**Suggested Answers**
**Problem set 8**
**Economics 60303**

Bill Evans
Spring 2014

1. Let $\Pr(x \le a) = \Phi[(a-\mu)/\sigma]$ and therefore, the PDF of the truncated distribution is

$$f(x \mid x > a) = \frac{1}{\sigma}\phi\left(\frac{x}{\sigma}\right) / \left(1-\Phi\left((a-\mu)/\sigma\right)\right)$$

Following the methods we used in class, it is straightforward to show that

$$E[x \mid x > a] = \mu + \frac{\sigma\phi\left((a-\mu)/\sigma\right)}{1-\Phi((a-\mu)/\sigma)}$$

2. We know that the conditional mean of y given x is

$$E[y \mid x] = \mu_y + \rho\left(\frac{\sigma_y}{\sigma_x}\right)(x-\mu_x)$$

And we also know from question 1 that we can write the truncated mean of x as

$$E[x \mid x > a] = \mu_x + \frac{\sigma_x\phi\left((a-\mu_x)/\sigma_x\right)}{1-\Phi((a-\mu_x)/\sigma_x)}$$

We also know that by the law of iterated expectations

$$E[y \mid x > a] = E[E[y \mid x] \mid x > a] = E\left[\mu_y + \rho\left(\frac{\sigma_y}{\sigma_x}\right)(x-\mu_x) \mid x > a\right]$$

Which equals

$$\mu_y + \rho\left(\frac{\sigma_y}{\sigma_x}\right)(E\left[E[(x-\mu_x) \mid y] \mid x > a\right]) = \mu_y + \rho\left(\frac{\sigma_y}{\sigma_x}\right)\left(\mu_x + \frac{\sigma_x\phi\left((a-\mu_x)/\sigma_x\right)}{1-\Phi((a-\mu_x)/\sigma_x)}\right)$$

$$= \mu_y + \left(\frac{\rho\sigma_y\phi\left((a-\mu_x)/\sigma_x\right)}{1-\Phi((a-\mu_x)/\sigma_x)}\right)$$

3. Consider a model where

$$y_i = \alpha + \varepsilon_i$$

Uncensored observations have a PDF of the form

$$\frac{1}{\sigma_\varepsilon}\phi\left(\frac{\varepsilon_i}{\sigma_\varepsilon}\right) = \frac{1}{\sigma_\varepsilon}\phi\left(\frac{y_i - \alpha}{\sigma_\varepsilon}\right)$$

The probability an observation is topcoded is

$$\Pr(y_i > y^t) = \Pr[z_i = (y-u)/\sigma > (y^t - u)/\sigma] = 1 - \Phi[(y^t - u)/\sigma] = \Phi[(u/\sigma) - (y^t/\sigma)]$$

In this model, redefine $u = \alpha$ and let $\sigma = \sigma_\varepsilon$. Therefore, the likelihood function for observation i is

$$L_i = \left[\frac{1}{\sigma}\phi\left(\frac{y_i - u}{\sigma}\right)\right]^{T_i}\left[\Phi[(u/\sigma) - (y^t/\sigma)]\right]^{1-T_i}$$

In this case, the coefficient on the constant in the model gives the ratio of the mean of the uncensored distribution divided by the standard deviation of the uncensored distribution.

The results from the topit specification are of the form

```
Tobit regression                                  Number of obs   =      19906
                                                  LR chi2(0)      =       0.00
                                                  Prob > chi2     =          .
Log likelihood = -16862.062                       Pseudo R2       =     0.0000

------------------------------------------------------------------------------
ln_weekly_~n |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   6.084551   .0039005  1559.96   0.000     6.076906    6.092196
-------------+----------------------------------------------------------------
      /sigma |   .5469946   .0029041                      .5413022    .5526869
------------------------------------------------------------------------------
  Obs. summary:          0  left-censored observations
                     18474     uncensored observations
                      1432 right-censored observations at ln_weekly_~n>=6.906755
```

In this case, the constant represents μ and sigma represents σ so the unconditional distribution of ln(weekly earnings) is normal with a mean of 6.084 and a standard deviation of 0.547. Notice that the sample descriptive statistics of the censored variable are

```
. sum ln_weekly_earn
```

```
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
ln_weekly_~n |     19906    6.067307     .513047   4.094345   6.906755
```

