(\frac{1}{nh})$$

$\text{MSE} = E|\hat{g}(x) - g(x)|^2 = \text{Var}(x;h) + \text{Bias}^2(x;h)$

$\text{AIMSE} = \int \text{MSE}(x;h) dx = D_k \frac{1}{nh} + \frac{1}{4}\int g''(x)^2 dx\, C_k^2 h^4 + o(\frac{1}{nh}) + o(h^4)$

$h^0 = \underset{h}{\text{argmin}}\ \text{AIMSE}(h)$: $h^0 = n^{-\frac{1}{5}}\left(\frac{D_k}{C_k^2 \int g''(x)^2 dx}\right)^{\frac{1}{5}}$

$\text{AIMSE}(h^0) \propto n^{-\frac{4}{5}}$

Bandwith selection

① Rule of thumb : (Silverman's rule) : theoretical h^0

② Plug-in & iteration: $h^0 \Rightarrow \hat{g}^0(x) \Rightarrow \int \hat{g}^{0''}(x)^2 dx$
$\Rightarrow \text{AIMSE}^0(h) \Rightarrow h^1$

③ Cross-validation.

$$\widehat{\text{AIMSE}}(h) = \int [\hat{g}(x) - g(x)]^2 dx$$
$$= \int \hat{g}(x)^2 dx - 2E\hat{g}(x) + \int g(x)^2 dx \quad \text{Const.}$$
$$= \frac{1}{n^2h^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\int k(\frac{x_i-x}{h}) k(\frac{x_j-x}{h}) dx - 2\frac{1}{n}\sum_{i=1}^{n}\hat{g}_{-i}(x_i)$$
$$= \frac{1}{n^2h}\sum_{i=1}^{n}\sum_{j=1}^{n}\bar{k}(\frac{x_i-x_j}{h}) - 2\frac{1}{n}\sum_{i=1}^{n}\hat{g}_{-i}(x_i)$$

where convolution kernel $\bar{k}(v) = \int k(u) k(u-v) du$

leave-1-out $\hat{g}_{-i}(x_i) = \frac{1}{(n-1)h}\sum_{j \neq i} k(\frac{x_i - x_j}{h})$

$\hat{h}_{cv} = \underset{h}{\text{argmin}}\ \widehat{\text{AIMSE}}(h)$

Consistency : $(\hat{h}_{cv} - h^0)/h^0 \xrightarrow{P} 0$

Boundary problem $x \in [0, 0+h)$: $[\frac{-x}{h}, 1]$

Boundary kernel :

Multi-dimensional $\hat{g}(x) = \frac{1}{n\prod_j h_j}\sum_{i=1}^{n}\prod_{j=1}^{d} k(\frac{x_{ij} - x_j}{h_j})$

$\text{Bias}(h) = O(h^2)$

$\text{Var}(h) = O(\frac{1}{nh^d})$

$\text{AIMSE}(h^0) \propto n^{-\frac{4}{d+4}}$

Non-parametric regression

$y_i = r(x_i) + \varepsilon_i$. $E(\varepsilon_i|x_i) = 0$ $\text{Var}(\varepsilon_i|x_i) = \sigma^2(x_i)$

① N-W estimation : $\hat{r}(x) = \sum_{i=1}^{n} w_i y_i$

where $w_i = k(\frac{x_i - x}{h})[\sum_{i=1}^{n} k(\frac{x_i - x}{h})]^{-1}$

② Local constant : $\hat{r}(x) = \underset{\alpha}{\text{argmin}} \sum_{i=1}^{n} k(\frac{x_i - x}{h})(y_i - \alpha)^2$

$\hat{r}(x) = [\sum_{i=1}^{n} k(\cdot)]^{-1}\sum_{i=1}^{n} k(\cdot) y_i$

③ Local linear : $(\hat{r}(x), \hat{r}'(x)) = \underset{(\alpha,\beta)}{\text{argmin}} \sum_{i=1}^{n} k(\frac{x_i - x}{h})[y_i - \alpha - \beta(x_i - x)]^2$

$z_i(x) = \begin{bmatrix} 1 \\ x_i - x \end{bmatrix}$ $\begin{bmatrix} \hat{r}(x) \\ \hat{r}'(x) \end{bmatrix} = [\sum_{i=1}^{n} k(\cdot) z_i z_i^T]^{-1}[\sum_{i=1}^{n} k(\cdot) z_i y_i]$

$$\text{Bias}(x;h) = E\hat{r}(x) - r(x)$$
$$= E\{\sum_{i=1}^{n} w_i E(y_i|x_i)\} - r(x) = E\{\sum_{i=1}^{n} w_i r(x_i)\} - r(x)$$
$$= \int \frac{1}{\frac{1}{nh}\sum_{i=1}^{n} k(\frac{x_i - x}{h})}\frac{1}{h} k(\frac{x_i - x}{h}) r(x_i) \prod_{i=1}^{n} g(x_i) dx_i - r(x)$$
$$= \int [g(x) + O(h^2)]^{-1} k(u) r(x+uh) g(x+uh) du - r(x)$$

