Homework #2

Professor: Pat Kline

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

1. Standard Errors for Composite Statistics

Subproblem A

The Delta Method (or the So-Called "Delta Method," so called by Dr. James Powell) says that, if we have a random vector θ that is asymptotically normal with

$$\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Sigma)$$

then for some $g(\theta)$ that is continuously differentiable at $\theta = \theta_0$ with Jacobian matrix

$$G_0 = \frac{\partial g(\theta_0)}{\partial \theta'}$$

we have that $g(\theta)$ is also asymptotically normal with

$$\sqrt{N}(g(\hat{\theta}) - g(\theta_0) \xrightarrow{d} N(0, G_0' \Sigma G_0)$$

In the context of Fehr and Goette (2007), observe that

$$\theta = \begin{pmatrix} \bar{Y_A} \\ \bar{Y_B} \end{pmatrix}$$

that

$$\Sigma = \begin{bmatrix} \hat{\sigma}^2_{\bar{Y_A}} & \hat{\sigma}_{\bar{Y_A}} \hat{\sigma}_{\bar{Y_B}} \\ \hat{\sigma}_{\bar{Y_A}} \hat{\sigma}_{\bar{Y_B}} & \hat{\sigma}^2_{\bar{Y_B}} \end{bmatrix}$$

and that

$$g(\theta) = \frac{\bar{Y_A} - \bar{Y_B}}{\bar{Y_B}}$$

We can find many of these values in the reprinted table. $\bar{Y_A} = 4131.33$, $\bar{Y_B} = 3005.75$, $\hat{\sigma}^2_{\bar{Y_A}} = (2669.21)^2$, and $\hat{\sigma}^2_{\bar{Y_B}} = (2054.20)^2$. But we still have to do a little bit of work.

Differentiating g with respect to each of its arguments and evaluating them at θ_0 yields G.

$$G = \begin{pmatrix} \frac{\partial g}{\partial \bar{Y_A}} = \frac{1}{\bar{Y_B}} \\ \frac{\partial g}{\partial \bar{Y_B}} = \frac{-\bar{Y_A}}{\bar{Y_B}^2} \end{pmatrix} \Rightarrow G_0 = \begin{pmatrix} \frac{\partial g(\theta_0)}{\partial \bar{Y_A}} = \frac{1}{3005.75} \\ \frac{\partial g(\theta_0)}{\partial \bar{Y_B}} = \frac{-4131.33}{9034533.06} \end{pmatrix}$$

The table also gives us the standard error $\frac{\hat{\sigma}_{\bar{Y_A}-\bar{Y_B}}}{\sqrt{n}}$, which unlike the sample standard deviations, already accounts for the sample size. Combined with a simple variance identity, it can give us $\hat{\sigma}_{\bar{Y_A}}\hat{\sigma}_{\bar{Y_B}}$.

$$VAR(\bar{Y_A} - \bar{Y_B}) = VAR(\bar{Y_A}) + VAR(\bar{Y_B}) - 2COV(\bar{Y_A}, \bar{Y_B})$$

$$\Rightarrow COV(\bar{Y_A}, \bar{Y_B}) = \frac{VAR(\bar{Y_A}) + VAR(\bar{Y_B}) - VAR(\bar{Y_A} - \bar{Y_B})}{2}$$

$$\Rightarrow \hat{\sigma}_{\bar{Y_A}}\hat{\sigma}_{\bar{Y_B}} = \frac{(2669.21)^2 + (2054.20)^2 - (42)(519.72)^2}{2}$$

$$\Rightarrow \hat{\sigma}_{\bar{Y_A}}\hat{\sigma}_{\bar{Y_B}} = -76.67$$

Now we can set up the matrix algebra to use the Delta Method to compute an *estimate* of the standard error, S.

