Qualitative and Limited Dependent Variables

An Overview of Models for Non-Continuous Outcomes

Jake Anderson

March 3, 2026

Outline

The Problem: Non-Continuous Outcomes

Everything so far assumes y is continuous and unbounded. But many economic outcomes are not:

- **Binary:** work or not, default or not, buy or not
- **Unordered categories:** car / bus / train / bike
- **Ordered categories:** strongly disagree $\rightarrow$ strongly agree
- **Counts:** doctor visits, patents filed, arrests
- **Censored:** hours worked (piled up at zero)

The Problem: Non-Continuous Outcomes

Everything so far assumes y is continuous and unbounded. But many economic outcomes are not:

- **Binary:** work or not, default or not, buy or not
- **Unordered categories:** car / bus / train / bike
- **Ordered categories:** strongly disagree $\rightarrow$ strongly agree
- **Counts:** doctor visits, patents filed, arrests
- **Censored:** hours worked (piled up at zero)

$\Longrightarrow$ OLS is the wrong tool for all of these. This deck introduces the right ones.

OLS on a Binary Outcome: The Linear Probability Model

Suppose $y \in \{0, 1\}$ (e.g., drives to work or not). If we run OLS:

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

OLS on a Binary Outcome: The Linear Probability Model

Suppose $y \in \{0, 1\}$ (e.g., drives to work or not). If we run OLS:

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

This is the **Linear Probability Model (LPM)**. The fitted value $\hat{y}$ is interpreted as a probability. But there are problems:

OLS on a Binary Outcome: The Linear Probability Model

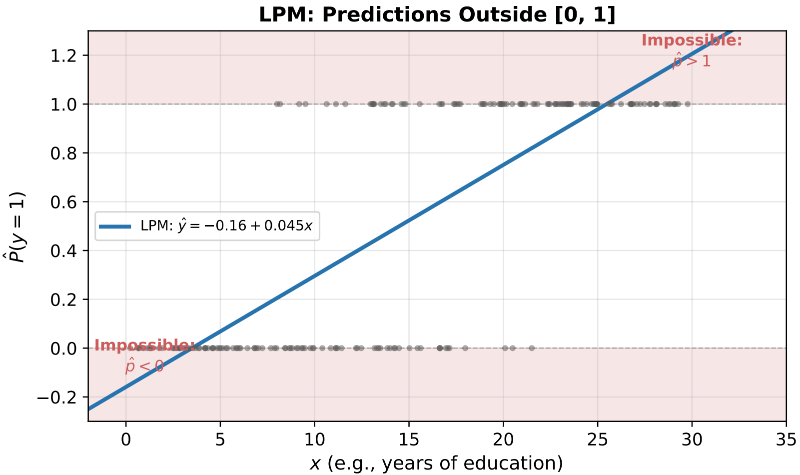
Suppose $y \in \{0, 1\}$ (e.g., drives to work or not). If we run OLS:

$$y_i = \beta_0 + \beta_1 x_i + e_i$$

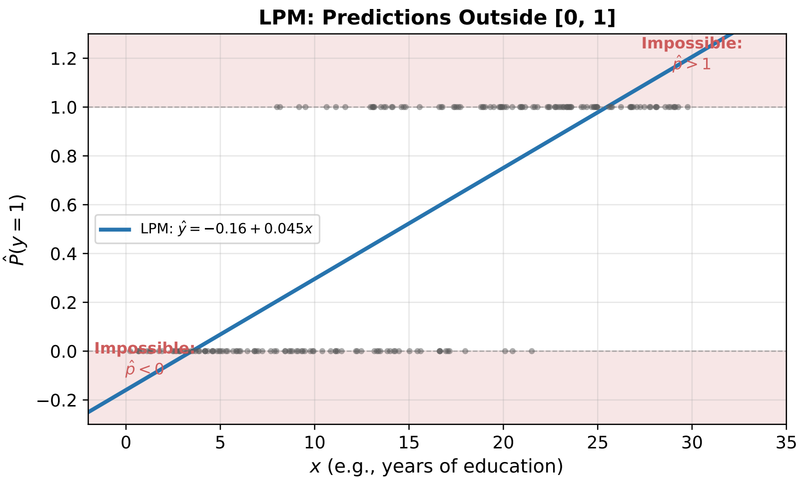
This is the **Linear Probability Model (LPM)**. The fitted value $\hat{y}$ is interpreted as a probability. But there are problems:

1. **Predictions outside** $[0, 1]$**:** OLS can predict $\hat{p} = -0.3$ or $\hat{p} = 1.4$
2. **Heteroskedasticity:** $\text{Var}(y \mid x) = p(1 - p)$ depends on x
3. **Constant marginal effects:** a one-unit change in x always changes probability by β_1, but probabilities are bounded

LPM: Predictions Outside $[0, 1]$



LPM: Predictions Outside $[0, 1]$



$\implies$ The LPM's linear structure cannot respect the $[0, 1]$ bounds. We need a function that maps $x'\beta$ into $[0, 1]$.

