Econometrics II

Censoring and Selection

Lasha Chochua

2026

Introduction

Motivation

- In many economic applications, the outcome variable Y is **not freely observed** for all units
- Two main problems:
  - **Censoring:** Y is observed but constrained to a boundary for some observations
  - **Selection:** Some observations are missing because of a non-random sampling process
- Under either problem, **OLS is biased** for the population parameters of interest
- Standard remedy: specialized estimators – **Tobit**, **CLAD**, **Heckman**

i Note

Key references: Maddala (1983), Amemiya (1985), Gourieroux (2000), Cameron and Trivedi (2005), Wooldridge (2010), Hansen (2022) Ch. 27.

A Concrete Example – Remittances

- Dataset: CHJ2004 – Filipino households
- Variable: **transfers from abroad** (*tabroad*), in thousands of pesos
- **80% of households** receive zero transfers
- For the remaining 20%, transfers are continuously distributed with a thick right tail

- **Question**: how do we proceed with regression analysis?
  - Use full sample including zeros?
  - Drop zeros and use subsample?
  - Something more principled?

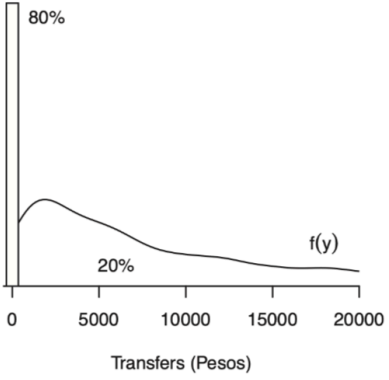

Figure 1: Transfers from Abroad

The bar at zero represents the 80% mass point. The line graph shows the continuous density for positive values.

Censored Regression: The Tobit Model

Three Distributions – Key Distinctions

Definitions: Uncensored, Censored, Truncated

Let Y^* be the latent (true) outcome.

- **Uncensored (Y^*):** the underlying continuous variable, fully observed in principle
- **Censored (Y):** we observe Y^* when it is positive, but record $Y = 0$ when $Y^* \leq 0$
- **Truncated ($Y^{\#}$):** observations with $Y = 0$ are **deleted** entirely from the sample

- These three objects have **different conditional means** – do not conflate them
- Running OLS on either the censored or truncated sample gives **biased** estimates of $\mathbb{E}[Y^*|X]$

The Tobit Model – Setup

- Proposed by **Tobin (1958)** to model household consumption of durable goods.
- The model posits a **latent** continuously-distributed variable Y^*:

$$\begin{aligned} Y^* &= X'\beta + e, \quad e \mid X \sim N(0, \sigma^2) \\ Y &= \max(Y^*, 0) \end{aligned} \tag{1}$$

- Y^* is **unobserved** (latent)
- Y is what we observe: positive values are uncensored, negative values are collapsed to zero
- Also called **Type 1 Tobit** or **censored regression**

> **Note**
>
> Y^* measures the **net benefit of an additional unit** relative to the status quo. Once it goes negative, the constraint $Y \geq 0$ (on *purchases*, not on *stock*) kicks in and censors the outcome.

The Censoring Process – Visually

- The latent Y^* has a normal density centered at $X'\beta$
- The portion where $Y^* > 0$ is kept as-is (uncensored)
- The portion where $Y^* < 0$ is **collapsed to a point mass** at zero

- As $X'\beta$ moves **right** $\Rightarrow$ less censoring
- As $X'\beta$ moves **left** $\Rightarrow$ more censoring

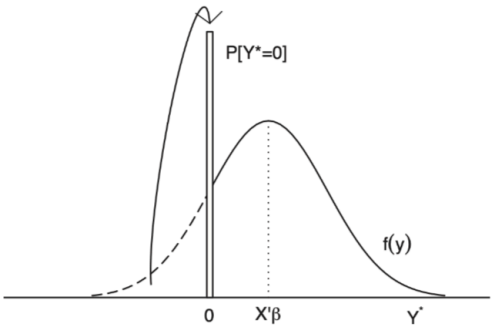

Figure 2: The Censoring Process

Assumptions of the Tobit Model

- The Tobit model (1) rests on three assumptions:

Assumption	Content	Criticality
Linearity	$\mathbb{E}[Y^*\|X] = X'\beta$	Not critical – $X'\beta$ can approximate a flexible function
Independence	$e \perp X$	**Critical** – heteroskedasticity changes the censoring process
Normality	$e \mid X \sim N(0, \sigma^2)$	**Critical** – hard to justify from first principles

Note

Because normality is critical but untestable from the censored data alone, robust alternatives (CLAD) are important in practice.

Censored Regression Functions

Three Conditional Means

- Under model (1), we can derive closed-form expressions for the three conditional means.
- Let $\phi(\cdot)$ and $\Phi(\cdot)$ denote the standard normal PDF and CDF, respectively.

$$m^*(X) = \mathbb{E}[Y^* \mid X] = X'\beta \tag{2}$$

$$m(X) = \mathbb{E}[Y \mid X] = X'\beta \cdot \Phi\left(\frac{X'\beta}{\sigma}\right) + \sigma\phi\left(\frac{X'\beta}{\sigma}\right) \tag{3}$$

$$m^{\#}(X) = \mathbb{E}[Y^{\#} \mid X] = X'\beta + \sigma\lambda\left(\frac{X'\beta}{\sigma}\right) \tag{4}$$

where $\lambda(x) = \phi(x)/\Phi(x)$ is the **inverse Mills ratio**.

Derivation – Truncated Mean $m^{\#}(X)$

- We want $\mathbb{E}[Y^{\#} \mid X] = \mathbb{E}[Y^* \mid Y^* > 0, X]$. Since $Y^* = X'\beta + e$:

$$\mathbb{E}[Y^* \mid Y^* > 0, X] = X'\beta + \mathbb{E}[e \mid e > -X'\beta]$$

- By definition of conditional expectation:

$$\mathbb{E}[e \mid e > -X'\beta] = \frac{\int_{-X'\beta}^{\infty} e \cdot f(e)\, de}{P(e > -X'\beta)}$$

Derivation – Truncated Mean $m^{\#}(X)$ (Cont.)

