240D Winter 2011 Solutions to Final Exam

1.(a) We have

$$\begin{aligned}\ln f(y_i) &= -(y_i - \mathbf{x}_i'\boldsymbol{\beta}) - e^{-(y_i-\mathbf{x}_i'\boldsymbol{\beta})}) \\ \ln L(\boldsymbol{\beta}) &= \sum_{i=1}^N \{-y_i + \mathbf{x}_i'\boldsymbol{\beta} - e^{-(y_i-\mathbf{x}_i'\boldsymbol{\beta})}\} \\ \partial \ln L(\boldsymbol{\beta})/\partial\boldsymbol{\beta} &= \sum_{i=1}^N \{\mathbf{x}_i - e^{-(y_i-\mathbf{x}_i'\boldsymbol{\beta})}\mathbf{x}_i\} = \sum_{i=1}^N \{1 - e^{-(y_i-\mathbf{x}_i'\boldsymbol{\beta})}\}\mathbf{x}_i = \mathbf{0}.\end{aligned}$$

(b) Since $\partial^2 \ln L(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}' = -\sum_{i=1}^N e^{-(y_i-\mathbf{x}_i'\boldsymbol{\beta})}\mathbf{x}_i\mathbf{x}_i'$ we have $\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} \mathcal{N}[\mathbf{0}, -\mathbf{A}_0^{-1}]$ by the information matrix equality $\mathbf{A}_0 = \text{plim } N^{-1}\partial^2 \ln L(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}\big|_{\boldsymbol{\beta}_0} = -\text{ plim } \frac{1}{N}\sum_{i=1}^N e^{-(y_i-\mathbf{x}_i'\boldsymbol{\beta}_0)}\mathbf{x}_i\mathbf{x}_i'\}$.

(c) For consistency need $\text{E}[e^{-(y_i-\mathbf{x}_i'\boldsymbol{\beta})}|\mathbf{x}_i] = 1$ so that $\text{E}[\partial \ln L(\boldsymbol{\beta})/\partial\boldsymbol{\beta}|_{\boldsymbol{\beta}_0}] = -\sum_{i=1}^N\{1-e^{-(y_i-\mathbf{x}_i'\boldsymbol{\beta})}\}\mathbf{x}_i = \mathbf{0}$.
This is unlikely to be the case, and is not implied by $\text{E}[y_i|\mathbf{x}_i] = c + \mathbf{x}_i'\boldsymbol{\beta}$.
Most likely inconsistent.

(d) For method of moments use $\text{E}[y_i|\mathbf{x}_i] = c + \mathbf{x}_i'\boldsymbol{\beta} \quad \Longrightarrow \quad \text{E}[\mathbf{x}_i(y_i - c - \mathbf{x}_i'\boldsymbol{\beta})] = \mathbf{0}$.
Method of moments $\widehat{\boldsymbol{\beta}}$ solves $\sum_i \mathbf{x}_i(y_i - c - \mathbf{x}_i'\boldsymbol{\beta}) = \mathbf{0}$.
$\widehat{\boldsymbol{\beta}} = (\sum_i \mathbf{x}_i\mathbf{x}_i')^{-1}\sum_i \mathbf{x}_i(y_i - c)$.

(e) But this is just the OLS estimator, except for the intercept (with coefficient β_1) OLS estimates $c + \beta_1$ rather than β_1. Can get distribution using the usual OLS theory: $\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} \mathcal{N}[\mathbf{0}, \mathbf{A}_0^{-1}\mathbf{B}_0\mathbf{A}_0^{-1}]$ where $\mathbf{A}_0 = \text{plim } \frac{1}{N}\sum_{i=1}^N \mathbf{x}_i\mathbf{x}_i'$ and $\mathbf{A}_0 = \text{plim } \frac{1}{N}\sum_{i=1}^N (y_i - c - \mathbf{x}_i'\boldsymbol{\beta})^2\mathbf{x}_i\mathbf{x}_i'$.