Using the results from problem

$$E[x\,|\,x>a]=\mu+\frac{\sigma\phi\left((a-\mu)/\sigma\right)}{1-\Phi((a-\mu)/\sigma)}=6.084+0.547\left(\frac{\phi(\ln(999)-6.084/0.547)}{1-\Phi(\ln(999)-6.084/0.547)}\right)=6.084+0.547(1.94)=7.15$$

Notice that exp(7.15) is $1269.5 ,  If you replace topcoded values with 7.15, the OLS estimate of that model is

```
gen earnwkl5=earnwkl;

. replace earnwkl5=7.15 if topcode==1;
(1432 real changes made)

. reg earnwkl5 age age2 educ black hispanic union;

      Source |       SS       df       MS              Number of obs =   19906
-------------+------------------------------           F(  6, 19899) = 1478.79
       Model |  1820.31808     6  303.386347           Prob > F      =  0.0000
    Residual |   4082.4587 19899  .205158988           R-squared     =  0.3084
-------------+------------------------------           Adj R-squared =  0.3082
       Total |  5902.77678 19905  .296547439           Root MSE      =  .45294

------------------------------------------------------------------------------
    earnwkl5 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0704456   .0021266    33.13   0.000     .0662772    .0746139
        age2 |  -.0006958    .000026   -26.78   0.000    -.0007467   -.0006449
        educ |   .0754998   .0012033    62.74   0.000     .0731411    .0778584
       black |  -.2199837   .0117719   -18.69   0.000    -.2430576   -.1969099
    hispanic |  -.1069377   .0141308    -7.57   0.000    -.1346351   -.0792402
       union |   .1179767   .0077448    15.23   0.000     .1027962    .1331572
       _cons |   3.503225   .0418847    83.64   0.000     3.421128    3.585323
------------------------------------------------------------------------------
```

Which are amazing similar to the tobit estimates

```
Tobit regression                                  Number of obs   =      19906
                                                  LR chi2(6)      =    7309.06
                                                  Prob > chi2     =     0.0000
Log likelihood = -13207.534                       Pseudo R2       =     0.2167

------------------------------------------------------------------------------
     earnwkl |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0703864     .00214    32.89   0.000     .0661919     .074581
        age2 |  -.0006948   .0000262   -26.55   0.000    -.0007461   -.0006435
        educ |   .0757658   .0012172    62.25   0.000       .07338    .0781515
       black |  -.2200147    .011795   -18.65   0.000    -.2431339   -.1968954
    hispanic |  -.1058161   .0141638    -7.47   0.000    -.1335783   -.0780539
       union |   .1191111   .0077791    15.31   0.000     .1038634    .1343588
```

```
       _cons |   3.499009   .0421806    82.95   0.000     3.416332    3.581686
-------------+----------------------------------------------------------------
      /sigma |   .4530426   .0023983                      .4483418    .4577434
------------------------------------------------------------------------------
  Obs. summary:          0  left-censored observations
                     18474     uncensored observations
                      1432 right-censored observations at earnwkl>=6.906755

is virtually identical
```

4. See Evans, Oates and Schwab, Journal of Political Economy, 1992. Let $g(\varepsilon_i, v_i)$ be the bivariate normal PDF described in the problem. Note that when

$y_i = 1$ we have $g(\varepsilon_i \geq -\beta_0 - x_i\beta_1 - w_i\beta_2, v_i = x_i - \pi_0 - z_i\pi_1 - w_i\pi_2)$ and when

$y_i = 0$ we have $g(\varepsilon_i < -\beta_0 - x_i\beta_1 - w_i\beta_2, v_i = x_i - \pi_0 - z_i\pi_1 - w_i\pi_2)$

For simplicity, write $\beta_0 + x_i\beta_1 + w_i\beta_2 + \varepsilon_i$ as $x_{1i}\beta + \varepsilon_i$ and $\pi_0 + z_i\pi_1 + w_i\pi_2 + v_i$ as $z_{1i}\pi + v_i$. The likelihood function for person i is

$$L_i = \left[g(\varepsilon_i \geq -x_{1i}\beta, v_i = x_i - z_{1i}\pi)\right]^{y_i=1}\left[g(\varepsilon_i < -x_{1i}\beta, v_i = x_i - z_{1i}\pi)\right]^{y_i=0}$$

To estimate the values of the loglikelihood, we exploit Baye's theorem where

$\Pr(A|B) = \Pr(A \cap B)/\Pr(B)$ so $\Pr(A \cap B) = \Pr(A|B)\Pr(B)$. Let A=$v_i$ and B=$\varepsilon_i$

Note that $\varepsilon_i | v_i \sim N[\rho v_i/\sigma_v, (1-\rho^2)]$ and the PDF for $v_i$ is $f(v_i) = \dfrac{1}{\sigma_v}\phi\left(\dfrac{x_i - z_{1i}\pi}{\sigma_v}\right)$.