The Latent Variable Framework

Many binary outcomes reflect an underlying continuous quantity we cannot observe. Call it y^*:

$$y_i^* = x_i'\beta + e_i$$

The Latent Variable Framework

Many binary outcomes reflect an underlying continuous quantity we cannot observe. Call it y^*:

$$y_i^* = x_i' \beta + e_i$$

We observe:

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

The Latent Variable Framework

Many binary outcomes reflect an underlying continuous quantity we cannot observe. Call it y^*:

$$y_i^* = x_i'\beta + e_i$$

We observe:

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

The probability of $y = 1$ depends on the distribution of e_i:

$$P(y_i = 1) = P(e_i > -x_i'\beta) = 1 - F(-x_i'\beta)$$

The Latent Variable Framework

Many binary outcomes reflect an underlying continuous quantity we cannot observe. Call it y^*:

$$y_i^* = x_i'\beta + e_i$$

We observe:

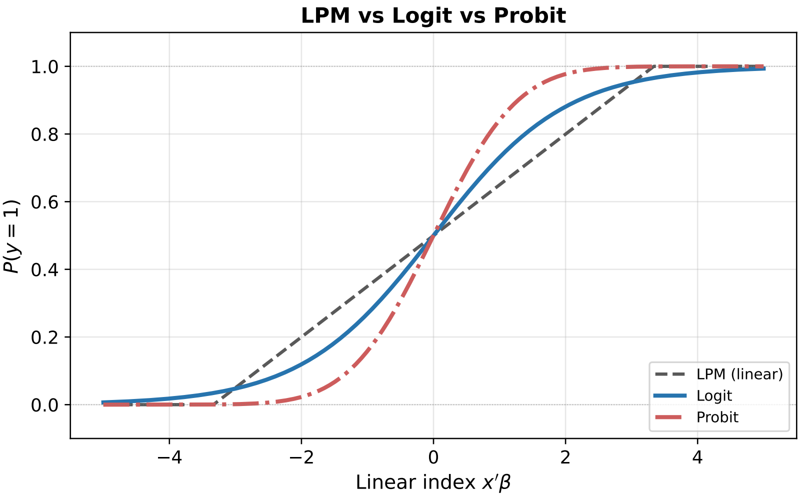
$$y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* \leq 0 \end{cases}$$

The probability of $y = 1$ depends on the distribution of e_i:

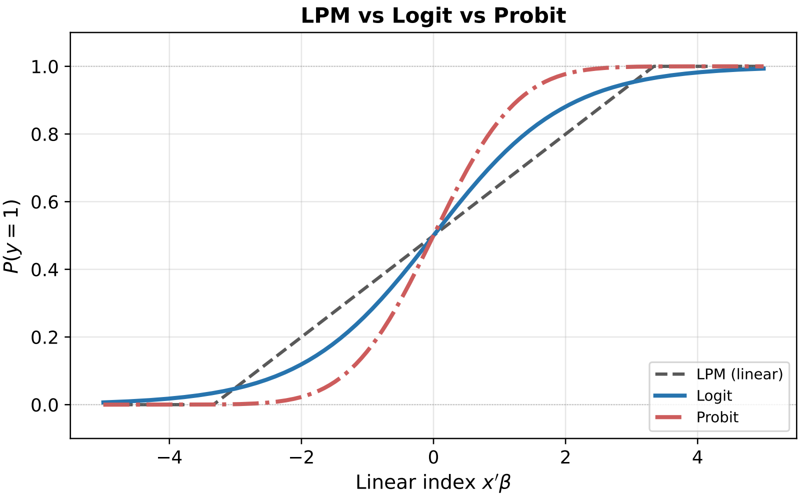
$$P(y_i = 1) = P(e_i > -x_i'\beta) = 1 - F(-x_i'\beta)$$

- If $e_i \sim N(0,1)$: $P(y = 1) = \Phi(x'\beta) \implies$ **Probit**
- If $e_i \sim$ Logistic: $P(y = 1) = \Lambda(x'\beta) = \dfrac{e^{x'\beta}}{1 + e^{x'\beta}} \implies$ **Logit**

The S-Curve: Logit and Probit vs LPM



The S-Curve: Logit and Probit vs LPM



Both logit and probit guarantee $\hat{p} \in [0, 1]$. The two S-curves are nearly identical in practice. Logit has

Estimation: Maximum Likelihood

We cannot use OLS for probit/logit. Instead, we use **Maximum Likelihood Estimation (MLE)**: find the β that makes the observed data least surprising.

Estimation: Maximum Likelihood

We cannot use OLS for probit/logit. Instead, we use **Maximum Likelihood Estimation (MLE)**: find the β that makes the observed data least surprising.

For each observation:

$$f(y_i) = [\Phi(x_i'\beta)]^{y_i}\ [1 - \Phi(x_i'\beta)]^{1-y_i}$$

Estimation: Maximum Likelihood

We cannot use OLS for probit/logit. Instead, we use **Maximum Likelihood Estimation (MLE)**: find the β that makes the observed data least surprising.

For each observation:

$$f(y_i) = [\Phi(x_i'\beta)]^{y_i}\ [1 - \Phi(x_i'\beta)]^{1-y_i}$$

Log-likelihood for the whole sample:

$$\ln L = \sum_{i=1}^{N} \left[y_i \ln \Phi(x_i'\beta) + (1 - y_i) \ln(1 - \Phi(x_i'\beta)) \right]$$

Estimation: Maximum Likelihood

We cannot use OLS for probit/logit. Instead, we use **Maximum Likelihood Estimation (MLE)**: find the β that makes the observed data least surprising.

For each observation:

$$f(y_i) = [\Phi(x_i'\beta)]^{y_i}\,[1 - \Phi(x_i'\beta)]^{1-y_i}$$

Log-likelihood for the whole sample:

$$\ln L = \sum_{i=1}^{N} \left[y_i \ln \Phi(x_i'\beta) + (1 - y_i) \ln(1 - \Phi(x_i'\beta)) \right]$$

$\implies$ MLE picks the β that maximizes this. In large samples, MLE is consistent, asymptotically normal, and efficient.

Marginal Effects: Why Coefficients Are Not Enough

In probit/logit, the coefficient β_k is **not** the marginal effect.

Marginal Effects: Why Coefficients Are Not Enough

In probit/logit, the coefficient β_k is **not** the marginal effect.

Probit:

$$\frac{\partial P}{\partial x_k} = \phi(x'\beta) \cdot \beta_k$$

Marginal Effects: Why Coefficients Are Not Enough

In probit/logit, the coefficient β_k is **not** the marginal effect.

