TA Session for Econometrics II 2025

5: Limited Dependent Variable Model

Jukina HATAKEYAMA

The University of Osaka, Department of Economics

September 28, 2025

Limited Dependent Variable Models

- Limited dependent variable (LDV) models are used when the observed outcome variable does not take values freely across the entire real line.
- Examples include:
  - Binary outcomes (e.g. participation vs non-participation).
  - Ordered outcomes (e.g. satisfaction levels).
  - Censored or truncated outcomes (e.g. expenditure, working hours).
- These models address the fact that standard linear regression is not appropriate when the dependent variable is restricted or only partially observed.

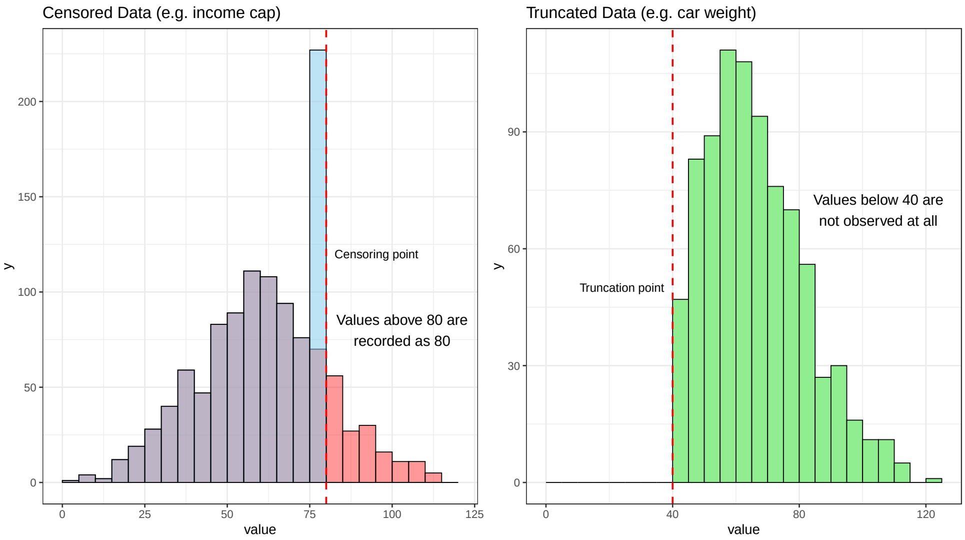
Censoring vs Truncation

- **Censoring:** The dependent variable is only partially observed beyond a certain threshold.
  - Example: Incomes above a reporting cap are recorded as "above £ 100,000" rather than their true value.
  - The observation remains in the sample, but the information is limited.
- **Truncation:** Observations are entirely excluded if they fall outside a certain range.
  - Example: A survey on labour supply that only includes individuals who are employed, excluding the unemployed.
  - The distribution of the sample is systematically altered.

Why LDV Models?

- Standard Ordinary Least Squares (OLS) would yield biased and inconsistent estimates under censoring or truncation.
- LDV models such as:
  - **Probit/Logit models** for binary outcomes,
  - **Ordered probit/logit** for ordinal data,
  - **Tobit models** for censored outcomes,

  provide consistent estimators tailored to the limited nature of the dependent variable.


Censored Data (e.g. income cap)
Censoring point
Values above 80 are recorded as 80
y
value
0
50
100
150
200
0
25
50
75
100
125
Truncated Data (e.g. car weight)
Truncation point
Values below 40 are not observed at all
y
value
0
30
60
90
0
40
80
120

Truncated Normal: density

Let $X \sim N(\mu, \sigma^2)$. Consider truncation from below at a (we observe X only when $X > a$).

Denote the standard normal density and c.d.f. by $\phi(\cdot)$ and $\Phi(\cdot)$ respectively, and set

$$\alpha \equiv \frac{a - \mu}{\sigma}.$$

The conditional (truncated) probability density function of X given $X > a$ is

$$f_{X|X>a}(x) = \frac{f_X(x)}{P(X > a)} = \frac{\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)}{1 - \Phi(\alpha)}, \qquad x > a,$$

and $f_{X|X>a}(x) = 0$ for $x \leq a$.

