Problem set 8
Economics 60303
(Due, Friday, April 25, 2013)

Bill Evans
Spring 2014

1. Suppose $x \sim N(\mu_x, \sigma_x^2)$. What is $E[x \mid x > a]$?

2. Suppose x and y are distributed as a bivariate normal where

$$\begin{bmatrix} x \\ y \end{bmatrix} \sim N\left(\begin{bmatrix} u_x \\ u_y \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix} \right)$$

What is $E[y \mid x > a]$?

3. Suppose y_i=ln(weekly earnings$_i$) for person i and y_i is topcoded at y^t. Assume that $y_i \sim N(\mu, \sigma^2)$. Suppose one were to estimate a Tobit model and the only covariate in the model is a constant. Show that with this model, you can estimate the underlying values of μ and σ^2. Using cpc87.dta, generate ln_weekly_earn and run a tobit model

```
tobit ln_weekly_earn, ul
```

what are the estimates for μ and σ^2 in this case.

Next, using the results from question 1, replace the topcoded value (ln(999)) with $E[x \mid x > a]$. Estimate a regression similar to the one for the Tobit section handouts (covariates include age, age2, education, plus dummies for black, Hispanic, and union). How do these results compare to the estimates from the tobit specification?

4. Consider the following limited dependent variable model. Suppose a dichotomous outcome y_i is a function of exogenous characteristics w_i and a potentially endogenous but continuous variable x_i. The decision to engage in activity y is described by the equation

$$\begin{aligned} y_i &= 1 \quad if \quad y_i^* = \beta_0 + x_i\beta_1 + w_i\beta_2 + \varepsilon_i \geq 0 \\ y_i &= 0 \quad if \quad y_i^* = \beta_0 + x_i\beta_1 + w_i\beta_2 + \varepsilon_i < 0 \end{aligned}$$

The concern in this model is that *x* and *y* are linked through the errors where

$$x_i = \pi_0 + z_i\pi_1 + w_i\pi_2 + v_i$$

The variable *z* is a scalar instrument and the joint error term is defined as a bivarate normal distribution where

$$\begin{bmatrix} \varepsilon_i \\ v_i \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{bmatrix} \right)$$

What is the log likelihood function for this maximum likelihood problem?

5. Consider a regression model that is censored from below. Let y_i^* be the true value of the outcome and if perfectly measured, the equation describing y_i^* would be

$$y_i^* = \beta_0 + x_i\beta_1 + \varepsilon_i$$

Where x_i is a scalar and we anticipate and define the estimate for $\hat{\beta}_1 > 0$. Suppose that for some reason, the data is censored from below. Define y_i to be the measured outcome and the measured value is

$$y_i = \begin{Bmatrix} y_i^* \ if \ y_i^* > y^t \\ y^t \ if \ y_i^* \le y^t \end{Bmatrix}$$

The OLS estimate of β_1 using censored data is $\hat{\beta}_1^c = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$.

In this case, does censoring bias the estimates of $\hat{\beta}_1^c$ up or down? Show all your work.

6. Suppose y is a dummy outcome (0 or 1) and the equation that generates an outcome of 1 is

$$y_i = 1 \quad if \quad y_i^* = \beta_0 + x_i\beta_1 + w_i\beta_2 + \varepsilon_i \ge 0$$

Where x is also a (0-1 dummy variable) where the equation that generates an outcome of 1is described by the equation

$$x_i = 1 \quad if \quad x_i^* = \pi_0 + z_i\pi_1 + w_i\pi_2 + v_i \ge 0$$

The equations are linked by assuming the errors are bivariate normal

$$\begin{bmatrix} \varepsilon_i \\ v_i \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right)$$

The likelihood function for person i is given by

$$L_i = \Pr(y_i = 1, x_i = 1)^{y_i x_i} \Pr(y_i = 1, x_i = 0)^{y_i(1-x_i)} \Pr(y_i = 0, x_i = 1)^{(1-y_i)x_i} \Pr(y_i = 0, x_i = 0)^{(1-y_i)(1-x_i)}$$

Let $g(\varepsilon_i, v_i; \rho)$ be the bivariate normal pdf and the CDF is defined as

$$G(a,b) = \int_{-\infty}^{a} \int_{-\infty}^{b} g(\varepsilon_i, v_i; \rho) dv_i d\varepsilon_i$$

Note that $(y_i = 0, x_i = 0)$ will only occur if $\varepsilon_i < -\beta_0 - x_i\beta_1 - w_i\beta_2 = a_i^c$ and $v_i < -\pi_0 - z_i\pi_1 - w_i\pi_2 = b_i^c$.