Probit:

$$\frac{\partial P}{\partial x_k} = \phi(x'\beta) \cdot \beta_k$$

Logit:

$$\frac{\partial P}{\partial x_k} = \Lambda(x'\beta)(1 - \Lambda(x'\beta)) \cdot \beta_k$$

Marginal Effects: Why Coefficients Are Not Enough

In probit/logit, the coefficient β_k is **not** the marginal effect.

Probit:

$$\frac{\partial P}{\partial x_k} = \phi(x'\beta) \cdot \beta_k$$

Logit:

$$\frac{\partial P}{\partial x_k} = \Lambda(x'\beta)(1 - \Lambda(x'\beta)) \cdot \beta_k$$

$\Longrightarrow$ The marginal effect depends on where you are on the S-curve:

- Near $p = 0.5$ (middle): large effect
- Near $p = 0$ or $p = 1$ (tails): small effect

Marginal Effects: Why Coefficients Are Not Enough

In probit/logit, the coefficient β_k is **not** the marginal effect.

Probit:

$$\frac{\partial P}{\partial x_k} = \phi(x'\beta) \cdot \beta_k$$

Logit:

$$\frac{\partial P}{\partial x_k} = \Lambda(x'\beta)(1 - \Lambda(x'\beta)) \cdot \beta_k$$

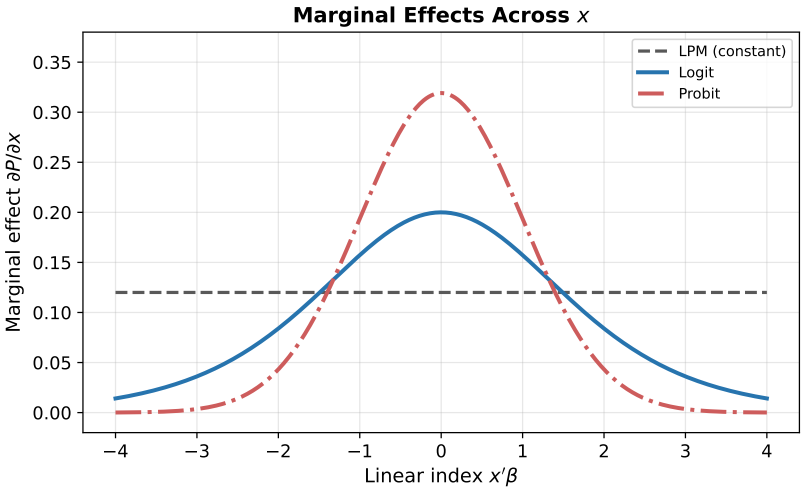
$\Longrightarrow$ The marginal effect depends on where you are on the S-curve:

- Near $p = 0.5$ (middle): large effect
- Near $p = 0$ or $p = 1$ (tails): small effect

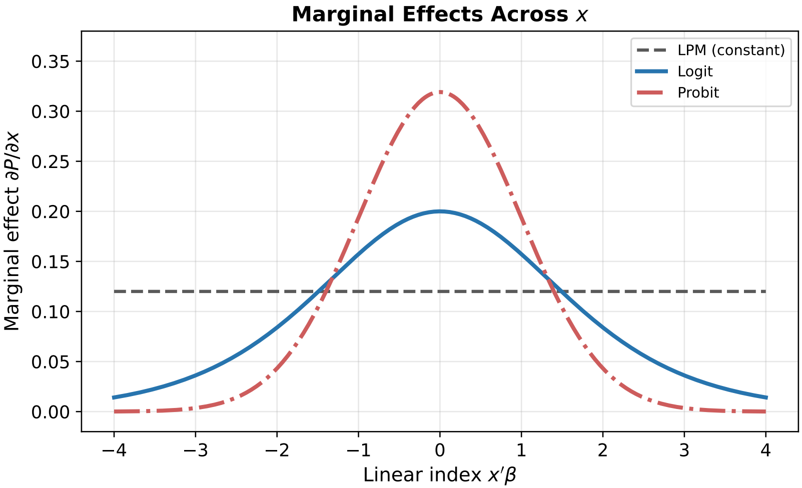
Common practice: report the **Average Marginal Effect (AME)**:

$$\widehat{AME} = \frac{1}{N} \sum_{i=1}^{N} \phi(\hat{\beta}_0 + \hat{\beta}_1 x_i) \cdot \hat{\beta}_1$$

Marginal Effects: LPM vs Logit/Probit



Marginal Effects: LPM vs Logit/Probit



The LPM assumes the same marginal effect everywhere. Logit/probit capture the fact that a change in

Comparing LPM, Probit, and Logit

Coefficient scaling (approximate):

$$\hat{\beta}_{\text{Logit}} \approx 4\,\hat{\beta}_{\text{LPM}}, \qquad \hat{\beta}_{\text{Probit}} \approx 2.5\,\hat{\beta}_{\text{LPM}}, \qquad \hat{\beta}_{\text{Logit}} \approx 1.6\,\hat{\beta}_{\text{Probit}}$$

Comparing LPM, Probit, and Logit

Coefficient scaling (approximate):

$$\hat{\beta}_{\text{Logit}} \approx 4\,\hat{\beta}_{\text{LPM}}, \qquad \hat{\beta}_{\text{Probit}} \approx 2.5\,\hat{\beta}_{\text{LPM}}, \qquad \hat{\beta}_{\text{Logit}} \approx 1.6\,\hat{\beta}_{\text{Probit}}$$

	LPM	**Probit**	**Logit**
Estimation	OLS	MLE	MLE
$\hat{p} \in [0, 1]$?	No	Yes	Yes
Marginal effects	Constant	Vary with x	Vary with x
Interpretation	Direct	Via ϕ	Via odds ratio

Comparing LPM, Probit, and Logit

Coefficient scaling (approximate):

$$\hat{\beta}_{\text{Logit}} \approx 4\,\hat{\beta}_{\text{LPM}}, \qquad \hat{\beta}_{\text{Probit}} \approx 2.5\,\hat{\beta}_{\text{LPM}}, \qquad \hat{\beta}_{\text{Logit}} \approx 1.6\,\hat{\beta}_{\text{Probit}}$$

	LPM	**Probit**	**Logit**
Estimation	OLS	MLE	MLE
$\hat{p} \in [0,1]$?	No	Yes	Yes
Marginal effects	Constant	Vary with x	Vary with x
Interpretation	Direct	Via ϕ	Via odds ratio

$\Longrightarrow$ In practice, all three give similar predicted probabilities and AMEs. Use probit/logit when you need predictions in $[0,1]$; use LPM as a quick baseline.

