240D Fall 2004 Solutions to Midterm Exam

1.(a) The objective function

$$\begin{aligned} Q_N(\boldsymbol{\beta}) &= \frac{1}{N}\sum\nolimits_i (y_i - \mathbf{x}_i{}'\boldsymbol{\beta})^4 \\ &= \frac{1}{N}\sum\nolimits_i (y_i - \mathbf{x}_i{}'\boldsymbol{\beta}_0 + \mathbf{x}_i{}'\boldsymbol{\beta}_0 - \mathbf{x}_i{}'\boldsymbol{\beta})^4 \\ &= \frac{1}{N}\sum\nolimits_i [u_i + \mathbf{x}_i{}'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})]^4 \qquad \text{given } y_i - \mathbf{x}_i{}'\boldsymbol{\beta}_0 = u_i \\ &= \frac{1}{N}\sum\nolimits_i \left(u_i^4 + u_i^3[\mathbf{x}_i{}'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})] + u_i^2[\mathbf{x}_i{}'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})]^2 + u_i[\mathbf{x}_i{}'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})]^3 + [\mathbf{x}_i{}'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})]^4\right), \end{aligned}$$

where use $(x+y)^4 = x^4 + 4x^3y + 6x^2y^2 + 4xy^3 + y^4$ [use Pascal's triangle].

(b) Then assuming a LLN can be applied so plim = limE

$$\begin{aligned} Q_0(\boldsymbol{\beta}) &= \operatorname{plim} Q_N(\boldsymbol{\beta}) \\ &= \frac{1}{N}\lim\sum\nolimits_i \mathrm{E}[u_i^4] + \lim\frac{1}{N}\sum\nolimits_i \mathrm{E}[4u_i^3\mathbf{x}_i'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})] + \lim\frac{1}{N}\sum\nolimits_i \mathrm{E}[6u_i^2[\mathbf{x}_i'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})]^2] \\ &\quad + \lim\frac{1}{N}\sum\nolimits_i \mathrm{E}[4u_i\mathbf{x}_i'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})]^3] + \lim\frac{1}{N}\sum\nolimits_i \mathrm{E}[[\mathbf{x}_i'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})]^4] \\ &= 3\sigma^4 + 0 + \lim\frac{1}{N}\sum_{i=1}^N 6\sigma^2[\mathbf{x}_i'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})]^2 + 0 + \lim\frac{1}{N}\sum_{i=1}^N [\mathbf{x}_i'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})]^4, \end{aligned}$$

using nonstochastic $\mathbf{x}$ and $\mathrm{E}[u] = 0$, $\mathrm{E}[u^2] = \sigma^2$, $\mathrm{E}[u^3] = 0$ and $\mathrm{E}[u^4] = 3\sigma^4$.

(c) Differentiate wrt $\boldsymbol{\beta}$ (not $\boldsymbol{\beta}_0$)

$$\begin{aligned} \frac{\partial Q_0(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \lim\frac{1}{N}\sum_{i=1}^N -12\sigma^2[\mathbf{x}_i'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})]\mathbf{x}_i - \lim\frac{1}{N}\sum_{i=1}^N 4[\mathbf{x}_i'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})]^3\mathbf{x}_i \\ &= \mathbf{0} \qquad \text{when } \boldsymbol{\beta} = \boldsymbol{\beta}_0. \end{aligned}$$

[and as stated no need to check that $\partial^2 Q_0(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'$ is positive definite at $\boldsymbol{\beta}_0$.]
Since $\operatorname{plim} Q_N(\boldsymbol{\beta})$ attains a local maximum at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, conclude that $\widehat{\boldsymbol{\beta}} = \arg\max Q_N(\boldsymbol{\beta})$ is consistent for $\boldsymbol{\beta}_0$.

(d) Apply LLN to $N^{-1}\sum_i 4u_i[\mathbf{x}_i'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})]^3$. i.e. The X_i in notes is $4u_i[\mathbf{x}_i'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})]^3$.
This is average of $4u_i[\mathbf{x}_i'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})]^3$.
Here $\mathrm{E}[4u_i[\mathbf{x}_i'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})]^3] = 0$ since $\mathrm{E}[u_i = 0]$
and $\mathrm{V}[4u_i[\mathbf{x}_i'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})]^3] = 16\sigma^2[\mathbf{x}_i'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})]^6$ since $\mathrm{E}[u_i^2] = \sigma^2$ and $\mathbf{x}_i$ is fixed (nonstochastic).