Note that the normalising factor $1 - \Phi(\alpha) = P(X > a)$ ensures the density integrates to one on (a, ∞).

Mean of a Truncated Normal: derivation

We compute the conditional mean by the definition

$$\mathbb{E}[X \mid X > a] \;=\; \int_a^\infty x\, f_{X|X>a}(x)\, dx \;=\; \frac{1}{1-\Phi(\alpha)} \int_a^\infty x\, f_X(x)\, dx.$$

Substitute $f_X(x) = \frac{1}{\sigma}\phi(\frac{x-\mu}{\sigma})$ and change variable $z = \frac{x-\mu}{\sigma}$ (so $x = \mu + \sigma z$, $dx = \sigma dz$):

$$\int_a^\infty x\, f_X(x)\, dx \;=\; \int_\alpha^\infty (\mu + \sigma z)\, \phi(z)\, dz.$$

Split the integral:

$$\int_\alpha^\infty (\mu + \sigma z)\, \phi(z)\, dz = \mu \int_\alpha^\infty \phi(z)\, dz \;+\; \sigma \int_\alpha^\infty z\, \phi(z)\, dz.$$

Use $\int_\alpha^\infty \phi(z)\, dz = 1 - \Phi(\alpha)$ and the identity $\int_\alpha^\infty z\, \phi(z)\, dz = \phi(\alpha)$ (since $\frac{d}{dz}[-\phi(z)] = z\phi(z)$ and evaluate bounds):

$$\int_a^\infty x\, f_X(x)\, dx = \mu\big(1 - \Phi(\alpha)\big) + \sigma\, \phi(\alpha).$$

Therefore

$$\mathbb{E}[X \mid X > a] = \frac{\mu\big(1 - \Phi(\alpha)\big) + \sigma\,\phi(\alpha)}{1 - \Phi(\alpha)} = \mu + \sigma\,\frac{\phi(\alpha)}{1 - \Phi(\alpha)}.$$

In compact form,

$$\boxed{\mathbb{E}[X \mid X > a] = \mu + \sigma\,\lambda(\alpha)}, \qquad \lambda(\alpha) \equiv \frac{\phi(\alpha)}{1 - \Phi(\alpha)} \quad \text{(inverse Mills ratio).}$$

Interpretation: truncation at a raises the conditional mean by $\sigma\lambda(\alpha)$; the shift depends on the position of a relative to μ (through α).

Truncated Normal Error Term

Consider the regression model:

$$y_i = X_i\beta + u_i, \quad u_i \sim N(0, \sigma^2),$$

observed only if $y_i > a$.

- This implies that u_i is observed only if

$$u_i > a - X_i\beta.$$

- Let

$$\alpha_i = \frac{a - X_i\beta}{\sigma}.$$

- Standardise: define $z_i = u_i/\sigma \sim N(0,1)$, observed if $z_i > \alpha_i$.

Expectation of the Truncated Error Term

- For the standard normal variable $z \sim N(0, 1)$:

$$\mathbb{E}[z \mid z > \alpha] = \frac{\phi(\alpha)}{1 - \Phi(\alpha)} \equiv \lambda(\alpha),$$

  where $\phi(\cdot)$ is the standard normal density and $\Phi(\cdot)$ is the distribution function.
- Thus, for $u_i = \sigma z_i$,

$$\mathbb{E}[u_i \mid u_i > a - X_i\beta] = \sigma\lambda(\alpha_i).$$

Expectation of y_i under Truncation

- Recall

$$y_i = X_i\beta + u_i.$$

- Therefore, conditional on $y_i > a$,

$$\mathbb{E}[y_i \mid y_i > a] = X_i\beta + \mathbb{E}[u_i \mid u_i > a - X_i\beta].$$

- Substituting from before,

$$\mathbb{E}[y_i \mid y_i > a] = X_i\beta + \sigma\lambda(\alpha_i).$$

Variance of y_i under Truncation

$$\mathrm{Var}(y_i \mid y_i > a) = \sigma^2 \Big[1 - \lambda(\alpha_i) \big(\lambda(\alpha_i) - \alpha_i \big) \Big].$$

- Truncation shifts the mean upward.
- Variance is reduced because lower tail observations are excluded.
- These properties distinguish truncated regression models from censored (Tobit) models.

