Micro-econometrics: Notes on Week 8 Exercise

Julia Giese
Julia.Giese@Nuffield.ox.ac.uk

Notes on suggested solutions:[1]

1.1

$$y_i = \begin{cases} x_i\beta + u_i & \text{if} \quad x_i\beta + u_i > 0, \\ 0 & \text{if} \quad x_i\beta + u_i \leq 0. \end{cases} \tag{1}$$

We are asked to explain intuitively why OLS estimators will be inconsistent whether we use all the observations, or we use only those where $y_i > 0$.

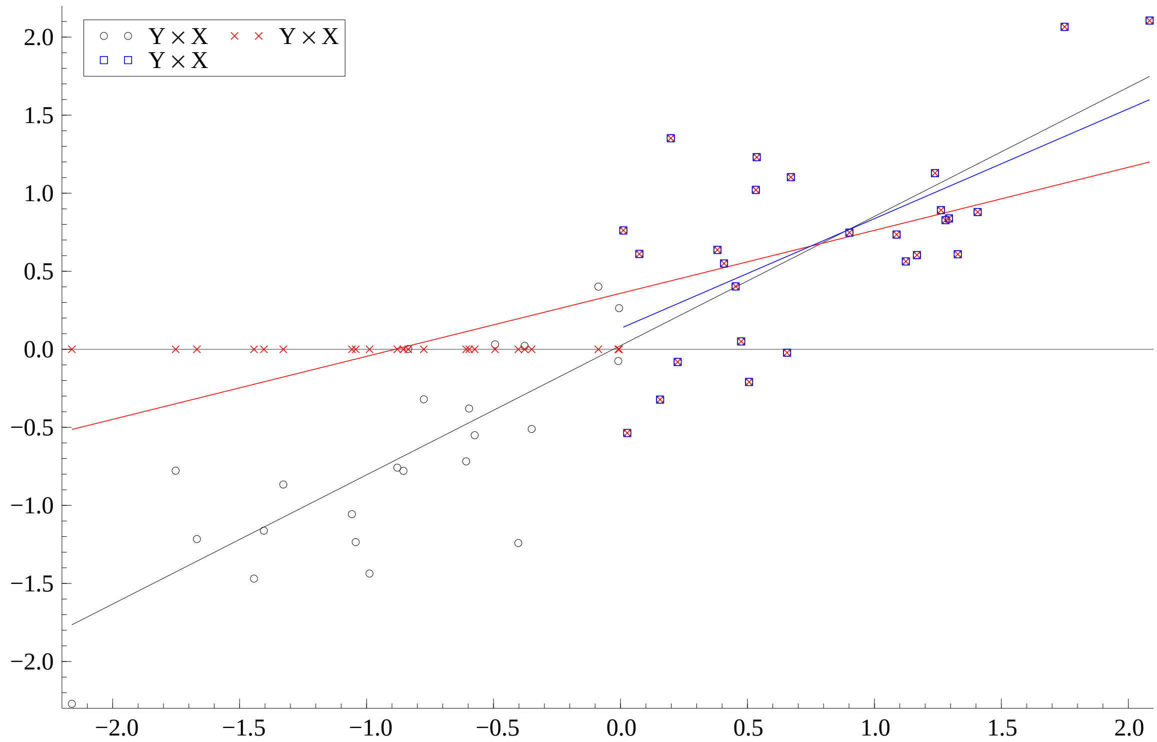

Figure 1: Plot of generated censored data, with OLS regression lines for truth (black), estimation with only $y_i > 0$ (blue), and estimation of all observations (red).

Firstly, we have a **censored** data problem, but it is useful to think about **truncated** data as well:

Censored data We do not observe values of the dependent (y_i) variable for a range of values of the independent (x_i) variables.

[1] Many thanks to James Reade for providing notes from last year.

Truncated data We do not observe values of both dependent (y_i) and independent (x_i) variables for a range of values.