Note: different observations with different fixed $\mathbf{x}_i$ have different variance.
So will need to use Markov LLN as $4u_i[\mathbf{x}_i'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})]^3$ is inid.
We need the $(1+\delta)$ side assumption. Here let $\delta = 1$ so that $1 + \delta = 2$.
Then $\sum_i\{\mathrm{E}[|4u_i[\mathbf{x}_i'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})]^3] - 0|^2/i^2\} = \sum_i\{16\sigma^2[\mathbf{x}_i'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})]^6/i^2\} < \infty$
if $\mathbf{x}_i$ are bounded and the parameter space is bounded.

2.(a) For $Q_N(\boldsymbol{\beta}) = N^{-1}\sum_i (y_i - \mathbf{x}_i{}'\boldsymbol{\beta})^4$ [This is easier to work with than the re-expression of $Q_N(\boldsymbol{\beta})$]

$$\begin{aligned}
\frac{\partial Q_N}{\partial \boldsymbol{\beta}} &= \frac{1}{N}\sum\nolimits_i -4(y_i - \mathbf{x}_i{}'\boldsymbol{\beta})^3\mathbf{x}_i \\
\sqrt{N}\frac{\partial Q_N}{\partial \boldsymbol{\beta}}\bigg|_{\beta_0} &= \frac{1}{\sqrt{N}}\sum\nolimits_i -4u_i{}^3\mathbf{x}_i \qquad \text{since } y_i - \mathbf{x}_i{}'\boldsymbol{\beta}_0 = u_i \\
&\xrightarrow{d} \mathcal{N}\left[\mathbf{0}, \mathbf{B}(\boldsymbol{\beta}_0) = \lim \mathrm{E}[\frac{1}{N}\sum\nolimits_i (-4u_i{}^3\mathbf{x}_i)(-4u_i{}^3\mathbf{x}_i)']\right] \\
&\xrightarrow{d} \mathcal{N}\left[\mathbf{0}, \mathbf{B}(\boldsymbol{\beta}_0) = \lim \frac{1}{N}\sum\nolimits_i 16\mathrm{E}[u_i{}^6]\mathbf{x}_i\mathbf{x}_i{}'\right] \\
&\xrightarrow{d} \mathcal{N}\left[\mathbf{0}, \mathbf{B}(\boldsymbol{\beta}_0) = \lim \frac{1}{N}\sum\nolimits_i 240\sigma^6\mathbf{x}_i\mathbf{x}_i{}'\right], \quad \text{since } E[u_i{}^3] = 15\sigma^6.
\end{aligned}$$

Note that the last expression is lim not limE since here $\mathbf{x}$ is nonstochastic.

(b) And

$$\begin{aligned}
\frac{\partial^2 Q_N}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} &= \frac{1}{N}\sum\nolimits_i 12(y_i - \mathbf{x}_i{}'\boldsymbol{\beta})^2\mathbf{x}_i\mathbf{x}_i' \\
\frac{\partial^2 Q_N}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'}\bigg|_{\beta_0} &= \frac{1}{N}\sum\nolimits_i 12u_i{}^2\mathbf{x}_i\mathbf{x}_i' \qquad \text{since } y_i - \mathbf{x}_i{}'\boldsymbol{\beta}_0 = u_i \\
&\xrightarrow{p} \mathbf{A}(\boldsymbol{\beta}_0) = \lim \mathrm{E}[\frac{1}{N}\sum\nolimits_i 12u_i{}^2\mathbf{x}_i\mathbf{x}_i'] \\
\xrightarrow{p} \mathbf{A}(\boldsymbol{\beta}_0) &= \lim \frac{1}{N}\sum\nolimits_i 12\sigma^2\mathbf{x}_i\mathbf{x}_i', \quad \text{since } E[u_i{}^2] = \sigma^2.
\end{aligned}$$