Estimation: Maximum Likelihood

For the truncated regression model

$$y_i = X_i\beta + u_i, \quad u_i \sim N(0,\sigma^2), \quad \text{observed only if } y_i > a,$$

the conditional density of an observation is

$$f(y_i \mid y_i > a, X_i) \;=\; \frac{\frac{1}{\sigma}\,\phi\!\left(\frac{y_i - X_i\beta}{\sigma}\right)}{1-\Phi\!\left(\frac{a - X_i\beta}{\sigma}\right)}, \quad y_i > a,$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the standard normal pdf and cdf.

The likelihood function for n independent observations is

$$L(\beta,\sigma^2) = \prod_{i=1}^{n} \frac{\frac{1}{\sigma}\,\phi\!\left(\frac{y_i - X_i\beta}{\sigma}\right)}{1-\Phi\!\left(\frac{a - X_i\beta}{\sigma}\right)}.$$

Taking logs gives the log-likelihood:

$$\ell(\beta, \sigma^2) = \sum_{i=1}^{n} \left\{ -\log \sigma + \log \phi\left(\frac{y_i - X_i\beta}{\sigma}\right) - \log\left[1 - \Phi\left(\frac{a - X_i\beta}{\sigma}\right)\right] \right\}.$$

Estimation: maximise $\ell(\beta, \sigma^2)$ with respect to β and σ^2.

Examples of Truncated Regression Models

1. **Buying a Car**
   Model: $y_i = x_i\beta + u_i$
   y_i: expenditure on a car
   x_i: income, price of the car, and other covariates
   Only individuals who purchased a car are observed; non-buyers are excluded.

2. **Working Hours of Wives**
   y_i: working hours of a wife
   x_i: number of children, age, education, husband's income, etc.
   Observed data exclude households where the wife does not work.

3. **Stochastic Frontier Model**
   $y_i = f(K_i, L_i) + u_i$
   y_i: production; K_i: capital stock; L_i: labour input
   Constraint: $y_i \leq f(K_i, L_i)$, i.e. $u_i \leq 0$
   $f(K_i, L_i)$ represents the maximum attainable output for given inputs.

Censored Regression Model

Model

$$y_i = \begin{cases} X_i\beta + u_i, & \text{if } y_i > a, \\ a, & \text{otherwise.} \end{cases}$$

Probability Mass at a

$$P(y_i = a) = P(y_i \leq a) = F(a) \equiv \int_{-\infty}^{a} f(x)\,dx,$$

where $f(\cdot)$ and $F(\cdot)$ denote the pdf and cdf of y_i.

Likelihood Function

$$L(\beta, \sigma^2) = \prod_{i=1}^{n} F(a)^{I(y_i=a)} \times f(y_i)^{1-I(y_i=a)},$$

where $I(y_i = a)$ is the indicator function.

Tobit Model (Normality Assumption)

Assumption

$$u_i \sim N(0, \sigma^2).$$

Likelihood Function

$$L(\beta, \sigma^2) = \prod_{i=1}^{n} \left(\int_{-\infty}^{a} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (y_i - X_i\beta)^2 \right) \, dy_i \right)^{I(y_i = a)}$$
$$\times \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (y_i - X_i\beta)^2 \right) \right)^{1 - I(y_i = a)} .$$

Estimation

Maximize $L(\beta, \sigma^2)$ w.r.t. β, σ^2.

Review: Nonlinear Optimisation (brief)

- Many likelihood problems require numerical maximisation.
- **Newton–Raphson** update:

$$\beta^{(j+1)} = \beta^{(j)} - \left[\frac{\partial^2 \ell(\beta^{(j)})}{\partial \beta \partial \beta'}\right]^{-1} \frac{\partial \ell(\beta^{(j)})}{\partial \beta}.$$

- **Method of scoring** replaces the Hessian by its expectation:

$$\beta^{(j+1)} = \beta^{(j)} - \left(E\left[\tfrac{\partial^2 \ell}{\partial \beta \partial \beta'}\right]\right)^{-1} \frac{\partial \ell(\beta^{(j)})}{\partial \beta}.$$

- Intuition: use local quadratic approximation of the log-likelihood and step towards stationary point.