Consider Figure 1. A dataset was generated according to $y_i = 0.9x_i + u_i$, where $u_i \sim \mathsf{IN}[0, 0.5]$. We see the problem of censoring. When x_i is below zero, we only observe $y_i = 0$, not the true value for y_i. In the graph, the true values of y_i are plotted to show the effect. Hence we actually see the red data points in our sample. So if we run an OLS regression on this observed data, we'll get a line which is too flat, biased. The black line is the true regression line. If we were to instead just run OLS on the datapoints where $y_i > 0$ (truncated case), we would again get a biased coefficient estimate, as much of the information is omitted in doing this.

1.2 Next, we are given the assumption that errors are normal, and asked how to consistently estimate β.

Firstly we could consider OLS, only to show it is not sufficient. We consider the moments of a censored normal distribution for this. OLS gives us the expected value of y_i given the data, hence it could be written:

$$\mathsf{E}(y\,|y>0) = \int_0^\infty y_i f(y\,|y>0)dy_i. \tag{2}$$

Now let us change variables here from u_i to ϵ_i, so that $y_i = x_i\beta + \sigma\epsilon_i$, where $\epsilon_i \sim \mathsf{IN}[0, 1]$. Then:

$$\begin{aligned}
\mathsf{E}(y\,|y>0) &= \int_{-x_i\beta}^{\infty} (x_i\beta + \sigma\epsilon_i) f(x_i\beta + \sigma\epsilon_i\,|x_i\beta + \sigma\epsilon_i > 0)\sigma d\epsilon_i & (3)\\
&= \int_{-x_i\beta}^{\infty} (x_i\beta + \sigma\epsilon_i) f(x_i\beta + \sigma\epsilon_i\,|\sigma\epsilon_i > -x_i\beta)\sigma d\epsilon_i & (4)\\
&= \int_{-x_i\beta}^{\infty} (x_i\beta + \sigma\epsilon_i) \frac{f(x_i\beta + \sigma\epsilon_i, \sigma\epsilon_i > -x_i\beta)}{\mathsf{P}(\sigma\epsilon_i > -x_i\beta)}\sigma d\epsilon_i. & (5)
\end{aligned}$$

Now given the range of values that the integral is evaluated over, then we can see that $f(x_i\beta + \sigma\epsilon_i, \sigma\epsilon_i > -x_i\beta) = f(x_i\beta + \sigma\epsilon_i)$ so:

$$\mathsf{E}(y\,|y>0) = \int_{-x_i\beta/\sigma}^{\infty} (x_i\beta + \sigma\epsilon_i) \frac{f(x_i\beta + \sigma\epsilon_i)}{\mathsf{P}(\sigma\epsilon_i > -x_i\beta)}\sigma d\epsilon_i. \tag{6}$$

Further, we can manipulate the distribution function for $f(y_i)$ a little:

$$\begin{aligned}
f(y_i) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-1}{2\sigma^2}(y_i - x_i\beta)^2\right\} & (7)\\
f(x_i\beta + \sigma\epsilon_i) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{\frac{-1}{2\sigma^2}(\sigma\epsilon_i)^2\right\} & (8)\\
&= \frac{1}{\sigma}\frac{1}{\sqrt{2\pi}} \exp\left\{\frac{-1}{2}(\epsilon_i)^2\right\} & (9)\\
&= \frac{1}{\sigma}\phi(\epsilon). & (10)
\end{aligned}$$

So:

$$\mathsf{E}(y\,|y>0) = \int_{-x_i\beta/\sigma}^{\infty} (x_i\beta + \sigma\epsilon_i) \frac{\frac{1}{\sigma}\phi(\epsilon)}{\mathsf{P}(\sigma\epsilon_i > -x_i\beta)}\sigma d\epsilon_i. \tag{11}$$

Let us also deal with the denominator:

$$\mathsf{P}\left(\sigma\epsilon_i > -x_i\beta\right) = \mathsf{P}\left(\epsilon_i > -\frac{x_i\beta}{\sigma}\right) = 1 - \Phi\left(-\frac{x_i\beta}{\sigma}\right) = \Phi\left(\frac{x_i\beta}{\sigma}\right). \tag{12}$$