$$S^2 = \frac{1}{n} G_0' \Sigma G_0 = \frac{1}{n} \begin{pmatrix} \frac{\partial g}{\partial \bar{Y_A}} & \frac{\partial g}{\partial \bar{Y_B}} \end{pmatrix} \begin{bmatrix} \hat{\sigma}^2_{\bar{Y_A}} & \hat{\sigma}_{\bar{Y_A}}\hat{\sigma}_{\bar{Y_B}} \\ \hat{\sigma}_{\bar{Y_A}}\hat{\sigma}_{\bar{Y_B}} & \hat{\sigma}^2_{\bar{Y_B}} \end{bmatrix} \begin{pmatrix} \frac{\partial g}{\partial \bar{Y_A}} \\ \frac{\partial g}{\partial \bar{Y_B}} \end{pmatrix}$$

$$\Rightarrow S^2 = \frac{1}{n}[(\frac{\partial g}{\partial \bar{Y_A}})[(\frac{\partial g}{\partial \bar{Y_A}})(\hat{\sigma}^2_{\bar{Y_A}}) + (\frac{\partial g}{\partial \bar{Y_B}})(\hat{\sigma}_{\bar{Y_A}}\hat{\sigma}_{\bar{Y_B}})] + (\frac{\partial g}{\partial \bar{Y_B}})[(\frac{\partial g}{\partial \bar{Y_A}})(\hat{\sigma}_{\bar{Y_A}}\hat{\sigma}_{\bar{Y_B}}) + (\frac{\partial g}{\partial \bar{Y_B}})(\hat{\sigma}^2_{\bar{Y_B}})]]$$

$$\Rightarrow S^2 = \frac{1}{42}[\frac{1}{3005.75}[(\frac{1}{3005.75})((2669.21)^2) + (\frac{-4131.33}{9034533.06})(-76.67)] + \frac{-4131.33}{9034533.06}[(\frac{1}{3005.75})(-76.67) + (\frac{-4131.33}{9034533.06})((2054.20)^2)]]$$

$$\Rightarrow S = .20$$

Subproblem B

Let's rename g to be η. Now we know that

$$\hat{\eta} \sim N(\eta, S^2)$$

So by the rules of variance,

$$\frac{\hat{\eta}}{.25} \sim N(\frac{\eta}{.25}, \frac{S^2}{(.25)^2})$$

$$\Rightarrow \frac{\hat{\eta}}{.25} \sim N(\frac{\eta}{.25}, \frac{(.20)^2}{(.25)^2})$$

$$\Rightarrow \frac{\hat{\eta}}{.25} \sim N(\frac{\eta}{.25}, .64)$$

The experiment's realization of $\hat{\eta}$ is

$$\hat{\eta} = \frac{\bar{Y_A} - \bar{Y_B}}{\bar{Y_B}} = \frac{4131.33 - 3005.75}{3005.75} = .37$$

Hence producing a 95% confidence interval for $\frac{\hat{\eta}}{.25}$ is straightforward.

$$\frac{\eta}{.25} \in \left[\frac{.37}{.25} \pm 1.96(\sqrt{.64})\right]$$

$$\Rightarrow \frac{\eta}{.25} \in [.23, 2.37]$$

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 sum of the individual 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

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

the interpretation here is that the "residual", that is, $E[Y_i] - \Lambda(X_i'\beta)$, is uncorrelated with X_i. $\Lambda(X_i'\beta)$ can be thought of as the "predicted value" for Y_i, since $P(Y_i = 1|X_i) = \Lambda(X_i'\beta)$.

c) The moment conditions identifying β_{NLLS} are

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

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, and $\Lambda(X_i'\beta)$ is the actual conditional expectation function of Y_i given X_i.

e) Under proper specification of the model, the asymptotic variance of the estimator is the additive inverse of the inverse of the expectation of the Hessian of the log likelihood.[1] As was derived in part a), the score (i.e. the gradient of the log-likelihood) is

$$s(X_i, \beta) = \left(Y_i - \frac{\exp(X_i'\beta)}{1+\exp(X_i'\beta)}\right)X_i'$$

Thus, the Hessian matrix is

$$\begin{aligned} \nabla_\beta s(X_i, \beta) &= -\frac{\exp(X_i'\beta)(1+\exp(X_i'\beta)) - \exp(X_i'\beta)^2}{(1+\exp(X_i'\beta))^2}X_iX_i' \\ &= -\frac{\exp(X_i'\beta)}{(1+\exp(X_i'\beta))^2}X_iX_i' \end{aligned}$$