More Than Two Choices

What if the dependent variable has three or more **unordered** categories?

- Transportation mode: car, bus, train, bike
- College choice: no college, 2-year, 4-year
- Insurance: none, public, public + add-on

More Than Two Choices

What if the dependent variable has three or more **unordered** categories?

- Transportation mode: car, bus, train, bike
- College choice: no college, 2-year, 4-year
- Insurance: none, public, public + add-on

The **multinomial logit** extends binary logit to J categories. With one category as the base (say $j = 1$):

$$P(y_i = j \mid x_i) = \frac{e^{x_i' \beta_j}}{\sum_{k=1}^{J} e^{x_i' \beta_k}}$$

More Than Two Choices

What if the dependent variable has three or more **unordered** categories?

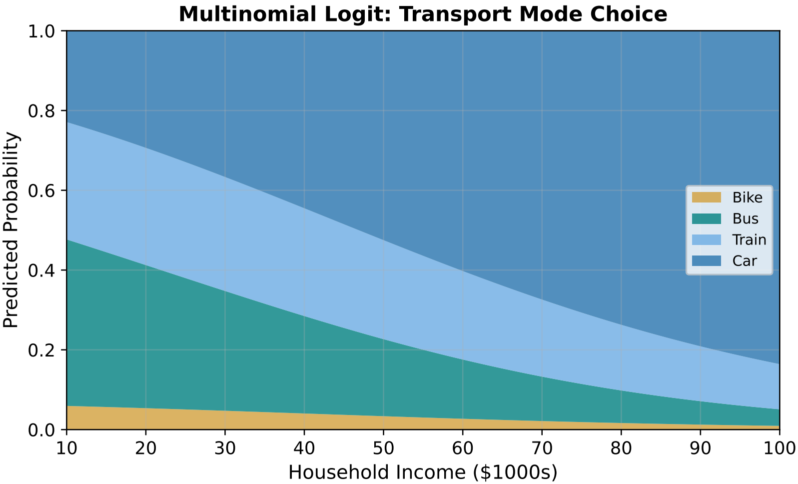
- Transportation mode: car, bus, train, bike
- College choice: no college, 2-year, 4-year
- Insurance: none, public, public + add-on

The **multinomial logit** extends binary logit to J categories. With one category as the base (say $j = 1$):

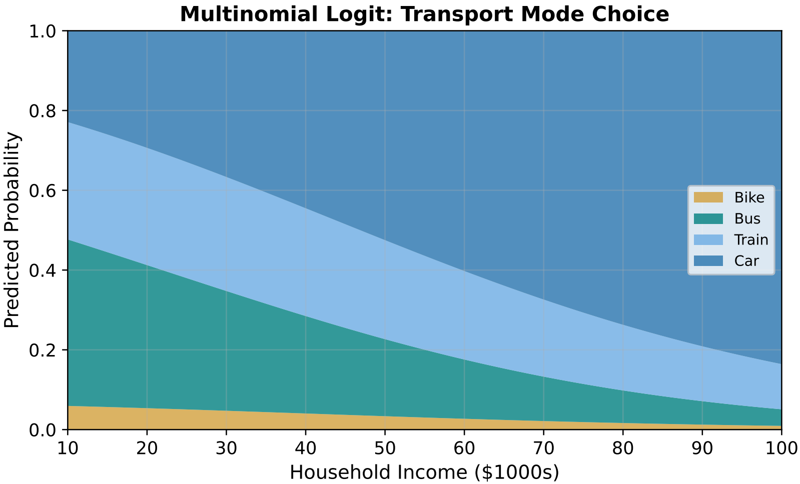
$$P(y_i = j \mid x_i) = \frac{e^{x_i' \beta_j}}{\sum_{k=1}^{J} e^{x_i' \beta_k}}$$

- Estimate $J - 1$ sets of coefficients (one per non-base category)
- Coefficients show the effect on the log-odds relative to the base
- Marginal effects are not the raw coefficients

Multinomial Logit: Predicted Probabilities



Multinomial Logit: Predicted Probabilities



As income rises, predicted choice shares shift from bus/bike toward car. The probabilities always sum to 1 across alternatives.

Independence of Irrelevant Alternatives (IIA)

Multinomial logit assumes the ratio of probabilities for any two choices does not depend on what other alternatives are available.

Independence of Irrelevant Alternatives (IIA)

Multinomial logit assumes the ratio of probabilities for any two choices does not depend on what other alternatives are available.

The red bus / blue bus problem:

- Initially: $P(\text{car}) = 0.5$, $P(\text{red bus}) = 0.5$
- Add an identical blue bus

Independence of Irrelevant Alternatives (IIA)

Multinomial logit assumes the ratio of probabilities for any two choices does not depend on what other alternatives are available.

The red bus / blue bus problem:

- Initially: $P(\text{car}) = 0.5$, $P(\text{red bus}) = 0.5$
- Add an identical blue bus

IIA predicts: $P(\text{car}) = P(\text{red bus}) = P(\text{blue bus}) = 0.33$

Independence of Irrelevant Alternatives (IIA)

Multinomial logit assumes the ratio of probabilities for any two choices does not depend on what other alternatives are available.