Then:

$$\mathsf{E}\left(y\,|y>0\,\right) = \frac{1}{\Phi\left(\frac{x_i\beta}{\sigma}\right)}\int_{-x_i\beta/\sigma}^{\infty}\left(x_i\beta+\sigma\epsilon_i\right)\frac{1}{\sigma}\phi\left(\epsilon_i\right)\sigma d\epsilon_i. \tag{13}$$

$$= \frac{1}{\Phi\left(\frac{x_i\beta}{\sigma}\right)}\left[\int_{-x_i\beta/\sigma}^{\infty} x_i\beta\frac{1}{\sigma}\phi\left(\epsilon_i\right)\sigma d\epsilon_i + \int_{-x_i\beta/\sigma}^{\infty}\sigma\epsilon_i\frac{1}{\sigma}\phi\left(\epsilon_i\right)\sigma d\epsilon_i\right]. \tag{14}$$

Next tidy up expression and use the fact that $\phi'(\epsilon_i) = -\epsilon_i\phi(\epsilon_i)$, and evaluate:

$$\mathsf{E}\left(y\,|y>0\,\right) = \frac{1}{\Phi\left(\frac{x_i\beta}{\sigma}\right)}\left[x_i\beta\int_{-x_i\beta/\sigma}^{\infty}\phi\left(\epsilon_i\right)d\epsilon_i + \sigma\int_{-x_i\beta/\sigma}^{\infty}\epsilon_i\phi\left(\epsilon_i\right)d\epsilon_i\right] \tag{15}$$

$$= \frac{1}{\Phi\left(\frac{x_i\beta}{\sigma}\right)}\left[x_i\beta\int_{-x_i\beta/\sigma}^{\infty}\phi\left(\epsilon_i\right)d\epsilon_i - \sigma\int_{-x_i\beta/\sigma}^{\infty}\phi'\left(\epsilon_i\right)d\epsilon_i\right] \tag{16}$$

$$= \frac{1}{\Phi\left(\frac{x_i\beta}{\sigma}\right)}\left\{x_i\beta\left[\Phi(\epsilon_i)\right]_{-x_i\beta}^{\infty} - \sigma\left[\phi(\epsilon_i)\right]_{-x_i\beta}^{\infty}\right\} \tag{17}$$

$$= x_i\beta + \sigma\frac{\phi\left(\frac{x_i\beta}{\sigma}\right)}{\Phi\left(\frac{x_i\beta}{\sigma}\right)}. \tag{18}$$

This is the (inverse) Mills ratio, which tells us that were we to run OLS, we would get:

$$y_i = x_i\beta + \sigma\frac{\phi\left(\frac{x_i\beta}{\sigma}\right)}{\Phi\left(\frac{x_i\beta}{\sigma}\right)} + \nu_i, \tag{19}$$

where $\mathsf{E}\left(\nu\,|x_i, y_i > 0\,\right) = 0$. But the prevalence of x_i in the second term, the Mills ratio, means the regressors will be correlated, and that the estimator of β will be inconsistent. This is because of omitted variable bias.

However, **maximum likelihood** delivers consistent estimates. Firstly we recall that the likelihood function for any given sample is the joint density of each observation:

$$L(y_i;\theta) = f(y_1, y_2, \ldots, y_N) = \prod_{i=1}^{N} f(y_i), \tag{20}$$

where the last step is achieved through an independence assumption. Hence we can consider each observation, and we note there are two types of observation:

- ones that are censored, hence y_i not observed, $x_i\beta + u_i \leq 0$.
- ones that are not censored, hence y_i observed, and $x_i\beta + u_i > 0$.

So considering the first type, we can consider the density, or probability of this observation:

$$\mathsf{P}\left(y_i = 0\right) = \mathsf{P}\left(x_i\beta + u_i \leq 0\right) \tag{21}$$

$$= \mathsf{P}\left(u_i \leq -x_i\beta\right) \tag{22}$$

$$= \mathsf{P}\left(\frac{u_i}{\sigma} \leq -\frac{x_i\beta}{\sigma}\right) \tag{23}$$

$$= \Phi\left(-\frac{x_i\beta}{\sigma}\right) = 1 - \Phi\left(\frac{x_i\beta}{\sigma}\right), \tag{24}$$

since normality is assumed, and $u_i/\sigma \sim \mathsf{IN}\,[0,1]$.

