PROBLEM SET 5

Problem 1 (The type I Tobit model with neglected heterogeneity)

Consider the following model of censoring

$$Y^* = X\beta + W\gamma + \epsilon,$$
$$Y = \max\left(Y^*, 0\right),$$

where $(X, W) \perp \epsilon$, $X \perp W$, and $\epsilon \sim \mathcal{N}(0, 1)$, where $\perp$ denotes stochastic independence.

(a) Suppose a random sample of observations of (Y, X, W) is observed. Can the parameters β, γ be consistently estimated, and if so, how?

(b) Suppose now that W is not observed, but rather only a random sample of (Y, X) is observed. Assume that $W \sim \mathcal{N}\left(0, \sigma_w^2\right)$. This is now a model of censoring with unobserved heterogeneity (W is the unobserved heterogeneity). Can β be consistently estimated in this model? Explain.

Problem 2 (A variation of the Tobit model)

Consider the model:

$$Y = \begin{cases} Y^* & \text{if } 0 < Y^* \leq 1 \\ 0 & \text{if } Y^* \leq 0 \\ 1 & \text{if } Y^* > 1 \end{cases},$$
$$Y^* = X\beta + U,$$

where U and X are independent and $U \sim \mathcal{N}\left(0, \sigma^2\right)$.

(a) Write the log-likelihood of an iid sample $\{(y_i, x_i) :, i = 1, ..., N\}$.

(b) What are the asymptotic properties of the Maximum Likelihood estimator?

(c) How could you perform a specification test of this model?

Problem 3 (Applied Tobit)

In the following we estimate alcohol consumption. Use the 'tobacco.csv' data set to solve the following problems:

1. Estimate a linear model of alcohol consumption:

$$\begin{aligned} alcohol_i &= \beta_0 + \beta_1 age + \beta_2 nadults + \beta_3 nkids + \beta_4 nkids2 \\ &\quad + \beta_5 \ln x + \beta_6 \ln x * age + \beta_7 \ln x * nadults + u_i \end{aligned}$$

2. In the estimation above are we dealing with problems due to data censoring or a corner solution model?
3. Estimate the model from question 1 using tobit.
4. Calculate meaningfull marginal effects, that is partial effects on both $E(Y|X)$ and $E(Y|X, Y > 0)$ and compare your results from the tobit estimation to the results from the OLS estimation.
5. Interpret the marginal effects. In particular, what is the effect of log(total expenditure) on alcohol consumption?
6. Extend the tobit model including the square of log(total expenditure) as well as age squared.
7. Test whether the quadratic terms are relevant using both a Wald test and a likelihood ratio test.
8. Estimate a probit model on the probability of consuming any alcohol. Compare the parameter estimates to the tobit model.
9. *For the more adventurous:* Perform a LM test for heteroskedasticity in the tobit model related to age (age) and the number of adults ($nadults$).

Problem 4 (RECAP + Probit Selection Model) [Is not being presented in tutorial]

(26 marks total) Consider the model

$$
\begin{aligned}
Y_1^* &= X_1\beta + \delta Y_2 + U_1, \\
Y_2^* &= X_1\gamma_1 + X_2\gamma_2 + U_2,
\end{aligned}
$$

where furthermore (Y_1, Y_2) are outcome variables that are determined by the realizations of (Y_1^*, Y_2^*). Below parts a-c of this question examine how estimation of model parameters can be performed under various rules for the determination of these variables. Estimation is to be based on a random sample of observations of (Y_1, Y_2, X_1, X_2), denoted $\{(y_{1i}, y_{2i}, x_{1i}, x_{2i}) : i = 1, ..., n\}$ throughout. X is used as shorthand for the exogenous variables (X_1, X_2), $\gamma \equiv (\gamma_1', \gamma_2')'$ denotes the parameters of the equation for Y_2^*, and $\theta \equiv (\beta', \delta')'$ denotes the parameters of the equation for Y_1^*. For every estimator you propose, be sure to provide any additional assumptions you require to ensure your estimator is consistent for full credit.

(a) Suppose that $Y_1 = Y_1^*$ and $Y_2 = Y_2^*$.

  i. Write down the OLS estimator for γ and explain what conditions are required for it to be consistent.

  ii. Suppose $E[U_1|X] = 0$. Provide a consistent estimator for θ and explain why it is consistent.

(b) Assume now and for part c as well that (U_1, U_2) are bivariate normally distributed with mean zero and variance matrix Ω.

For parts b(i) and b(ii) only suppose that $Y_1 = Y_1^*$ and $Y_2 = 1[Y_2^* > 0]$.

  i. Provide a consistent estimator for γ.

  ii. Provide a consistent estimator for θ and explain why it is consistent.

(c) For this part assume that $Y_1 = 1[Y_1^* > 0]$ and $Y_2 = Y_2^*$.

  i. Provide a consistent estimator for γ.

  ii. Describe an estimation method that will produce a consistent estimator for $\theta \equiv (\beta', \delta')'$ and explain why it is consistent.