Problem Set #12

Danny Edgel
Econ 710: Economic Statistics and Econometrics II
Spring 2021

April 22, 2021

Discussed and/or compared answers with Sarah Bass, Emily Case, Katherine Kwok, Michael Nattinger, and Alex Von Hafften

Exercise 27.1

The latent variable model is:

$$Y^* = X'\beta + e$$

We can derive the conditional means of the censored ($m(X)$) and truncated ($m^{\#}(X)$) variables by assuming that $e|X \sim \mathcal{N}(0, \sigma^2)$. Then, for the censored conditional mean,

$$m(X) = \mathbb{E}\left[Y|X\right] = \mathbb{E}\left[Y^* \mathbb{1}\left\{Y^* > 0\right\}|X\right] = X'\beta\Phi\left(\frac{X'\beta}{\sigma}\right) + \sigma\phi\left(\frac{X'\beta}{\sigma}\right)$$

And for the truncated conditional mean,

$$m^{\#}(X) = \mathbb{E}\left[Y|X\right] = \mathbb{E}\left[Y^*|X, Y^* > 0\right] = X'\beta + \sigma\lambda\left(\frac{X'\beta}{\sigma}\right)$$

Exercise 27.2

No, the OLS estimate of β is biased downward in this model.

Exercise 27.4

An NLLS estimator for the conditional mean of the model in (27.2) is:

$$\min_{\beta,\sigma} \left(Y - X'\beta\Phi\left(\frac{X'\beta}{\sigma}\right) - \sigma\phi\left(\frac{X'\beta}{\sigma}\right)\right)^2$$

Exercise 27.8

The latent variable model for (27.7) is:

$$\begin{aligned} Y^* &= X'\beta + e \\ S^* &= Z'\gamma + uS \qquad\qquad\qquad\qquad = \mathbb{1}\{S^* > 0\} \\ Y &= \begin{cases} Y^*, & S = 1 \\ \text{missing}, & S = 0 \end{cases} \end{aligned}$$

Assume:

$$\begin{pmatrix} e \\ u \end{pmatrix} \sim \mathcal{N}\left(0, \begin{pmatrix} \sigma^2 & \sigma_{21} \\ \sigma 21 & 1 \end{pmatrix}\right)$$

Then,

$$\begin{aligned} \mathbb{E}[Y|X,Z,S=1] &= \mathbb{E}[Y^*|X,Z,S=1] = \mathbb{E}[X'\beta + e|X,Z,S=1] \\ &= X'\beta + \mathbb{E}[e|X,Z,S=1] = X'\beta + \mathbb{E}[e|X\gamma + u > 0] \\ \mathbb{E}[e|X\gamma + u > 0] &= \mathbb{E}[e|u > -X\gamma] = Cov(e,u)\frac{\phi(Z'\gamma)}{\Phi(Z'\gamma)} = \sigma_{21}\lambda(Z'\gamma) \\ \mathbb{E}[Y|X,Z,S=1] &= X'\beta + \sigma_{21}\lambda(Z'\gamma) \end{aligned}$$

Exercise 27.9

The results of the OLS regression (using robust standard errors) are displayed in the table at the end of this exercise. Absent of misspecification, these results would suggest that the replationship between transfers and income is linear and perfectly symmetric about \$1,000. However, the dependent variable is censored at zero, so this model is misspecified. The share of observations that are censored is 11.98% . I would expect censoring bias to be a problem in this example.

VARIABLES	(1) (a)	(2) (c)	(3) (d)	(4) (d)	(5) (e)
Income ('000)	-42.67***	-42.82***	-42.67***		-47.40***
	(3.827)	(3.803)	(2.632)		(0.378)
(Income-1)$\mathbb{1}$ {Income > 1}	42.67***	42.86***	42.67***		47.39***
	(3.827)	(3.804)	(2.633)		(0.378)
var(e.tinkind)				384.4***	
				(5.834)	
Constant	49.72***	50.53***	49.72***		49.00***
	(3.810)	(3.774)	(2.609)		(0.372)
Observations	8,684	6,734	8,684	8,684	8,140
R-squared	0.030	0.033			

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Exercise 28.12

Table 28.1 is reproduced below for the subsample of Hispanic women, reporting only AIC and BIC.[1] The results show wide variation in the returns to education from model choice. Using BIC, I would select model 2. Using AIC, I would select model 8. If this were my analysis (and I for some reason wasn't concerned about endogeneity), I would certainly use a spline model, as those get both low AIC and BIC values, and I would use model 5 as my primary model, as it is extremely close to both the best AIC and BIC estimates.

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Return	25%	35%	35%	37%	47%	47%	38%	47%	47%
s.e.	4	4	4	5	5	5	7	7	7
BIC	4814	**4435**	4468	4817	4437	4470	4825	4445	4478
AIC	4766	4381	4390	4757	4371	4380	4759	**4373**	4382
Education	College	Spline	Dummy	College	Spline	Dummy	College	Spline	Dummy
Experience	2	2	2	4	4	4	6	6	6

[1] The AIC and BIC below are substantially higher than those from Table 28.1. To ensure that my methodology was correct, I first coded the whole result using a subsample of Asian women. I got all of the same results as in Table 28.1 *except* for the AIC and BIC. I cannot explain this difference but am nonetheless confident in my results.