# Ordered Probit / Ordered Logit

## Modeling Outcomes That Have a Ranking but Not a Scale

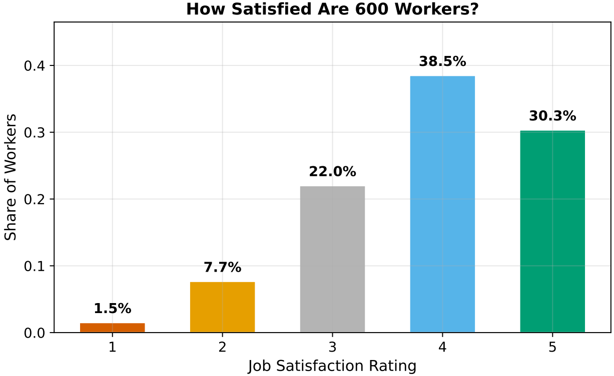
Jake Anderson

May 16, 2026

# Outline

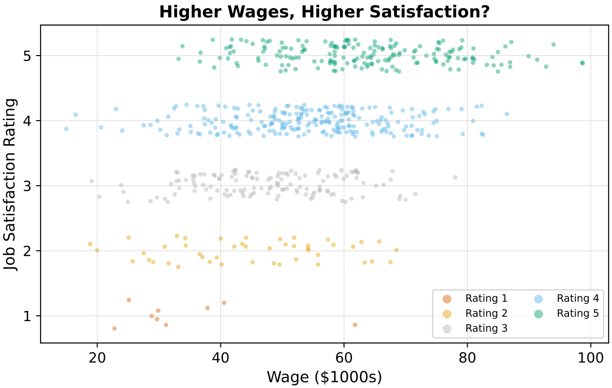
# The Data

A firm surveys **600 workers** about job satisfaction. Each worker reports a rating from 1 (very dissatisfied) to 5 (very satisfied).




The outcome is **ordinal**: $5 > 4 > 3 > 2 > 1$, but the distances between categories are not meaningful. Is the gap from 1 to 2 the same as 4 to 5?

# Does Wage Predict Satisfaction?




Higher-paid workers tend to report higher ratings. But the outcome takes only five discrete values. How should we model this?

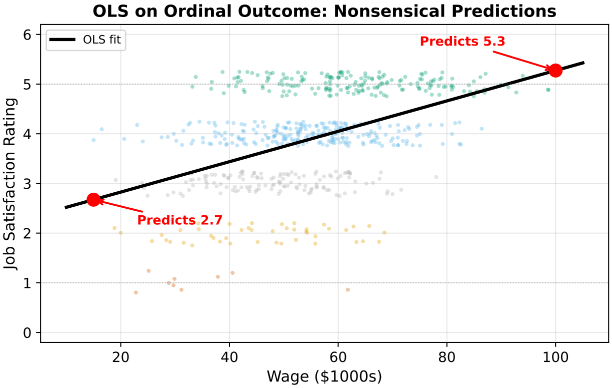
# First Instinct: Run OLS

Treat the rating as a continuous number and regress it on wage:

$$\text{Rating}_i = \beta_0 + \beta_1 \text{Wage}_i + \varepsilon_i$$

This is fast and gives a slope you can interpret. What could go wrong?

# OLS on Ordinal Outcomes: The Failure




At high wages, OLS predicts **5.3**. But the scale only goes to 5. At low wages, it predicts non-integer values between categories. OLS treats ordinal categories as if they were continuous and equally spaced.

# What Goes Wrong with OLS on Ordinal Data

**Three problems:**

1. **Predictions outside the valid range.** OLS can predict 0.5 or 5.3 when the outcome only takes values 1–5
2. **Equal-spacing assumption.** OLS treats the jump from 1 to 2 as identical to the jump from 4 to 5. The numerical labels are arbitrary; replacing 1–5 with 2, 4, 6, 8, 10 would change the slope
3. **Constant marginal effects.** OLS forces each $1k wage increase to add the same amount to the predicted rating, regardless of where on the scale the worker starts

$\Longrightarrow$ We need a model that respects the ordinal nature of the outcome: categories have a ranking, but the distances between them are unknown.

