Homework #2

Professor: Pat Kline

Students: Christina Brown, Sam Leone, Peter McCrory, Preston Mui

2. Logit MLE

a) Derive the score of the logit-likelihood: First, noting that

$$\frac{\partial \Lambda(X_i'\beta)}{\partial \beta} = \frac{\exp(X_i'\beta)X_i'(1+X_i'\beta) - \exp(X_i'\beta)\exp(X_i'\beta)X_i'}{(1+\exp(X_i'\beta))^2} = \frac{\exp(X_i'\beta)X_i'}{(1+\exp(X_i'\beta))^2}$$

The score of the logit log-likelihood is therefore

$$\begin{aligned} s(\beta) &= \sum_i \frac{Y_i \exp(X_i'\beta)X_i'}{\Lambda(X_i'\beta)(1+\exp(X_i'\beta))^2} - \frac{(1-Y_i)\exp(X_i'\beta)X_i'}{(1-\Lambda(X_i'\beta))(1+\exp(X_i'\beta))^2} \\ &= \sum_i \Big(Y_i - (1-Y_i)\exp(X_i'\beta)\Big)\frac{X_i'}{1+\exp(X_i'\beta)} \\ &= \sum_i \Big(Y_i(1+\exp(X_i'\beta)) - \exp(X_i'\beta)\Big)\frac{X_i'}{1+\exp(X_i'\beta)} \\ &= \sum_i \Big(Y_i - \frac{\exp(X_i'\beta)}{1+\exp(X_i'\beta)}\Big)X_i' \end{aligned}$$

b) The moment condition identifying β_{ML} is

$$E\Bigg[\Big(Y_i - \frac{\exp(X_i'\beta)}{1+\exp(X_i'\beta)}\Big)X_i'\Bigg] = 0$$
$$E\Bigg[E[Y_i - \frac{\exp(X_i'\beta)}{1+\exp(X_i'\beta)}|X_i]X_i'\Bigg] = 0$$

c) The moment conditions identifying β_{NLLS} are

$$\begin{aligned} 0 &= E\Bigg[-2(Y_i - \Lambda(X_i'\beta))\frac{\exp(X_i'\beta)}{(1+\exp(X_i'\beta))^2}X_i'\Bigg] \\ &= E\Bigg[E[(Y-\Lambda(X_i'\beta))|X_i]\frac{\exp(X_i'\beta)}{(1+\exp(X_i'\beta))^2}X_i'\Bigg] \end{aligned}$$

d) Under which conditions does β_{NLLS} coincide with β_{ML}? The moment conditions will coincide when $E[Y_i - \Lambda(X_i'\beta)|X_i'] = 0$ for all X_i; that is, they coincide when the logit model is correctly specified.

3. Matlab Probit DGP

a) (Matlab program attached)

b) The ML point estimates of $\hat{\beta}^{con} = (\hat{\alpha}, \hat{b})$ are $(-0.0175, 0.9159)$. The standard errors are $(0.3030, 0.3607)$, respectively.

c) Score test: Following Wooldridge (2010), page 570, the LM statistic is the ESS from the following regression

$$\frac{\hat{u}_i}{\sqrt{\hat{G}_i(1-\hat{G}_i)}} = \alpha \frac{\hat{g}_i}{\sqrt{\hat{G}_i(1-\hat{G}_i)}} x_i + \gamma \frac{\hat{g}_i}{\sqrt{\hat{G}_i(1-\hat{G}_i)}} z_i$$

where x_i is the regressor matrix in the unconstrained regression (constant and X) and z_i is the vector of X_i^2. The ESS from this regression was 0.2963, which has a p-value of 0.58621. So, one does not reject the null that $b_2 = 0$.

d) The unrestricted model yields point estimates of $\hat{\beta}^{unc} = (\hat{\alpha}, \hat{b}, \hat{b_2}) = (-0.0435, 0.9175, 0.0428)$. The Wald test will test the null $g(\beta) = 0$ where

$$g(\beta) \equiv (0,0,1) \cdot \beta = b_2$$
$$G(\beta) \equiv \frac{\partial g(\beta)}{\partial \beta} = (0,0,1)$$

and the Wald statistic is given by

$$N \cdot \hat{b_2} \cdot \left(G \cdot \frac{1}{\sqrt{N}} H^{-1} \cdot G' \right)^{-1} \cdot \hat{b_2}$$

where H is the average Hessian at the ML estimate. Because we know that the model is correctly specified, I use the inverse Hessian instead of the sandwich estimator. The Wald evaluates to 7.4592, which has a p-value of 0.0063 under the χ^2 distribution with d.f. 1, so one rejects the null that $b_2 = 0$. This is starkly different from the score test result, which did not reject the null. This makes sense, as the Wald tends to reject more than the LM test. If one bumps the number of observations up (say, to 5000), both tests reject the null.