$\text{ABias}(x;h) = \frac{1}{2} r''(x) C_k h^2$

$$\text{Var}(x;h) = E[\hat{r}(x) - E\hat{r}(x)]^2$$
$$= E\{E[(\sum_{i=1}^{n} w_i y_i - E\sum_{i=1}^{n} w_i r(x_i))^2 | x_i]\}$$
$$= \sum_{i=1}^{n} w_i y_i - \sum_{i=1}^{n} w_i r(x_i) + \sum_{i=1}^{n} w_i r(x_i) - E\sum_{i=1}^{n} w_i r(x_i)$$
$$= E\{\sum_{i=1}^{n} w_i^2 E((y_i - r(x_i))^2 | x_i) + (\sum_{i=1}^{n} w_i r(x_i) - E\sum_{i=1}^{n} w_i r(x_i))^2\}$$

$\text{AVar}(x;h) = \frac{1}{nh}\frac{\sigma^2(x)}{g(x)} D_k$

$\sqrt{nh}[\hat{r}(x) - r(x) - \text{ABias}(x;h)] \xrightarrow{d} N(0, \frac{\sigma^2(x) D_k}{g(x)})$

Bandwidth selection P31

① Plug-in

② Cross-validation

Semi-parametric Estimation

1. Partially linear model : $E(u_i|x_i, z_i) = 0$

$y_i = x_i^T\beta + g(z_i) + u_i$ $\text{Var}(u_i|x_i, z_i) = \sigma^2(x_i, z_i)$

Robinson : ① NP Reg :

$y_i = \hat{E}(y_i|z_i) + \hat{e}_{yi} \Rightarrow \hat{y}_i = y_i - \hat{E}(y_i|z_i)$

$x_i = \hat{E}(x_i|z_i) + \hat{e}_{xi}$ $\hat{x}_i = x_i - \hat{E}(x_i|z_i)$

② OLS : $\hat{y}_i = \hat{x}_i^T\hat{\beta} + \hat{u}_i$

③ NP Reg : $y_i - x_i^T\hat{\beta} = g(z_i) + u_i$

P35

2. Semi-parametric single index $y_i = g(x_i^T\beta) + u_i$

Ichimura ① $\hat{\beta} = \underset{\beta}{\text{argmin}} \sum_{i=1}^{n}[y_i - \underbrace{\hat{g}_{-i}(x_i^T\beta)}_{\text{Leave-1-out}}]^2 1_i(b)$

② NP Reg.

3. Varying coefficient model

$y_i = x_i^T\beta(z_i) + u_i$ $E(u_i|x_i, z_i) = 0$

High dimensional regression

$Y = X\beta + e. \quad X_{n\times p}.$

$\hat{\beta}_{OLS} = (X^TX)^{-1}X^TY$

$\hat{\beta}_r = \underset{\beta}{\operatorname{argmin}} \|Y - X\beta\|_2^2 + \lambda \|\beta\|_r^r \quad \begin{cases} \sum_{j=1}^p 1\{\beta_j \neq 0\} \\ \sum_{j=1}^p |\beta_j| \\ \sum_{j=1}^p \beta_j^2 \end{cases}$

Ridge: $\hat{\beta} = \underset{\beta}{\operatorname{argmin}} (Y - X\beta)^T(Y - X\beta) + \lambda \beta^T\beta$

FOC: $-2X^T(Y - X\beta) + 2\lambda\beta = 0$

$\Rightarrow \hat{\beta}_{ridge} = (X^TX + \lambda I_p)^{-1}X^TY$

① $Var(\hat{\beta}_{OLS}|X) - Var(\hat{\beta}_{ridge}|X) \geq 0.$

$X^TX = Q\Lambda Q^T. \quad QQ^T = Q^TQ = I \qquad \Sigma = E(\varepsilon\varepsilon^T|X)$

$Var(\hat{\beta}_{OLS}|X) = E((X^TX)^{-1}X^T\varepsilon\varepsilon^TX(X^TX)^{-1}|X)$

$= Q\, diag\{\sigma^2\lambda_i^{-1}\}Q^T$

$Var(\hat{\beta}_{ridge}|X) = Q\, diag\{\sigma^2(\lambda_i + \lambda)^{-1}\}Q^T$

② $\hat{\beta}_{OLS,j} \neq 0 \Rightarrow \hat{\beta}_{ridge,j} \neq 0$

Lasso: $\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \frac{1}{2}(Y - X\beta)^T(Y - X\beta) + \lambda\sum_{j=1}^p|\beta_j|$

FOC: $-X^T(Y - X\beta) + \lambda\eta(\beta) = 0$ subgrediant

$\Rightarrow \hat{\beta}_{Lasso} = (X^TX)^{-1}(X^TY - \lambda\eta(\beta))$

If $X^TX = I_p$:

$\hat{\beta}_{Lasso,j} = \hat{\beta}_{OLS,j} - \lambda\, sign(\hat{\beta}_{OLS,j})$. if $|\hat{\beta}_{OLS,j}| > \lambda$

BSS $\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \frac{1}{2}(Y - X\beta)^T(Y - X\beta) + \lambda\sum_j 1\{\beta_j \neq 0\}$

FOC: $-X^T(Y - X\beta) = 0$ $\quad S_0$: # of $\beta_j \neq 0$

$\Rightarrow \hat{\beta}_{bss} = (X^TX)^{-1}(X^TY - \lambda S_0).$