The conditional probability $\Pr(y_i = 1 | v_i)$ is

$$\Pr(y_i = 1 | v_i) = \Pr(\varepsilon_i \geq -x_{1i}\beta | v_i) = 1 - \Phi\left[\frac{-x_{1i}\beta - \rho v_i/\sigma_v}{\left(1-\rho^2\right)^{0.5}}\right] = \Phi\left[\frac{x_{1i}\beta + \rho v_i/\sigma_v}{\left(1-\rho^2\right)^{0.5}}\right]$$

$$= \Phi\left[\frac{x_{1i}\beta + \rho(x_i - z_{1i}\pi)/\sigma_v}{\left(1-\rho^2\right)^{0.5}}\right]$$

Likewise, it is easy to show that

$$\Pr(y_i = 0 \mid v_i) = 1 - \Phi\left[\frac{x_{1i}\beta + \rho(x_i - z_{1i}\pi)/\sigma_v}{\left(1-\rho^2\right)^{0.5}}\right]$$

Putting these together, the likelihood function is the

$$L_i = \left[\Phi\left[\frac{x_{1i}\beta + \rho(x_i - z_{1i}\pi)/\sigma_v}{\left(1-\rho^2\right)^{0.5}}\right] f(v_i)\right]^{y_i} \left[\left(1 - \Phi\left[\frac{x_{1i}\beta + \rho(x_i - z_{1i}\pi)/\sigma_v}{\left(1-\rho^2\right)^{0.5}}\right]\right) f(v_i)\right]^{1-y_i}$$

Taking logs and summing over all i

$$\ell = \sum_{i=1}^{n}\left\{ y_i\left(\ln\Phi\left[\frac{x_{1i}\beta + \rho(x_i - z_{1i}\pi)/\sigma_v}{\left(1-\rho^2\right)^{0.5}}\right]\right) + (1-y_i)y_i \ln\left(1 - \Phi\left[\frac{x_{1i}\beta + \rho(x_i - z_{1i}\pi)/\sigma_v}{\left(1-\rho^2\right)^{0.5}}\right]\right) + \ln\left[f(x_i - z_{1i}\pi)\right]\right\}$$

Notice that is $\rho = 0$ the model collapses to

$$\ell = \sum_{i=1}^{n}\left\{ y_i\left(\ln\Phi\left[x_{1i}\beta\right]\right) + (1-y_i)y_i \ln\left(1-\Phi\left[x_{1i}\beta\right]\right) + \ln\left[f(x_i - z_{1i}\pi)\right]\right\}$$

Which is a probit and an OLS – maximized separately.

5. When in doubt, draw a graph. In Figure 1 below, we have an uncensored scatter plot of x and y. In Figure 2, we censor the data from below. Given the mass of observations are the point of censoring, the estimated OLS line is now much flatter than before.