- **Denominator:** since $e \sim N(0, \sigma^2)$:

$$P(e > -X'\beta) = \Phi\left(\frac{X'\beta}{\sigma}\right)$$

- **Numerator:** substitute $u = e/\sigma$, so $e = \sigma u$, $de = \sigma\, du$:

$$\int_{-X'\beta}^{\infty} e \cdot \frac{1}{\sigma}\phi\left(\frac{e}{\sigma}\right) de = \int_{-X'\beta/\sigma}^{\infty} \sigma u \cdot \phi(u)\, du$$

- Use the identity $u \cdot \phi(u) = -\phi'(u)$ (**Why?**):

$$= \sigma \int_{-X'\beta/\sigma}^{\infty} -\phi'(u)\, du = \sigma \left[-\phi(u)\right]_{-X'\beta/\sigma}^{\infty} = \sigma\, \phi\left(\frac{X'\beta}{\sigma}\right)$$

The Identity $u \cdot \phi(u) = -\phi'(u)$

- Recall the standard normal pdf:

$$\phi(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$$

- Differentiate with respect to u (chain rule on the exponent):

$$\phi'(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \cdot (-u) = -u \cdot \phi(u)$$

- Rearranging:

$$u \cdot \phi(u) = -\phi'(u)$$

Derivation – Truncated Mean $m^{\#}(X)$ (Cont.)

- **Combining:**

$$\mathbb{E}[e \mid e > -X'\beta] = \frac{\sigma\,\phi(X'\beta/\sigma)}{\Phi(X'\beta/\sigma)} = \sigma\lambda\left(\frac{X'\beta}{\sigma}\right)$$

- Therefore:

$$m^{\#}(X) = X'\beta + \sigma\lambda\left(\frac{X'\beta}{\sigma}\right)$$

Derivation – Censored Mean $m(X)$

- Split $\mathbb{E}[Y \mid X]$ by whether $Y^* > 0$ or not:

$$\mathbb{E}[Y \mid X] = \mathbb{E}[Y^* \mid Y^* > 0, X] \cdot \Pr[Y^* > 0 \mid X] + 0 \cdot \Pr[Y^* \leq 0 \mid X]$$

- Substituting $\Pr[Y^* > 0 \mid X] = \Phi(X'\beta/\sigma)$ and the truncated mean from above:

$$m(X) = \left[X'\beta + \sigma\lambda\left(\tfrac{X'\beta}{\sigma}\right)\right] \Phi\left(\tfrac{X'\beta}{\sigma}\right)$$

- Expanding and using the key identity $\lambda(x)\Phi(x) = \phi(x)$:

$$m(X) = X'\beta\, \Phi\left(\frac{X'\beta}{\sigma}\right) + \sigma\phi\left(\frac{X'\beta}{\sigma}\right)$$

The Inverse Mills Ratio

Definition: Inverse Mills Ratio

$$\lambda(x) = \frac{\phi(x)}{\Phi(x)}$$

where ϕ is the standard normal PDF and Φ is the standard normal CDF.

- $\lambda(x) > 0$ for all x, and $\lambda'(x) < 0$ (decreasing)
- Captures the **selection effect**: among observations with $Y^* > 0$, the error e is truncated and its mean is positive
- Appears in both the truncated mean (4) and the Heckman selection correction

The Inverse Mills Ratio (Cont.)

Note

Intuition: if we only observe Y^* when it is positive, then the average error e in this subsample is positive – the inverse Mills ratio measures exactly how much.

Ranking of Conditional Means

- Since $Y^* \leq Y \leq Y^\#$ (in the sense of their distributions), we have:

$$m^*(x) \leq m(x) \leq m^\#(x)$$

with **strict inequality** whenever the censoring probability is positive.

Ranking of Conditional Means (Cont.)

- **Censored mean** $m(x)$**:** biased upward relative to $m^*(x)$
- **Truncated mean** $m^{\#}(x)$**:** even more biased – highest of the three
- Bias is **larger** when censoring probability is higher

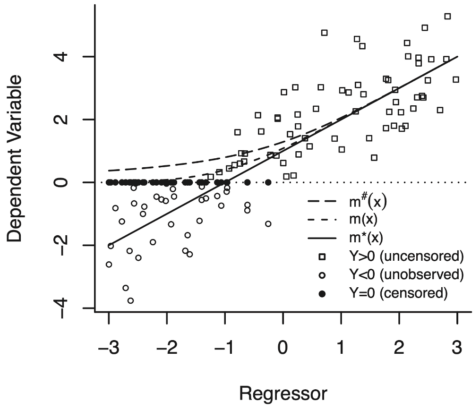

Figure 3: Scatter of (Y^*, X)

$m^*(x)$ *solid,* $m(x)$ *dashed,* $m^{\#}(x)$ *long dashes. Open squares* $= Y > 0$*, filled circles* $= Y = 0$ *(censored)*

Censoring Probability

- The conditional probability that an observation is censored:

$$\Pr[Y^* < 0 \mid X] = \Pr[e < -X'\beta \mid X] = \Phi\left(\frac{-X'\beta}{\sigma}\right)$$

Censoring Probability (Cont.)

- Censoring probability **decreases** as $X'\beta$ increases
- Example (from Figure 3, $Y^* \mid X \sim N(1 + X, 1)$):
- $X = -3$: censoring probability $\approx 98\%$
- $X = -1$: censoring probability $\approx 50\%$
- $X = 1$: censoring probability $\approx 2\%$

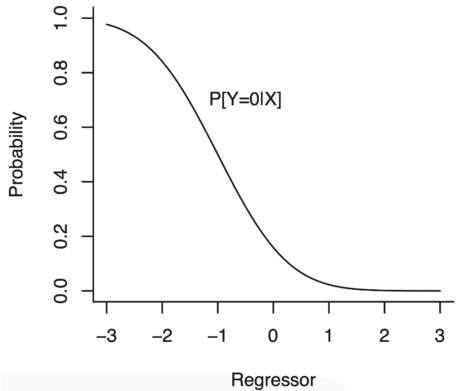

Figure 4: Censoring Probability

Bias of OLS

Why OLS Fails

- If data are generated by the censored model (1), running OLS on:
  - **Full sample** (including zeros): estimates $m(x)$, not $m^*(x)$ – **biased upward**
  - **Truncated sample** (dropping zeros): estimates $m^{\#}(x)$, not $m^*(x)$ – **even more biased**

- The intuition: we are fitting a linear model to a **nonlinear** conditional mean
- Even a consistent nonparametric estimator would recover $m(x)$ or $m^{\#}(x)$, not $m^*(x)$

! Important

Key insight: dropping the censored observations does *not* fix the problem – it makes things worse. The truncated mean $m^{\#}(x)$ is more biased than the censored mean $m(x)$.