The red bus / blue bus problem:

- Initially: $P(\text{car}) = 0.5$, $P(\text{red bus}) = 0.5$
- Add an identical blue bus

IIA predicts: $P(\text{car}) = P(\text{red bus}) = P(\text{blue bus}) = 0.33$
But realistically: $P(\text{car}) = 0.5$, $P(\text{red bus}) = P(\text{blue bus}) = 0.25$

Independence of Irrelevant Alternatives (IIA)

Multinomial logit assumes the ratio of probabilities for any two choices does not depend on what other alternatives are available.

The red bus / blue bus problem:

- Initially: $P(\text{car}) = 0.5$, $P(\text{red bus}) = 0.5$
- Add an identical blue bus

IIA predicts: $P(\text{car}) = P(\text{red bus}) = P(\text{blue bus}) = 0.33$
But realistically: $P(\text{car}) = 0.5$, $P(\text{red bus}) = P(\text{blue bus}) = 0.25$

$\Longrightarrow$ Adding a clone of an existing option should not steal share from a completely different option.
Test IIA with the Hausman-McFadden test; if it fails, consider nested logit or mixed logit.

When Categories Have a Natural Ranking

Sometimes the categories are ordered but the distances between them are unknown:

- Survey responses: strongly disagree $\rightarrow$ strongly agree
- Health satisfaction: low, medium, high
- Bond ratings: AAA, AA, A, BBB, ...

When Categories Have a Natural Ranking

Sometimes the categories are ordered but the distances between them are unknown:

- Survey responses: strongly disagree $\rightarrow$ strongly agree
- Health satisfaction: low, medium, high
- Bond ratings: AAA, AA, A, BBB, ...

The **ordered probit/logit** model assumes an underlying latent variable y^*:

$$y_i^* = x_i'\beta + e_i$$

When Categories Have a Natural Ranking

Sometimes the categories are ordered but the distances between them are unknown:

- Survey responses: strongly disagree $\rightarrow$ strongly agree
- Health satisfaction: low, medium, high
- Bond ratings: AAA, AA, A, BBB, ...

The **ordered probit/logit** model assumes an underlying latent variable y^*:

$$y_i^* = x_i'\beta + e_i$$

The observed outcome depends on where y^* falls relative to threshold parameters (**cutpoints**) $\mu_1, \mu_2, \ldots$:

$$y_i = \begin{cases} 1 & \text{if } y_i^* \leq \mu_1 \\ 2 & \text{if } \mu_1 < y_i^* \leq \mu_2 \\ 3 & \text{if } y_i^* > \mu_2 \end{cases}$$

When Categories Have a Natural Ranking

Sometimes the categories are ordered but the distances between them are unknown:

- Survey responses: strongly disagree → strongly agree
- Health satisfaction: low, medium, high
- Bond ratings: AAA, AA, A, BBB, . . .

The **ordered probit/logit** model assumes an underlying latent variable y^*:

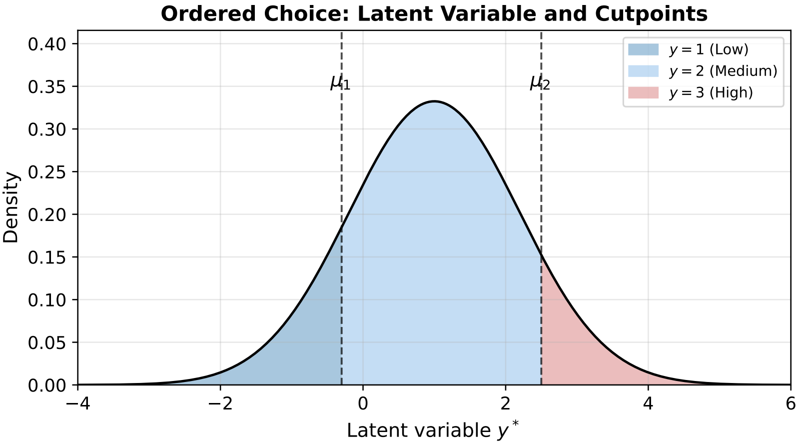
$$y_i^* = x_i'\beta + e_i$$

The observed outcome depends on where y^* falls relative to threshold parameters (**cutpoints**) $\mu_1, \mu_2, \ldots$:

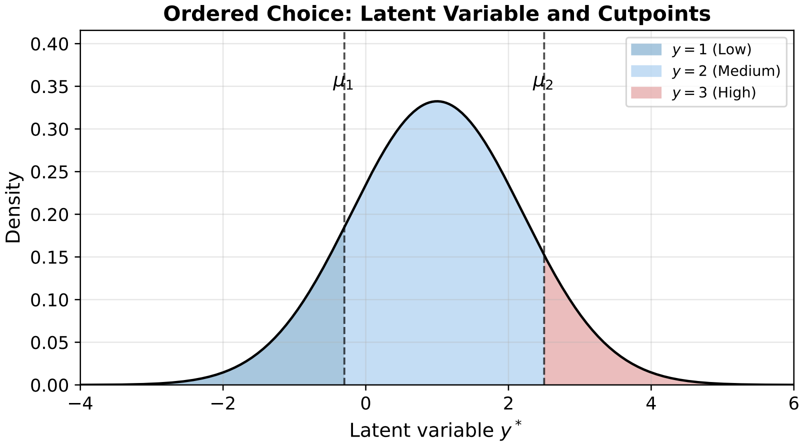
$$y_i = \begin{cases} 1 & \text{if } y_i^* \leq \mu_1 \\ 2 & \text{if } \mu_1 < y_i^* \leq \mu_2 \\ 3 & \text{if } y_i^* > \mu_2 \end{cases}$$

The cutpoints μ are estimated along with β.

Ordered Choice: The Latent Variable



Ordered Choice: The Latent Variable



A change in x shifts the entire distribution of y^*, simultaneously changing the probability of every category. The marginal effects must sum to zero across categories.

