2006 Fall Part 2

Note that all x_i's are $1 \times k$ vector in Part 2.

1. Logit model
Y_i has a binary distribution.

$$\Pr(y_i = 1|x_i) = \frac{\exp(x_i\beta)}{1+\exp(x_i\beta)}$$

so the log likelihood function is

$$\begin{aligned}
\mathcal{L} = \sum_{i=1}^n \log f(y_i|x_i) &= \sum_{i=1}^n y_i \log \frac{\exp(x_i\beta)}{1+\exp(x_i\beta)} + \sum_{i=1}^n (1-y_i)\log\frac{1}{1+\exp(x_i\beta)} \\
&= \sum_{i=1}^n y_i x_i \beta - \sum_{i=1}^n y_i \log[1+\exp(x_i\beta)] - \sum_{i=1}^n (1-y_i)\log[1+\exp(x_i\beta)] \\
&= \sum_{i=1}^n y_i x_i \beta - \sum_{i=1}^n \log[1+\exp(x_i\beta)]
\end{aligned}$$

FOC is

$$\frac{\partial \mathcal{L}}{\partial \beta} = \sum_{i=1}^n x_i' y_i - \sum_{i=1}^n \frac{\exp(x_i\beta)}{1+\exp(x_i\beta)} x_i' = \sum_{i=1}^n x_i' \left[y_i - \frac{\exp(x_i\beta)}{1+\exp(x_i\beta)} \right] = 0$$

Since

$$\frac{\partial^2 \mathcal{L}}{\partial \beta^2} = \sum_{i=1}^n \frac{[\exp(x_i\beta)]^2}{[1+\exp(x_i\beta)]^2} x_i' x_i - \sum_{i=1}^n \frac{\exp(x_i\beta)}{1+\exp(x_i\beta)} x_i' x_i = -\sum_{i=1}^n \frac{\exp(x_i\beta)}{[1+\exp(x_i\beta)]^2} x_i' x_i$$

Fisher Information is

$$i(\beta) = \frac{1}{n} I(\beta) = E\left(\frac{\exp(x_i\beta)}{[1+\exp(x_i\beta)]^2} x_i' x_i \right)$$

With this, consider the test of nonlinear hypothesis $\gamma(\beta) = 0^{p\times 1}$. Denote simply MLE of β by $\widehat{\beta}$ and restricted MLE by $\widetilde{\beta}$.

For Wald,

$$\sqrt{n}\left(\widehat{\beta} - \beta\right) \xrightarrow{d} N\left(0, i(\beta)^{-1}\right)$$

$$\sqrt{n}\left(\gamma\left(\widehat{\beta}\right) - \gamma(\beta)\right) \xrightarrow{d} N\left(0, \frac{\partial \gamma(\beta)}{\partial \beta'} i(\beta)^{-1} \frac{\partial \gamma(\beta)'}{\partial \beta}\right)$$

Under H_0, $\gamma(\beta) = 0$, and $\widehat{\beta} \xrightarrow{p} \beta$, so

$$\sqrt{n}\gamma\left(\widehat{\beta}\right) \xrightarrow{d} N\left(0, \frac{\partial \gamma(\beta)}{\partial \beta'} i(\beta)^{-1} \frac{\partial \gamma(\beta)'}{\partial \beta}\right)$$

$$n\gamma\left(\widehat{\beta}\right)' \left[\frac{\partial \gamma\left(\widehat{\beta}\right)}{\partial \beta'} i\left(\widehat{\beta}\right)^{-1} \frac{\partial \gamma\left(\widehat{\beta}\right)'}{\partial \beta} \right]^{-1} \gamma\left(\widehat{\beta}\right) \xrightarrow{d} \chi^2_p$$

For LM, according to LLN and the definition of Fisher Information,

$$\frac{1}{\sqrt{n}} \frac{\partial \mathcal{L}(\beta)}{\partial \beta} \xrightarrow{d} N\left(0, i(\beta)\right)$$

Under H_0, $\widetilde{\beta} \xrightarrow{p} \beta$, and thus

$$\frac{1}{\sqrt{n}} \frac{\partial \mathcal{L}\left(\widetilde{\beta}\right)}{\partial \beta} \xrightarrow{d} N\left(0, i(\beta)\right)$$

$$\frac{1}{n} \frac{\partial \mathcal{L}\left(\widetilde{\beta}\right)}{\partial \beta} i\left(\widetilde{\beta}\right)^{-1} \frac{\partial \mathcal{L}\left(\widetilde{\beta}\right)}{\partial \beta} \xrightarrow{d} \chi_p^2$$

$$-\frac{\partial \mathcal{L}\left(\widetilde{\beta}\right)}{\partial \beta} \left[\frac{\partial^2 \mathcal{L}\left(\widetilde{\beta}\right)}{\partial \beta \partial \beta'}\right]^{-1} \frac{\partial \mathcal{L}\left(\widetilde{\beta}\right)}{\partial \beta} \xrightarrow{d} \chi_p^2$$

For LR,

$$2\left[\mathcal{L}\left(\widehat{\beta}\right) - \mathcal{L}\left(\widetilde{\beta}\right)\right] \xrightarrow{d} \chi_p^2$$

2. NLS model (a)
It's neither unbiased nor consistent.

For biasedness, consider $g(t) = t$, i.e., the linear model $y_i = x_i \beta_0 + \varepsilon_i$.

$$\widehat{\beta} = \beta_0 + \left(\sum_{i=1}^n x_i' x_i\right)^{-1} \sum_{i=1}^n x_i' \varepsilon_i$$

Here we have

$$E\left(\widehat{\beta}\right) = \beta_0 + E\left[\left(\sum_{i=1}^n x_i' x_i\right)^{-1} \sum_{i=1}^n x_i' \varepsilon_i\right]$$

whose second term is not equal to 0 with $E[x_i \varepsilon_i] = 0$ only.