Greene's Formula – Quantifying the Bias

- **Greene (1981):** when regressors $X \sim N(0, \Sigma)$, the OLS slope estimand satisfies:

$$\beta^{BLP} = \beta(1 - \pi) \tag{5}$$

  where $\pi = \Pr[Y = 0]$ is the **censoring proportion**.

- OLS slope coefficients are **shrunk toward zero** proportionally to π
- If $\pi = 0.8$ (transfers example): OLS slope is only 20% of the true value!
- π is easily estimated from the sample – provides a quick bias diagnostic

> **Note**
>
> **Rule of thumb:** if estimated censoring bias is less than $\sim 5\%$ (i.e., π is small), OLS on the full sample may be acceptable. Otherwise, use censoring-corrected estimators.

Deriving Greene's Formula – Sketch

- The key trick (Goldberger 1981): note $Y^* \sim N(\alpha, \sigma_Y^2)$ with $\sigma_Y^2 = \sigma^2 + \beta'\Sigma\beta$.
- Using moments of the truncated normal distribution:

$$\mathbb{E}[(Y^* - \alpha)Y^* \mid Y^* > 0] = \sigma_Y^2$$

- By **joint normality** of X and Y^*, the cross-moment is preserved under truncation:

$$\mathbb{E}[XY^* \mid Y^* > 0] = \mathbb{E}[XY^*] = \Sigma\beta$$

- Assembling the BLP formula:

$$\beta^{BLP} = \mathbb{E}[XX']^{-1}\mathbb{E}[XY] = \mathbb{E}[XX']^{-1}\mathbb{E}[XY^* \mid Y^* > 0](1-\pi) = \beta(1-\pi)$$

which is (5).

The Tobit Estimator

Tobit – Likelihood Construction

- The censored variable Y has a **mixed continuous/discrete** conditional distribution:

$$F(y \mid x) = \begin{cases} 0 & y < 0 \\ \Phi\left(\dfrac{y - x'\beta}{\sigma}\right) & y \geq 0 \end{cases}$$

- The density (with respect to a mixed measure) is:

$$f(y \mid x) = \Phi\left(\frac{-x'\beta}{\sigma}\right)^{\mathbf{1}\{y=0\}} \cdot \left[\sigma^{-1}\phi\left(\frac{y - x'\beta}{\sigma}\right)\right]^{\mathbf{1}\{y>0\}}$$

- First component: **probit-like** (probability of censoring)
- Second component: **normal regression** density (for uncensored observations)

Tobit – Deriving the Density

- Y has a **mixed distribution**: point mass at zero + continuous part for $Y > 0$
- **At $Y = 0$:** this happens whenever $Y^* \leq 0$, so:

$$\Pr[Y = 0 \mid x] = \Pr[Y^* \leq 0 \mid x] = \Pr[e \leq -x'\beta] = \Phi\left(\frac{-x'\beta}{\sigma}\right)$$

- **At $Y > 0$:** the censoring constraint is not binding, so $Y = Y^*$ and the density is just the normal regression density:

$$f(Y \mid x, Y > 0) = \sigma^{-1}\phi\left(\frac{y - x'\beta}{\sigma}\right)$$

The Tobit Likelihood – Uncensored Observations

- At $Y > 0$: the censoring constraint is not binding, so $Y = Y^*$
- The density is the $N(x'\beta, \sigma^2)$ pdf. Substituting $z = (y - x'\beta)/\sigma$:

$$f(y) = \phi(z) \cdot \frac{dz}{dy} = \phi\left(\frac{y - x'\beta}{\sigma}\right) \cdot \frac{1}{\sigma}$$

- The σ^{-1} is the **Jacobian** of the transformation – without it the density does not integrate to 1:

$$\int_{-\infty}^{\infty} \frac{1}{\sigma}\phi\left(\frac{y - x'\beta}{\sigma}\right) dy = \int_{-\infty}^{\infty} \phi(z)\, dz = 1 \checkmark$$

- Therefore for uncensored observations:

$$f(y \mid x, Y > 0) = \frac{1}{\sigma}\phi\left(\frac{y - x'\beta}{\sigma}\right)$$

Tobit – Deriving the Density (Cont.)

- Combining both cases into a single expression using indicator exponents:

$$f(y \mid x) = \underbrace{\Phi\left(\frac{-x'\beta}{\sigma}\right)^{\mathbf{1}\{y=0\}}}_{\text{discrete part}} \cdot \underbrace{\left[\sigma^{-1}\phi\left(\frac{y - x'\beta}{\sigma}\right)\right]^{\mathbf{1}\{y>0\}}}_{\text{continuous part}}$$

- When $\mathbf{1}\{y = 0\} = 1$: only the first factor survives; when $\mathbf{1}\{y > 0\} = 1$: only the second survives

Tobit – Deriving the Log-Likelihood

- Take $\log$ of the density $f(y \mid x)$ for each observation:

$$\log f(y_i \mid x_i) = \mathbf{1}\{y_i = 0\} \log \Phi\left(\frac{-x_i'\beta}{\sigma}\right) + \mathbf{1}\{y_i > 0\} \log\left[\sigma^{-1}\phi\left(\frac{y_i - x_i'\beta}{\sigma}\right)\right]$$

- Sum over all n observations and split into two groups:

$$\ell_n = \sum_{Y_i=0} \log \Phi\left(\frac{-X_i'\beta}{\sigma}\right) + \sum_{Y_i>0} \log\left[\sigma^{-1}\phi\left(\frac{Y_i - X_i'\beta}{\sigma}\right)\right]$$

