Answer Keys to Problem Set 3

Microeconometrics I with Joan Llull
IDEA, Fall 2024

TA: Conghan Zheng

October 16, 2024

**Note: 1. This document is not intended to be a complete solution, but rather to provide some examples of interpretations and key points for the answer. 2. For the computational part, please refer to the code scripts.*

Tobit

- Consider the regression with log wages as the dependent variable.
- The coefficient estimate for Region 2 is negative, indicating that wages in Region 2 are lower than in the base region (Region 1).
- However, if we look at the positive coefficient estimate for its interaction with the time trend, we can say that average wages in Region 2 have declined more slowly over time than those in Region 1 (due to the negative general time trend).
- Note that the interaction `region` $2\times$`time` also has a "base category": `region` $1\times$`time`.

Correction for Convexity

- When calculating predicted wages from Tobit, we must take into account the convexity of the exponential function.
- $\ln(y) = x\beta + \varepsilon, \varepsilon \sim N(0, \sigma^2)$ implies that $\ln(y) \sim N(x\beta, \sigma^2)$. Therefore, y is log-normally distributed. A log-normal is left-skewed, its mean is not simply e^{μ} (where μ is $x\beta$ from the normal of $\ln(y)$), but $e^{\mu+\frac{1}{2}\sigma^2}$, where $e^{-\frac{1}{2}\sigma^2}$ is called the *correction for convexity*.
- The intuition for this correction in its moments is: the exponential function is a strictly convex function, after the convex transformation the positive deviations from the mean are enlarged and the negative deviations from the mean are reduced[1].
- Consequently, there is a positive shift in the mean of the transformed random variable. Unlike the normal distribution, where the mean doesn't affect the shape of the distribution. For more on this correction in the truncated means predicted by various models, see the table in the TA slides.

[1] Consider the Jensen inequality $\mathbb{E}(f(X)) \geq f(\mathbb{E}(X))$

Marginal Effects

- Marginal effects of `time` on the left-censored mean:

$$\frac{\partial \mathbb{E}(y|x)}{\partial \texttt{time}} = \Phi\left(\frac{x\beta}{\sigma}\right) \cdot \beta_{\texttt{time}}$$

	Delta-method dy/dx	std. err.	z	P>|z|	[95% conf.	interval]
time	-.0175196	.000627	-27.94	0.000	-.0187485	-.0162907

- The estimate for the at-average marginal effect of `time` on the left-censored mean of log wages is negative, which implies that (regardless of whether wages are positive) the real wage of an individual with average characteristics is decreasing over time. Nominal wages do not keep up with inflation.

Two-Part Model

- The log likelihood of the Tobit model is smaller than the log likelihood of the two-part model, which is the sum of the log likelihoods of the two parts (Probit and OLS). The two-part model fits the data significantly better, even when AIC or BIC is used to panelize the two-part model for its additional parameters.
- Comparing the estimates from the Tobit model with those from part II of the two-part model allows us to identify the channel through which the regressors affect wages: participation, outcome, or both.

Heckman Two-Step Estimation (LIML)

- Heckman Two-Step Approach:

$$\text{Step 1}: \quad \mathbb{P}(w_i^* > \underline{w}_i) = \Phi\left(x_{1i}'\beta_1\right), \;\; \varepsilon_1 \sim \mathcal{N}(0,1) \quad \rightarrow \hat{\beta}_1$$

Theoretically, only $\frac{\beta_1}{\sigma_1}$ is identified, so we choose to normalize σ_1 to 1.

$$\text{Step 2}: \quad w_i = x_{2i}'\beta_2 + \delta_{12}\lambda(x_{1i}'\hat{\beta}_1) + \nu_i, \;\; \varepsilon_2 = \delta_{12}\varepsilon_1 + \eta \quad \rightarrow \hat{\beta}_2, \hat{\delta}_{12}$$

- We include exclusion restrictions by making $x_1 \neq x_2$:

