Nova School of Business and Economics
Microeconometrics
Final Exam
May 15th, 2023

PLEASE ANSWER FOUR OF THE FOLLOWING FIVE QUESTIONS

Question 1:

1. Consider a data set with sample means $\bar{y} = 0.8$ and $\bar{x} = 0.3$. Logit regression of y on the scalar x leads to an estimated intercept equal to 0.4 (with standard error of 0.1) and slope equal to 0.2 (with standard error of 0.05). Provide a numerical interpretation of the marginal impact of a change in x.

2. Derive the probit model based on observing $y_i = 1$ if $y_i^* = \mathbf{x}_i'\beta + u_i > 0$, where u_i is normally distributed.

3. Consider estimation of a three-outcome model with $\Pr[y_i = j] = F_j[\mathbf{x}_i'\beta_j]$, j=1, 2, 3. Give the log-likelihood function for this model.

4. Derive the probabilities for a three-outcome ordered probit model.

5. State the probabilities for a three-outcome multinomial logit model.

6. Present the additive random utility model in the case of three alternative model.

Question 2:

Suppose that y given $\boldsymbol{x}$ follows a two-limit Tobit model, with limit points $a_1 < a_2$:
$y^* = \boldsymbol{X\beta} + u \quad u|\boldsymbol{x} \sim Normal(0, \sigma^2)$
$y = a_1 \quad \text{if } y^* \leq a_1$
$y = y^* \quad \text{if } a_1 < y^* < a_2$
$y = a_2 \quad \text{if } y* \geq a_2$

1. Find $P(y = a_1)$ and $P(y = a_2)$ in terms of the standard normal cumulative density function, $\boldsymbol{x}$, $\boldsymbol{\beta}$, and σ.

2. Find the density of y given $\boldsymbol{x}$ for $a_1 < y < a_2$.

3.) If $z \sim Normal(0,1)$, it can be shown that $E(z|c_1 < z < c_2) = \{\phi(c_1) - \phi(c_2)\}/\{\Phi(c_2) - \Phi(c_1)\}$. Use this fact to find $E(y|\boldsymbol{x}, a_1 < y < a_2)$ and $E(y|\boldsymbol{x})$).

4. Write the log-likelihood function for observation i.

5. How would you estimate $E(y|\boldsymbol{x}, a_1 < y < a_2)$ and $E(y|\boldsymbol{x})$)?

6. Derive the marginal effect of x_j for $E(y|\boldsymbol{x})$).

7. Suppose that you obtain $\hat{\gamma}$ from a standard OLS regression of y_i on $\boldsymbol{x_i}$, using all observations. Would you compare $\hat{\gamma}$ to the two-limit estimate $\hat{\beta}_j$? What would be a sensible comparison?

Question 3:

Consider a Poisson regression model where the nonnegative integer-valued random variable y^* has density

$$f^*(y^*) = \frac{e^{-\mu}\mu^y}{y^*!}, y^* = 0, 1, 2, ...$$

Suppose that μ_i depends on regressors x according to $\mu_i = E[y_i|\mathbf{x}_i] = \exp(\mathbf{x}_i'\beta)$. Data are independent over i.

Suppose we only observe $y = y^*$ when $y^* \geq 1$ (meaning that when $y^* = 0$ we do not observe anything).

1. Obtain the density function $f(y|\mathbf{x})$ of the observed y.

2. Derive the log-likelihood function for β. Hence obtain the first-order conditions for the MLE of β.

3. Now suppose we observe only

$$y_i = 2 \text{ if } y_i^* \geq 2$$
$$y_i = 1 \text{ if } y_i^* = 1$$
$$y_i = 0 \text{ if } y_i^* = 0$$

Derive the log-likelihood function for β.

4. Explain the notion of over- and under-dispersion. How would you test the presence of over-dispersion? What options would you consider to overcome the problem of over-dispersion?

Question 4: Consider an exponential duration distribution with hazard rate λ.

1. Derive the cumulative hazard function.
2. Obtain the expressions for the corresponding Survival and Probability Density functions.
3. Develop an appropriate expression for the joint likelihood of N independent observations when durations arrive from a flow sampling plan and where all durations all are complete.
4. How can you show that mean duration is equal to $1/\lambda$?
5. Now obtain the expression of the likelihood function for the case of a sampling plan corresponding to a stock sample (all durations are incomplete and there is an oversampling of long durations).
6. What do you conclude from the comparison of your answers to 3 and 5?

Question 5:

Consider now the estimation results provided in table5 (in the next page) in the study on the sources of the union wage gap by Addison, Portugal, and Vilares (Journal of Econometrics, 2022).

1. Explain in simple words the meaning of the quantile regression parameters in the first line of the table, where the regressor is the fraction of unionized workers at the workplace (union membership).
2. What is the connection between the maximum score estimator, the LAD estimator and the median regression.

3. Explain the role of the check function in the estimation process of the quantile regression.

4. What is the main computational problem estimating linear regression models with high-dimensional fixed effects?