- Expand the second term using $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$:

$$\log\left[\sigma^{-1}\phi\left(\frac{Y_i - X_i'\beta}{\sigma}\right)\right] = -\log\sigma + \log\phi\left(\frac{Y_i - X_i'\beta}{\sigma}\right) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{(Y_i - X_i'\beta)^2}{2\sigma^2}$$

- Substituting back gives (6) – the **probit-like** first sum plus the **normal regression** second sum

Tobit – Log-Likelihood

- So:

$$\ell_n(\beta, \sigma^2) = \sum_{Y_i=0} \log \Phi\left(\frac{-X_i'\beta}{\sigma}\right) - \frac{1}{2} \sum_{Y_i>0} \left[\log(2\pi\sigma^2) + \frac{(Y_i - X_i'\beta)^2}{\sigma^2}\right] \tag{6}$$

- The MLE $(\hat{\beta}, \hat{\sigma}^2)$ maximizes (6)
- Asymptotic normality: **Amemiya (1973)**
- The nickname "**Tobit**" was coined by Goldberger – by analogy with probit and logit

Tobit – Reparameterization for Computation

- **Olsen (1978):** reparameterize as $\gamma = \beta/\sigma$, $\nu = 1/\sigma$. Then:

$$\ell_n(\gamma, \nu) = \underbrace{\sum_{Y_i=0} \log \Phi(-X_i'\gamma)}_{\text{concave in } \gamma} + \underbrace{\sum_{Y_i>0} \log(\nu/\sqrt{2\pi})}_{\text{concave in } \nu} - \frac{1}{2} \underbrace{\sum_{Y_i>0} (Y_i\nu - X_i'\gamma)^2}_{\text{concave in } (\gamma,\nu)}$$

- Each term is **globally concave** in (γ, ν)
- Sum is globally concave $\Rightarrow$ **unique global maximum**, Newton algorithms converge reliably

> **Note**
>
> **Software:** `tobit` in Stata; `tobit()` in the `AER` package in R.

James Tobin (1918–2002)

- One of the leading macroeconomists of the mid-twentieth century
- **Nobel Prize in Economics, 1981**
- His 1958 paper introduced censored regression and the MLE that now bears his name

Identification in Tobit Regression

Nonparametric Censored Regression

- Relax Tobit's normality assumption. Consider the general model:

$$Y^* = m(X) + e, \quad \mathbb{E}[e] = 0, \quad e \sim F \text{ independent of } X, \quad Y = \max(Y^*, 0)$$

  where both $m(\cdot)$ and $F(\cdot)$ are **unknown**.

- **What is identified?**

! Theorem (Nonparametric Identification)

If $m(X)$ has **unbounded support** on $\mathbb{R}$ (e.g., X has an unbounded distribution), then:
- $m(x)$ is nonparametrically identified
- $F(e)$ is nonparametrically identified

provided that X and e are **independent**.

Identification – Intuition

- When $m(X)$ has full support, there exist regions $\mathcal{X}$ where $\Pr[Y = 0 \mid X = x] \approx 0$
- In those regions, the censoring constraint is **not binding** – we essentially observe Y^* directly
- This identifies $m(x)$ for $x \in \mathcal{X}$, which then identifies $F(e)$
- Global identification of $m(x)$ follows from identification of F and the censoring probability

Identification – Intuition (Cont.)

- **When does identification fail?**
  - If $m(X)$ is bounded above ($m(X) \leq \bar{m}$ with $\Pr[Y = 0 \mid X] > 0$ everywhere): $F(e)$ is not identified for $e \leq -\bar{m} \Rightarrow$ intercept is not identified
  - If e is **not independent** of X: the censoring probability does not identify $m(x)$

Quantile Regression

Quantile Regression – Short Intro

- OLS estimates the **conditional mean** $\mathbb{E}[Y \mid X]$ – sensitive to outliers and skewness
- **Quantile regression** estimates the **conditional quantile** $Q_\tau[Y \mid X]$ for $\tau \in (0, 1)$:

$$Q_\tau[Y \mid X] = \inf\{y : \Pr[Y \leq y \mid X] \geq \tau\}$$

- $\tau = 0.5$: conditional **median** regression; $\tau = 0.9$: conditional **90th percentile**
- The linear quantile regression model assumes $Q_\tau[Y \mid X] = X'\beta_\tau$ – different τ gives different β_τ

Conditional Quantile – Unpacking the Definition

- The conditional quantile is defined as:

$$Q_\tau[Y \mid X] = \inf\{y : \Pr[Y \leq y \mid X] \geq \tau\}$$

- Reading piece by piece:
  - $\Pr[Y \leq y \mid X] \geq \tau$ – the CDF of Y given X, evaluated at y; we want it to reach at least τ
  - $\{y : \Pr[Y \leq y \mid X] \geq \tau\}$ – the set of all y values where the CDF has already reached τ
  - $\inf\{\cdot\}$ – take the left edge of that set – the smallest such y

Quantile Regression – Example

- Let $Y =$ hourly wage, $X =$ years of education. Among college graduates ($X = 16$):

Quantile τ	**Interpretation**	$Q_\tau[Y \mid X = 16]$
0.1	10% earn less than	\$12/hr
0.5	50% earn less than	\$22/hr
0.9	90% earn less than	\$45/hr

Quantile Regression – Example (Cont.)

- Add one more year of education ($X = 17$). The **quantile slopes** β_τ are:

Quantile τ	$Q_\tau[Y \mid X = 17]$	β_τ
0.1	\$13/hr	$+\$1$
0.5	\$24/hr	$+\$2$
0.9	\$52/hr	$+\$7$

- OLS gives one average slope $\beta^{OLS} = 3$ – masking large heterogeneity across the distribution

i Note

$\beta_{0.9} \gg \beta_{0.1}$: education stretches the top and compresses the bottom – **increasing wage dispersion**. Quantile regression reveals this; OLS averages it away.

Quantile Regression – Short Intro (Cont.)