And so, the asymptotic variance of β_{ML} under correct specification is given by

$$\sqrt{N}(\beta_{ML} - \beta) \xrightarrow{d} \mathcal{N}(0, H(\beta)^{-1})$$

with

$$H(\beta) = E\left[\frac{\exp(X_i'\beta)}{(1+\exp(X_i'\beta))^2}X_iX_i'\right]$$

[1] I just wanted to write that all out.

f) Let $V_s = E\left[s(X_i,\beta)s(X_i,\beta)'\right]$, the outer product of the score. Under mispecification, the asymptotic variance of β_{ML} is as follows:

$$\sqrt{N}(\beta_{ML}-\beta) \xrightarrow{d} \mathcal{N}\left(0, H(\beta)^{-1}V_s H(\beta)^{-1}\right)$$

with $H(\beta)$ defined as in the previous problem and

$$V_s = E\left[\left(Y_i - \frac{\exp(X_i'\beta)}{1+\exp(X_i'\beta)}\right)^2 X_i X_i'\right]$$

g) Suppose the data are independent across but not necessarily within clusters. Propose a cluster robust estimator of the asymptotic variance of β_{ML}

One idea is to construct a similar estimator to the one we have for cluster robust standard errors for OLS:

$$\frac{J}{J-K}\frac{1}{N}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\exp(X_i'\beta)}{(1+\exp(X_i'\beta))^2}X_i'X_i\right)^{-1}\left(\frac{1}{N}\sum_{j=1}^{J}X_j u_j u_j' X_j'\right)\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\exp(X_i'\beta)}{(1+\exp(X_i'\beta))^2}X_i'X_i\right)^{-1}$$

where

$$u_j = \left[\left(Y_s - \frac{\exp(X_s'\beta)}{(1+\exp(X_s'\beta))^2}\right)\right]_{s\in j}.$$

That is, u_j is the vector of "residuals" for observations s in cluster j.

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.

4. Clustered DGP in Stata

a) Regressing Y_{ic} on D_{ic}, we get the results in table 1.

Table 1: Clustered DGP

	(1) Y_ic, Regular SE	(2) Y_ic, Robust SE	(3) Y_ic, Clustered SE
d_ic	1.010***	1.010***	1.010***
	(0.03237)	(0.03244)	(0.09508)
Constant	0.0394*	0.0394*	0.0394
	(0.02278)	(0.02037)	(0.10761)
R^2	0.089	0.089	0.089
Observations	10000	10000	10000

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01

The basic SE are 0.03237. The standard errors are slightly larger (0.03244) when we correct for heteroskedasticity in column 2 and significantly larger (0.09508) when we account for the intracluster correlation of observations from the same cluster.

b) Changing σ_η^2 to 0. Regressing Y_{ic} on D_{ic}, we get the results in table :

The standard errors are fairly similar for columns 1 and 2 to what we had in A. They are 0.02732 for the regular SE, 0.027731 for robust SE. The clustered SE (0.02420) are similar to the magnitude of the non-clustered version.

c) Comment on your differences in the answers to a) and b).
The SE are fairly similar in columns 1 and 2 for parts a) and b) though they are slightly smaller in part b as a result of the lower variance of y_{ic} in part b. In part b since there is no intra-cluster correlation in the coefficient on d_{ic} the clustered standard errors are of a similar magnitude to the non-clustered version (0.02420).

Table 2: Clustered DGP

	(1) Y_ic, Regular SE	(2) Y_ic, Robust SE	(3) Y_ic, Clustered SE
d_ic	0.992***	0.992***	0.992***
	(0.02732)	(0.02731)	(0.02420)
Constant	-0.00105	-0.00105	-0.00105
	(0.01911)	(0.01923)	(0.09650)
R^2	0.117	0.117	0.117
Observations	10000	10000	10000

Standard errors in parentheses

* p<0.10, ** p<0.05, *** p<0.01

d) In table 3, we see that the simulation rejects the null that the coefficient on d_{ic}=1 57% of the time when we don't cluster our standard errors and about 5% (which we would expect) when we cluster. So the non-clustered version is rejecting too often and the clustered version is rejecting at about the rate of p which we set.