6. Random Coefficient Binary Choice Model:

a) As was derived in the lecture notes, the simulated log likelihood can be derived as follows. Denote the Logistic cdf with Λ and its pdf with λ.

First observe that

$$Pr(Y_i = 1|X_i, b_i) = Pr(\epsilon_i > -b_i x_i) = 1 - Pr(\epsilon_i < -b_i x_i) = 1 - \Lambda(-b_i x_i) = \Lambda(b_i x_i)$$

Thus, assuming that b_i and ϵ_i are independent:

$$Pr(Y_i = 1|X_i) = \int \Lambda(bx_i) \frac{1}{\sigma} \phi(\frac{b_i - \mu}{\sigma}) db$$

The likelihood of this model for a single observation is

$$L(Y_i, X_i; \mu, \sigma) = \left(\int \Lambda(bx_i) \frac{1}{\sigma} \phi(\frac{b_i - \mu}{\sigma}) db \right)^{Y_i} \left(1 - \int \Lambda(bx_i) \frac{1}{\sigma} \phi(\frac{b_i - \mu}{\sigma}) db \right)^{1-Y_i}$$

Thus, for **fixed** random draws $\{u_i\}_{m=1}^{M}$ from $\mathcal{N}(0,1)$, the simulated likelihood is

$$\hat{L}_M(Y_i, X_i, \mu, \sigma) = \left(\frac{1}{M} \sum_{m=1}^{M} \Lambda(\underbrace{(\mu + \sigma u_{im})}_{\equiv b} X_i) \right)^{Y_i} \left(1 - \frac{1}{M} \sum_{m=1}^{M} \Lambda(\underbrace{(\mu + \sigma u_{im})}_{\equiv b} X_i) \right)^{1-Y_i}$$

Clearly, taking logs (of the product of all the likelihoods) allows us to recover the simulated log-likelihood of the full sample.

Let's differentiate with respect to μ and σ!

For μ:

$$\begin{aligned}
&\frac{\partial}{\partial \mu} \frac{1}{N} \sum_{i}^{N} Y_i log \left(\frac{1}{M} \sum_{m=1}^{M} \Lambda((\mu + \sigma u_{im}) X_i) \right) + (1 - Y_i) log \left(1 - \frac{1}{M} \sum_{m=1}^{M} \Lambda((\mu + \sigma u_{im}) X_i) \right) \\
&= \frac{1}{N} \sum_{i}^{N} \frac{Y_i \frac{1}{M} \sum_{m=1}^{M} \lambda((\mu + \sigma u_{im}) X_i) X_i}{\frac{1}{M} \sum_{m=1}^{M} \Lambda((\mu + \sigma u_{im}) X_i)} \\
&\quad - (1 - Y_i) \frac{1}{N} \sum_{i}^{N} \frac{(1 - Y_i) \frac{1}{M} \sum_{m=1}^{M} \lambda((\mu + \sigma u_{im}) X_i) X_i}{1 - \frac{1}{M} \sum_{m=1}^{M} \Lambda((\mu + \sigma u_{im}) X_i)}
\end{aligned}$$

For σ:

$$\begin{aligned}
&\frac{\partial}{\partial \sigma} \frac{1}{N} \sum_{i}^{N} Y_i log \left(\frac{1}{M} \sum_{m=1}^{M} \Lambda((\mu + \sigma u_{im}) X_i) \right) + (1 - Y_i) log \left(1 - \frac{1}{M} \sum_{m=1}^{M} \Lambda((\mu + \sigma u_{im}) X_i) \right) \\
&= \frac{1}{N} \sum_{i}^{N} \frac{Y_i \frac{1}{M} \sum_{m=1}^{M} \lambda((\mu + \sigma u_{im}) X_i) X_i u_{im}}{\frac{1}{M} \sum_{m=1}^{M} \Lambda((\mu + \sigma u_{im}) X_i)} \\
&\quad - (1 - Y_i) \frac{1}{N} \sum_{i}^{N} \frac{(1 - Y_i) \frac{1}{M} \sum_{m=1}^{M} \lambda((\mu + \sigma u_{im}) X_i) X_i u_{im}}{1 - \frac{1}{M} \sum_{m=1}^{M} \Lambda((\mu + \sigma u_{im}) X_i)}
\end{aligned}$$

b)