- Estimated by minimizing the **asymmetrically weighted** check function:

$$\hat{\beta}_\tau = \text{argmin}_\beta \frac{1}{n} \sum_{i=1}^{n} \rho_\tau(Y_i - X_i'\beta), \quad \rho_\tau(u) = u(\tau - \mathbf{1}\{u < 0\})$$

- **All observations** are used – quantile regression does not subset the data
- The check function penalizes **underpredictions** with weight τ and **overpredictions** with weight $1 - \tau$
- For $\tau = 0.9$: heavy penalty on underpredictions $\Rightarrow$ fitted line is pushed up until exactly 90% of observations lie below it
- For $\tau = 0.5$: symmetric penalties $\Rightarrow$ equivalent to LAD (least absolute deviations)

Quantile Regression – The Check Function $\rho_\tau(u)$

- The check function penalizes positive and negative residuals **asymmetrically**:

$$\rho_\tau(u) = u(\tau - \mathbf{1}\{u < 0\}) = \begin{cases} \tau \cdot u & u \geq 0 \\ (1-\tau) \cdot |u| & u < 0 \end{cases}$$

Quantile τ	**Above line** ($u > 0$)	**Below line** ($u < 0$)	**Effect**
$\tau = 0.1$	cheap: rate 0.1	costly: rate 0.9	line pushed **down**
$\tau = 0.5$	rate 0.5	rate 0.5	symmetric – estimates **median**
$\tau = 0.9$	costly: rate 0.9	cheap: rate 0.1	line pushed **up**

- At optimum, the fraction of observations **below** the line equals τ:

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}[Y_i < X_i'\hat{\beta}_\tau] = \tau$$

Quantile Regression – Recap (Cont.)

Note

Key property – equivariance: for any monotone increasing function $h(\cdot)$:

$$Q_\tau[h(Y) \mid X] = h(Q_\tau[Y \mid X])$$

This is what Powell exploits for censored quantile regression.

Quantile Regression Estimation – Example ($\tau = 0.1$)

- 5 observations: $(X, Y) = (1, 2), (2, 5), (3, 3), (4, 8), (5, 6)$. Minimize $\sum_i \rho_{0.1}(Y_i - \hat{b}X_i)$

			$\hat{b} = 1.0$ **(optimal)**		$\hat{b} = 1.5$	
i	X_i	Y_i	u_i	Penalty	u_i	Penalty
1	1	2	$+1.00$	$0.1 \times 1.00 = 0.10$	$+0.50$	$0.1 \times 0.50 = 0.05$
2	2	5	$+3.00$	$0.1 \times 3.00 = 0.30$	$+2.00$	$0.1 \times 2.00 = 0.20$
3	3	3	0.00	0	-1.50	$0.9 \times 1.50 = 1.35$
4	4	8	$+4.00$	$0.1 \times 4.00 = 0.40$	$+2.00$	$0.1 \times 2.00 = 0.20$
5	5	6	$+1.00$	$0.1 \times 1.00 = 0.10$	-1.50	$0.9 \times 1.50 = 1.35$
Total				**0.90**		**3.15**

Quantile Regression Estimation – Example ($\tau = 0.1$)

Note

$\widehat{b} = 1.0$ wins: all residuals are non-negative – the line sits at the very bottom of the data, leaving 0% of observations below it. $\widehat{b} = 1.5$ is severely penalized because observations 3 and 5 fall below the line, each triggering the costly 0.9 rate.

Powell's Quantile Approach – Relaxing Independence

- **Powell (1984, 1986):** replace independence with a **conditional quantile restriction**:

$$Y^* = q_\tau(X) + e_\tau, \quad Q_\tau[e_\tau \mid X] = 0, \quad Y = \max(Y^*, 0)$$

- By **equivariance of quantiles** to monotone transformations:

$$Q_\tau[Y \mid X = x] = \max(q_\tau(x), 0)$$

- $Q_\tau[Y \mid X = x]$ is identified from data
- Therefore $q_\tau(x)$ is identified for any x such that $q_\tau(x) > 0$

Powell's Quantile Approach – Relaxing Independence (Cont.)

! Important

Key insight: quantiles, not means, are nonparametrically identified from censored distributions – and without requiring $e \perp X$.

Limitation: $q_\tau(x)$ is only identified in the sub-population where censoring does not exceed $\tau\%$.

Censored Quantile Functions – Illustration

- Example:
  $Y^* \mid X \sim N\left(\sqrt{X} - \frac{3}{2},\ 2 + X\right)$
- **Solid lines:** portions of $q_\tau(x)$ above zero (identified from censored data)
- **Dashed lines:** portions below zero (not identified)

- $q_{0.9}(x)$: identified for **all** x
- $q_{0.3}(x)$: identified for **no** x
- $q_{0.7}(x)$, $q_{0.5}(x)$: identified for a **subset** of x

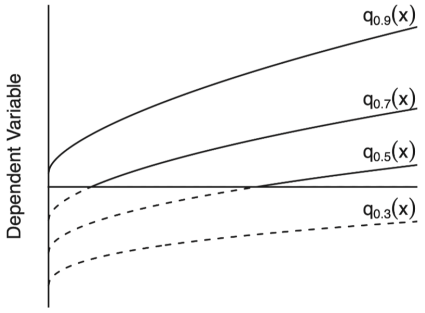

Figure 5: Conditional quantile functions for $\tau = 0.3, 0.5, 0.7, 0.9$, solid above zero and dashed below zero

CLAD and CQR Estimators

Censored LAD – Model and Criterion

- **Powell (1984):** censored median regression. The model:

$$Y^* = X'\beta + e, \quad \text{med}[e \mid X] = 0, \quad Y = \max(Y^*, 0)$$

- By equivariance of the median:

$$\text{med}[Y \mid X] = \max(X'\beta, 0)$$

Equivariance of the Median

- **Equivariance property:** for any monotone increasing function $h(\cdot)$:

$$\text{med}[h(Y^*) \mid X] = h(\text{med}[Y^* \mid X])$$

- $\max(y, 0)$ is monotone increasing in y, so applying it to Y^*:

$$\text{med}[Y \mid X] = \text{med}[\max(Y^*, 0) \mid X] = \max(\text{med}[Y^* \mid X], 0) = \max(X'\beta, 0)$$

- **Why does equivariance hold for medians but not means?**
  - Median is defined by a **ranking condition** – applying $h(\cdot)$ re-ranks observations monotonically, so the 50th percentile point moves with h
  - Mean depends on **actual values** – $\mathbb{E}[\max(Y^*, 0)] \neq \max(\mathbb{E}[Y^*], 0)$ in general

Censored LAD – Model and Criterion (Cont.)