Table 3: Monte Carlo Simulations

	(1)				
	count	mean	sd	min	max
reject	1000	.567	.4957386	0	1
reject_cluster	1000	.055	.2280943	0	1

e) Here we have that the regression with non-clustered standard errors rejects the test that the coefficient equals the average of the cluster level η's 3% of the time and the clustered standard errors are never rejecting, so in this case clustering is causing us not reject often enough.

Table 4: Monte Carlo Simulations

	(1)				
	count	mean	sd	min	max
reject	1000	.03	.1706726	0	1
reject_cluster	1000	0	0	0	0

f) When treatment is at the cluster level like in the case of the set up for part d then clustering standard errors will produce the accurate standard errors. However, in e, when the treatment is at the individual level but those individuals are within a cluster, for example a school or village then clustering the errors will actually cause our standard errors to be larger than we would want, leading to us to not reject enough. For the purpose of the internal validity of the study, using non-clustered standard errors when treatment is at the individual level is appropriate.

5. CPS and WLS

a) See results in column 1 of table 5.

Table 5: Wage Regression (Weighted)

	(1) Log wage	(2) Log wage (weighted by cell size)	(3) Log wage (weighted by 1/cell var)
agecat==2	0.314***	0.314***	0.314***
	(0.00746)	(0.00073)	(0.00047)
agecat==3	0.458***	0.458***	0.458***
	(0.00731)	(0.00071)	(0.00046)
agecat==4	0.494***	0.494***	0.494***
	(0.00771)	(0.00075)	(0.00048)
agecat==5	0.472***	0.472***	0.472***
	(0.00966)	(0.00094)	(0.00061)
agecat==6	0.295***	0.295***	0.295***
	(0.01651)	(0.00161)	(0.00104)
educcat==2	0.265***	0.265***	0.265***
	(0.00754)	(0.00073)	(0.00047)
educcat==3	0.415***	0.415***	0.415***
	(0.00764)	(0.00074)	(0.00048)
educcat==4	0.771***	0.771***	0.771***
	(0.00783)	(0.00076)	(0.00049)
Sex	-0.261***	-0.261***	-0.261***
	(0.00458)	(0.00045)	(0.00029)
Constant	2.153***	2.153***	2.153***
	(0.00995)	(0.00097)	(0.00063)
R^2	0.297	0.978	0.978
Observations	56182	56182	135056

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

b) Collapsed results used in analysis below.

c) See column 2 of table 5. The coefficients are identical (as we would expect) but the standard errors are much smaller when we weight by bin size.

d) If wages are *iid* within a cell, then the variance of each cell is σ^2/N_c where σ^2 is the population variance and N_c is the number of observations in the cell.

e) Formula above is used to calculate the variance of the cell means.

f) See results in column 3 of table 5.

g) With an RSS of 0.133 and 38 degrees of freedom we are not able to reject this model at the 5% level. In other words we cannot reject our model explains the data up to a sampling error.

h) See results in table 7.

i) With an RSS of 0.009 and 15 degrees of freedom we cannot reject this goodness of fit of this model at the 5% level either.

j) See fig. 1. The model does a very nice job of fitting the data. The place it has a bit of trouble fitting is for older more educated individuals. The model overestimates the wages of 65+ males with some college and underestimates the wages for males with a BA or more.

k) A simple model which passes the chi-squared test is just lnwage on sex. We cannot reject that this model explains all of the variation in lnwage up to a sampling error.

