240D Fall 2005 Solutions to Final Exam

1.(a) Suppose for model $y_i = \exp(\mathbf{x}_i'\boldsymbol{\beta}) + u_i$ we have $\mathrm{E}[u_i|\mathbf{x}_i] \neq 0$ but $\mathrm{E}[u_i|\mathbf{z}_i] = 0$.
Then $\mathrm{E}[\mathbf{z}_i u_i] = \mathrm{E}[\mathbf{z}_i(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))] = \mathbf{0}$.
There are m equations in only k unknowns (as $m > k$) so the model is overidentified.
This is a GMM setting so we minimize a quadratic form.
For this question the weighting matrix is the identity matrix and we minimize $Q_N(\boldsymbol{\beta})$ given in the question.

(b) Differentiating yields

$$\begin{aligned}
\frac{\partial Q_N(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= 2\left(\frac{\partial}{\partial \boldsymbol{\beta}}\left(\frac{1}{N}\sum\nolimits_i \mathbf{z}_i(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))'\right)\left(\frac{1}{N}\sum\nolimits_i \mathbf{z}_i(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))\right)\right) \\
&= 2 \times \frac{1}{N}\sum\nolimits_i \frac{\partial(\mathbf{z}_i(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))')}{\partial \boldsymbol{\beta}} \times \frac{1}{N}\sum\nolimits_i \mathbf{z}_i(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta})) \\
&= 2 \times \frac{1}{N}\sum\nolimits_i \frac{\partial(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))}{\partial \boldsymbol{\beta}}\mathbf{z}_i' \times \frac{1}{N}\sum\nolimits_i \mathbf{z}_i(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta})) \\
&= -2 \times \frac{1}{N}\sum\nolimits_i \exp(\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i\mathbf{z}_i' \times \frac{1}{N}\sum\nolimits_i \mathbf{z}_i(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta})).
\end{aligned}$$

Cancelling out the scalar multiple -2 gives the f.o.c. in part (b).

(c) Part (b) upon multiplication by $\sqrt{N}$ becomes

$$\begin{aligned}
\frac{1}{N}\sum\nolimits_i \exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})\mathbf{x}_i\mathbf{z}_i' \times \sqrt{N}\frac{1}{N}\sum\nolimits_i \mathbf{z}_i(y_i - \exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})) &= \mathbf{0} \\
\text{or } \mathbf{G}_N(\widehat{\boldsymbol{\beta}}) \times \sqrt{N}\mathbf{g}_N(\widehat{\boldsymbol{\beta}}) &= \mathbf{0},
\end{aligned}$$

where $\mathbf{G}_N(\boldsymbol{\beta})' = N^{-1}\sum_i \exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})\mathbf{x}_i\mathbf{z}_i'$ and $\mathbf{g}_N(\boldsymbol{\beta}) = N^{-1}\sum_i \mathbf{z}_i(y_i - \exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})$.

(d) By an exact first-order Taylor series expansion

$$\begin{aligned}
\mathbf{g}_N(\widehat{\boldsymbol{\beta}}) &= \mathbf{g}_N(\boldsymbol{\beta}_0) + \left.\frac{\partial \mathbf{g}_N(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'}\right|_{\boldsymbol{\beta}^+}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\
&= \mathbf{g}_N(\boldsymbol{\beta}_0) + \mathbf{G}_N(\boldsymbol{\beta}^+)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0).
\end{aligned}$$

(e) Substitute this back in (c).

$$\begin{gathered}
\mathbf{G}_N(\widehat{\boldsymbol{\beta}})' \times \left[\sqrt{N}\mathbf{g}_N(\boldsymbol{\beta}_0) + \mathbf{G}_N(\boldsymbol{\beta}^+)\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\right] = \mathbf{0} \\
\Rightarrow \sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \left[\mathbf{G}_N(\widehat{\boldsymbol{\beta}})'\mathbf{G}_N(\boldsymbol{\beta}^+)\right]^{-1}\mathbf{G}_N(\widehat{\boldsymbol{\beta}})'\sqrt{N}g_N(\boldsymbol{\beta}_0) \\
\xrightarrow{d} [\mathbf{G}_0'\mathbf{G}_0]^{-1} \times \mathcal{N}[\mathbf{0}, \mathbf{S}_0] \\
\xrightarrow{d} \mathcal{N}[[\mathbf{G}_0'\mathbf{G}_0]^{-1}\mathbf{G}_0'\mathbf{S}_0\mathbf{G}_0[\mathbf{G}_0'\mathbf{G}_0]^{-1}]
\end{gathered}$$

where we assume $\sqrt{N}g_N(\boldsymbol{\beta}_0) \xrightarrow{d} \mathcal{N}[\mathbf{0}, \mathbf{S}_0]$ and plim $\mathbf{G}_N(\widehat{\boldsymbol{\beta}}) = \mathbf{G}_0$.