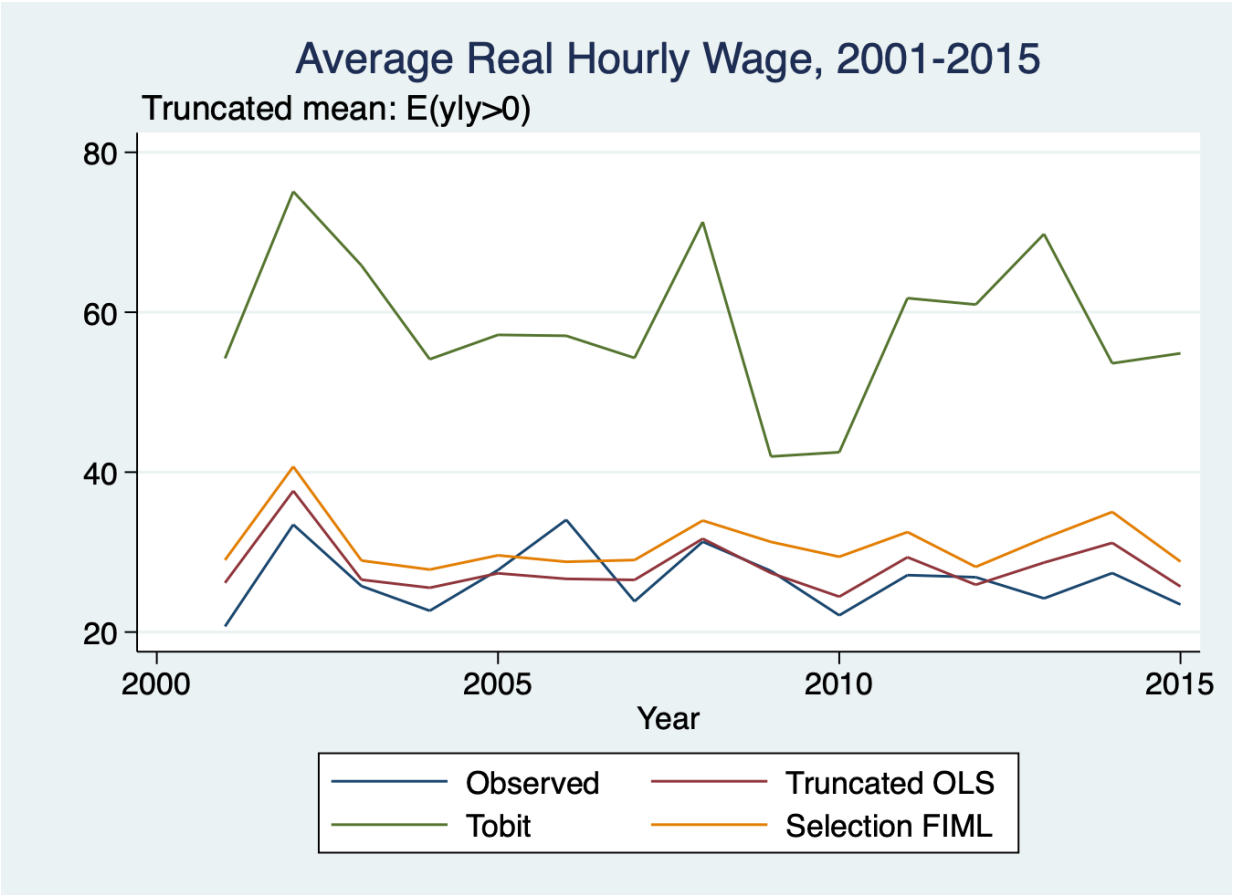
$$x_1 = \{x_2, \texttt{educ\_sp}, \texttt{benefits}\}$$
$$x_2 = \text{all regressors required in Exercise 1.1}$$

- Under the null that selection is exogenous, we expect a zero δ_{12}, which can be tested by the t-statistic of its estimate. A significantly negative parameter estimate of the Mill's ratio in our regression results provides strong evidence of selection.

LIML vs. FIML

- LIML assumes a univariate normal ε_1 and a linear dependence of ε_2 on ε_1.
- FIML assumes a bivariate normal on ε_1 and ε_2. FIML is more restrictive, more efficient, more accurate, but less robust than LIML.
- FIML can be computationally demanding in some applications. Sometimes FIML is not even feasible due to the complexity of the model (not globally concave).
- From our regression results, the FIML has higher precision and lower predicted wage compared to the LIML.

Compare Truncated Means



Truncated means

- The gap between the Tobit truncated means and the other three curves shows the role of selection in driving the Tobit truncated means higher. And the selection is higher in some years (for example, the gap is larger around 2008).

Indentification Considerations

Unemployment Benefits

1. Collinearity
   - Unemployment benefits are one of the most important determinants of an individual's willingness to participate in the labor market, as they provide an alternative source of income.
   - In the United States, they are calculated on the basis of year, marital status, number of children, spouse's employment status, and earnings.
   - It is important to note that most of these determinants are already included in other regressors in x_2 (and hence x_1), which indicates a strong collinearity between the explanatory variables in the estimation of the participation equation.
   - This collinearity is within the participation equation, it's true even if there is enough non-linearity provided by the Mill's ratio, or enough exogenous variation provided by the other exclusion restriction variable `educ_sp`.
2. Unjustified Exclusion Restrictions
   - We are not sure that benefits are uncorrelated with final wages (benefits could affect wages through past wages), which might suggest that benefits should also be included in the outcome equation.
   - And in the US case, benefits may be low and not variable.