Then considering the second type:

$$f(y_i, y_i > 0) = f(y_i)\mathsf{P}\,(y_i > 0|y_i)\,, \tag{25}$$

by Bayes' theorem. Now $\mathsf{P}\,(y_i > 0) = 1$ if $y_i > 0$ and 0 otherwise so

$$f(y_i, y_i > 0) = f(y_i) = f(x_i\beta + u_i),$$

and:

$$L(\theta; y_i) = \prod_{y_i>0} f(x_i\beta + u_i) \prod_{y_i=0} \left[1 - \Phi\left(\frac{x_i\beta}{\sigma}\right)\right]. \tag{26}$$

Given that $f(y_i) \sim \mathsf{IN}\,[x_i'\beta, \sigma_\epsilon^2]$ then substituting in $y_i = x_i'\beta + \epsilon_i$ gives $f(x_i'\beta + \epsilon_i) = (1/\sigma_\epsilon)\phi(\epsilon_i/\sigma_\epsilon)$, and so the log-likelihood function is:

$$L(\theta; y_i) = \prod_{y_i>0} \frac{1}{\sigma_\epsilon}\phi\left(\frac{\epsilon_i}{\sigma_\epsilon}\right) \prod_{y_i=0} \left[1 - \Phi\left(\frac{x_i\beta}{\sigma}\right)\right]. \tag{27}$$

1.3

$$y_i = \begin{cases} x_i\beta + u_i & \text{if} \quad 0 < x_i\beta + u_i < 40, \\ 0 & \text{if} \quad x_i\beta + u_i \leq 0, \\ 40 & \text{if} \quad x_i\beta + u_i \geq 40. \end{cases} \tag{28}$$

So now there are three contributions to the likelihood:

- y_i is observed, so we get $f(y_i, 0 < x_i\beta + u_i < 40)$, and a similar process to above will ensure all we need from here is $f(y_i)$.
- y_i censored below, so we get the same as in (24).
- y_i censored above, so we get:

$$\begin{aligned} \mathsf{P}\,(y_i = 40) &= \mathsf{P}\,(x_i\beta + u_i \geq 40) && (29)\\ &= \mathsf{P}\,(u_i \geq 40 - x_i\beta) && (30)\\ &= \mathsf{P}\left(\frac{u_i}{\sigma} \geq \frac{40 - x_i\beta}{\sigma}\right) && (31)\\ &= \Phi\left(\frac{40 - x_i\beta}{\sigma}\right). && (32) \end{aligned}$$

So the likelihood in this case becomes:

$$L(\theta; y_i) = \prod_{y_i=0} \left[1 - \Phi\left(\frac{x_i\beta}{\sigma}\right)\right] \prod_{0<y_i<40} f(x_i\beta + u_i) \prod_{y_i=40} \left[\Phi\left(\frac{40 - x_i\beta}{\sigma}\right)\right]. \tag{33}$$

The log-likelihood can be found from here simply by taking logs:

$$\log L(\theta; y_i) = \sum_{y_i=0} \log\left[1 - \Phi\left(\frac{x_i\beta}{\sigma}\right)\right] \sum_{0<y_i<40} \log f(x_i\beta + u_i) \sum_{y_i=40} \log\left[\Phi\left(\frac{40 - x_i\beta}{\sigma}\right)\right]. \tag{34}$$

The normalisation we require here is $\sigma = 1$, because with data x_i, we only observe the estimated coefficient $\widehat{\beta/\sigma}$, and hence we impose $\sigma = 1$.