(c) Combining $\mathbf{A}(\boldsymbol{\beta}_0)^{-1}\mathbf{B}(\boldsymbol{\beta}_0)\mathbf{A}(\boldsymbol{\beta}_0)^{-1} = ??$ Does it simplify? Here yes.

$$\begin{aligned}
&\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
&\xrightarrow{d} \mathcal{N}\left[\mathbf{0}, \mathbf{A}(\boldsymbol{\beta}_0)^{-1}\mathbf{B}(\boldsymbol{\beta}_0)\mathbf{A}(\boldsymbol{\beta}_0)^{-1}\right] \\
&\xrightarrow{d} \mathcal{N}\left[\mathbf{0}, \left(\lim \frac{1}{N}\sum\nolimits_i 12\sigma^2\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\left(\lim \frac{1}{N}\sum\nolimits_i 240\sigma^6\mathbf{x}_i\mathbf{x}_i'\right)\left(\lim \frac{1}{N}\sum\nolimits_i 12\sigma^2\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\right] \\
&\xrightarrow{d} \mathcal{N}\left[\mathbf{0}, \frac{5\sigma^2}{3}\left(\lim \frac{1}{N}\sum\nolimits_i \mathbf{x}_i\mathbf{x}_i'\right)^{-1}\right], \quad \text{using } 240\sigma^6/(12\sigma^2 \times 12\sigma^2) = 20\sigma^2/12 = 5\sigma^2/3.
\end{aligned}$$

(d) Use $\widehat{\mathrm{V}}[\widehat{\boldsymbol{\beta}}] = (5s^2/3)[\sum_i \mathbf{x}_i\mathbf{x}_i']^{-1}$ where e.g. $s^2 = N^{-1}\sum_i (y_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}})^2$.

3.(a) This answer is longer than needed as I try to explain in detail what is going on.
Now using $\mathrm{E}[u_i] = 0$ and u_i iid so that $\mathrm{E}[u_i^k] = \mathrm{E}[u^k]$ we have

$$\begin{aligned}
Q_0(\boldsymbol{\beta}) &= \mathrm{E}[u^4] + 4\mathrm{E}[u^3]\lim \frac{1}{N}\sum\nolimits_i \mathbf{x}_i'(\boldsymbol{\beta}_0 - \boldsymbol{\beta}) + 6\mathrm{E}[u^2]\lim \frac{1}{N}\sum\nolimits_i [\mathbf{x}_i'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})]^2 \\
&\quad + 0 + \lim \frac{1}{N}\sum\nolimits_i [\mathbf{x}_i'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})]^4 \\
\frac{\partial Q_0(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= -4\mathrm{E}[u^3]\lim \frac{1}{N}\sum\nolimits_i \mathbf{x}_i'\boldsymbol{\beta} - 12\mathrm{E}[u^2]\lim \frac{1}{N}\sum_{i=1}^N [\mathbf{x}_i'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})]\mathbf{x}_i - 4\lim \frac{1}{N}\sum_{i=1}^N [\mathbf{x}_i'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})]^3\mathbf{x}_i.
\end{aligned}$$

Thus in general

$$\frac{\partial Q_0}{\partial \boldsymbol{\beta}}\bigg|_{\beta_0} = -4\mathrm{E}[u^3]\lim \frac{1}{N}\sum\nolimits_i \mathbf{x}_i \neq \mathbf{0}$$

so in general $\widehat{\boldsymbol{\beta}}$ will be inconsistent.
However, we were given information that u_i is symmetric, so $\mathrm{E}[u^3] = 0$.
Given this information $\partial Q_0/\partial \boldsymbol{\beta}|_{\boldsymbol{\beta}_0} = 0$ and hence the estimator is actually consistent.

(b) The simplifications $\mathrm{E}[{u_i}^3] = 15\sigma^6$ and $\mathrm{E}[{u_i}^2] = \sigma^2$ used to get $\mathbf{A}(\boldsymbol{\beta}_0)$ and $\mathbf{B}(\boldsymbol{\beta}_0)$ are no longer possible.
So use the sandwich estimator $N\widehat{\mathbf{A}}^{-1}\widehat{\mathbf{B}}\widehat{\mathbf{A}}^{-1}$ where

$$\begin{aligned}\widehat{\mathbf{A}} &= \frac{1}{N}\sum\nolimits_i 12{\widehat{u}_i}^2\mathbf{x}_i\mathbf{x}_i' \\ \widehat{\mathbf{B}} &= \frac{1}{N}\sum\nolimits_i 16{\widehat{u}_i}^6\mathbf{x}_i{\mathbf{x}_i}',\end{aligned}$$

where $\widehat{u}_i = \widehat{y}_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}}$.

