240D Winter 2012 Solutions to Final Exam

1.(a) Here $\ln L(\boldsymbol{\beta},\alpha)=\sum_i \ln f(y_i)=\sum_i \left\{(\alpha-1)\ln y_i-\frac{y_i}{\exp(\mathbf{x}_i'\boldsymbol{\beta})}-\alpha\mathbf{x}_i'\boldsymbol{\beta}-\ln\Gamma(\alpha)\right\}$

(b) Differentiation yields

$$\begin{aligned}\frac{\partial \ln L}{\partial \boldsymbol{\beta}} &= \sum\nolimits_i \left(\frac{y_i}{\exp(\mathbf{x}_i'\boldsymbol{\beta})}\mathbf{x}_i-\alpha\mathbf{x}_i\right)=\sum\nolimits_i\left(\frac{y_i-\alpha\exp(\mathbf{x}_i'\boldsymbol{\beta})}{\exp(\mathbf{x}_i'\boldsymbol{\beta})}\mathbf{x}_i\right)=\mathbf{0}.\\ \frac{\partial \ln L}{\partial \alpha} &= \sum\nolimits_i\left(\ln y_i-\mathbf{x}_i'\boldsymbol{\beta}-\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}\right)=0.\end{aligned}$$

(c) Easiest to derive the outer product of the gradient estimate $\widehat{\mathbf{B}}^{-1}$. This yields for $\boldsymbol{\theta}=[\boldsymbol{\beta}'\ \alpha]'$.

$$\widehat{\mathbf{V}}[\widehat{\boldsymbol{\theta}}]=\begin{bmatrix}\sum_i\left(\frac{y_i-\widehat{\alpha}\exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})}{\exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})}\right)^2\mathbf{x}_i\mathbf{x}_i' & \sum_i\left(\ln y_i-\mathbf{x}_i'\widehat{\boldsymbol{\beta}}-\frac{\Gamma'(\widehat{\alpha})}{\Gamma(\widehat{\alpha})}\right)\left(\frac{y_i-\widehat{\alpha}\exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})}{\exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})}\right)\mathbf{x}_i\\ \sum_i\left(\frac{y_i-\widehat{\alpha}\exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})}{\exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})}\right)\left(\ln y_i-\mathbf{x}_i'\widehat{\boldsymbol{\beta}}-\frac{\Gamma'(\widehat{\alpha})}{\Gamma(\widehat{\alpha})}\right)\mathbf{x}_i & \sum_i\left(\ln y_i-\mathbf{x}_i'\widehat{\boldsymbol{\beta}}-\frac{\Gamma'(\widehat{\alpha})}{\Gamma(\widehat{\alpha})}\right)^2\end{bmatrix}^{-1}$$

Or can use Hessian which $-\widehat{\mathbf{A}}^{-1}$ yields after some algebra yields

$$\widehat{\mathbf{V}}[\widehat{\boldsymbol{\theta}}]=\begin{bmatrix}\sum_i\left(\frac{y_i}{\exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})}\right)\mathbf{x}_i\mathbf{x}_i' & \sum_i\mathbf{x}_i\\ \sum_i\mathbf{x}_i & \sum_i\left(\frac{\Gamma''(\widehat{\alpha})}{\Gamma(\widehat{\alpha})}-\frac{\Gamma'(\widehat{\alpha})^2}{\Gamma(\widehat{\alpha})^2}\right)\end{bmatrix}^{-1}$$

Note: In general we use $\left(-\mathrm{E}\left[\frac{\partial^2\ln L}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\right]\right)^{-1}$. Here

$$\left(\begin{bmatrix}\mathrm{E}\left[\frac{\partial^2\ln L}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right] & \mathrm{E}\left[\frac{\partial^2\ln L}{\partial\alpha\partial\boldsymbol{\beta}'}\right]\\ \mathrm{E}\left[\frac{\partial^2\ln L}{\partial\boldsymbol{\beta}\partial\boldsymbol{\alpha}}\right] & \mathrm{E}\left[\frac{\partial^2\ln L}{\partial\alpha^2}\right]\end{bmatrix}\right)^{-1}\neq-\begin{bmatrix}\left(\mathrm{E}\left[\frac{\partial^2\ln L}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right]\right)^{-1} & \left(\mathrm{E}\left[\frac{\partial^2\ln L}{\partial\alpha\partial\boldsymbol{\beta}'}\right]\right)^{-1}\\ \left(\mathrm{E}\left[\frac{\partial^2\ln L}{\partial\boldsymbol{\beta}\partial\boldsymbol{\alpha}}\right]\right)^{=1} & \left(\mathrm{E}\left[\frac{\partial^2\ln L}{\partial\alpha^2}\right]\right)^{-1}\end{bmatrix}$$

except in the special case that $\mathrm{E}\left[\frac{\partial^2\ln L}{\partial\boldsymbol{\beta}\partial\boldsymbol{\alpha}}\right]=\mathbf{0}$.