Ordered Choice: Marginal Effects

For a continuous variable in a three-category model:

Ordered Choice: Marginal Effects

For a continuous variable in a three-category model:

$$\frac{\partial P(y=1)}{\partial x_k} = -\phi(\mu_1 - x'\beta) \cdot \beta_k$$

$$\frac{\partial P(y=2)}{\partial x_k} = [\phi(\mu_1 - x'\beta) - \phi(\mu_2 - x'\beta)] \cdot \beta_k$$

$$\frac{\partial P(y=3)}{\partial x_k} = \phi(\mu_2 - x'\beta) \cdot \beta_k$$

Ordered Choice: Marginal Effects

For a continuous variable in a three-category model:

$$\frac{\partial P(y=1)}{\partial x_k} = -\phi(\mu_1 - x'\beta) \cdot \beta_k$$

$$\frac{\partial P(y=2)}{\partial x_k} = [\phi(\mu_1 - x'\beta) - \phi(\mu_2 - x'\beta)] \cdot \beta_k$$

$$\frac{\partial P(y=3)}{\partial x_k} = \phi(\mu_2 - x'\beta) \cdot \beta_k$$

$\implies$ The sign of β_k tells you the direction for the highest and lowest categories, but the middle categories could go either way.

Ordered Choice: Marginal Effects

For a continuous variable in a three-category model:

$$\frac{\partial P(y=1)}{\partial x_k} = -\phi(\mu_1 - x'\beta) \cdot \beta_k$$

$$\frac{\partial P(y=2)}{\partial x_k} = [\phi(\mu_1 - x'\beta) - \phi(\mu_2 - x'\beta)] \cdot \beta_k$$

$$\frac{\partial P(y=3)}{\partial x_k} = \phi(\mu_2 - x'\beta) \cdot \beta_k$$

$\implies$ The sign of β_k tells you the direction for the highest and lowest categories, but the middle categories could go either way.

For binary variables: compute the **discrete difference** (change in each category's probability when the dummy goes from 0 to 1).

Counts: Non-Negative Integers

Some outcomes are counts: doctor visits, patents, arrests. Counts are non-negative integers, often right-skewed with many zeros.

Counts: Non-Negative Integers

Some outcomes are counts: doctor visits, patents, arrests. Counts are non-negative integers, often right-skewed with many zeros.

The **Poisson model** assumes:

$$P(Y = y) = \frac{e^{-\mu}\,\mu^{y}}{y!}, \qquad y = 0, 1, 2, \ldots$$

where $\mu = E(Y) = \mathrm{Var}(Y)$.

Counts: Non-Negative Integers

Some outcomes are counts: doctor visits, patents, arrests. Counts are non-negative integers, often right-skewed with many zeros.

The **Poisson model** assumes:

$$P(Y = y) = \frac{e^{-\mu} \mu^y}{y!}, \qquad y = 0, 1, 2, \ldots$$

where $\mu = E(Y) = \text{Var}(Y)$.

We model the conditional mean as:

$$\mu_i = \exp(x_i' \beta)$$

Counts: Non-Negative Integers

Some outcomes are counts: doctor visits, patents, arrests. Counts are non-negative integers, often right-skewed with many zeros.

The **Poisson model** assumes:

$$P(Y = y) = \frac{e^{-\mu} \mu^y}{y!}, \qquad y = 0, 1, 2, \dots$$

where $\mu = E(Y) = \mathrm{Var}(Y)$.

We model the conditional mean as:

$$\mu_i = \exp(x_i' \beta)$$

$\Longrightarrow$ The exponential ensures $\mu > 0$. A one-unit increase in x_k multiplies the expected count by e^{β_k}.

The Overdispersion Problem

Poisson assumes $E(Y) = \text{Var}(Y)$ (**equidispersion**). In real data, the variance almost always exceeds the mean (**overdispersion**).

The Overdispersion Problem

Poisson assumes $E(Y) = \text{Var}(Y)$ (**equidispersion**). In real data, the variance almost always exceeds the mean (**overdispersion**).

If overdispersion is present:

- Poisson coefficient estimates are still **consistent**
- But standard errors are **too small**
- Hypothesis tests become unreliable

The Overdispersion Problem

Poisson assumes $E(Y) = \text{Var}(Y)$ (**equidispersion**). In real data, the variance almost always exceeds the mean (**overdispersion**).

If overdispersion is present:

- Poisson coefficient estimates are still **consistent**
- But standard errors are **too small**
- Hypothesis tests become unreliable

The **negative binomial model** relaxes equidispersion:

$$\text{Var}(Y) = \mu + \alpha\mu^2$$

The Overdispersion Problem

Poisson assumes $E(Y) = \text{Var}(Y)$ (**equidispersion**). In real data, the variance almost always exceeds the mean (**overdispersion**).

If overdispersion is present:

- Poisson coefficient estimates are still **consistent**
- But standard errors are **too small**
- Hypothesis tests become unreliable

The **negative binomial model** relaxes equidispersion:

$$\text{Var}(Y) = \mu + \alpha\mu^2$$

When $\alpha = 0$: reduces to Poisson. When $\alpha > 0$: allows overdispersion.

The Overdispersion Problem

Poisson assumes $E(Y) = \text{Var}(Y)$ (**equidispersion**). In real data, the variance almost always exceeds the mean (**overdispersion**).