4. Do Newton-Raphson $(\widehat{\boldsymbol{\beta}}_{s+1} - \widehat{\boldsymbol{\beta}}_s) = -\mathbf{H}_s^{-1}\mathbf{g}_s$, where

$$\begin{aligned}\mathbf{g}_s &= \left.\frac{\partial Q_N(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}\right|_{\widehat{\boldsymbol{\beta}}_s} = N^{-1}\sum_{i=1}^{N} -4(y_i - {\mathbf{x}_i}'\widehat{\boldsymbol{\beta}})^3\mathbf{x}_i \\ \mathbf{H}_s &= \left.\frac{\partial^2 Q_N(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right|_{\widehat{\boldsymbol{\beta}}_s} = N^{-1}\sum_{i=1}^{N} 12(\widehat{y}_i - \mathbf{x}_i'\widehat{\boldsymbol{\beta}})^2\mathbf{x}_i\mathbf{x}_i'.\end{aligned}$$

Alternatively method of scoring uses $\mathbf{E}\mathbf{H}_s = N^{-1}\sum_i 12s^2\mathbf{x}_i\mathbf{x}_i'$ where s^2 is estimate of σ^2.
Obvious starting value is the OLS estimator, which is consistent here.

5.(a) Since $p_i = \Lambda(\mathbf{x}_i'\boldsymbol{\beta}) = \exp\mathbf{x}_i'\boldsymbol{\beta}/(1+\exp(\mathbf{x}_i'\boldsymbol{\beta}))$ we have

$$\begin{aligned}\ln\mathcal{L}_N &= \sum_i \ln f(y_i) = \sum_i \ln[p_i^{y_i}(1-p_i)^{1-y_i}] \\ &= \sum_i y_i\ln p_i + (1-y_i)\ln(1-p_i) \\ &= \sum_i y_i\ln\Lambda(\mathbf{x}_i'\boldsymbol{\beta}) + (1-y_i)\ln(1-\Lambda(\mathbf{x}_i'\boldsymbol{\beta}) \\ &= \sum_i y_i\{\mathbf{x}_i'\boldsymbol{\beta} - \ln(1+\exp(\mathbf{x}_i'\boldsymbol{\beta}))\} + (1-y_i)\ln(1+\exp(\mathbf{x}_i'\boldsymbol{\beta}))\end{aligned}$$

(b)

$$\frac{\partial p_i}{\partial\mathbf{x}_i} = \frac{\partial\Lambda(\mathbf{x}_i'\boldsymbol{\beta})}{\partial\mathbf{x}_i}\boldsymbol{\beta} = \Lambda'(\mathbf{x}_i'\boldsymbol{\beta})\boldsymbol{\beta} = [\Lambda(\mathbf{x}_i'\boldsymbol{\beta})(1-\Lambda(\mathbf{x}_i'\boldsymbol{\beta})]\boldsymbol{\beta} = \frac{e^{\mathbf{x}_i'\boldsymbol{\beta}}}{(1+e^{\mathbf{x}_i'\boldsymbol{\beta}})^2}\boldsymbol{\beta},$$

where some algebra has been skipped.

The course grade will be based on a curve from the combined scores of midterm 1 (35%), final (50%) and assignments (15%).
The curve for this exam is only a guide to give you a rough idea of how you are doing.

