

1.(a) Since $f(y_i|\mathbf{x}_i) = \Lambda(\mathbf{x}'_i\boldsymbol{\beta})^{y_i}(1 - \Lambda(\mathbf{x}'_i\boldsymbol{\beta}))^{1-y_i}$, the log-likelihood

$$\ln L(\boldsymbol{\beta}) = \sum_i \ln f(y_i|\mathbf{x}_i) = \sum_i y_i \ln \Lambda(\mathbf{x}'_i\boldsymbol{\beta}) + (1 - y_i) \ln(1 - \Lambda(\mathbf{x}'_i\boldsymbol{\beta})).$$

Then using $\partial\Lambda(z)/\partial z = \Lambda(z)(1 - \Lambda(z))$ for $\Lambda(z) = e^z/(1 + e^z)$ we obtain

$$\begin{aligned} \frac{\partial \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \sum_i \left(y_i \frac{\Lambda'(\mathbf{x}'_i\boldsymbol{\beta})}{\Lambda(\mathbf{x}'_i\boldsymbol{\beta})} \mathbf{x}_i - (1 - y_i) \frac{\Lambda'(\mathbf{x}'_i\boldsymbol{\beta})}{1 - \Lambda(\mathbf{x}'_i\boldsymbol{\beta})} \mathbf{x}_i \right) \\ &= \sum_i \left(\frac{y_i}{\Lambda(\mathbf{x}'_i\boldsymbol{\beta})} - \frac{(1 - y_i)}{1 - \Lambda(\mathbf{x}'_i\boldsymbol{\beta})} \right) \Lambda'(\mathbf{x}'_i\boldsymbol{\beta}) \mathbf{x}_i \\ &= \sum_i \left(\frac{y_i - y_i\Lambda(\mathbf{x}'_i\boldsymbol{\beta}) - \Lambda(\mathbf{x}'_i\boldsymbol{\beta}) + y_i\Lambda(\mathbf{x}'_i\boldsymbol{\beta})}{\Lambda(\mathbf{x}'_i\boldsymbol{\beta})(1 - \Lambda(\mathbf{x}'_i\boldsymbol{\beta}))} \right) \Lambda(\mathbf{x}'_i\boldsymbol{\beta})(1 - \Lambda(\mathbf{x}'_i\boldsymbol{\beta})) \mathbf{x}_i \\ &= \sum_i (y_i - \Lambda(\mathbf{x}'_i\boldsymbol{\beta})) \mathbf{x}_i = \mathbf{0}. \end{aligned}$$

(b) We have

$$\begin{aligned} \sum_i (y_i - \Lambda(\mathbf{x}'_i\hat{\boldsymbol{\beta}})) \mathbf{x}_i &= \sum_i (y_i - \Lambda(\mathbf{x}'_i\boldsymbol{\beta}_0)) \mathbf{x}_i - \sum_i \frac{\partial\Lambda(\mathbf{x}'_i\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} \mathbf{x}_i \Big|_{\boldsymbol{\beta}_+} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \mathbf{0}. \\ \Rightarrow \quad (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) &= \left(\sum_i \frac{\partial\Lambda(\mathbf{x}'_i\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} \mathbf{x}_i \Big|_{\boldsymbol{\beta}_+} \right)^{-1} \sum_i (y_i - \Lambda(\mathbf{x}'_i\boldsymbol{\beta}_0)) \mathbf{x}_i \\ \Rightarrow \quad \sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) &= \left(\frac{1}{N} \sum_i \Lambda(\mathbf{x}'_i\boldsymbol{\beta}^+)(1 - \Lambda(\mathbf{x}'_i\boldsymbol{\beta}^+)) \mathbf{x}_i \mathbf{x}'_i \right)^{-1} \frac{1}{\sqrt{N}} \sum_i (y_i - \Lambda(\mathbf{x}'_i\boldsymbol{\beta}_0)) \mathbf{x}_i \\ &\xrightarrow{d} \left(\frac{1}{N} \sum_i \Lambda(\mathbf{x}'_i\boldsymbol{\beta}_0)(1 - \Lambda(\mathbf{x}'_i\boldsymbol{\beta}_0)) \mathbf{x}_i \mathbf{x}'_i \right)^{-1} \times \mathcal{N}[0, \mathbf{B}_0] \end{aligned}$$