**Figure 1: Plot of X and Y1**

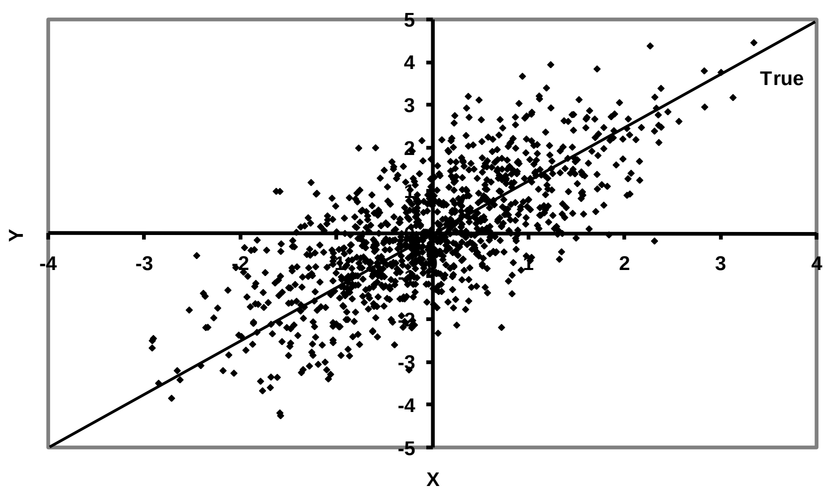
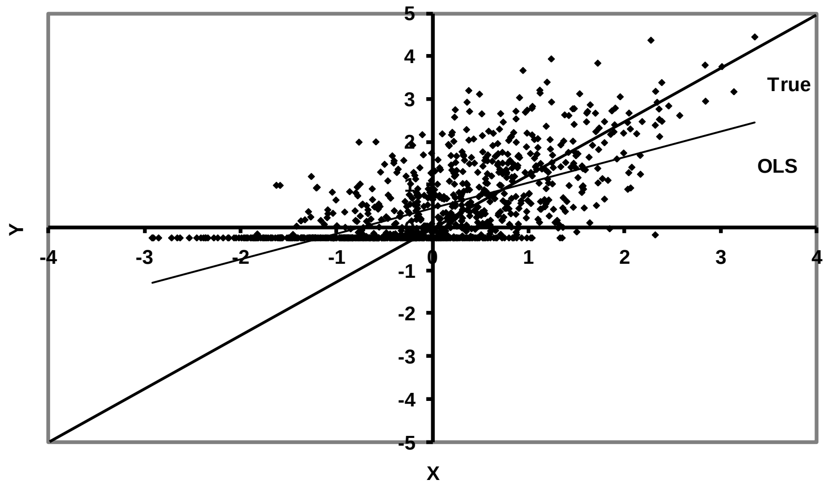


**Figure 2: Plot of X and Y Censored from Below**




To show that censoring from below generates attenuation in the estimate for the slope, let $D_i = 1$ is a variable is not censored and $D_i = 0$ if it is censored. The OLS estimate of $\beta_1$ when the data is not censored is given as

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})y_i^*}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\sum_{i=1}^{n}\left(D_i(x_i - \overline{x})y_i^* + (1-D_i)(x_i - \overline{x})y_i^*\right)}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{a+b}{d}$$

The OLS estimate of $\beta_1$ when the data is censored is given as

$$\hat{\beta}_1^c = \frac{\sum_{i=1}^{n}(x_i - \overline{x})y_i}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\sum_{i=1}^{n}\left(D_i(x_i - \overline{x})y_i^* + (1-D_i)(x_i - \overline{x})y^t\right)}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{a+c}{d}$$

Note that since the data is censored from below, $y^f > y_i^*$. Note as well that since y is censored from below we anticipate that we are on the left hand side of the distribution for x and in most cases, $(1-D_i)(x_i - \overline{x})y^t$ <0 because $(x_i - \overline{x})$ <0. Therefore, $(1-D_i)(x_i - \overline{x})y^f < (1-D_i)(x_i - \overline{x})y_i^* < 0$
or c<b<0. Since a>0 and construction d>0, a+c<a+b, this is because c is subtracting a larger value off of (a) than is (b), generating some attenuation bias in the estimate for $\hat{\beta}_1^c$.

8. There are three pictures below that help answer the problem. IN the pictures, I have graphed the contours of the bivariate normal PDF with a positive covariance. In the first figure, we graphically illustrate the area we want to calculate with the bivariate normal CDF which is $\Pr(\varepsilon_i < a_i^c, v_i < b_i^c) = G(a_i^c, b_i^c)$.

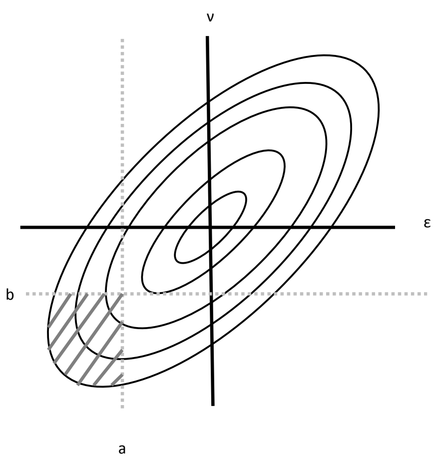
In the second graph, we consider the area for $\Pr(y_i = 0, x_i = 1) = \Pr(\varepsilon_i < a_i^c, v_i \geq b_i^c)$. Note that the single cross hatched area give us $\Pr(\varepsilon_i < a_i^c) = \Phi(\varepsilon_i < a_i^c)$ but this area has too much area. To get the appropriate area, we must subtract off $\Pr(\varepsilon_i < a_i^c, v_i < b_i^c) = G(a_i^c, b_i^c)$ so