2.1

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \sim \mathsf{IN}\left[\begin{pmatrix} x'\beta \\ z'\gamma \end{pmatrix}, \begin{pmatrix} \sigma_{y1}^2 & \rho\sigma_{y1}\sigma_{y2} \\ \rho\sigma_{y1}\sigma_{y2} & \sigma_{y2}^2 \end{pmatrix}\right] \tag{35}$$

Firstly we are asked to look at the conditional moments of the distribution. We have seen this before (in Maths crash course and in Jennie's problem sets), but it is useful to know proof. We have Bayes' theorem, which tells us that:

$$f(y_1, y_2) = f(y_1 \,|y_2\,)f(y_2). \tag{36}$$

So let us write the bivariate normal density, and try to split it up into the constituent parts. Then we can simply read off what we want.

$$\begin{aligned}\phi(y_1, y_2) = {} & \frac{1}{2\pi\sigma_{y1}\sigma_{y2}\sqrt{1-\rho^2}} \\ & \times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y_1 - x'\beta}{\sigma_{y1}}\right)^2 - 2\rho\left(\frac{y_2 - z'\gamma}{\sigma_{y2}}\right)\left(\frac{y_1 - x'\beta}{\sigma_{y1}}\right) + \left(\frac{y_2 - z'\gamma}{\sigma_{y2}}\right)^2\right]\right\}\end{aligned} \tag{37}$$

$$\begin{aligned} = {} & \frac{1}{2\pi\sigma_{y1}\sigma_{y2}\sqrt{1-\rho^2}} \\ & \times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left[\left(\frac{y_1 - x'\beta}{\sigma_{y1}}\right) - \rho\left(\frac{y_2 - z'\gamma}{\sigma_{y2}}\right)\right]^2 \right.\right. \\ & \left.\left. -\rho^2\left(\frac{y_2 - z'\gamma}{\sigma_{y2}}\right)^2 + \left(\frac{y_2 - z'\gamma}{\sigma_{y2}}\right)^2\right]\right\}\end{aligned} \tag{38}$$

$$\begin{aligned} = {} & \frac{1}{\sqrt{2\pi\sigma_{y1}^2(1-\rho^2)}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y_1 - x'\beta}{\sigma_{y1}}\right) - \rho\left(\frac{y_2 - z'\gamma}{\sigma_{y2}}\right)\right]^2\right\} \\ & \times \frac{1}{\sqrt{2\pi\sigma_{y2}^2}} \exp\left\{-\frac{1}{2}\left(\frac{y_2 - z'\gamma}{\sigma_{y2}}\right)^2\right\}\end{aligned} \tag{39}$$

$$\begin{aligned} = {} & \frac{1}{\sqrt{2\pi\sigma_{y1}^2(1-\rho^2)}} \exp\left\{-\frac{1}{2\sigma_{y1}^2(1-\rho^2)}\left[y_1 - x'\beta - \frac{\rho\sigma_{y1}}{\sigma_{y2}}\left(y_2 - z'\gamma\right)\right]^2\right\} \\ & \times \frac{1}{\sqrt{2\pi\sigma_{y2}^2}} \exp\left\{-\frac{1}{2}\left(\frac{y_2 - z'\gamma}{\sigma_{y2}}\right)^2\right\}\end{aligned} \tag{40}$$

$$= \phi(y_1 \,|y_2\,)\phi(y_2). \tag{41}$$

So reading off, we get that:

$$\mathsf{E}\left(y_1 \,|y_2\,\right) = x'\beta + \frac{\rho\sigma_{y1}}{\sigma_{y2}}\left(y_2 - z'\gamma\right). \tag{42}$$

The second result can now be stated in the simpler way since we've proved how we get conditional means. Using that $y_1 \sim \mathsf{IN}\left[x_i'\beta, \sigma_{y1}\right]$, then $y_1 = x_i'\beta + \epsilon_{y1}$, with $\epsilon_{y1} \sim \mathsf{IN}\left[0, \sigma_{y1}\right]$, and the equivalent for y_2, then:

$$\begin{aligned} \mathsf{E}\left(y_1 \left| y_2 > a\right.\right) &= \mathsf{E}\left(x'\beta + \epsilon_{y1} \left| z'\gamma + \epsilon_{y2} > a\right.\right) && (43)\\ &= x'\beta + \mathsf{E}\left(\epsilon_{y1} \left| z'\gamma + \epsilon_{y2} > a\right.\right), && (44)\end{aligned}$$

and here we use (42), the form of the conditional mean, to argue the form for ϵ_{y1} in terms of ϵ_{y2}, to get:[2]