Scores out of	35		A	28 and above
75th percentile	29	(83%)	A-	23 and above
Median	25	(71%)	B+	18 and above
25th percentile	22	(63%)	B	13 and above

Explanations and Common Errors

1.(a) I had not intended tricky math here. Thought it was straightforward.
(b)-(c) The key here is that $\widehat{\boldsymbol{\beta}}$ appears nowhere in the answer !!!!
We find $Q_0(\boldsymbol{\beta}) = \text{plim}\, Q_N(\boldsymbol{\beta})$ and determine whether $Q_0(\boldsymbol{\beta})$ is maximized at $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. As simple as that. This is one of two things I expect you to understand by the end of the course.
(d) "Formally apply a LLN" means just that.
Here the X_i of the theorem is $X_i = 4u_i[\mathbf{x}_i'(\boldsymbol{\beta}_0 - \boldsymbol{\beta})]^3$. Need to start with this.
Then this is inid since $\mathbf{x}_i$ is nonstochastic and different $\mathbf{x}_i$ values will change the distribution of X_i.
[If instead both $\mathbf{x}_i$ and u_i were iid then unconditional on both X_i would be iid].

2. The first-order Taylor series expansion of the f.o.c. etcetera yields

$$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = -\left(\left. \frac{\partial^2 Q_N(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} \right|_{\boldsymbol{\beta}_0} \right)^{-1} \sqrt{N} \left. \frac{\partial Q_N(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} \right|_{\boldsymbol{\beta}_0}.$$

(a)-(b) Then

$$\sqrt{N} \left. \frac{\partial Q_N(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} \right|_{\boldsymbol{\beta}_0} \xrightarrow{d} \mathcal{N}\left[\mathbf{0}, \mathbf{B}_0 = \text{plim}\left[N \left. \frac{\partial Q_N(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} \frac{\partial Q_N(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}'} \right|_{\boldsymbol{\beta}_0}\right] = \lim \mathrm{E}\left[N \left. \frac{\partial Q_N(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} \frac{\partial Q_N(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}'} \right|_{\boldsymbol{\beta}_0}\right]\right]$$

and

$$\mathbf{A}_0 = \text{plim} \left. \frac{\partial^2 Q_N(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} \right|_{\boldsymbol{\beta}_0} = \lim \mathrm{E}\left[\left. \frac{\partial^2 Q_N(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'} \right|_{\boldsymbol{\beta}_0} \right].$$

(c) Combining gives

$$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} \mathcal{N}[\mathbf{0}, \mathbf{A}_0^{-1}\mathbf{B}_0\mathbf{A}_0^{-1}],$$

(d) and in practice

$$\widehat{\boldsymbol{\beta}} \overset{a}{\sim} \mathcal{N}[\boldsymbol{\beta}_0, N\mathbf{A}_0^{-1}\mathbf{B}_0\mathbf{A}_0^{-1}],$$

and to implement we use $N\widehat{\mathbf{A}}^{-1}\widehat{\mathbf{B}}\widehat{\mathbf{A}}^{-1}$ where $\widehat{\mathbf{A}}$ and $\widehat{\mathbf{B}}$ **do not depend on unknowns.**
The preceding is the other thing I expect you to understand by the end of the course.

Note that the final answer was 5/3 times the variance of the OLS estimator. For this problem with normal errors the MLE = OLS. So the estimator of question 1 is inefficient compared to the fully efficient OLS.

3.(a) A detailed answer is given above. Some explanation / intuition follows.
In general if model is misspecified then the estimator will be inconsistent.
In special cases: OLS and LEF models correct specification of mean is enough (i.e. $\mathrm{E}[u] = 0$).
For the estimator in question 1 what is needed for consistency is both zero mean error and third moment of error zero: (i.e. $\mathrm{E}[u] = 0$ and (i.e. $\mathrm{E}[u^3] = 0$)]. So more than just mean correct, but less than fully correct distribution. Symmetry of error will do.

3.(b) Many people got the right start - use sandwich errors. But then failed to answer for question 1 estimator. Instead gave result for OLS which is not relevant here.

4. Many people gave NR but then ailed to give details for estimator of question 1.
Very few suggested OLS as good starting estimate for this particular problem.

5.(b) We want $\partial \Pr[y_i|\mathbf{x}_i]/\partial\mathbf{x}_i$ or $\partial \Pr[y|\mathbf{x}]/\partial\mathbf{x}$ for all of vector $\mathbf{x}$
or $\partial \Pr[y_i|\mathbf{x}_{ij}]/\partial x_{ij}$ or $\partial \Pr[y|\mathbf{x}]/\partial x_j$ for the j^{th} component.