If overdispersion is present:

- Poisson coefficient estimates are still **consistent**
- But standard errors are **too small**
- Hypothesis tests become unreliable

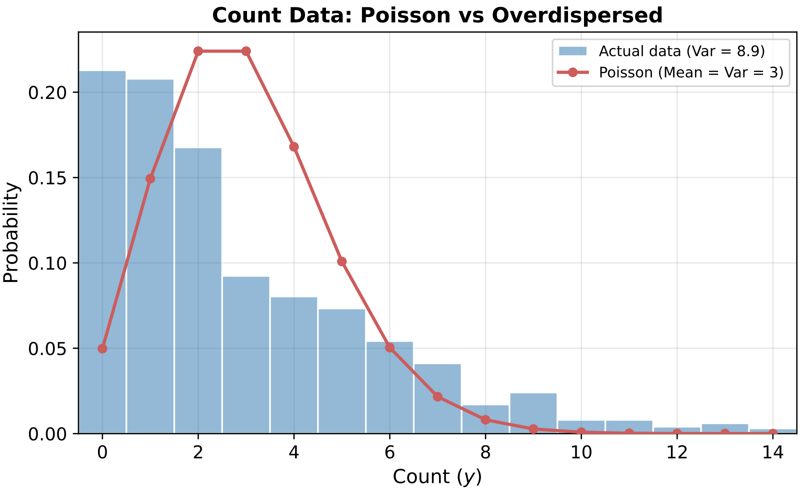
The **negative binomial model** relaxes equidispersion:

$$\text{Var}(Y) = \mu + \alpha\mu^2$$

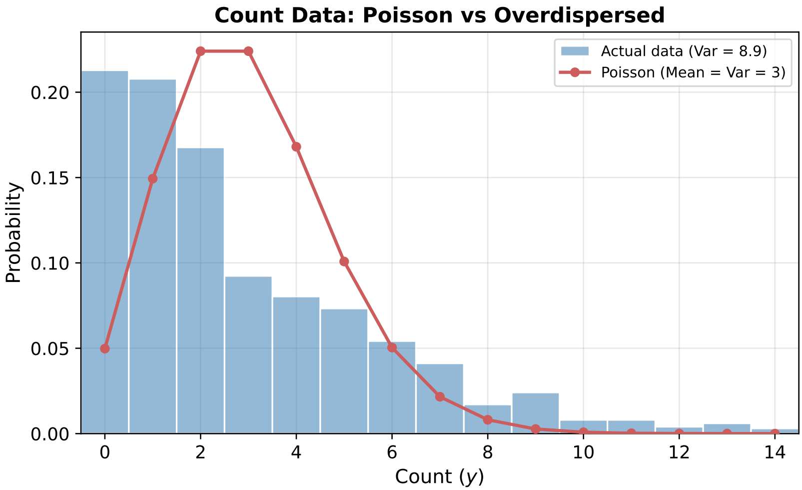
When $\alpha = 0$: reduces to Poisson. When $\alpha > 0$: allows overdispersion.

$\implies$ Test for overdispersion by testing $H_0 : \alpha = 0$.

Count Data: Poisson vs Overdispersed



Count Data: Poisson vs Overdispersed



The actual data has a longer right tail and more zeros than Poisson predicts. The negative binomial [illegible]

Censored vs Truncated Data

Censored: everyone is in the sample, but some values are "clipped."

- Hours worked: we observe 0 for non-workers, but their "desired hours" might be negative

Censored vs Truncated Data

Censored: everyone is in the sample, but some values are "clipped."

- Hours worked: we observe 0 for non-workers, but their "desired hours" might be negative

Truncated: some observations are excluded entirely.

- If we only survey earners above the poverty line, we never see anyone below it

Censored vs Truncated Data

Censored: everyone is in the sample, but some values are "clipped."

- Hours worked: we observe 0 for non-workers, but their "desired hours" might be negative

Truncated: some observations are excluded entirely.

- If we only survey earners above the poverty line, we never see anyone below it

The observed censored variable:

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

Censored vs Truncated Data

Censored: everyone is in the sample, but some values are "clipped."

- Hours worked: we observe 0 for non-workers, but their "desired hours" might be negative

Truncated: some observations are excluded entirely.

- If we only survey earners above the poverty line, we never see anyone below it

The observed censored variable:

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

$\Longrightarrow$ Censoring creates a pile-up at zero. OLS on the censored data attenuates the slope toward zero (similar to measurement error bias).

The Tobit Model

The **Tobit model** handles censored data. It combines a probit (for whether $y > 0$) with a linear regression (for the magnitude when positive):

$$y_i^* = x_i'\beta + e_i, \qquad e_i \sim N(0, \sigma^2)$$

The Tobit Model

The **Tobit model** handles censored data. It combines a probit (for whether $y > 0$) with a linear regression (for the magnitude when positive):

$$y_i^* = x_i'\beta + e_i, \qquad e_i \sim N(0, \sigma^2)$$

A change in x affects the outcome through two channels:

1. **Extensive margin:** changing $P(y > 0)$
2. **Intensive margin:** changing $E(y \mid y > 0)$

The Tobit Model

The **Tobit model** handles censored data. It combines a probit (for whether $y > 0$) with a linear regression (for the magnitude when positive):

$$y_i^* = x_i'\beta + e_i, \qquad e_i \sim N(0, \sigma^2)$$

A change in x affects the outcome through two channels:

1. **Extensive margin:** changing $P(y > 0)$
2. **Intensive margin:** changing $E(y \mid y > 0)$

Limitation: Tobit assumes the same β governs both margins. If the decision to participate depends on different factors than the amount, Tobit is misspecified.