(d) In general the MLE for both $\boldsymbol{\beta}$ and α will be inconsistent.
Here there is some hope that MLE for $\boldsymbol{\beta}$ may be consistent, since $\mathrm{E}[\partial\ln L/\partial\boldsymbol{\beta}]=\mathbf{0}$ requires only correct specification of the mean (then $\mathrm{E}\left[\sum_i\frac{y_i-\alpha\exp(\mathbf{x}_i'\boldsymbol{\beta})}{\exp(\mathbf{x}_i'\boldsymbol{\beta})}\mathbf{x}_i\right]=0$). [Half credit for saying this].
But $\mathrm{E}[\partial\ln L/\partial\alpha]=0$ requires the much stronger assumption that $\mathrm{E}[\ln y_i]=\mathbf{x}_i'\boldsymbol{\beta}+\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}$
(then $\mathrm{E}\left[\sum_i\left(\ln y_i-\mathbf{x}_i'\boldsymbol{\beta}-\frac{\Gamma'(\alpha)}{\Gamma(\alpha)}\right)\right]=0$).
This fails and the two equations jointly estimated will yield inconsistent estimates.
One way to see this is that $\widehat{\alpha}$ inconsistent then contaminates $\boldsymbol{\beta}$ that solves $\sum_i\left(\frac{y_i-\widehat{\alpha}\exp(\mathbf{x}_i'\boldsymbol{\beta})}{\exp(\mathbf{x}_i'\boldsymbol{\beta})}\mathbf{x}_i\right)=\mathbf{0}$.
More formally, the information matrix is not block-diagonal as $\mathrm{E}[\partial^2\ln L/\partial\boldsymbol{\beta}\partial\alpha]\neq\mathbf{0}$ and estiamtion of α effects etimation of $\boldsymbol{\beta}$.

(e) Two possible methods are based on $\mathrm{E}[y_i|\mathbf{x}_i]=\exp(\mathbf{x}_i'\boldsymbol{\beta})$ are
NLS of y_i on $\exp(\mathbf{x}_i'\boldsymbol{\beta})$ which minimizes $\sum_i(y_i-\exp(\mathbf{x}_i'\boldsymbol{\beta}))^2$.
MM estimation based on $\mathrm{E}[(y_i-\exp(\mathbf{x}_i'\boldsymbol{\beta}))\mathbf{x}_i]=\mathbf{0}$ which solves $\sum_i(y_i-\exp(\mathbf{x}_i'\boldsymbol{\beta}))\mathbf{x}_i=\mathbf{0}$.

(f) Here $\mathrm{E}[y]=\alpha\lambda$ and $\mathrm{V}[y]=\alpha\lambda^2$.
So $\mathrm{E}[\mathbf{x}(y-\alpha\exp(\mathbf{x}'\boldsymbol{\beta}))]=0$ and $\mathrm{E}[(y-\alpha\exp(\mathbf{x}'\boldsymbol{\beta}))^2-1]=0$.
Let $h(y_i,\mathbf{x}_i,\alpha,\boldsymbol{\beta})=[(\mathbf{x}_i(y_i-\alpha\exp(\mathbf{x}_i'\boldsymbol{\beta})))'\quad((y_i-\alpha\exp(\mathbf{x}_i'\boldsymbol{\beta}))^2-1)]'$.
The GMM estimator minimizes

$$Q_N(\alpha,\boldsymbol{\beta})=\frac{1}{N}\left(\sum\nolimits_i h(y_i,\mathbf{x}_i,\alpha,\boldsymbol{\beta})\right)'\mathbf{W}_N\left(\sum\nolimits_i h(y_i,\mathbf{x}_i,\alpha,\boldsymbol{\beta})\right),$$

where any full rank weighting matrix will do since this is just-identified.

