Advanced Econometrics

University of Warsaw

Jerzy Mycielski

Doctoral School of Social Sciences, 2024

Limited dependent and qualitative variables

- It is often the case (especially in the case of micro level data) that the set of possible values of the dependent variable domain is not equal to $\mathbb{R}$ (set of real numbers)
- Such dependent variable cannot be represented by standard linear model:

$$Y = X\beta + \varepsilon, \ \varepsilon \sim IID$$

  without imposing restrictions on the sets of possible values of X, and β and ε.
- The examples of such a case are:
  - binary response variable (e.g. employment status: working, not working)
  - qualitative, ordered response variables (e.g. education: primary, secondary, higher)
  - qualitative unordered response variable (e.g. choice of transportation mode: car, bus, tram)
  - limited depended response variable (e.g. spendings for food $\geq$ 0)

Latent variable models

- One common solution to define the correct model for limited or qualitative dependent variable
- Latent variable cannot be observed directly but influences the variable which is observable.
- Usually we assume that latent variable id generated from linear model:

$$Y^* = X\beta + \varepsilon, \; \varepsilon \sim IID$$

- Relationship between latent variable depends on the model.

Qualitative depend variable

- Binary choice data

$$Y = \begin{cases} 1 & \text{if } Y^* > 0 \\ 0 & \text{if } Y^* \leq 0 \end{cases}$$

depending on the choice of distribution of ε: Linear Probability Model, logit, probit

- Qualitative, ordered response data

$$Y = \begin{cases} 0 & \text{if } Y^* < a_1 \\ 1 & \text{if } a_1 < Y^* \leq a_2 \\ \vdots & \\ J & \text{if } a_{J-1} < Y^* \leq a_J \end{cases}$$

where $a_1, \ldots, a_J$ are additional parameters to be estimated

- Depending on the choice of distribution of ε: ordered probit, ordered logit (Proportonal Odds Model - POM)

Example: binary regression

Table 25.1: Binary Choice Regressions for Marriage

	Logit		Probit	
	Coefficient	AME	Coefficient	AME
age	0.217	0.044	0.132	0.045
	(0.006)	(0.001)	(0.003)	(0.001)
education	0.014	0.003	0.009	0.003
	(0.010)	(0.002)	(0.006)	(0.002)
Black	−0.767	−0.156	−0.454	−0.153
	(0.092)	(0.018)	(0.054)	(0.018)
Asian	0.033	0.007	0.025	0.008
	(0.103)	(0.021)	(0.063)	(0.021)
Hispanic	−0.084	−0.017	−0.048	−0.017
	(0.063)	(0.013)	(0.038)	(0.013)
MidWest	0.272	0.056	0.165	0.056
	(0.074)	(0.011)	(0.045)	(0.015)
South	0.338	0.069	0.203	0.069
	(0.070)	(0.014)	(0.043)	(0.014)
West	0.383	0.078	0.228	0.077
	(0.072)	(0.015)	(0.044)	(0.015)
Intercept	−6.45		−3.93	
	(0.21)		(0.12)	

Source: Hansen (2022)

Identification and normalization of dispersion parameter

- Assume that in original model $Var(\varepsilon) = \sigma^2$:

$$Y = \begin{cases} Y = 0 \text{ for } X'\beta + \varepsilon < 0 \\ Y = 1 \text{ dla } X'\beta + \varepsilon \geq 0 \end{cases}$$

- Normalize this equation by dividing β and ε by σ:

$$y_i = \begin{cases} Y = 0 \text{ dla } X'\frac{\beta}{\sigma} + \frac{\varepsilon_i}{\sigma} < 0 \\ Y = 1 \text{ dla } X'\frac{\beta}{\sigma} + \frac{\varepsilon_i}{\sigma} \geq 0 \end{cases}$$

- Notice that these two models are observationally equivalent! Parameter σ is not identified.
- This model can only be estimated for normalized $\sigma = 1$:

$$y_i = \begin{cases} y_i = 0 \text{ dla } x_i\beta^* + \varepsilon_i^* < 0 \\ y_i = 1 \text{ dla } x_i\beta^* + \varepsilon_i^* \geq 0 \end{cases}$$

Qualitative depend variable

- Qualitative unordered response variable

$$Y_i^* = X_i\beta_i + \varepsilon_i,\ \varepsilon_i \sim IID$$

$$Y = \max_i \{Y_1^*, \ldots, Y_J^*\}$$

special cases:

  - $\beta_i = \beta$ conditional logit, data consist the characteristics of all the choices, we observe which choice was made.
    - This model is closely realated to theory of consumer/firm choice. Y^* can be intepreted as utility/profit.
    - If price is included as one of the attributes we can estimate Willigness To Pay for this attribute as

$$WTP = -\frac{\beta_k}{\beta_P}$$

    - This property of conditional logit is often used experimental economics in Discrete Choice Experiments (DCE)
  - $X_i = X$ multinominal logit, data consist only the attributes of agents making decisions.
- ε_i has extreme value type 1 distribution

Independence on Irrelevant Alternatives

- The assumption that ε_i are independent across alternatives important - together with the choice of distribution it implies that relative probabilities of two choices (Odds) only dependent on differences between characteristics of these choices and not other choices (Independence on Irrelevant Alternatives - IIA)
- For similar choice IIA can result in counterintuitive predictions of the model ("red bus/blue bus puzzle")
- This limitation of the conditional logit model resulted in many modifications of this model
  - mixed logit: β parameter is random and varies across individuals
  - hierarchical logit: ε_i are not independent and given by GEV (Generalized Extreme Value) distribution
  - conditional probit: ε_i is normaly distributed and correlated across alternatives

- Limited depended response variable

$$Y = \begin{cases} 0 & \text{if } Y^* \leq a \\ Y^* & \text{if } Y^* > a \end{cases}$$

- This model is known as tobit model.
- Usually we assume that ε has normal distribution.
- Interval regression model

$$Y = \begin{cases} 0 & \text{if } Y^* < a_1 \\ 1 & \text{if } a_1 < Y^* \leq a_2 \\ \vdots & \\ J & \text{if } a_{J-1} < Y^* \leq a_J \end{cases}$$

  and a_j are known.
- It is also possible to define a model which is a mixture of the tobit and interval regression models

Count data models

- Sometimes Y has values which are natural numbers (counts).
- Such variable is quantitative however it is not continous and therefore cannot be represented with linear model with continously distributed error term.
- Assume Y has descrete random distribution with location parameter being a function of explanatory variables.
- Example: Poisson model

$$Pr\left(Y = y \middle| \mathbf{x}\right) = \frac{e^{-\lambda(X'\beta)}\left[\lambda\left(X'\beta\right)\right]^{y}}{y!}$$

as $\lambda\left(X'\beta\right) > 0$ it is usually assumed that $\lambda\left(X'\beta\right) = e^{X'\beta}$

Limitations of the Poisson model

- Notice that in Poisson model

$$\mathbb{E}\left(Y|\mathbf{x}\right) = var\left(Y|\mathbf{x}\right) = e^{X'\beta}$$

- We often observe underdispersion $(E\left(Y|\mathbf{x}\right) > Var\left(Y|\mathbf{x}\right))$ or overdispersion $(E\left(Y|\mathbf{x}\right) > Var\left(Y|\mathbf{x}\right))$
- In such cases instead of Poisson distribution, we could use negative binomial model
- It was however shown that even if $\mathbb{E}\left(Y|\mathbf{x}\right) \neq var\left(Y|\mathbf{x}\right)$ but $\mathbb{E}\left(Y|\mathbf{x}\right)$ is correctly specified ML estimators of β are consistent.
- The are cases when observations of zero is generated from different mechanism then other counts (e.g. number of children, number of cigarettes smoked) in these cases we use
    - Zero Inflated Poisson (ZIP)
    - hurdle models

Interpretation of the parameters

- Sometimes the effect on Y^* is the one of interest (e.g. partial observability of variable of interest)
- In such a case the regression derivative is:

$$\frac{\partial \mathbb{E}(Y^*|X)}{\partial X} = \beta$$

- In other cases, especially if Y^* do not have clear intepretation we are interested in regression derivative for probabilities of observable variable Y:

$$\frac{\partial Pr(Y = j|X)}{\partial X}$$

or in unconditional or conditional expected values of Y e.g. for tobit model

$$\frac{\partial \mathbb{E}(Y|X)}{\partial X}, \frac{\partial \mathbb{E}(Y|Y > 0, X)}{\partial X}$$

Generalized Linear Models (GLM)

- Many of the popular limited dependent, qualitative or count variable models can be considered as members of the class known as GLM.
- For GLM models we assume that there is a function (link function) for which

$$g\left[\mathbb{E}\left(y_i|\mathbf{x}_i\right)\right] = \mathbf{x}_i\beta \text{ or}$$

$$\mathbb{E}\left(y_i|\mathbf{x}_i\right) = g^{-1}\left(\mathbf{x}_i\beta\right)$$

- We also assume that y_i exponential family of distributions e.g.: binomial, Poisson, negative binomial, gaussian, inverted gaussian, gamma
- The popular link functions are: identity $g(\mu) = \mu$, log: $g(\mu) = \ln(\mu)$, logit: $g(\mu) = \ln\left(\frac{\mu}{1-\mu}\right)$, probit: $g(\mu) = \Phi^{-1}(\mu)$, power: $g(\mu) = \mu^n$, negative binomial $g(\mu) = \ln\left(\frac{\mu}{\mu+k}\right)$, loglog: $g(\mu) = -\ln(-\ln\mu)$
- Only some of the combinations of link functions and distributions make sense