Table 6: Wage Regression (Weighted)

	(1)
	Log wage
agecat==2	0.301***
	(0.00103)
agecat==3	0.414***
	(0.00103)
agecat==4	0.424***
	(0.00103)
agecat==5	0.407***
	(0.00106)
agecat==6	0.355***
	(0.00068)
educcat==2	0.109***
	(0.00076)
educcat==3	0.238***
	(0.00083)
educcat==4	0.757***
	(0.00081)
Sex	-0.192***
	(0.00054)
<=25 × Dropout	0
	(.)
<=25 × HS	0.0576***
	(0.00078)
<=25 × Some College	0.00573***
	(0.00084)
<=25 × BA+	-0.135***
	(0.00087)
25-35 × Dropout	0.00900***
	(0.00084)
25-35 × HS	0.135***
	(0.00071)
25-35 × Some College	0.152***
	(0.00078)
25-35 × BA+	0
	(.)
35-45 × Dropout	-0.0677***
	(0.00084)
35-45 × HS	0.0784***
	(0.00070)
35-45 × Some College	0.154***
	(0.00078)
35-45 × BA+	0
	(.)
45-55 × Dropout	-0.0676***
	(0.00086)
45-55 × HS	0.112***
	(0.00071)
45-55 × Some College	0.158***
	(0.00079)
45-55 × BA+	0
	(.)
55-65 × Dropout	-0.00166*
	(0.00090)
55-65 × HS	0.0884***
	(0.00076)
R^2	0.998
Observations	135056

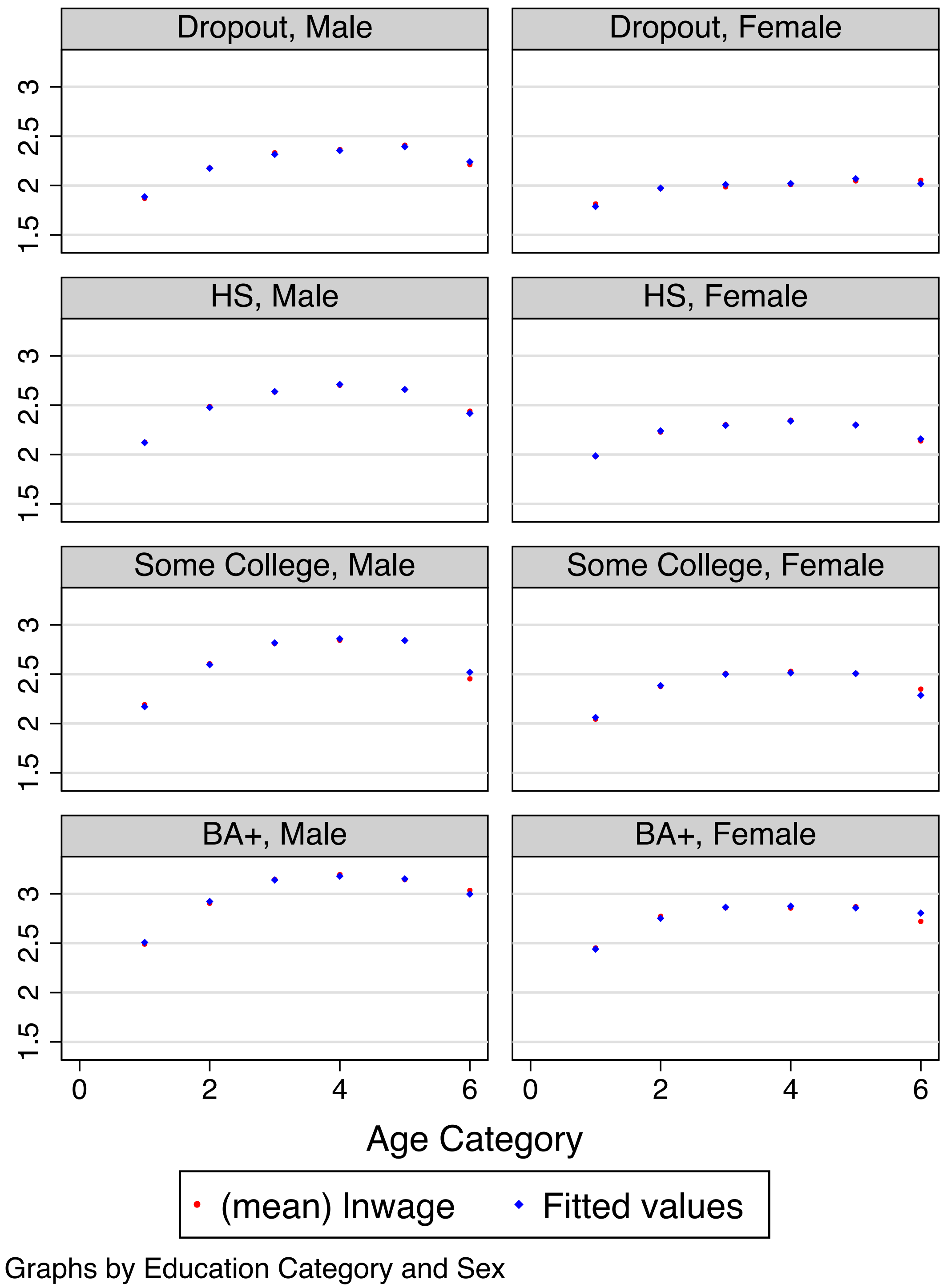
Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Table 7: Wage Regression (Weighted) cont