2.(a) Here $\Pr[y=0] = \Pr[y^*=0] = e^{-\mu} = \exp(-\exp(\mathbf{x}'\boldsymbol{\beta}))$. So

$$\Pr[y=1] = 1 - \Pr[y=0] = 1 - \exp(-\exp(\mathbf{x}'\boldsymbol{\beta})).$$

Estimate by binary MLE. $\widehat{\boldsymbol{\beta}}$ maximizes $L_N(\boldsymbol{\beta}) = \sum_i y_i \ln(1 - \exp(-\exp(\mathbf{x}'\boldsymbol{\beta})) + (1-y_i)\ln(\exp(-\exp(\mathbf{x}'\boldsymbol{\beta})))$.

(b) This is ordered model

$$\begin{aligned} p_0 &= \Pr[y=0] = \Pr[y^*=0] = e^{-\mu} = \exp(-\exp(\mathbf{x}'\boldsymbol{\beta})). \\ p_1 &= \Pr[y=1] = \Pr[y^*=1] = \mu e^{-\mu} = \exp(\mathbf{x}'\boldsymbol{\beta})\exp(-\exp(\mathbf{x}'\boldsymbol{\beta})). \\ p_2 &= \Pr[y=2] = 1 - p_0 - p_1. \end{aligned}$$

Estimate by multinomial MLE. $\widehat{\boldsymbol{\beta}}$ maximizes $L_N(\boldsymbol{\beta}) = \sum_i (y_{0i}\ln p_{0i} + y_{1i}\ln p_{1i} + y_{2i}\ln p_{2i})$ where $y_{0i} = 1$ if $y_i = 0$, $y_{1i} = 1$ if $y_i = 1$, $y_{2i} = 1$ if $y_i = 2$.

(c) For notational simplicity initially suppress conditioning on $\mathbf{x}$

$$f(y) = f(y^*|y^*\geq 1) = \frac{f(y^*)}{\Pr[y^*\geq 1]} = \frac{e^{-\mu}\mu^y/y^*!}{(1-\Pr[y^*=0])} = \frac{e^{-\mu}\mu^y/y^*!}{(1-e^{-\mu})}$$

So

$$\ln f(y|\mathbf{x}) = -\exp(\mathbf{x}_i'\boldsymbol{\beta}) + y_i\mathbf{x}_i'\boldsymbol{\beta} - \ln y_i! - \ln(1 - e^{-\exp(\mathbf{x}_i'\boldsymbol{\beta})}).$$

(d) Very few got this.

$$\begin{aligned} \mathrm{E}[y] &= \mathrm{E}[y^*|y^*\geq 1] \\ &= \sum\nolimits_{y^*=1}^{\infty} \frac{y^* f(y^*)}{\Pr[y^*\geq 1]} = \frac{1}{\Pr[y^*\geq 1]}\sum\nolimits_{y^*=1}^{\infty} y^* f(y^*) = \frac{1}{\Pr[y^*\geq 1]}\sum\nolimits_{y^*=0}^{\infty} y^* f(y^*) = \frac{1}{1-e^{-\mu}}\mu, \end{aligned}$$

using $\sum_{y^*=0}^{\infty} y^* f(y^*)$ is $\mathrm{E}[y^*]$ and we were told that for the Poisson that $\mathrm{E}[y^*] = \mu$.
(e) Since

$$\mathrm{E}[y_i|\mathbf{x}_i] = \frac{\exp(\mathbf{x}_i'\boldsymbol{\beta})}{1 - e^{-\exp(\mathbf{x}_i'\boldsymbol{\beta})}}$$

do nonlinear least squares regression of y_i on $\exp(\mathbf{x}_i'\boldsymbol{\beta})/(1 - e^{-\exp(\mathbf{x}_i'\boldsymbol{\beta})})$.
Or do MM based on $\sum_i \mathbf{x}_i(y_i - \exp(\mathbf{x}_i'\boldsymbol{\beta})/(1 - e^{-\exp(\mathbf{x}_i'\boldsymbol{\beta})})) = \mathbf{0}$.

3.(a) A sequence of random variables $\{b_N\}$ converges in probability to b if for any $\varepsilon > 0$ and $\delta > 0$, there exists $N^* = N^*(\varepsilon, \delta)$ such that for all $N > N^*$, $\Pr[|b_N - b| < \varepsilon] > 1 - \delta$.