2.(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) A sequence of random variables $\{b_N\}$ converges in distribution to a random variable b if $\lim_{N\to\infty} F_N = F$, at every continuity point of F, where F_N is the distribution of b_N, F is the distribution of b, and convergence is in the usual mathematical sense.

(c) $\mathbf{g}(\widehat{\boldsymbol{\beta}}) = \mathbf{g}(\boldsymbol{\beta}_0) + \partial\mathbf{g}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}|_{\boldsymbol{\beta}_0} \times (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ by a first-order Taylor series expansion.
So $\sqrt{N}(\mathbf{g}(\widehat{\boldsymbol{\beta}}) - \mathbf{g}(\boldsymbol{\beta}_0)) = \partial\mathbf{g}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}'\big|_{\boldsymbol{\beta}_0} \times (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} \partial\mathbf{g}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}|_{\boldsymbol{\beta}_0} \times \mathcal{N}[0, \text{V}_0]$
$\xrightarrow{d} \mathcal{N}[\mathbf{0}, \partial\mathbf{g}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}'\big|_{\boldsymbol{\beta}_0} \times \text{V}_0 \times \partial\mathbf{g}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}|_{\boldsymbol{\beta}_0}]$

(d),(e) $\Pr[y_i = 1|\mathbf{x}_i] = \Phi(\mathbf{x}_i'\boldsymbol{\beta})$. $\partial \Pr[y_i = 1|\mathbf{x}_i]/\partial\mathbf{x}_{ik} = \phi(\mathbf{x}_i'\boldsymbol{\beta}) \times \beta_k$, where $\phi(\cdot)$ is the standard normal density.
MEM: $\phi(\overline{\mathbf{x}}_i'\widehat{\boldsymbol{\beta}}) \times \widehat{\beta}_k$ use `mfx` or `margins, dydx(*) atmean`
AME: $\frac{1}{N}\sum_i \phi(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}) \times \widehat{\beta}_k$ use `margeff` or `margins, dydx(*)`

3.(a) This is logit since

$$\Pr[y=1] = \Pr[y^* \geq 0] = \Pr[\mathbf{x}'\boldsymbol{\beta} + u \geq 0] = \Pr[-u \leq \mathbf{x}'\boldsymbol{\beta}] = F(\mathbf{x}'\boldsymbol{\beta}) = 1/(1+\exp(-\mathbf{x}'\boldsymbol{\beta})).$$

Equivalently $\Pr[y=1] = \exp(\mathbf{x}'\boldsymbol{\beta})/(1+\exp(\mathbf{x}'\boldsymbol{\beta}))$.
Estimate by logit MLE. $\widehat{\boldsymbol{\beta}}$ maximizes $L_N(\boldsymbol{\beta}) = \sum_i y_i \ln F(\mathbf{x}_i'\boldsymbol{\beta}) + (1-y_i)\ln(1-F(\mathbf{x}_i'\boldsymbol{\beta}))$.

(b) This is ordered logit since

$$\begin{aligned}
\Pr[y &= 2] = \Pr[y^* > \alpha] = \Pr[\mathbf{x}'\boldsymbol{\beta} + u > \alpha] = \Pr[-u < \mathbf{x}'\boldsymbol{\beta} - \alpha] = F(\mathbf{x}'\boldsymbol{\beta} - \alpha) = 1/(1+\exp(\alpha - \mathbf{x}'\boldsymbol{\beta})). \\
\Pr[y &= 1] = \Pr[0 \leq y^* < \alpha] = \Pr[y^* > 0] - \Pr y^* \geq \alpha] = F(\mathbf{x}'\boldsymbol{\beta}) - F(\mathbf{x}'\boldsymbol{\beta} - \alpha). \\
\Pr[y &= 0] = 1 - \Pr[y=1] - \Pr[y=2]
\end{aligned}$$

Estimate by ordered logit MLE that maximizes the resulting log-likelihood function.