Poisson Distribution — definition and intuition

Poisson pmf

$$P(X=x)=\frac{e^{-\lambda}\lambda^{x}}{x!},\qquad x=0,1,2,\dots$$

- Expectation and variance: $E[X]=\lambda$, $\mathrm{Var}(X)=\lambda$. 直感的には「単位時間あたりの発生率」が λ。
- Typical examples: number of doctor visits, number of accidents at a junction, number of patents per firm.
- For regression, we set a positive mean λ_i that depends on covariates.

Poisson regression: model specification

$$E(y_i) = \lambda_i, \qquad \lambda_i = \exp(X_i \beta).$$

- Use the exponential link so that $\lambda_i > 0$ automatically.
- Equivalent form: $\log \lambda_i = X_i \beta$ (log-link).
- Intuition: covariates affect the log expected count additively; effects on counts are multiplicative.

Likelihood and log-likelihood (intuition)

$$L(\beta) = \prod_{i=1}^{n} \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, \qquad \lambda_i = \exp(X_i\beta).$$

$$\ell(\beta) = \sum_{i=1}^{n} \left(-\lambda_i + y_i \log \lambda_i - \log y_i!\right) = -\sum_i \exp(X_i\beta) + \sum_i y_i X_i\beta - \sum_i \log y_i!.$$

- The term $-\sum_i \exp(X_i\beta)$ penalises large predicted counts.
- The term $\sum_i y_i X_i\beta$ rewards fit where observed counts are large.
- $\sum_i \log y_i!$ does not depend on β (constant for optimisation).

Score function (first derivative) — derivation and intuition

Differentiate $\ell(\beta)$ w.r.t. β:

$$\frac{\partial \ell(\beta)}{\partial \beta} = -\sum_{i=1}^{n} X_i' \exp(X_i\beta) + \sum_{i=1}^{n} X_i' y_i = \sum_{i=1}^{n} X_i'(y_i - \mu_i),$$

where $\mu_i = \exp(X_i\beta)$.

- Set this to zero for the FOC: $\sum_i X_i'(y_i - \mu_i) = 0$. 直感的には「(重み付けされた）残差の総和がゼロ」。
- Nonlinearity: μ_i depends on β, so we require iterative solution.

Newton–Raphson / Scoring for Poisson

$$H(\beta) = \frac{\partial^2 \ell(\beta)}{\partial \beta \partial \beta'} = -\sum_{i=1}^{n} X_i' X_i \mu_i,$$

hence the Newton update is

$$\beta^{(j+1)} = \beta^{(j)} + \left(\sum_i X_i' X_i \mu_i^{(j)} \right)^{-1} \sum_i X_i' (y_i - \mu_i^{(j)}).$$

- For Poisson, the observed Hessian equals its expectation (so Newton–Raphson = Fisher scoring).
- The update has the form of an **IRLS** (iteratively reweighted least squares) algorithm:

$$(X'WX)\,\Delta\beta = X'(y-\mu),$$

  with weights $W_{ii} = \mu_i^{(j)}$.

Interpreting coefficients — incidence rate ratios (IRR)

- A coefficient β_k has multiplicative effect on the mean:

$$\frac{\partial \log \lambda_i}{\partial x_{ik}} = \beta_k.$$

- Therefore $\exp(\beta_k)$ is the **incidence rate ratio (IRR)**:

$$\text{IRR} = \frac{E[y \mid x_k + 1]}{E[y \mid x_k]} = \exp(\beta_k).$$

- **Numeric example:** if $\beta_k = 0.20$, $\exp(0.20) \approx 1.22 \rightarrow$ a one-unit increase in x_k raises expected count by $\approx$ 22%.