(f) No, unless $\mathbf{S}_0 = p\mathbf{I}_k$ for some multiple p.
Optimal GMM minimizes

$$\left(\frac{1}{N}\sum\nolimits_i \mathbf{z}_i(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))'\right)'\widehat{\mathbf{S}}^{-1}\left(\frac{1}{N}\sum\nolimits_i \mathbf{z}_i(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta}))'\right)$$

where $\widehat{\mathbf{S}} \xrightarrow{p} \mathbf{S}_0$. For this example and with heteroskedastic errors $\widehat{\mathbf{S}} = \frac{1}{N}\sum_i \widehat{u}_i^2\mathbf{z}_i\mathbf{x}_i'$.

2.(a) In general for binary data
$L_N(\boldsymbol{\beta}) = \prod_{i=1}^{N} f(y_i) = \sum_i \ln f(y_i) = \sum_i \ln[p_i^{y_i}(1-p_i)^{y_i}] = \sum_i [y_i \ln p_i + (1-y_i)\ln(1-p_i)]$.
Here $p_i = \Phi(\mathbf{x}_i'\boldsymbol{\beta}) = \Phi_i$ since probit model.
Substituting and using $\Phi'(\mathbf{x}_i'\boldsymbol{\beta}) = \boldsymbol{\phi}(\mathbf{x}_i'\boldsymbol{\beta}) = \boldsymbol{\phi}_i$ gives f.o.c.

$$\sum_i \frac{y_i}{\Phi_i}\phi_i \mathbf{x}_i + \frac{1-y_i}{1-\Phi_i}\phi_i \mathbf{x}_i = \mathbf{0}.$$

This is an adequate answer. It simplifies to $\sum_i \frac{y_i - \Phi_i}{\Phi_i(1-\Phi_i)}\phi_i \mathbf{x}_i = \mathbf{0}$.

(b) $\Pr(y=1) = \Phi(\beta_1 + \beta_2 x)$, so the marginal effect

$$\frac{\partial \Pr(y=1)}{\partial x} = \beta_2 \phi(\beta_1 + \beta_2 x) = 0.2\phi(0.4 + 0.2x) = 0.2\phi(1) = 0.2 \times [(1/\sqrt{2\pi})e^{-0.5}]0.048,$$

when evaluated at the sample mean $\bar{x} = 0.3$.

(c) $\Pr[y_i = 1] = \Pr[U_{i1} > U_{i0}] = \Pr[\mathbf{x}_i'\boldsymbol{\beta}_1 + \varepsilon_{i1} > \mathbf{x}_i'\boldsymbol{\beta}_0 + \varepsilon_{i0}] = \Pr[(\varepsilon_{i0} - \varepsilon_{i1}) < \mathbf{x}_i'(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)]$.
Here $\varepsilon_{i1} - \varepsilon_{i0} \sim \mathcal{N}[0,2]$ since ε_{i0} and ε_{i1} are iid $\mathcal{N}[0,1]$.
[Formally $\mathrm{V}[\varepsilon_{i1} - \varepsilon_{i0}] = \mathrm{V}[\varepsilon_{i1}] + \mathrm{V}[\varepsilon_{i0}] - 2\mathrm{Cov}[\varepsilon_{i0}, \varepsilon_{i1}] = 1 + 1 - 2 \times 0 = 2$].
So

$$\Pr[y_i = 1] = \Pr\left[\frac{(\varepsilon_{i1} - \varepsilon_{i0})}{\sqrt{2}} \leq \frac{\mathbf{x}_i'(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)}{\sqrt{2}}\right] = \Phi\left(\frac{\mathbf{x}_i'(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)}{\sqrt{2}}\right).$$