(c) This is a truncated model. MLE maximizes log-likelihood $\sum \ln f(y_i)$ where $f(y)$ is the density

$$\begin{aligned}
f(y) &= f^*(y^*)/\Pr(y^* \geq 0) \\
&= \{f_u(y - \mathbf{x}'\boldsymbol{\beta})\} \times / \Pr(-u \leq \mathbf{x}'\boldsymbol{\beta}) \\
&= f(y - \mathbf{x}'\boldsymbol{\beta})/F(\mathbf{x}'\boldsymbol{\beta}), \text{ where } f \text{ and } F \text{ are given in question.} \\
&= \left(\tfrac{\exp(y - \mathbf{x}'\boldsymbol{\beta})}{(1+\exp(-y+\mathbf{x}'\boldsymbol{\beta}))^2}\right) / \left(\tfrac{1}{1+\exp(-\mathbf{x}'\boldsymbol{\beta})}\right)
\end{aligned}$$

The second line uses change of variables result that $f(y^*)dy^* = g(u)|\frac{dy^*}{du}|\ du = g(u)du$ here since $|\frac{dy^*}{du}| = 1$.
The third line uses symmetry of F so $\Pr(-u \leq \mathbf{x}'\boldsymbol{\beta}) = \Pr(u \leq \mathbf{x}'\boldsymbol{\beta})$.

(d) Least squares is based on the conditional mean. Here

$$\begin{aligned}
\mathrm{E}[y] &= \mathrm{E}[y^*|y^* \geq 0] = \mathbf{x}'\boldsymbol{\beta} + \ \mathrm{E}[u|\mathbf{x}'\boldsymbol{\beta} + u \geq 0] \\
&= \mathbf{x}'\boldsymbol{\beta} + \ \mathrm{E}[u| - u \leq \mathbf{x}'\boldsymbol{\beta}] = \mathbf{x}'\boldsymbol{\beta} - \ \mathrm{E}[-u| - u \leq \mathbf{x}'\boldsymbol{\beta}] \\
&= \mathbf{x}'\boldsymbol{\beta} - \mathbf{x}'\boldsymbol{\beta} - \ln(1 - F(\mathbf{x}'\boldsymbol{\beta}))/F(\mathbf{x}'\boldsymbol{\beta}) \\
&= -\ln(1 - F(\mathbf{x}'\boldsymbol{\beta}))/F(\mathbf{x}'\boldsymbol{\beta}) \\
&= -\ln\left(1 - \tfrac{1}{1+\exp(-\mathbf{x}'\boldsymbol{\beta})}\right) / \left(\tfrac{1}{1+\exp(-\mathbf{x}'\boldsymbol{\beta})}\right)
\end{aligned}$$

So do NLS regression of y_i on $-\ln(1 - F(\mathbf{x}_i'\boldsymbol{\beta}))/F(\mathbf{x}_i'\boldsymbol{\beta})$.

4.(a) Use code similar to the following (here command `tobit` should return values close to 1)

```
set obs 100000
set seed 10101
generate x = rnormal()
generate ystar = 1 + 1*x + rnormal()
generate d = ystar > 0
generate y = ystar
replace y = 0 if ystar < 0
tobit y x
```

(b) First do probit on y_{1i} on x_{1i} to get $\widehat{\boldsymbol{\beta}}_1$ and hence $\lambda(\mathbf{x}_{1i}'\widehat{\boldsymbol{\beta}}_1)$.
Second do OLS for those with $y_{2i} > 0$ of y_{2i} on $\mathbf{x}_{2i}$ and $\lambda(\mathbf{x}_{1i}'\widehat{\boldsymbol{\beta}}_1)$.

(c) 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 perform both stages of the two-steep procedure getting estimates $\widehat{\boldsymbol{\beta}}_{1,b}$ and $\widehat{\boldsymbol{\beta}}_{2,b}$ at the b^{th} round.

Then $\widehat{\mathrm{V}}[\widehat{\boldsymbol{\beta}}_2] = \frac{1}{B-1}\sum_{b=1}^{B}(\widehat{\boldsymbol{\beta}}_2 - \overline{\widehat{\boldsymbol{\beta}}}_2)(\widehat{\boldsymbol{\beta}}_2 - \overline{\widehat{\boldsymbol{\beta}}}_2)'$.
Standard errors are the square root of the diagonal entries in $\widehat{\mathrm{V}}[\widehat{\boldsymbol{\beta}}_2]$.