- We have a **parametric but nonlinear** median regression for Y. The appropriate estimator is LAD:

Definition: CLAD Criterion

$$M_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} |Y_i - \max(X_i'\beta, 0)|$$

The **CLAD estimator** is $\hat{\beta}_{\text{CLAD}} = \operatorname{argmin}_\beta M_n(\beta)$.

Censored LAD – R Implementation (1)

```r
# CLAD: minimize (1/n) * sum|Y - max(X'b, 0)|
clad <- function(beta, Y, X) {
  mean(abs(Y - pmax(X %*% beta, 0)))
}

# Data generating process
set.seed(42)
n     <- 500
X     <- cbind(1, rnorm(n))
beta  <- c(1, 2)
Ystar <- X %*% beta + rnorm(n)
Y     <- pmax(Ystar, 0)
```

Censored LAD – R Implementation (2)

```r
# Fraction censored
cat("Censoring proportion:", mean(Y == 0), "\n")
```

```
Censoring proportion: 0.34
```

```r
# CLAD estimation
result <- optim(
  par    = c(0, 0),
  fn     = clad,
  Y      = Y,
  X      = X,
  method = "Nelder-Mead"
)

# OLS for comparison
ols <- lm(Y ~ X[, 2])
```

Censored LAD – R Implementation (3)

```
             Intercept   Slope
True beta:   1 2
CLAD:        1.048 1.931
OLS:         1.443 1.339
```

Censored Quantile Regression – CQR

- **Powell (1986):** generalize CLAD to arbitrary quantiles $\tau \in (0, 1)$.
- Model: $Y^* = X'\beta + e$, $Q_\tau[e \mid X] = 0$, $Y = \max(Y^*, 0)$.
- By equivariance: $Q_\tau[Y \mid X] = \max(X'\beta, 0)$.

Censored Quantile Regression – CQR (Cont.)

Definition: CQR Criterion

$$M_n(\beta;\tau) = \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(Y_i - \max(X_i'\beta, 0))$$

where $\rho_\tau(u) = u(\tau - \mathbf{1}\{u < 0\})$ is the **check function**. The CQR estimator: $\hat{\beta}_{\text{CQR}}(\tau) = \operatorname{argmin}_\beta M_n(\beta;\tau)$.

Note

$\tau = 0.5$ gives CLAD. For general τ, CQR traces the full conditional quantile function of Y^*.

Properties of CLAD and CQR

- Both estimators are **asymptotically normal** (Powell 1984, 1986) – by arguments analogous to quantile regression
- **Identification requirement:** a positive fraction of the population must satisfy $X'\beta > 0$; the relevant design matrix must be full rank in this sub-population

- **Important caveat:** $M_n(\beta)$ and $M_n(\beta;\tau)$ are **not globally convex**
  - Minimization algorithms may converge to a **local minimum**
  - Use multiple starting values in practice

Advantages of CLAD/CQR over Tobit

Property	Tobit	CLAD/CQR
Normality required	Yes	No
Homoskedasticity required	Yes	No
Globally convex criterion	Yes	No
Robust to heavy tails	No	Yes

Illustrating Censored Regression

Application – Transfers and Income

- **Data:** 8,684 Filipino households (CHJ2004). Dependent variable: transfers received (domestic + abroad + in-kind). Censoring proportion: **18%** – large enough to generate significant bias.

- **Specification:** linear spline in income with 5 knots + 15 control regressors. Four estimators compared:

Estimator	Notes
OLS	Full sample, ignores censoring
Tobit	MLE, assumes normality + homoskedasticity
LAD	Median regression, no censoring correction
CLAD	Robust to both censoring **and** non-normality

Application – Results

Key findings:

- All four estimators agree on the **shape**: slope ≈ -1 for low income, flat for high income, sharp break at 20,000 pesos
- But **levels differ substantially:**
  - OLS is several thousand pesos above the others
  - Tobit is shifted down vs. OLS – consistent with censoring bias (censoring probability is increasing in income, so OLS is positively biased)
  - LAD and CLAD similar – large level difference from OLS reflects skewed Y (mean 7,700 vs. median 1,200 pesos)

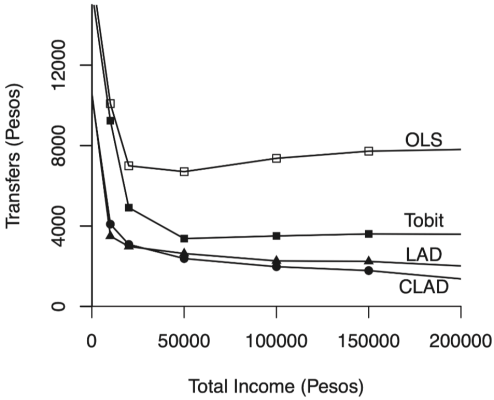

Figure 6: Effect of Income on Transfers

Sample Selection Bias

What is Sample Selection?

- **Censoring:** we observe the variable but it may be constrained
- **Selection:** the variable is simply **not observed** for part of the population – and the missing part is non-random

- **Classic examples:**
  - **Wage regression:** wages are only observed for employed workers – selection on employment may be correlated with unobserved ability
  - **Program evaluation:** participants self-select into training programs – volunteers may differ systematically from non-volunteers
  - **Surveys:** respondents are non-random – those who reply may have systematically different opinions
  - **Ratings:** customers who rate products are self-selected – unhappy customers more likely to leave reviews

What is Sample Selection? (Cont.)

Note

In all cases, the sample is **not a random draw** from the population. OLS on the selected sample estimates $\mathbb{E}[Y \mid X, S = 1]$, not $\mathbb{E}[Y \mid X]$.