(d) Let the ordered outcomes by $y_i = 1, 2$ and 3.
The latent variable is $y_i^* = \mathbf{x}_i'\boldsymbol{\beta} + u_i$, where u_i is logistic distributed.
Then $\ln L = \prod_{i=1}^{N} f(y_i) = \sum_i \ln f(y_i) = \sum_i \ln[p_{1i}^{y_{1i}} p_{2i}^{y_{2i}} p_{3i}^{y_{3i}}] = \sum_i y_{1i}\ln p_{1i} + y_{2i}\ln p_{2i} + y_{32i}\ln p_{3i}]$
where y_{1i}, $y_{2i}$and y_{3i} are indicator variables with
$p_{1i} = \Pr[y_i = 1] = \Pr[y_i^* \leq \alpha_1] = \Pr[\mathbf{x}_i'\boldsymbol{\beta} + u_i \leq \alpha_1] = \Pr[u_i < \alpha_1 - \mathbf{x}_i'\boldsymbol{\beta}] = \Lambda(\alpha_1 - \mathbf{x}_i'\boldsymbol{\beta})$
$p_{2i} = \Pr[y_i = 2] = \Pr[\alpha_1 < y_i^* \leq \alpha_{2i}] = \Pr[y_i^* \leq \alpha_{2i}] - \Pr[y_i^* \leq \alpha_{2i}] = \Lambda(\alpha_2 - \mathbf{x}_i'\boldsymbol{\beta}) - \Lambda(\alpha_1 - \mathbf{x}_i'\boldsymbol{\beta})$
$p_{3i} = 1 - p_{1i} - p_{2i} = 1 - \Lambda(\alpha_2 - \mathbf{x}_i'\boldsymbol{\beta})$
If an intercept is included in $\mathbf{x}_i$ then only one of α_1 and α_2 are identified.
If an intercept is not included in $\mathbf{x}_i$ then both α_1 and α_2 are identified.

3.(a) We have

$$\begin{aligned}
\mathrm{E}[y|d=1] &= \mathrm{E}[\mathbf{x}'\boldsymbol{\beta} + \alpha d + \varepsilon | d = 1] \\
&= \mathrm{E}[\mathbf{x}'\boldsymbol{\beta} + \alpha + \varepsilon | d = 1] \\
&= \mathbf{x}'\boldsymbol{\beta} + \alpha + \mathrm{E}[\varepsilon | d = 1] \\
&= \mathbf{x}'\boldsymbol{\beta} + \alpha + \mathrm{E}[\varepsilon | \mathbf{w}'\boldsymbol{\gamma} + v > 0] \\
&= \mathbf{x}'\boldsymbol{\beta} + \alpha + \mathrm{E}[\varepsilon | v > -\mathbf{w}'\boldsymbol{\gamma}].
\end{aligned}$$

(b) Now suppose $\varepsilon = \delta v + u$ where u is independent of v and has mean zero

$$\begin{aligned}
\mathrm{E}[\varepsilon | v > -\mathbf{w}'\boldsymbol{\gamma}] &= \mathrm{E}[\delta v + u | v > -\mathbf{w}'\boldsymbol{\gamma}] \\
&= \mathrm{E}[\delta v | v > -\mathbf{w}'\boldsymbol{\gamma}] + \mathrm{E}[u | v > -\mathbf{w}'\boldsymbol{\gamma}] \\
&= \delta \mathrm{E}[v | v > -\mathbf{w}'\boldsymbol{\gamma}] + \mathrm{E}[u] \text{ as } u \text{ is independent of } v \\
&= \delta \mathrm{E}[v | v > -\mathbf{w}'\boldsymbol{\gamma}] \text{ as } \mathrm{E}[u] = 0 \\
&= \delta \phi(\mathbf{w}'\boldsymbol{\gamma}) / \Phi(\mathbf{w}'\boldsymbol{\gamma})
\end{aligned}$$

where the last line uses the result given for censored standard normal. So

$$\mathrm{E}[y|d=1] = \mathbf{x}'\boldsymbol{\beta} + \alpha + \delta\phi(\mathbf{w}'\boldsymbol{\gamma}) / \Phi(\mathbf{w}'\boldsymbol{\gamma}).$$

(c) From probit of d on $\mathbf{w}$ we get $\widehat{\boldsymbol{\gamma}}$ and $\phi(\mathbf{w}'\widehat{\boldsymbol{\gamma}})/\Phi(\mathbf{w}'\widehat{\boldsymbol{\gamma}})$.
Then for those with $d = 1$ do OLS of y on $\mathbf{x}$ and $\phi(\mathbf{w}'\widehat{\boldsymbol{\gamma}})/\Phi(\mathbf{w}'\widehat{\boldsymbol{\gamma}})$.