$$\begin{aligned} \mathsf{E}\left(y_1 \left| y_2 > a\right.\right) &= x'\beta + \mathsf{E}\left(\left.\frac{\rho\sigma_{y1}}{\sigma_{y2}}\epsilon_{y2}\right| z'\gamma + \epsilon_{y2} > a\right) && (45)\\ &= x'\beta + \rho\sigma_{y1}\mathsf{E}\left(\left.\frac{\epsilon_{y2}}{\sigma_{y2}}\right| z'\gamma + \epsilon_{y2} > a\right) && (46)\\ &= x'\beta + \rho\sigma_{y1}\mathsf{E}\left(\left.\frac{\epsilon_{y2}}{\sigma_{y2}}\right| \frac{\epsilon_{y2}}{\sigma_{y2}} > \frac{a - z'\gamma}{\sigma_{y2}}\right) && (47)\\ &= x'\beta + \rho\sigma_{y1}\int_{\frac{a-z'\gamma}{\sigma_{y2}}}^{\infty} \frac{\epsilon_{y2}}{\sigma_{y2}}\phi\left(\left.\frac{\epsilon_{y2}}{\sigma_{y2}}\right| \frac{\epsilon_{y2}}{\sigma_{y2}} > \frac{a - z'\gamma}{\sigma_{y2}}\right) d\left(\frac{\epsilon_{y2}}{\sigma_{y2}}\right) && (48)\\ &= x'\beta + \rho\sigma_{y1}\int_{\frac{a-z'\gamma}{\sigma_{y2}}}^{\infty} \frac{\frac{\epsilon_{y2}}{\sigma_{y2}}\phi\left(\frac{\epsilon_{y2}}{\sigma_{y2}}, \frac{\epsilon_{y2}}{\sigma_{y2}} > \frac{a-z'\gamma}{\sigma_{y2}}\right)}{\mathsf{P}\left(\frac{\epsilon_{y2}}{\sigma_{y2}} > \frac{a-z'\gamma}{\sigma_{y2}}\right)} d\left(\frac{\epsilon_{y2}}{\sigma_{y2}}\right) && (49)\\ &= x'\beta + \rho\sigma_{y1}\int_{\frac{a-z'\gamma}{\sigma_{y2}}}^{\infty} \frac{\frac{\epsilon_{y2}}{\sigma_{y2}}\phi\left(\frac{\epsilon_{y2}}{\sigma_{y2}}\right)}{\mathsf{P}\left(\frac{\epsilon_{y2}}{\sigma_{y2}} > \frac{a-z'\gamma}{\sigma_{y2}}\right)} d\left(\frac{\epsilon_{y2}}{\sigma_{y2}}\right) && (50)\\ &= x'\beta + \frac{\rho\sigma_{y1}}{1 - \Phi\left(\frac{a-z'\gamma}{\sigma_{y2}}\right)}\int_{\frac{a-z'\gamma}{\sigma_{y2}}}^{\infty} \frac{\epsilon_{y2}}{\sigma_{y2}}\phi\left(\frac{\epsilon_{y2}}{\sigma_{y2}}\right) d\left(\frac{\epsilon_{y2}}{\sigma_{y2}}\right) && (51)\\ &= x'\beta + \frac{\rho\sigma_{y1}}{1 - \Phi\left(\frac{a-z'\gamma}{\sigma_{y2}}\right)}\left[-\int_{\frac{a-z'\gamma}{\sigma_{y2}}}^{\infty} \phi'\left(\frac{\epsilon_{y2}}{\sigma_{y2}}\right) d\left(\frac{\epsilon_{y2}}{\sigma_{y2}}\right)\right] && (52)\\ &= x'\beta + \rho\sigma_{y1}\frac{\phi\left(\frac{a-z'\gamma}{\sigma_{y2}}\right)}{1 - \Phi\left(\frac{a-z'\gamma}{\sigma_{y2}}\right)}. && (53)\end{aligned}$$

Line (49) makes use of Bayes' rule, and line (50) makes use of the fact the integral is only over the region $\frac{\epsilon_{y2}}{\sigma_{y2}} > \frac{a-z'\gamma}{\sigma_{y2}}$, hence the joint probability is of two identical events, and $\mathsf{P}(A, A) = \mathsf{P}(A)$. Then line (51) makes use of the fact that $\epsilon_{y2}/\sigma_{y2}$ is standard normally distributed and writes it in terms of the CDF of the standard normal.