Practical issues and diagnostics

- **Overdispersion:** If empirical $\mathrm{Var}(y_i) > E(y_i)$, the Poisson variance assumption fails. Remedies:
  - use robust (sandwich) standard errors; or
  - estimate a Negative Binomial model (extra parameter for dispersion).
- **Excess zeros:** if many zeros, consider zero-inflated Poisson (ZIP) or hurdle models.
- **Goodness of fit:** use deviance, Pearson chi-square, AIC/BIC, residual plots.
- **Software:** in R, `glm(formula, family=poisson)`; for negative binomial, `glm.nb` (MASS).

Simple applied example (intuition)

Count of doctor visits per year (y): covariates include age, chronic conditions, insurance status.

$$\log E[y_i] = \beta_0 + \beta_1 \text{age}_i + \beta_2 \text{chronic}_i + \beta_3 \text{insured}_i.$$

- If $\hat{\beta}_3 = 0.30$ then $\exp(0.30) \approx 1.35$: insured people have 35% higher expected visits, holding other covariates constant.
- Check residual deviance and compare Poisson vs Negative Binomial if overdispersion present.

Takeaways

- Poisson regression is a natural first model for non-negative integer responses; log-link enforces positivity and gives multiplicative interpretation.
- Estimation by MLE requires iterative optimisation; Newton-Raphson / scoring (IRLS) are standard and intuitive.
- Always check model fit: if variance $>$ mean or many zeros, consider NB, ZIP or hurdle models.

Motivation: Too Many Zeros

- In real-world data, the number of zeros can be much larger than what a standard Poisson distribution predicts.
- Examples:
  - Number of accidents in a year (many people may have zero accidents).
  - Number of doctor visits (a large fraction may have zero visits).
- $\Rightarrow$ We need to extend the Poisson model to account for these "excess zeros".

Basic Idea: Two Regimes

- Suppose there are two regimes:
  1. **Regime 1 (Always Zero)**: Some individuals always produce a zero outcome.
  2. **Regime 2 (Poisson)**: Others follow a Poisson distribution.
- Then the probability is a mixture of these two regimes:

$$\begin{aligned} P(y_i = j) = P(y_i = j \mid \text{Regime 1}) P(\text{Regime 1}) \\ + P(y_i = j \mid \text{Regime 2}) P(\text{Regime 2}). \end{aligned}$$

Assumptions

- Under Regime 1:

$$P(y_i = 0 \mid \text{Regime 1}) = 1, \quad P(y_i = j \mid \text{Regime 1}) = 0 \ (j = 1, 2, \dots)$$

- Let $F_i = F(Z_i\alpha) =$ probability of being in Regime 1.
- Under Regime 2:

$$P(y_i = j \mid \text{Regime 2}) = \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}, \quad \lambda_i = \exp(X_i\beta).$$

Distribution of y_i

- Combining the assumptions:

$$P(y_i = j) = F_i I_i + \frac{e^{-\lambda_i}\lambda_i^{y_i}}{y_i!}(1 - F_i), \quad j = 0, 1, 2, \dots$$

where

$$I_i = \begin{cases} 1 & \text{if } y_i = 0, \\ 0 & \text{if } y_i > 0. \end{cases}$$

- Interpretation:
  - If $y_i = 0$, it could come from either regime.
  - If $y_i > 0$, it must come from the Poisson regime.

Log-Likelihood Function

- The likelihood for n observations is:

$$L(\alpha, \beta) = \prod_{i=1}^{n} \left(F_i I_i + \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} (1 - F_i) \right).$$

- Taking logs:

$$\log L(\alpha, \beta) = \sum_{i=1}^{n} \log \left(F_i I_i + \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} (1 - F_i) \right).$$

- Estimation: maximise $\log L(\alpha, \beta)$ with respect to α, β.
- Methods: Newton – Raphson, method of scoring.

Takeaway

- The Zero-Inflated Poisson (ZIP) model handles excess zeros by introducing a two-regime framework.
- Intuition: Some zeros are "structural" (always zero), others are "random" (from Poisson).
- This makes the model much more flexible than the standard Poisson model.