Delta method

- We often need to find distributions of the functions of estimated parameters (e.g. to calculate confidence intervals of regression derivatives)
- Direct derivation of this distributions is usually infeasible
- However, we can obtain asymptotic aproximations of these distributions using delta method
- Assume that
$$\sqrt{n}\left(\widehat{\boldsymbol{\mu}}-\boldsymbol{\mu}\right) \xrightarrow{d} N\left(0,\boldsymbol{V}\right)$$
- Then if $g()$ is differentiable in the neighborhood of μ
$$\sqrt{n}\left(g\left(\widehat{\boldsymbol{\mu}}\right)-g\left(\boldsymbol{\mu}\right)\right) \xrightarrow{d} N\left(0,\mathbf{G}'\mathbf{V}\mathbf{G}\right)$$
where $G=\frac{\partial}{\partial u}g\left(u\right)'$.

Resampling methods

- Derivation of the final sample distributions is often practicaly impossible
- It also happens that even the derivation of the asymptotic distributions of the statistics is cumbersome
- In this cases resampling methods are often used as a tool to
    - obtain in a simple way asymptotically valid aproximation of the distributions of statistics
    - improve the quality of approximations relative to standard methods (asymptotic refinement)
- There are two main resampling methods used:
    - jackknife
    - bootstrap
- Jacknife is a simpler method but bootstrap is more general

Jackknife

- Jackknife sample is constructed by omitting observation i
- In this way we obtain n samples
- For every one these samples we obtain calculate estimator $\widehat{\theta}_{(-i)}$
- The Tukey's jackknife estimator of variance

$$\widehat{\mathbf{V}}_{\widehat{\theta}} = \frac{n-1}{n} \sum_{i=1}^{n} \left(\widehat{\theta}_{(-i)} - \overline{\widehat{\theta}}_{(-i)} \right) \left(\widehat{\theta}_{(-i)} - \overline{\widehat{\theta}}_{(-i)} \right)'$$

- This method can also be used for estimation of the variance of transformations of estimators $g\left(\widehat{\theta}\right)$
- When used for clustered samples, we omit one cluster rather than one observation
- It can be proven that under quite general conditions jackknife estimator of variance is equivalent (but not better) to one obtained with delta method but does not require calculation of derivatives

Bootstrap

- This method is based on drawing randomly *with replacement* the bootstrap sample from original sample.
- Notice that bootstrap sample contains some duplicates of original observations
- Number of bootstraped samples B can be made arbitrary large
- Bootstrap can be used to estimate variance of the estimator:

$$\widehat{\mathbf{V}}_{\widehat{\boldsymbol{\theta}}}^{boot} = \frac{n-1}{n} \sum_{i=1}^{n} \left(\widehat{\theta}^{*}(b) - \overline{\widehat{\theta}}^{*} \right) \left(\widehat{\theta}_{(-i)} - \overline{\widehat{\theta}}_{(-i)} \right)'$$

- It can be proven that $\widehat{\theta}^{*}(b)$ is converging in probability, and also in distribution to $\widehat{\theta}$.
- It was also proven that $\widehat{\mathbf{V}}_{\widehat{\boldsymbol{\theta}}}^{boot} \xrightarrow{p^{*}} \mathbf{V}_{\widehat{\boldsymbol{\theta}}}$
- It is suggested that trimmed (with extreme $\widehat{\theta}^{*}(b)$ deleted from the sample) bootstrap estimator of variance has batter properties

Bootstrap confidence intervals

- Bootstrap can also be used to construct confidence intervals e.g. percentile t-interval

$$T^* = \frac{\widehat{\theta}^* - \widehat{\theta}}{se\left(\widehat{\theta}^*\right)}$$

$$C^{pt} = \left[\widehat{\theta} - se\left(\widehat{\theta}\right)\widehat{q}^*_{1-\frac{\alpha}{2}}, \widehat{\theta} - se\left(\widehat{\theta}\right)\widehat{q}^*_{\frac{\alpha}{2}}\right]$$

- Percentile t-interval achieves an asymptotic refinement (converges to true values with the rate n rather than $\sqrt{n}$)
- Boostrap p-values

$$p^* = \frac{1}{B}\sum_{b=1}^{B}\mathbb{I}\left(\left|T^*\left(b\right)\right| > \left|T\right|\right)$$

Bibliography

Hansen, B. (2022). *Econometrics*. Princeton University Press. ISBN: 9780691235899.