2.2 We are asked to explain the sample selection problem inherent in the above model, and note what happens as $a \to -\infty$.

The sample selection problem describes the situation where the sample under consideration is *not* randomly selected from the population. The model given in the questions helps us out here:

$$y_{1i} = \begin{cases} x_i'\beta + \epsilon_i & \text{if} \quad z_i'\gamma + v_i \geq a, \\ a & \text{otherwise.} \end{cases} \qquad (54)$$

[2] So (42) suggests that $\epsilon_{y1} = \rho\frac{\sigma_{y1}}{\sigma_{y2}}\epsilon_{y2} + \xi$, where $\xi \sim \mathsf{IN}[0, \sigma_\xi]$.

Letting $w_i = z_i'\gamma + v_i$, in this model, the data are only observed, i.e. sampled, if $w_i > a$, some arbitrary value a. w_i depends on a set of individual characteristics, z_i also, which determine whether the individual is sampled or not. Selection is not random. It is potentially harmful too, if ϵ_i and v_i are correlated, and it is argued this is generally the case. Trying to work out what kind of car people like from observed sales of cars might be one such example, since the people in the sample are not random, but the ones who can actually afford the car.

As $a \to -\infty$ however, the problem is alleviated. In other words, less and less observations are censored. From (53) we see that as $a \to -\infty$, $\phi\left(\frac{a-z'\gamma}{\sigma_{y2}}\right) \to 0$ and $\Phi\left(\frac{a-z'\gamma}{\sigma_{y2}}\right) \to 0$, so the OLS estimator we discussed in Question 1 becomes consistent.

2.3 Derive the likelihood for this model, carefully explaining any required normalisations.

Firstly, it might help to write out what this model (54) means for our joint distribution. y_{2i} has been replaced by w_i, and ϵ_{y2} by v_i, so:

$$\begin{pmatrix} y_{1i} \\ w_i \end{pmatrix} \sim \mathsf{IN}\left[\begin{pmatrix} x'\beta \\ z'\gamma \end{pmatrix}, \begin{pmatrix} \sigma_{y1}^2 & \rho\sigma_{y1}\sigma_v \\ \rho\sigma_{y1}\sigma_v & \sigma_v^2 \end{pmatrix}\right] \tag{55}$$

As with last week's problem, we find that the scale of a variable is not observed. The variable this time is $w_i = z_i'\gamma + v_i$ — all we observe is whether or not it is above a. Hence we normalise this variance, $Var(v_i) = 1$. Two types of observations will contribute to the likelihood:

- y_{1i} observed so $w_i > a$. Then:

$$\mathsf{P}(y_{1i}, w_i > a) = \mathsf{P}(w_i > a\,|y_{1i})\, f(y_{1i}). \tag{56}$$

  So from before:

$$\begin{aligned} \mathsf{E}(v_i\,|\epsilon_i) &= \frac{\rho\sigma_v}{\sigma_{y1}}\epsilon_i && (57)\\ &= \frac{\rho\sigma_v}{\sigma_{y1}}(y_{1i} - x_i'\beta), && (58) \end{aligned}$$

  Now it might not be immediately clear, but here we are looking at the reverse conditional relationship (w_i on y_{1i}) to what we were looking at earlier (y_{1i} on w_i) so using (40) reversed we have that:

$$w_i|y_{1i} \sim \mathsf{IN}\left[z_i'\gamma + \frac{\rho\sigma_v}{\sigma_{y1}}(y_{1i} - x_i'\beta), (1-\rho^2)\sigma_v\right]. \tag{59}$$