Here

$$\begin{aligned} \mathbf{B}_0 &= \lim V \left[\frac{1}{\sqrt{N}} \sum_i (y_i - \Lambda(\mathbf{x}'_i\boldsymbol{\beta}_0)) \mathbf{x}_i \right] = \lim E \left[\frac{1}{N} \sum_i (y_i - \Lambda(\mathbf{x}'_i\boldsymbol{\beta}_0))^2 \mathbf{x}_i \mathbf{x}'_i \right] \\ &= \lim E \left[\frac{1}{N} \sum_i \Lambda(\mathbf{x}'_i\boldsymbol{\beta}_0)(1 - \Lambda(\mathbf{x}'_i\boldsymbol{\beta}_0)) \mathbf{x}_i \mathbf{x}'_i \right] \end{aligned}$$

So simplifies to $\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} \mathcal{N}[0, \mathbf{B}_0^{-1}]$.

(c) The essential assumption is $E[(y_i - \Lambda(\mathbf{x}'_i\boldsymbol{\beta}))|\mathbf{z}_i] = 0$ or $E[\mathbf{z}_i(y_i - \Lambda(\mathbf{x}'_i\boldsymbol{\beta}))] = \mathbf{0}$.

Then $\hat{\boldsymbol{\beta}}$ minimizes

$$\left(\sum_i (y_i - \Lambda(\mathbf{x}'_i\boldsymbol{\beta})) \mathbf{x}_i \right)' \mathbf{W} \left(\sum_i (y_i - \Lambda(\mathbf{x}'_i\boldsymbol{\beta})) \mathbf{z}_i \right)$$

where \mathbf{W} is symmetric full rank. e.g. $\mathbf{W} = (\mathbf{Z}'\mathbf{Z})^{-1}$.

(d) In the just-identified case the f.o.c. are

$$\sum_i (y_i - \Lambda(\mathbf{x}'_i\boldsymbol{\beta})) \mathbf{z}_i = \mathbf{0}.$$

Using method of part (b), with \mathbf{z}_i replacing \mathbf{x}_i , we get

$$\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} \left(\frac{1}{N} \sum_i \Lambda(\mathbf{x}'_i\boldsymbol{\beta}_0)(1 - \Lambda(\mathbf{x}'_i\boldsymbol{\beta}_0)) \mathbf{z}_i \mathbf{z}'_i \right)^{-1} \times \mathcal{N}[0, \mathbf{B}_0]$$

$$\mathbf{B}_0 = \lim V \left[\frac{1}{\sqrt{N}} \sum_i (y_i - \Lambda(\mathbf{x}'_i\boldsymbol{\beta}_0)) \mathbf{z}_i \right] = \lim E \left[\frac{1}{N} \sum_i (y_i - \Lambda(\mathbf{x}'_i\boldsymbol{\beta}_0))^2 \mathbf{z}_i \mathbf{z}'_i \right]$$

[or GMM with $\mathbf{W} = \mathbf{I}$ gives limit variance matrix $(\mathbf{G}'_0\mathbf{G}_0)^{-1}(\mathbf{G}'_0\mathbf{S}_0\mathbf{G}_0)(\mathbf{G}'_0\mathbf{G}_0)^{-1}$.

Since \mathbf{G}_0 is square in just-identified case: $= \mathbf{G}_0^{-1}(\mathbf{G}'_0)^{-1}(\mathbf{G}'_0\mathbf{S}_0\mathbf{G}_0)\mathbf{G}_0^{-1}(\mathbf{G}'_0)^{-1} = \mathbf{G}_0^{-1}\mathbf{S}_0(\mathbf{G}'_0)^{-1}$.]