Any quality programming language will provide you a function to evaluate the bivariate normal (BVN) CDF and hence, it would be trivial to calculate $\Pr(y_i = 0, x_i = 0)$ as

$$\Pr(\varepsilon_i < a_i^c, v_i < b_i^c) = G(a_i^c, b_i^c)$$

given candidate values for the parameters. However, given the BVN function $G(\bullet,\bullet)$ only evaluates the CDF, how would you evaluate

$$\Pr(y_i = 1, x_i = 0) = \Pr(\varepsilon_i \ge a_i^c, v_i < b_i^c) = ?$$
$$\Pr(y_i = 0, x_i = 1) = \Pr(\varepsilon_i < a_i^c, v_i \ge b_i^c) = ?$$
$$\Pr(y_i = 1, x_i = 1) = \Pr(\varepsilon_i \ge a_i^c, v_i \ge b_i^c) = ?$$

Computer portion of the problem set

Limited dependent variable models are thought to do a better job of describing the data generating process than least-squares models. In this problem, we are going to do some simulations that compare outcomes from a simultaneous equation system that explicitly models the limited nature of variables with a 2SLS model where we ignore the data generating process and estimate the model by least squares.

The equation of interest can be written in the following form:

$$y_i = \beta_0 + x_i\beta_1 + w_i\beta_2 + \varepsilon_i$$

where y_i and w_i are continuous, and x_i is a dummy variable. The concern is that $cov(x_i,\varepsilon_i) \neq 0$. The selection equation that describes the realization of the dichotomous variable x_i can be described as follows

$$x_i = 1 \quad if \quad x_i^* = \gamma_0 + z_i\gamma_1 + w_i\gamma_2 + v_i \ge 0$$
$$x_i = 0 \quad if \quad x_i^* = \gamma_0 + z_i\gamma_1 + w_i\gamma_2 + v_i < 0$$

where z_i is the identifying variable (e.g., the instrument in 2SLS models) that identifies selection into treatment that is not in the equation of interest. In this case, z_i is a dummy variable.

$$\begin{bmatrix} v_i \\ \varepsilon_i \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{bmatrix} \right)$$

This model can easily be estimated via maximum likelihood by the mis-measured treatment effect model in STATA

```
treatreg y w, treat(x=w z)
```

I have generated a data set called problemset8c.dta that contains 45,000 observations for w and z. What I want you to do is write a program that does the following

1. Given ρ and σ, generate draws to v_i and ε_i. In all these problems, let σ=22
2. Given v_i, produce a value for x_i^* that equals $x_i^* = -1.6 + (0.18)z_i + (0.04)w_i + v_i$
3. Given x_i^* produce $x_i = 1$ if $x_i^* \geq 0$ and $x_i = 0$ otherwise
4. Given x_i ε_i and w_i produce and estimate for y_i that equals $y_i = -3 - 6x_i + w_i + \varepsilon_i$
5. Estimate the equation for y_i three ways
    a. OLS
    b. 2SLS
    c. The treatment effect model by MLE
6. For each model in 5, save the coefficient on x and the t-statistic (or z-score), plus for model 5c, save rho.
7. For each ρ generate 100 synthetic samples
8. Do steps 1-7 for values of ρ equal to 0.5, 0.25, 0, -0.25, -0.5

I want you to fill in the following table with the averages across the 100 samples for the 5 sets of simulations.

	OLS		2SLS		Treatment effect model		
Value of ρ	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							
0.25							
0.00							
-0.25							
-0.50							

Given the results in this table, what does this say about the ability of 2SLS to produce the appropriate estimate of $\hat{\beta}_1$. In what cases is it more appropriate to estimate MLE versus 2SLS models?

Here are some helpful programming hints. First, how do you draw random errors from a bivariate normal distribution? Remember that given a random (k x 1) vector β with a (k x k) V/C matrix of Σ, a set of linear restrictions (q x k) R such that r=Rβ will produce a (q x q) V/C matrix for r of RΣR′. Note that any positive definite matrix can be written in lower Cholesky form $\Sigma=PP'$. Consider a (2x2) V/C matrix that can be written as

$$\begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} = \sum_1 = PP' = \begin{bmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} \\ 0 & l_{22} \end{bmatrix} = \begin{bmatrix} l_{11}^2 & l_{11}l_{21} \\ l_{11}l_{21} & l_{21}^2 + l_{22}^2 \end{bmatrix}$$

Consider two random variables that are standard normal but independent

$$\xi = \begin{bmatrix} \xi_{1i} \\ \xi_{2i} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right)$$

By definition, $P\xi$ has a mean of $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and a V/C matrix $PI_2P' = PP' = \sum_1$. Note then that $P\xi$ can be written as

$$P\xi = \begin{bmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{bmatrix} \begin{bmatrix} \xi_{1i} \\ \xi_{2i} \end{bmatrix} = \begin{bmatrix} l_{11}\xi_{1i} \\ l_{21}\xi_{1i} + l_{22}\xi_{2i} \end{bmatrix}$$

So, draw two independent standard normal random variables ξ_{1i} and ξ_{2i}. Let $\nu_i = l_{11}\xi_{1i}$ and $\varepsilon_i = l_{21}\xi_{1i} + l_{22}\xi_{2i}$. All you have to do is given the V/C matrix for this case $\begin{bmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{bmatrix}$ figure out what $l_{11}, l_{21}, and\, l_{22}$ are.

Second, when you estimate the treatment effect model, you can output the beta and z-score for x and the rho for the model via the following text

```
treatreg y w, treat(x=w z)
local b2_te=_b[x];
local t2_te = _b[x] / _se[x]
local rho=e(rho);
```

Just a note, on my computer it took about 35 minutes to do 100 draws for ρ=0.25, so give yourself enough time to estimate these models.