  So this allows us to write:

$$w_i = z_i'\gamma + \frac{\rho\sigma_v}{\sigma_{y1}}(y_{1i} - x_i'\beta) + \xi_i, \tag{60}$$

  and from this regression set-up we see that ξ_i is independent of y_{1i} and x_i and

z_i, and also $\xi_i \sim \mathsf{IN}[0,(1-\rho^2)]$ Hence we can use this to write:

$$\mathsf{P}(w_i > a|y_i) = \mathsf{P}\left(z_i'\gamma + \frac{\rho\sigma_v}{\sigma_{y1}}(y_{1i} - x_i'\beta) + \xi_i > a\right) \tag{61}$$

$$= \mathsf{P}\left(\xi_i > a - \left(z_i'\gamma + \frac{\rho\sigma_v}{\sigma_{y1}}(y_{1i} - x_i'\beta)\right)\right) \tag{62}$$

$$= \mathsf{P}\left(\frac{\xi_i}{(1-\rho^2)^{1/2}} > \frac{a - \left(z_i'\gamma + \frac{\rho\sigma_v}{\sigma_{y1}}(y_{1i} - x_i'\beta)\right)}{(1-\rho^2)^{1/2}}\right) \tag{63}$$

$$= 1 - \Phi\left(\frac{a - \left(z_i'\gamma + \frac{\rho\sigma_v}{\sigma_{y1}}(y_{1i} - x_i'\beta)\right)}{(1-\rho^2)^{1/2}}\right), \tag{64}$$

where the y_{1i} disappears because it's independent of ξ_i. Pulling this together:

$$\mathsf{P}(y_{1i}, w_i > a) = \mathsf{P}(w_i > a\,|y_{1i})\, f(y_{1i}) \tag{65}$$

$$= \left[1 - \Phi\left(\frac{a - \left(z_i'\gamma + \frac{\rho\sigma_v}{\sigma_{y1}}(y_{1i} - x_i'\beta)\right)}{(1-\rho^2)^{1/2}}\right)\right] f(y_{1i}) \tag{66}$$

$$= \left[1 - \Phi\left(\frac{a - \left(z_i'\gamma + \frac{\rho\sigma_v}{\sigma_{y1}}(y_{1i} - x_i'\beta)\right)}{(1-\rho^2)^{1/2}}\right)\right] \frac{1}{\sigma_{y1}}\phi\left(\frac{\epsilon_i}{\sigma_{y1}}\right). \tag{67}$$

- y_{1i} unobserved so $w_i < a$. Then:

$$\mathsf{P}(w_i < a) = \mathsf{P}(z_i'\gamma + v_i < a) = \Phi(a - z_i'\gamma) \tag{68}$$

These can then be combined in our likelihood.

$$L(\theta; y_{1i}) = \prod_{i=1}^{N} f(y_{1i}) = \prod_{y_{1i}^* \text{ observed}}^{N} \mathsf{P}(y_{1i}, w_i > a) \prod_{y_{1i}^* \text{ unobserved}}^{N} \mathsf{P}(w_i < a) \tag{69}$$

$$= \prod_{y_{1i}^* \text{ observed}}^{N} \left[1 - \Phi\left(\frac{a - \left(z_i'\gamma + \frac{\rho\sigma_v}{\sigma_{y1}}(y_{1i} - x_i'\beta)\right)}{(1-\rho^2)^{1/2}}\right)\right] \frac{1}{\sigma_{y1}}\phi\left(\frac{\epsilon_i}{\sigma_{y1}}\right) \tag{70}$$

$$\times \prod_{y_{1i}^* \text{ unobserved}}^{N} \Phi(a - z_i'\gamma) \tag{71}$$

Normalisations required here are $\sigma_v = 1$ because the scale of w_i is never observed, only whether it is positive or not.

A further remark:
If you are planning to do empirical time series modelling either in your DPhil or in your job, I would recommend a summer school at the University of Copenhagen, entitled: ***"Econometric Methodology and Macroeconomics Applications - the Cointegrated VAR Model"***. It is run by Søren Johansen, Katarina Juselius and Anders Rahbek and takes place from 4-24 August 2008. Besides technical lectures, you get the chance to work on your data set with the experts for three weeks – and have a paper written by the end of it!
For further information talk to me or visit `www.econ.ku.dk/Summerschool` .