2.(a) $\Pr(y = 1) = \Lambda(\beta_1 + \beta_2 x)$, so the marginal effect

$$\frac{\partial \Pr(y = 1)}{\partial x} = \beta_2 \Lambda'(\beta_1 + \beta_2 x) = 0.2 \Lambda'(0.4 + 0.2x).$$

We can evaluate at the sample mean $\bar{x} = 0.3$, giving $0.2 \times \Lambda'(1) = 0.2 \times \Lambda(1) \times (1 - \Lambda(1)) = 0.2 \times e/(1 - e)^2 = 0.0393$.

Alternatively we can use $\partial \Pr(y = 1)/\partial x \simeq \beta_2 \bar{y}(1 - \bar{y}) = 0.2 \times 0.8 \times 0.2 = 0.032$.

Or can use log-odds ratio interpretation: odds ratio up by 20%.

(b) For probit u_i is standard normal with c.d.f. $\Phi(\cdot)$ and

$$\Pr[y_i = 1 | \mathbf{x}_i] = \Pr[y_i^* > 0 | \mathbf{x}_i] = \Pr[\mathbf{x}_i \boldsymbol{\beta} + u_i > 0 | \mathbf{x}_i] = \Pr[u_i < -\mathbf{x}_i \boldsymbol{\beta} | \mathbf{x}_i] = \Pr[-u_i < \mathbf{x}_i \boldsymbol{\beta} | \mathbf{x}_i] = \Phi(\mathbf{x}_i \boldsymbol{\beta})$$

(c) Define three mutually exclusive dummies $y_{1i} = 1$ if $y_i = 1$ and $y_{1i} = 0$ if $y_i \neq 1$, and similarly for y_{2i} and y_{3i} . Then

Density for i^{th} observation is $F_1(\mathbf{x}'_i \boldsymbol{\beta}_1)^{y_{1i}} \times F_2(\mathbf{x}'_i \boldsymbol{\beta}_2)^{y_{2i}} \times F_3(\mathbf{x}'_i \boldsymbol{\beta}_3)^{y_{3i}}$.

Log-likelihood is $\sum_i [y_{1i} \ln F_1(\mathbf{x}'_i \boldsymbol{\beta}_1) + y_{2i} \ln F_2(\mathbf{x}'_i \boldsymbol{\beta}_2) + y_{3i} \ln F_3(\mathbf{x}'_i \boldsymbol{\beta}_3)]$.

Need a normalization such as $\boldsymbol{\beta}_1 = \mathbf{0}$.

(d) We suppose $y_i^* = \mathbf{x}'_i \boldsymbol{\beta} + u_i$ where u_i is standard normal and \mathbf{x}_i has no intercept.

We observe $y_i = 1$ if $y_i^* \leq \alpha_1$; $y_i = 2$ if $\alpha_1 < y_i^* \leq \alpha_2$; and $y_i = 3$ if $y_i^* > \alpha_2$.

Then

$$\Pr[y_i = 1] = \Pr[y_i^* \leq \alpha_1] = \Pr[\mathbf{x}'_i \boldsymbol{\beta} + u_i \leq \alpha_1] = \Pr[u_i \leq \alpha_1 - \mathbf{x}'_i \boldsymbol{\beta}] = \Phi(\alpha_1 - \mathbf{x}'_i \boldsymbol{\beta}).$$

$$\Pr[y_i = 2] = \Pr[\alpha_1 < y_i^* \leq \alpha_2] = \Pr[y_i^* \leq \alpha_2] - \Pr[y_i^* \leq \alpha_1] = \Phi(\alpha_2 - \mathbf{x}'_i \boldsymbol{\beta}) - \Phi(\alpha_1 - \mathbf{x}'_i \boldsymbol{\beta}).$$

$$\Pr[y_i = 3] = 1 - \Pr[y_i = 1] - \Pr[y_i = 2] = 1 - \Phi(\alpha_2 - \mathbf{x}'_i \boldsymbol{\beta}).$$

(e) We have

$$p_{ij} = \Pr[y_i = j] = \frac{e^{\mathbf{x}'_i \boldsymbol{\beta}_j}}{\sum_{l=1}^m e^{\mathbf{x}'_i \boldsymbol{\beta}_l}}, \quad j = 1, 2, 3.$$

(f) In general

$$U_{1i} = V_{1i} + \varepsilon_{1i}$$

$$U_{2i} = V_{2i} + \varepsilon_{2i}$$

$$U_{3i} = V_{3i} + \varepsilon_{3i}$$

where V_{ji} is deterministic and ε_{ji} is purely random.