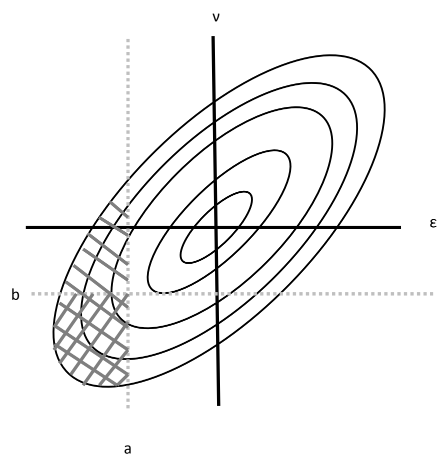
$$\Pr(y_i = 0, x_i = 1) = \Pr(\varepsilon_i < a_i^c, v_i \geq b_i^c) = \Phi(a_i^c) - G(a_i^c, b_i^c)$$

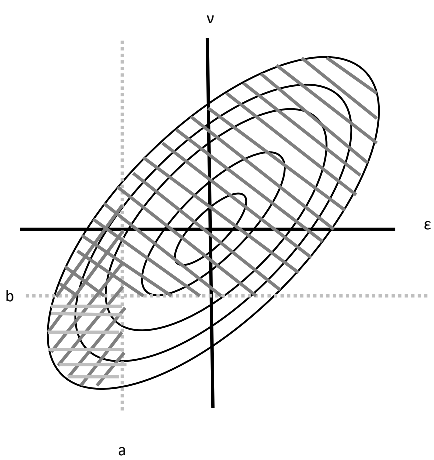
It should be no surprise that $\Pr(y_i = 1, x_i = 0) = \Pr(\varepsilon_i \geq a_i^c, v_i < b_i^c) = \Phi(b_i^c) - G(a_i^c, b_i^c)$

The area for $\Pr(y_i = 1, x_i = 1) = \Pr(\varepsilon_i \geq a_i^c, v_i \geq b_i^c)$ is given in the third picture. The single cross-hatched area is $\Pr(v_i \geq b_i^c) = 1 - \Phi(b_i^c)$. This area is of course too big. If we subtract off $\Pr(\varepsilon_i < a_i^c) = \Phi(\varepsilon_i < a_i^c)$ we are subtracting off too much so we must add back $G(a_i^c, b_i^c)$ so

$$\Pr(y_i = 1, x_i = 1) = \Pr(\varepsilon_i \geq a_i^c, v_i \geq b_i^c) = 1 - \Phi(b_i^c) - \Phi(a_i^c) + G(a_i^c, b_i^c)$$



**Computer portion of the problem set**

Below is my table. Note four things. First, as ρ goes from large negative to a large positive number, the OLS estimate of $\hat{\beta}_1$ goes from a large positive to a large negative number. Second, in each case, the 2SLS model does fine job of replicating the treatment effect estimate. Third, note that the average t-statistic is the same in all 2SLS models. This is because the t-statistic squared in the 2SLS is roughly equal to the t-statistic squared in the reduced-form. That relationship does not change at all when we move from ρ being a large negative to a large positive number. Fourth, for ρ small, the efficiency of the treatment over the 2SLS is marginal. However, as ρ increases in value, there is a clear efficiency gain in using the treatment effect model.

| | OLS | | 2SLS | | Treatment effect model | | |
|---|---|---|---|---|---|---|---|
| Value of $P$ | Average $\hat{\beta}_1$ | Average t-stat on $\hat{\beta}_1$ | Average $\hat{\beta}_1$ | Average t-stat on $\hat{\beta}_1$ | Average $\hat{\beta}_1$ | Average z-score on $\hat{\beta}_1$ | Average $\hat{\rho}$ |
| 0.50 | 11.75 | 59.7 | -6.39 | -2.01 | -5.94 | -4.38 | 0.50 |
| 0.25 | 2.90 | 13.8 | -6.30 | -2.01 | -6.08 | -2.51 | 0.25 |
| 0.00 | -5.96 | -27.8 | -6.20 | -2.00 | -6.16 | -2.10 | 0.01 |
| -0.25 | -14.8 | -70.6 | -6.08 | -2.00 | -6.26 | -2.36 | -0.24 |
| -0.50 | -23.7 | -120.2 | -5.95 | -1.99 | -5.96 | -4.29 | -0.50 |