(b) Remarkably dew got this completely correct. Simplest is Lindeberg-Levy CLT.
Let $\{X_i\}$ be iid with $\mathrm{E}[X_i] = \mu$ and $\mathrm{V}[X_i] = \sigma^2$. Then $Z_N = \frac{\bar{X}_N - \mu}{\sigma/\sqrt{N}} \xrightarrow{d} \mathcal{N}[0,1]$.
[Other CLT's can be given].

(c) $y^* = 1 + 2x + u$ where $x \sim \mathrm{N}[0,1]$ and $u \sim \mathrm{N}[0, x^2]$
We observe $y = 1$ if $y^* > 0$ and $y = 0$ if $y^* \leq 0$.

(d) In (c) I had meant to generate y from a Tobit model but mistekenly generated a binary variable. So the natural thing would be to try probit estimation. Tobit is inappropriate.
But I gave fiull vredit if you thought Tobit was still apropriate, but then noted that the Tobit MLE of y on x will be **inconsistent** for $\boldsymbol{\beta}$ as the error here is heteroskedastic. It is not enough to say that standard errors will be wrong. Inconsistency is the most serious problem.

(e) I had intended the question to be about the sample selection model, but if you answered correctly for the Tobit model you also got full credit. The sample selection model is

$$\begin{aligned} y_1^* &= \mathbf{x}_1'\boldsymbol{\beta}_1 + \varepsilon_1 \\ y_2^* &= \mathbf{x}_2'\boldsymbol{\beta}_2 + \varepsilon_2, \end{aligned}$$

and we observe $y_1 = \begin{cases} 1 & \text{if } y_1^* > 0 \\ 0 & \text{if } y_1^* \leq 0, \end{cases}$ and $y_2 = \begin{cases} y_2^* & \text{if } y_1^* > 0 \\ - & \text{if } y_1^* \leq 0. \end{cases}$

The errors $(\varepsilon_1, \varepsilon_2)$ have means $(0,0)$, variances $(1, \sigma_2^2)$ and covariance $\rho\sigma_2^2$.
ε_1 is standard normal. If the MLE is used $(\varepsilon_1, \varepsilon_2)$ are joint normal.

(f) B times do the following.
- Completely resample with replacement all the data $\{(y_{1i}, y_{2i}, \mathbf{x}_{1i}, \mathbf{x}_{2i}), i = 1, ..., N\}$
- For each resample get estimate $\widehat{\boldsymbol{\beta}}_b$ and form $\widehat{\text{ME}}_b = \exp(\bar{\mathbf{x}}_b' \widehat{\boldsymbol{\beta}}_b)$.

Standard error is the standard deviation of the B s $\widehat{\text{ME}}_b' s$.

(g) This is optimal two-step GMM. Minimize

$$Q_N(\boldsymbol{\theta}) = \frac{1}{N} \left(\sum_i h(\mathbf{w}_i, \boldsymbol{\theta}) \right)' \widehat{\mathbf{S}}^{-1} \left(\sum_i h(\mathbf{w}_i, \boldsymbol{\theta}) \right),$$

where $\widehat{\mathbf{S}} = \sum_{i=1}^N h(\mathbf{w}_i, \widetilde{\boldsymbol{\theta}}) h(\mathbf{w}_i, \widetilde{\boldsymbol{\theta}})'$ and $\widetilde{\boldsymbol{\theta}}$ is a consistent initial estimate such as first-step GMM.

4.(a) No. The default se's assume independence of $u_{it}$and u_{is}. But the error u_{it} is likely to be positively correlated with u_{is}, $i \neq s$, decreasing the informational content of the data. Panel robust se's adjust for this.

(b) Yes. The RE-GLS does control for clustering so might expoect the two to be similar. The difference is due to the wrong model for clustered errors (equicorrelation) or heteroskedasticity.

(c) $y_{it} = \alpha_i + \mathbf{x}_{it}' \boldsymbol{\beta} + u_{it} \Rightarrow (y_{it} - \bar{y}_i) = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \boldsymbol{\beta} + (u_{it} - \bar{u}_i)$.
So do OLS of $(y_{it} - \bar{y}_i)$ on $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$. (Other methods are possible).

(d) `xtreg y x, vce(robust)` or `xtreg y x, vce(Cluster id)`.

(e) That the RE estimator is fully efficient under H_0. This requires that the error $y_{it} = \alpha_i + \varepsilon_{it}$ where both α_i and ε_{it} are i.i.d.