We choose $y_i = j$ if $U_{ji} \geq U_{ki}$ for all k .

Setting $V_{ji} = \exp(\mathbf{x}'_i \boldsymbol{\beta}_j)$ and assuming ε_{ji} are i.i.d. type 1 extreme value yields multinomial logit.

3.(a) For notational simplicity 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

$$f(y|\mathbf{x}) = \frac{e^{-\exp(\mathbf{x}'_i \boldsymbol{\beta})} \exp(\mathbf{x}'_i \boldsymbol{\beta})^{y_i} / y_i^*!}{(1 - e^{-\exp(\mathbf{x}'_i \boldsymbol{\beta})})}$$

(b) We have

$$\begin{aligned} \ln L(\boldsymbol{\beta}) &= \sum_i (-\exp(\mathbf{x}'_i \boldsymbol{\beta}) + y_i \mathbf{x}'_i \boldsymbol{\beta} - \ln y_i! - \ln(1 - e^{-\exp(\mathbf{x}'_i \boldsymbol{\beta})})) . \\ \Rightarrow \frac{\partial \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \sum_i \left(-\exp(\mathbf{x}'_i \boldsymbol{\beta}) \mathbf{x}_i + y_i \mathbf{x}_i - \frac{\exp(\mathbf{x}'_i \boldsymbol{\beta}) e^{-\exp(\mathbf{x}'_i \boldsymbol{\beta})}}{(1 - e^{-\exp(\mathbf{x}'_i \boldsymbol{\beta})})} \right) = \mathbf{0} . \end{aligned}$$

(c) This is tricky

$$E[y] = E[y^* | y^* \geq 1] = \sum_{y^*=1}^{\infty} \frac{y^* f(y^*)}{\Pr[y^* \geq 1]} = \frac{1}{\Pr[y^* \geq 1]} \sum_{y^*=1}^{\infty} y^* f(y^*) = \frac{1}{\Pr[y^* \geq 1]} \sum_{y^*=0}^{\infty} y^* f(y^*)$$

(d) Since

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

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

(e) We have

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

and we observe

$$y_1 = \begin{cases} 1 & \text{if } y_1^* > 0 \\ 0 & \text{if } y_1^* \leq 0, \end{cases}$$

and

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

Then if $(\varepsilon_1, \varepsilon_2)$ are joint normal with means $(0, 0)$, variances $(1, \sigma_2^2)$ and covariance $\rho \sigma_2^2$ (or less restrictively ε_1 is standard normal and $\varepsilon_2 = \rho \sigma_2 \varepsilon_1 + \nu$ where ν is independent of ε_1), we have

$$E[y_2 | y_1 = 1] = E[y_2^2 | y_1^* > 0] = \mathbf{x}'_2 \boldsymbol{\beta}_2 + \rho \sigma \lambda(\mathbf{x}'_1 \boldsymbol{\beta}_1).$$

First stage probit of y_1 on \mathbf{x}_1 gives $\widehat{\boldsymbol{\beta}}_1$ and $\lambda(\mathbf{x}'_1 \widehat{\boldsymbol{\beta}}_1) = \phi(\mathbf{x}'_1 \widehat{\boldsymbol{\beta}}_1) / \Phi(\mathbf{x}'_1 \widehat{\boldsymbol{\beta}}_1)$.

Second stage OLS for positive y_2 of y_2 on \mathbf{x}_2 $\lambda(\mathbf{x}'_1 \widehat{\boldsymbol{\beta}}_1)$.

4.(a) This part of the question was in error.