The Tobit Model

The **Tobit model** handles censored data. It combines a probit (for whether $y > 0$) with a linear regression (for the magnitude when positive):

$$y_i^* = x_i'\beta + e_i, \qquad e_i \sim N(0, \sigma^2)$$

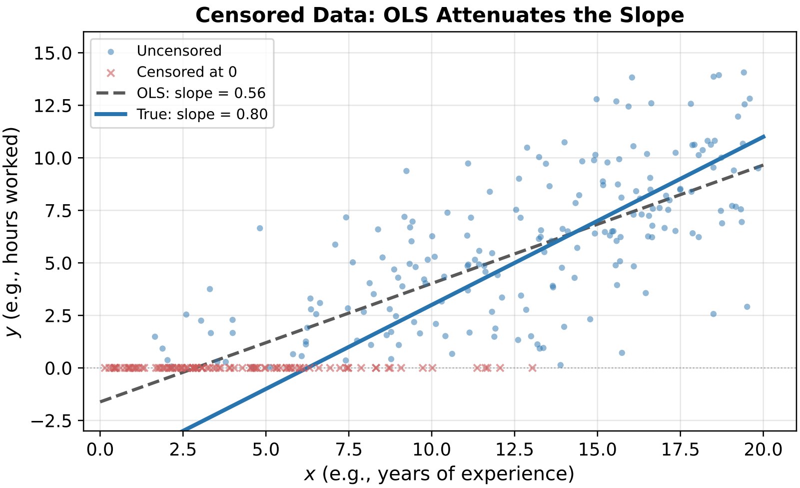
A change in x affects the outcome through two channels:

1. **Extensive margin:** changing $P(y > 0)$
2. **Intensive margin:** changing $E(y \mid y > 0)$

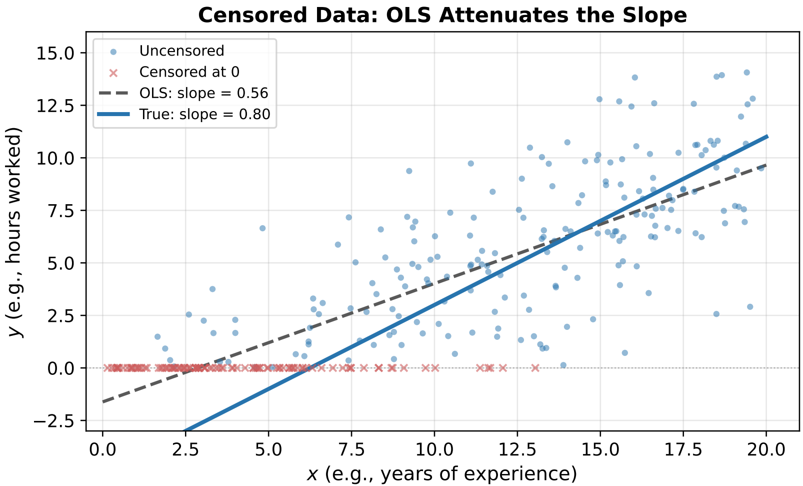
Limitation: Tobit assumes the same β governs both margins. If the decision to participate depends on different factors than the amount, Tobit is misspecified.

$\Longrightarrow$ The Heckman selection model relaxes this by allowing separate equations for the two stages.

Censored Data: OLS vs the True Relationship

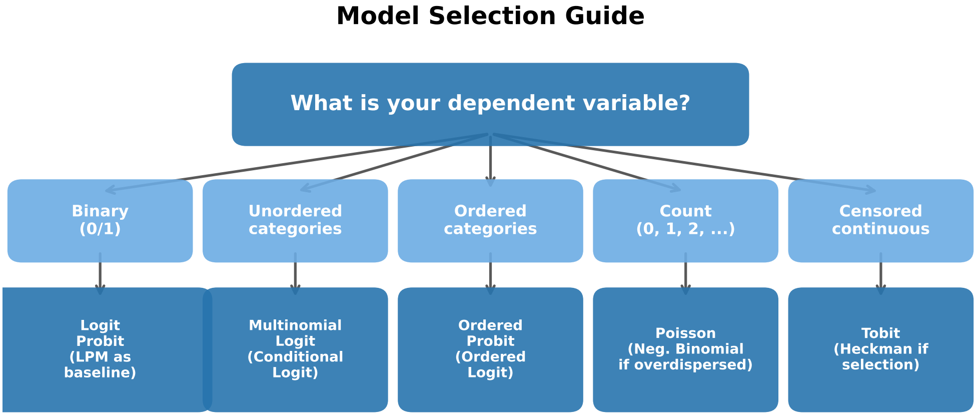


Censored Data: OLS vs the True Relationship



OLS pulls the slope toward zero because it treats the censored zeros as genuine low values. Tobit recovers the steeper true slope.

Choosing the Right Model



All estimated by Maximum Likelihood (except LPM, which uses OLS)

Interpret coefficients through marginal effects, not raw values

Model Selection: Summary Table

Dependent Variable	Model	Estimation
Binary (0/1)	LPM, Probit, Logit	OLS / MLE
Unordered categories	Multinomial Logit	MLE
Ordered categories	Ordered Probit/Logit	MLE
Count $(0, 1, 2, \ldots)$	Poisson, Neg. Binomial	MLE
Censored continuous	Tobit	MLE
Selected sample	Heckman Selection	Two-step / MLE

Model Selection: Summary Table

Dependent Variable	Model	Estimation
Binary (0/1)	LPM, Probit, Logit	OLS / MLE
Unordered categories	Multinomial Logit	MLE
Ordered categories	Ordered Probit/Logit	MLE
Count (0, 1, 2, . . .)	Poisson, Neg. Binomial	MLE
Censored continuous	Tobit	MLE
Selected sample	Heckman Selection	Two-step / MLE

$\Longrightarrow$ The common thread: match the model to the structure of y. In all cases, interpret results through **marginal effects**, not raw coefficients.

Model Selection: Summary Table

Dependent Variable	Model	Estimation
Binary (0/1)	LPM, Probit, Logit	OLS / MLE
Unordered categories	Multinomial Logit	MLE
Ordered categories	Ordered Probit/Logit	MLE
Count (0, 1, 2, . . .)	Poisson, Neg. Binomial	MLE
Censored continuous	Tobit	MLE
Selected sample	Heckman Selection	Two-step / MLE

$\Longrightarrow$ The common thread: match the model to the structure of y. In all cases, interpret results through **marginal effects**, not raw coefficients.

$\Longrightarrow$ For all MLE models: goodness of fit is measured by pseudo-R^2 and percent correctly predicted, not R^2.

Thank you!

jakeanderson@g.ucla.edu