(f) Usual Hausman test is H$= (\widehat{\boldsymbol{\theta}}_{\text{FE}} - \widetilde{\boldsymbol{\theta}}_{\text{RE}})' (\widehat{\text{V}}[\widehat{\boldsymbol{\theta}}_{\text{FE}}] - \widehat{\text{V}}[\widetilde{\boldsymbol{\theta}}_{\text{RE}}])^{-1} (\widehat{\boldsymbol{\theta}}_{\text{FE}} - \widetilde{\boldsymbol{\theta}}_{\text{RE}}) \overset{a}{\sim} \chi^2(q)$.
$\widehat{\beta}_{\text{FE}} = 0.17$ with default standard error 0.03 and $\widehat{\beta}_{\text{RE}} = 0.12$ with default standard error 0.02.
Note that if indeed the RE is fully efficient then the default standard errors are correct and we would use these.
H$= (0.17 - 0.12)^2 / (0.03^2 - 0.02^2) = .0025/.0005 = 5 > \chi^2_{0.05}(1) = 3.84$.
Reject H_0. Conclude that there is a difference so FE is the model.

(g) Now H$= (\widehat{\boldsymbol{\theta}}_{\text{FE}} - \widetilde{\boldsymbol{\theta}}_{\text{RE}})' (\widehat{\text{V}}[\widehat{\boldsymbol{\theta}}_{\text{FE}}] + \widehat{\text{V}}[\widetilde{\boldsymbol{\theta}}_{\text{RE}}] - 2 + \widehat{\text{Cov}}[\widetilde{\boldsymbol{\theta}}_{\text{RE}}, \widehat{\boldsymbol{\theta}}_{\text{FE}}])^{-1} (\widehat{\boldsymbol{\theta}}_{\text{FE}} - \widetilde{\boldsymbol{\theta}}_{\text{RE}}) \overset{a}{\sim} \chi^2(q)$.
$\widehat{\beta}_{\text{FE}} = 0.17$ with robust s.e. 0.08, $\widehat{\beta}_{\text{RE}} = 0.12$ with robust s.e. 0.05, and $\widehat{\text{Cov}}[\widetilde{\beta}_{\text{RE}}, \widehat{\beta}_{\text{FE}}] = 0.02^2$.
H$= (0.17 - 0.12)^2 / (0.08^2 + 0.05^2 - 2 \times 0.02^2) = .0025/.0081 = 0.31 < \chi^2_{0.05}(1) = 3.84$.
Reject H_0. Conclude that there is no difference so RE is the model.

(h) Stacking we have $\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{u}_i$, where $\mathbf{y}_i$ and $\mathbf{u}_i$ are $T \times 1$ and $\mathbf{X}_i$ is $T \times k$ with i^{th} row $\mathbf{x}_i'$.
Then $\widehat{\boldsymbol{\beta}} = (\sum_i \mathbf{X}_i' \mathbf{X}_i)^{-1} \sum_i \mathbf{X}_i' \mathbf{y}_i = \boldsymbol{\beta} + (\sum_i \mathbf{X}_i' \mathbf{X}_i)^{-1} \sum_i \mathbf{X}_i' \mathbf{u}_i$.
The asymptotic variance is $(\sum_i \mathbf{X}_i' \mathbf{X}_i)^{-1} \text{Var}(\sum_i \mathbf{X}_i' \mathbf{u}_i) (\sum_i \mathbf{X}_i' \mathbf{X}_i)^{-1}$.
Given independence over i and $\text{E}[\mathbf{u}_i | \mathbf{x}_i] = 0$ this becomes $(\sum_i \mathbf{X}_i' \mathbf{X}_i)^{-1} (\sum_i \text{E}[\mathbf{X}_i' \mathbf{u}_i \mathbf{u}_i' \mathbf{X}_i) (\sum_i \mathbf{X}_i' \mathbf{X}_i)^{-1}$.
So use $(\sum_i \mathbf{X}_i' \mathbf{X}_i)^{-1} (\sum_i \mathbf{X}_i' \widehat{\mathbf{u}}_i \widehat{\mathbf{u}}_i' \mathbf{X}_i) (\sum_i \mathbf{X}_i' \mathbf{X}_i)^{-1}$ where $\widehat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i \widehat{\boldsymbol{\beta}}$.

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

Scores out of	50	
$75th$ percentile	38	(76%)
Median	31.5	(63%)
$25th$ percentile	26	(52%)

A	36 and above
A-	30 and above
B+	24 and above