The test statistic given in the output as `Wald chi2(1)` is a test of overall fit, i.e. that all slope coefficients are zero. Here this is just $\beta_{\text{chronic}} = 0$ since there is only one regressor.

If default standard errors were used, `Wald chi2(1)` would be an LR test.

In fact robust standard errors were used, so `Wald chi2(1)` is a Wald test with value $3.64 < 1.96^2 = 3.84$ and $p = 0.0565$.

So there is not enough information for a Wald test. This part is graded leniently.

(b) The bootstrap standard error is the standard deviation of `bchronic` which is 0.5404076.

This yields $t = 0.9833014/0.5404076 = 1.820 < z_{.025} = 1.96$. Do not reject H_0 .

(c) Here $t_b^* = (\hat{\beta}_b^* - \hat{\beta})/se[\hat{\beta}_b^*]$ has 2.5 and 97.5 percentiles are -2.756 and 2.569.

The original t-statistic of 1.91 is inside these percentiles. So do not reject H_0 .

5.(a) Here

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

OLS of $y_{it} - \bar{y}_i$ on $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$ gives

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

(b) If fixed effects are present then FE estimator is consistent but RE is inconsistent. If random effects are present then both are consistent. Do Hausman test.

$H_0 : \alpha_i$ is random versus $H_a : \alpha_i$ is fixed effect. For $\boldsymbol{\beta}_1$ coefficients of time-varying regressors

$$\mathbf{H} = \left(\tilde{\boldsymbol{\beta}}_{1,\text{RE}} - \hat{\boldsymbol{\beta}}_{1,\text{W}} \right)' \left[\hat{\mathbf{V}}[\hat{\boldsymbol{\beta}}_{1,\text{W}}] - \hat{\mathbf{V}}[\tilde{\boldsymbol{\beta}}_{1,\text{RE}}] \right]^{-1} \left(\tilde{\boldsymbol{\beta}}_{1,\text{RE}} - \hat{\boldsymbol{\beta}}_{1,\text{W}} \right),$$

and reject at level α if $\mathbf{H} > \chi^2_{\alpha}(\dim[\boldsymbol{\beta}_1])$.

(c) In general H test uses $\hat{\mathbf{V}}[\hat{\boldsymbol{\beta}}_{1,\text{W}} - \tilde{\boldsymbol{\beta}}_{1,\text{RE}}]$. If $\tilde{\boldsymbol{\beta}}_{1,\text{RE}}$ is fully efficient this simplifies to answer in (b). This is the case if $\varepsilon_{it} \sim \mathcal{N}[0, \sigma_{\varepsilon}^2]$. If this assumption cannot be made need more general form for H. Implement by bootstrap.

(d) Given $y_{it} = \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}$, for the i^{th} individual stack all T observations, so

$$\begin{bmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \alpha_i + \begin{bmatrix} \mathbf{x}'_{i1} \\ \vdots \\ \mathbf{x}'_{iT} \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iT} \end{bmatrix}$$

or

$$\begin{aligned} \mathbf{y}_i &= \mathbf{e}\alpha_i + \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i \\ \implies \mathbf{Q}\mathbf{y}_i &= \mathbf{Q}\mathbf{e}\alpha_i + \mathbf{Q}\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Q}\boldsymbol{\varepsilon}_i \\ \implies \mathbf{Q}\mathbf{y}_i &= \mathbf{Q}\mathbf{X}_i\boldsymbol{\beta} + \mathbf{Q}\boldsymbol{\varepsilon}_i \end{aligned}$$

as $\mathbf{Q}\mathbf{e} = \mathbf{I}_T - T^{-1}\mathbf{e}\mathbf{e}'\mathbf{e} = \mathbf{e} - T^{-1}\mathbf{e}\mathbf{e}'\mathbf{e} = \mathbf{e} - T^{-1}\mathbf{e}\mathbf{T} = \mathbf{e} - \mathbf{e} = \mathbf{0}$.

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

Scores out of	50	A	36 and above
75th percentile	40 (80%)	A-	30 and above
Median	33 (66%)	B+	24 and above
25th percentile	29.5 (59%)	B	18 and above