To see inconsistency, consider $g(t) = t^2$ and x_i: scalar, i.e., $y_i = x_i^2 \beta_0^2 + \varepsilon_i$.

$$\widehat{\beta}^2 = \beta_0^2 + \left(\frac{1}{n} \sum_{i=1}^n x_i^4\right)^{-1} \frac{1}{n} \sum_{i=1}^n x_i^2 \varepsilon_i$$

Under proper assumptions, the second term converges to

$$\left[E\left(x_i^4\right)\right]^{-1} E\left(x_i^2 \varepsilon_i\right)$$

which is not equal to 0 as long as $E\left(x_i^2 \varepsilon_i\right) \neq 0$. So $\widehat{\beta}^2 \xrightarrow{p} \beta_0^2$ is not true, and thus $\widehat{\beta} \xrightarrow{p} \beta_0$ is not true, either.

(b)
It's not unbiased but consistent under the assumption that ε_i is stochastically independent of $x_1, \cdots, x_n$.

For biasedness, consider $g(t) = t^{-1}$ and x_i: scalar.

$$y_i = \frac{1}{x_i \beta_0} + \varepsilon_i$$

$$\frac{1}{\widehat{\beta}} = \frac{1}{\beta_0} + \left(\sum_{i=1}^n \frac{1}{x_i^2}\right)^{-1} \sum_{i=1}^n \frac{\varepsilon_i}{x_i}$$

Clearly,

$$E\left(\frac{1}{\widehat{\beta}} \middle| x\right) = \frac{1}{\beta_0}$$

But this doesn't mean $E\left(\widehat{\beta}\middle| x\right) = \beta_0$. For example, if $f(z) = 1$ for $z \in (0,1)$, $E(z) = 0.5$, but $E\left(z^{-1}\right) \neq 2$.

To see consistency,

$$\min \frac{1}{2n}\sum_{i=1}^{n}\left[y_i - g(x_i\beta)\right]^2$$

FOC is

$$-\frac{1}{n}\sum_{i=1}^{n}\left[y_i - g(x_i\beta)\right] g'(x_i\beta)x_i' = 0$$

LHS converges in probability to

$$E\left[y_i - g(x_i\beta)\right] g'(x_i\beta)x_i'$$

At $\beta = \beta_0$,

$$E\left(\ [y_i - g(x_i\beta_0)]\, g'(x_i\beta_0)x_i' \ \right) = E\left[\ \varepsilon_i g'(x_i\beta_0)x_i' \ \right] = E\left[\ E(\varepsilon_i|x_i)g'(x_i\beta_0)x_i' \ \right] = 0$$

Only $\beta = \beta_0$ satisfies the above condition as long as $g(x_i\beta) \neq g(x_i\beta_0)$ for any $\beta \neq \beta_0$. Therefore $\widehat{\beta}$ converges in probability to β_0.

(d)

$$\frac{1}{n}\sum_{i=1}^{n}\left[y_i - g\left(x_i\widehat{\beta}\right)\right] g'\left(x_i\widehat{\beta}\right)x_i' = 0$$

Applying mean value theorem around β_0 on LHS,

$$\frac{1}{n}\sum_{i=1}^{n}\left[y_i - g(x_i\beta_0)\right] g'(x_i\beta_0)x_i' + \left(\frac{1}{n}\sum_{i=1}^{n}\left[y_i - g\left(x_i\widetilde{\beta}\right)\right] g''\left(x_i\widetilde{\beta}\right)x_i'x_i - \frac{1}{n}\sum_{i=1}^{n}\left[g'\left(x_i\widetilde{\beta}\right)\right]^2 x_i'x_i\right)\left(\widehat{\beta} - \beta_0\right) = 0$$

where $\widetilde{\beta}$ is between β_0 and $\widehat{\beta}$. So $\widehat{\beta} \xrightarrow{p} \beta_0$ and $\widetilde{\beta} \xrightarrow{p} \beta_0$. Therefore,

$$\sqrt{n}\left(\widehat{\beta} - \beta_0\right) \xrightarrow{d} N\left(0, B_0^{-1}\Omega_0 B_0^{-1}\right)$$

where

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left[y_i - g\left(x_i\widetilde{\beta}\right)\right] g''\left(x_i\widetilde{\beta}\right)x_i'x_i - \frac{1}{n}\sum_{i=1}^{n}\left[g'\left(x_i\widetilde{\beta}\right)\right]^2 x_i'x_i &\xrightarrow{p} B_0 \equiv E\left(\left[g'(x_i\beta_0)\right]^2 x_i'x_i\right) \\
\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i g'(x_i\beta_0)x_i' &\xrightarrow{d} N\left(0, \Omega_0\right) \\
\Omega_0 &= E\left(\varepsilon_i^2\left[g'(x_i\beta_0)\right]^2 x_i'x_i\right)
\end{aligned}$$