(d) Assume $\mathbf{z}_i$ satisfies $\mathrm{E}[\mathbf{z}_i(y_i - \Lambda(\mathbf{x}_i'\boldsymbol{\beta}))] = \mathbf{0}$.
$\widehat{\boldsymbol{\beta}}$ minimizes $\left[\sum_{i=1}^{N}\mathbf{z}_i(y_i - \Lambda(\mathbf{x}_i'\boldsymbol{\beta}))\right]' \mathbf{W} \left[\sum_{i=1}^{N}\mathbf{z}_i(y_i - \Lambda(\mathbf{x}_i'\boldsymbol{\beta}))\right]$ where e.g. $\mathbf{W} = \left[\sum_{i=1}^{N}\mathbf{z}_i\mathbf{z}_i'\right]^{-1}$.

(e) In general $\mathbf{W} = \left(\mathrm{V}\left[\sum_{i=1}^{N}\mathbf{z}_i(y_i - \Lambda(\mathbf{x}_i'\boldsymbol{\beta}))\right]\right)^{-1}$.
Given independence $\mathbf{W} = \left(\left[\sum_{i=1}^{N}\mathrm{E}[(y_i - \Lambda(\mathbf{x}_i'\boldsymbol{\beta}))^2\mathbf{z}_i\mathbf{z}_i'\right]\right)^{-1}$.
So use $\mathbf{W} = \left(\left[\sum_{i=1}^{N}(y_i - \Lambda(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}))^2\mathbf{z}_i\mathbf{z}_i'\right]\right)^{-1}$ where $\widehat{\boldsymbol{\beta}}$ is consistent for $\boldsymbol{\beta}$.

5.(a) FE model: $y_{it} = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + u_{it}$.
Three methods (not exhaustive) are (1) OLS of $(y_{it} - \bar{y}_i)$ on $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$; (2) OLS of y_{it} on $\mathbf{x}_{it}$ and $\bar{\mathbf{x}}_i$; and (3) OLS of y_{it} on $\mathbf{x}_{it}$ and a complete set of individual dummies.

(b) $y_{it} = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + u_{it}$ where α_i i.i.d. $(\alpha, \sigma_\alpha^2)$ and u_{it} i.i.d. (α, σ_u^2).

(c) Usual Hausman test is H$=(\widehat{\boldsymbol{\theta}}_{\mathrm{FE}} - \widetilde{\boldsymbol{\theta}}_{\mathrm{RE}})'(\widehat{\mathrm{V}}[\widehat{\boldsymbol{\theta}}_{\mathrm{FE}}] - \widehat{\mathrm{V}}[\widetilde{\boldsymbol{\theta}}_{\mathrm{RE}}])^{-1}(\widehat{\boldsymbol{\theta}}_{\mathrm{FE}} - \widetilde{\boldsymbol{\theta}}_{\mathrm{RE}}) \overset{a}{\sim} \chi^2(q)$.
Weakness is that this requires $\widetilde{\boldsymbol{\theta}}_{\mathrm{RE}}$ to be fully efficient which requires the assumptions in part (b).
In practice these assumptions of homoskedasticity and equicorrelation are unlikely to be met.

(d) 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 $(\sum_i \mathbf{X}_i'\mathbf{X}_i)^{-1}\mathrm{Var}(\sum_i \mathbf{X}_i'\mathbf{u}_i)(\sum_i \mathbf{X}_i'\mathbf{X}_i)^{-1}$.
Given independence over i and $\mathrm{E}[\mathbf{u}_i|\mathbf{x}_i] = 0$ this becomes $(\sum_i \mathbf{X}_i'\mathbf{X}_i)^{-1}(\sum_i \mathrm{E}[\mathbf{X}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{X}_i)(\sum_i \mathbf{X}_i'\mathbf{X}_i)^{-1}$.
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	
75*th* percentile	38	(77%)
Median	34	(68%)
25*th* percentile	29	(58%)

A	38 and above
A-	31 and above
B+	24 and above