(d) No. $\mathbf{x}$ will also include an intercept with coefficient β_1, say. So cannot separately distinguish between β_1 and α.
[Aside: However, if we also have data on $d = 1$ then we can get similar expression for $\mathrm{E}[y|d = 0]$ and estimate α using data on both those with $d = 1$ and those with $d = 0$.]

4.(a) The usual t or z test statistic is 5.664/1.741 = 3.253. This exceeds $z_{.025} = 1.96$ in absolute value. So reject H_0.

(b) The bootstrap standard error is the standard deviation of $\widehat{\beta}^*$ which is 1.939.
This yields $t = 5.664/1.939 = 2.921$. This exceeds $z_{.025} = 1.96$ in absolute value. So reject H_0.

(c) The bootstrap refinement is based on $t^* = (\widehat{\beta}^* - \widehat{\beta})/s_{\widehat{\beta}} = (\widehat{\beta}^*_b - 5.664)/1.741$.
The 2.5 and 97.5 percentiles are -2.183 and 2.066 .
The original t-statistic of 3.253 falls outside these percentiles. So reject H_0.

5.(a)

$$\begin{aligned} y_{it} &= \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it} \\ \Rightarrow \quad \bar{y}_i &= \alpha_i + \bar{\mathbf{x}}_i'\boldsymbol{\beta} + \bar{\varepsilon}_i \\ \Rightarrow \quad y_{it} - \bar{y}_i &= (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_i) \end{aligned}$$

Do OLS on this equation

$$\widehat{\boldsymbol{\beta}}_{\mathrm{W}} = \left[\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)'\right]^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(y_{it} - \bar{y}_i).$$

(b) Stack model as

$$\widetilde{\mathbf{y}}_i = \widetilde{\mathbf{X}}_i\boldsymbol{\beta} + \widetilde{\boldsymbol{\varepsilon}}_i,$$

where $\widetilde{\mathbf{y}}_i$, $\widetilde{\mathbf{X}}_i$ and $\widetilde{\boldsymbol{\varepsilon}}_i$ have i^{th} entries $y_{it} - \bar{y}_i$ and $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$ and $(\varepsilon_{it} - \bar{\varepsilon}_i)$.
Then

$$\widehat{\boldsymbol{\beta}}_{\mathrm{W}} = \left[\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i\right]^{-1}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{y}}_i.$$

(c) By generalization of White (1980)

$$\widehat{\mathrm{V}}\left[\widehat{\boldsymbol{\beta}}_{\mathrm{W}}\right] = \left[\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i\right]^{-1}\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\widehat{\widetilde{\boldsymbol{\varepsilon}}}_i\widehat{\widetilde{\boldsymbol{\varepsilon}}}_i'\widetilde{\mathbf{X}}_i\left[\sum_{i=1}^{N}\widetilde{\mathbf{X}}_i'\widetilde{\mathbf{X}}_i\right]^{-1},$$

where $\widehat{\widetilde{\boldsymbol{\varepsilon}}}_i$ is $\widetilde{\boldsymbol{\varepsilon}}_i$ evaluated at $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}_{\mathrm{W}}$.

(d) No. It is consistent. But it is not fully efficient as this is the situation where the random effects estimator is fully efficient.

(e) Yes. It is consistent. Furthermore it can be shown to be fully efficient in this situation.

The curve for this exam is only a guide. Course grade based on course score.

Scores out of	50	
$75th$ percentile	42	(84%)
Median	39	(78%)
$25th$ percentile	38	(76%)

A	40 and above
A-	34 and above
B+	28 and above

For course the curve is

A+	92 and above
A	79 and above
A-	66 and above
B+	53 and above

Comments on Exam

2. Very few got part 2(d) on ordered logit. I realize now that I did not cover this model in the class, but it was on the reading list and it was part of assignment 4.
To not have too much weight placed on the ordered logit, I graded 2(d) out of 2 only; question 2 out of 10 not 12; and then added 2 points to get a score out of 50.

4. There was a typo. t_b^* should have been $t_b^* = (\widehat{\beta}_b^* - 5.664)/s_{\widehat{\beta}_b^*}$.