The Selection Bias Formula – Why Bias Arises

- OLS on the selected sample estimates $X'\beta$ **plus a bias term** – the error e is not mean-zero in the selected sample:

$$\mathbb{E}[Y \mid X, S = 1] = X'\beta + \underbrace{\mathbb{E}[e \mid u > -X'\gamma]}_{\text{selection bias}}$$

- **Why?** If e and u are correlated – same unobservables drive both Y and selection – this conditional expectation is nonzero

The Selection Bias Formula – Projection Argument

- **Project** e onto u: write $e = \rho u + \varepsilon$ where $\rho = \mathrm{Cov}(e,u)/\mathrm{Var}(u)$ and $\varepsilon \perp\!\!\!\perp u$:

$$\mathbb{E}[e \mid u > -X'\gamma] = \rho\, \mathbb{E}[u \mid u > -X'\gamma] + \mathbb{E}[\varepsilon \mid u > -X'\gamma]$$

- The second term vanishes since $\varepsilon \perp\!\!\!\perp u$, leaving $\rho\, g(X'\gamma)$
- **Under normality** $u \sim N(0,1)$, the truncated mean formula gives $g(X'\gamma) = \lambda(X'\gamma)$:

$$\mathbb{E}[Y \mid X, S = 1] = X'\beta + \rho\, \lambda(X'\gamma) \tag{7}$$

The Selection Bias Formula – Interpreting Each Term

Term	Interpretation
$X'\beta$	structural effect of X on Y – what we want
ρ	correlation between outcome error e and selection error u – if $\rho = 0$, no bias
$\lambda(X'\gamma)$	how selective the sample is at each X – large λ means severe selection
$\rho\,\lambda(X'\gamma)$	omitted variable bias – what OLS attributes to β but is due to selection

Note

OLS omits $\rho\,\lambda(X'\gamma)$. Since $\lambda(X'\gamma)$ is correlated with X, this is omitted variable bias. Heckman's two-step fixes this by estimating $\lambda(X'\hat{\gamma})$ in a first stage and including it as a control in the second stage.

Heckman's Model

Heckman's Setup

- **Heckman (1979):** selection bias can be corrected if we observe **both** selected and non-selected units (i.e., we observe S for everyone, even when Y is missing).

- The model (observe $\{Y_i, X_i, Z_i\}$ for a random sample):

$$Y^* = X'\beta + e$$

$$S^* = Z'\gamma + u, \quad S = \mathbf{1}\{S^* > 0\}, \quad Y = \begin{cases} Y^* & \text{if } S = 1 \\ \text{missing} & \text{if } S = 0 \end{cases}$$

$$\begin{pmatrix} e \\ u \end{pmatrix} \sim N\left(0, \begin{pmatrix} \sigma^2 & \sigma_{21} \\ \sigma_{21} & 1 \end{pmatrix}\right)$$

- Variance of u normalized to 1 (not separately identified)
- σ_{21} is the key covariance driving selection bias

Heckman's Model – Classic Example

- Y^*: potential log-wage (what a person would earn if employed)
- S^*: latent utility of employment; $S = 1$ means the person is employed
- Y: observed wage (missing for the non-employed)

Object	Interpretation
β	Wage equation coefficients (returns to education, experience)
γ	Employment equation coefficients (e.g., non-labor income, family structure)
e	Unobserved ability and other wage determinants
u	Unobserved factors in employment decision
σ_{21}	Correlation of unobserved ability with employment propensity

The Selected-Sample CEF under Heckman's Model

- Under joint normality of (e, u), the conditional expectation of Y given selection:

$$\mathbb{E}[Y \mid X, Z, S = 1] = X'\beta + \sigma_{21}\,\lambda(Z'\gamma) \tag{8}$$

- The selection correction term $\sigma_{21}\lambda(Z'\gamma)$ is an **omitted variable** in the naive OLS
- If $\sigma_{21} = 0$: selection is exogenous and OLS is consistent
- Testing $H_0 : \sigma_{21} = 0$ provides a formal test of exogenous selection

Note

Compare (8) to the truncated mean (4): same mathematical structure! The Tobit model is the special case where $\gamma = \beta/\sigma$ and $\sigma_{21} = \sigma$.

Heckman Two-Step Estimator (Heckit)

Step 1: Estimate the **selection equation** by probit:

$$S_i \sim \text{Probit}(Z_i'\gamma) \quad \Rightarrow \quad \hat{\gamma}$$

Step 2: Construct the **inverse Mills ratio** for each selected observation:

$$\hat{\lambda}_i = \lambda(Z_i'\hat{\gamma}) = \frac{\phi(Z_i'\hat{\gamma})}{\Phi(Z_i'\hat{\gamma})}$$

Step 3: Run **OLS** of Y_i on $(X_i, \hat{\lambda}_i)$ using only the selected sub-sample $(S_i = 1)$:

$$Y_i = X_i'\beta + \sigma_{21}\hat{\lambda}_i + \text{residual}$$

Heckman Two-Step Estimator (Heckit) (Cont.)

Note

$\widehat{\lambda}_i$ is a **generated regressor** (Section 12.26 in Hansen) – conventional standard errors are inconsistent. Use the Heckman (1979) corrected covariance formula, or bootstrap the full two-step procedure.

Heckman – Identification and Exclusion Restriction

- **Key identification requirement:** Z should contain at least one variable that:
  1. **Affects selection** S^* (i.e., appears in $Z'\gamma$ with non-zero coefficient)
  2. **Does not directly affect** Y^* (i.e., excluded from $X'\beta$)
- This is an **exclusion restriction** – the same concept as in IV regression.
- **Why?** Without an exclusion restriction, identification relies entirely on the functional form of $\lambda(\cdot)$ (its nonlinearity). This is weak in practice – high collinearity between X and $\lambda(X'\hat{\gamma})$.

! Important

Classic example (wage equation): non-labor income or number of young children affect the employment decision (S^*) but arguably not the wage offer (Y^*). These serve as exclusion restrictions.