	(1) Log wage
55-65 × Some College	0.167***
	(0.00084)
55-65 × BA+	0
	(.)
>65 × Dropout	0
	(.)
>65 × HS	0
	(.)
>65 × Some College	0
	(.)
>65 × BA+	0
	(.)
<=25 × Male	0
	(.)
<=25 × Female	0.124***
	(0.00055)
25-35 × Male	-0.0212***
	(0.00054)
25-35 × Female	0
	(.)
35-45 × Male	0.0836***
	(0.00054)
35-45 × Female	0
	(.)
45-55 × Male	0.112***
	(0.00054)
45-55 × Female	0
	(.)
55-65 × Male	0.101***
	(0.00058)
55-65 × Female	0
	(.)
>65 × Male	0
	(.)
>65 × Female	0
	(.)
Dropout × Male	0
	(.)
Dropout × Female	-0.0302***
	(0.00026)
HS × Male	0.0672***
	(0.00020)
HS × Female	0
	(.)
Some College × Male	0.0410***
	(0.00020)
Some College × Female	0
	(.)
BA+ × Male	0
	(.)
BA+ × Female	0
	(.)
Constant	2.078***
	(0.00057)
R^2	0.998
Observations	135056

Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01

Figure 1: Predicted and actual log wages by subgroup



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:

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

For σ:

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

b) See *max_sml.m* file.

c) See *max_sml.m* file.

d) These results are invariant to initial conditions:

$$\begin{bmatrix} \mu \\ ln(\sigma) \end{bmatrix} = \begin{bmatrix} 1.6011 \\ 0.5139 \end{bmatrix}$$

e) Let $\hat{\theta}_{MSL}$ denote the estimated parameters from the method of simulated likelihood and the true parameters $\theta = [\mu, \sigma]'$. Under proper specification, we know that

$$\sqrt{N}(\hat{\theta}_{MSL} - \theta) \rightarrow \mathcal{N}(\mathbf{0}, \mathbf{H}(\theta)^{-1})$$

By the delta-method, with $g(\theta) \equiv [\mu, ln(\sigma)]'$, we have that

$$\sqrt{N}(g(\hat{\theta}_{MSL}) - g(\theta)) \rightarrow \mathcal{N}(\mathbf{0}, \mathbf{G}\mathbf{H}(\theta)^{-1}\mathbf{G})$$

where

$$\mathbf{G} \equiv \frac{\partial g(\theta)}{\partial \theta} = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{\sigma} \end{bmatrix}$$

When we calculate the square root of the diagonal elements of $\frac{1}{N}\mathbf{H}(\hat{\theta})^{-1}$, we derive the standard errors of our estimates of μ and $\ln(\sigma)$: $(0.2633, 0.3884)$.

f) To calculate the standard errors for the incorrect model, we replace $\mathbf{H}(\theta)^{-1}$ with $\mathbf{H}(\theta)^{-1}\mathbf{V_s}\mathbf{H}(\theta)^{-1}$, where $\mathbf{V_s}$ is the population variance-covariance matrix of the scores—defined by the the criterion function we are maximizing.

Again, we use the analogy principle to replace the asymptotic variance terms with the simulated likelihood, sample averages. This yields standard errors of μ and $ln(\sigma)$: $(0.4839, 0.6417)$.