Heckman – MLE vs. Two-Step

Full MLE: maximize the joint log-likelihood:

$$\ell_n = \sum_{S_i=0} \log \Phi(-Z_i'\gamma) + \sum_{S_i=1} \left[\log \Phi\left(\frac{Z_i'\gamma + \frac{\sigma_{21}}{\sigma^2}(Y_i - X_i'\beta)}{\sqrt{1 - \sigma_{21}/\sigma^2}} \right) - \frac{1}{2}\log(2\pi\sigma^2) - \frac{(Y_i - X_i'\beta)^2}{2\sigma^2} \right]$$

	Two-Step (Heckit)	**Full MLE**
Efficiency	Less efficient	Asymptotically efficient
Computation	Simple	More demanding
Preferred for	Preliminary analysis	Final reporting
SEs	Need correction	Automatic

Heckman MLE – Deriving the Joint Density (1) (Optional)

- We need the joint density $f(s, y \mid x, z)$ for both selected and non-selected observations
- **Non-selected ($S = 0$):** Y is missing, only $S = 0$ is observed:

$$\Pr[S = 0 \mid x, z] = \Pr[Z'\gamma + u \leq 0] = 1 - \Phi(Z'\gamma)$$

- **Selected ($S = 1$):** both $S = 1$ and $Y = y$ are observed. Factor the joint density:

$$f(y, S = 1 \mid x, z) = \int_0^\infty f(y, s^* \mid x, z)\, ds^*$$

- Factor inside the integral using the conditional:

$$= \int_0^\infty f(s^* \mid y, x, z)\, f(y \mid x, z)\, ds^* = \Pr[S^* > 0 \mid y, x, z] \cdot f(y \mid x, z)$$

Heckman MLE – Deriving the Joint Density (2) (Optional)

- The marginal density of Y is normal: $f(y \mid x, z) = \sigma^{-1}\phi\left(\frac{y - x'\beta}{\sigma}\right)$
- Need the conditional distribution of $S^* = Z'\gamma + u$ given $Y = y$
- Since (e, u) are jointly normal, $u \mid e$ is normal:

$$u \mid e \sim N\left(\frac{\sigma_{21}}{\sigma^2}e,\ 1 - \frac{\sigma_{21}^2}{\sigma^2}\right)$$

Heckman MLE – Deriving the Joint Density (3) (Optional)

- Given that $e = y - x'\beta$ when $Y = y$, substituting:

$$S^* \mid Y = y \sim N\left(Z'\gamma + \frac{\sigma_{21}}{\sigma^2}(y - x'\beta),\ 1 - \frac{\sigma_{21}^2}{\sigma^2}\right)$$

- Therefore:

$$\Pr[S^* > 0 \mid y, x, z] = \Phi\left(\frac{Z'\gamma + \frac{\sigma_{21}}{\sigma^2}(y - x'\beta)}{\sqrt{1 - \sigma_{21}^2/\sigma^2}}\right)$$

- Combining both cases into the joint density:

$$f(s, y \mid x, z) = [1 - \Phi(Z'\gamma)]^{1-s}\left[\Phi\left(\frac{Z'\gamma + \frac{\sigma_{21}}{\sigma^2}(y - x'\beta)}{\sqrt{1 - \sigma_{21}^2/\sigma^2}}\right) \cdot \frac{1}{\sigma}\phi\left(\frac{y - x'\beta}{\sigma}\right)\right]^s$$

Nonparametric Selection

Relaxing Normality in the Selection Model

- Heckman's model assumes joint normality of (e, u). What if this fails?

- **Nonparametric selection model** (unknown joint distribution of (e, u)):

$$Y^* = m(X) + e, \quad S^* = g(Z) + u, \quad S = \mathbf{1}\{S^* > 0\}, \quad Y = Y^* \text{ if } S = 1$$

- Normalize $u \sim N(0, 1)$ (not identified separately from g)
- **Propensity score:** $p(Z) = \Pr[S = 1 \mid Z] = \Phi(Z'\gamma)$ – nonparametrically identified

- The CEF of Y given selection takes three equivalent forms:

$$\mathbb{E}[Y \mid X, Z, S = 1] = X'\beta + h_1(Z'\gamma) = X'\beta + h_2(p(Z)) = X'\beta + h_3(\lambda(Z'\gamma))$$

- h_1, h_2, h_3 are **unknown functions** to be estimated nonparametrically (series expansion)

Two-Step Nonparametric Estimation

Each representation suggests a two-step estimator, all sharing the same first step:

- **Step 1:** Estimate $\hat{\gamma}$ by probit regression of S on Z.
- **Step 2 (three variants):**
  - **Based on $Z'\hat{\gamma}$:** regress Y on X and a polynomial in $Z_i'\hat{\gamma}$
  - **Based on propensity score $\hat{p}$:** regress Y on X and a polynomial in $\hat{p}_i = \Phi(Z_i'\hat{\gamma})$
    - Most interpretable: regression adjusts for selection probability directly
  - **Based on $\widehat{\lambda}$:** regress Y on X and a polynomial in $\widehat{\lambda}_i = \lambda(Z_i'\hat{\gamma})$
    - First-order accurate under near-normality
- **Das, Newey, and Vella (2003):** complete asymptotic theory for this class.

Two-Step Nonparametric Estimation (Cont.)

Note

All three are **generated-regressor** two-step estimators. Use bootstrap (full two-step) for valid standard errors.

Summary

Summary (1)

Problem	Structure	Estimator	Key Assumption
Censoring	$Y = \max(Y^*, 0)$, observe Y for all i	Tobit (MLE)	Normality + $e \perp X$
Censoring	$Y = \max(Y^*, 0)$, observe Y for all i	CLAD/CQR	No normality, no homoskedasticity
Selection	Y missing when $S = 0$; observe S always	Heckman (Heckit or MLE)	Normality + exclusion restriction
Selection	Y missing when $S = 0$	Nonparametric (Das et al.)	Exclusion restriction; unknown F

Summary (2)

Note

Practical recommendation: always start by computing the censoring proportion $\hat{\pi}$. If $\hat{\pi} > 5$–10%, OLS is likely to be substantially biased. When in doubt, compare Tobit and CLAD – if they agree, you are in good shape; if not, trust CLAD.

Required Reading

- Hansen, B. (2022). *Econometrics*. Princeton University Press. **Chapter 27: Censoring and Selection** (pp. 853–869).

- **Optional:**
  - Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, 26(1), 24–36.
  - Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica*, 47(1), 153–161.
  - Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, 25(3), 303–325.
  - Powell, J. L. (1986). Censored regression quantiles. *Journal of Econometrics*, 32(1), 143–155.