# What Would a Better Model Look Like?

A proper model for ordinal outcomes should:

1. **Produce valid probabilities.** $P(\text{Rating} = j) \in (0, 1)$, summing to 1.
2. **Respect the ordering.** Use $5 > 4 > 3 > 2 > 1$ without assuming equal spacing.
3. **Allow flexible marginal effects.** Effect of a wage increase can depend on where the worker sits on the scale.

$\Longrightarrow$ Where can we find a model with these properties?

# The Latent Continuous Variable Behind Ordered Ratings

Think about what the 1–5 scale really represents.

Satisfaction is probably **continuous** in a worker's mind. A worker who reports "4" and one who reports "5" might differ by a hair; another pair of 4 and 5 reporters might differ enormously.

The ratings are a **coarse discretization** of something continuous. What if we modeled that continuous variable directly and let the discrete ratings emerge from it?

$\Longrightarrow$ This is the latent variable approach: posit an unobserved continuous satisfaction level, then map it to the observed categories through thresholds.

# Outline

# The Idea: A Continuous Variable Behind Discrete Ratings

Imagine each worker has a latent (unobserved) satisfaction level $y_i^*$ that is continuous:

$$y_i^* = \beta_1 \,\text{Wage}_i + \beta_2 \,\text{Hours}_i + \beta_3 \,\text{Support}_i + \varepsilon_i$$

(no intercept; it is absorbed into the cutpoints, explained shortly)

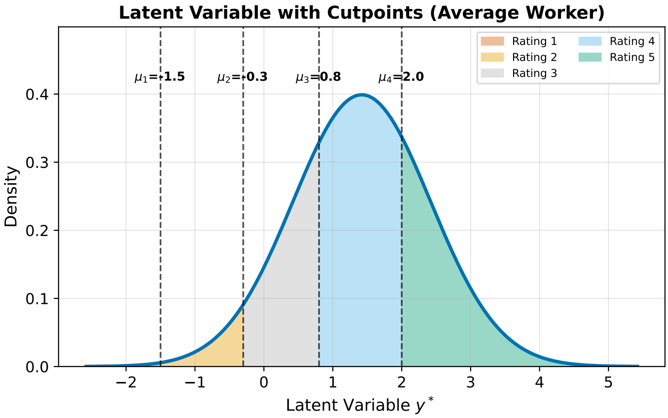
- $y_i^*$ can take any real value (no boundary problems)
- We never observe $y_i^*$ directly
- We observe only which **category** it falls into

The mapping from $y_i^*$ to the observed rating uses **cutpoints** $\mu_1 < \mu_2 < \mu_3 < \mu_4$:

$$\text{Rating}_i = j \quad \Longleftrightarrow \quad \mu_{j-1} < y_i^* \leq \mu_j$$

where $\mu_0 = -\infty$ and $\mu_5 = +\infty$.

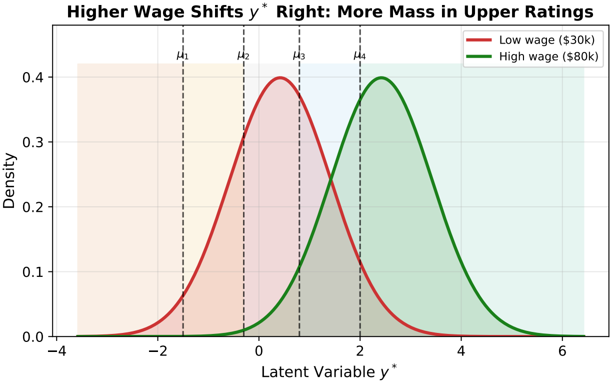
# Cutpoints Partition the Latent Variable




The density of $y_i^*$ is divided into five regions by four cutpoints. The area in each region equals the probability of that rating.

# How Covariates Shift the Distribution

Define the **linear index**: $\text{XB}_i = \beta_1 \, \text{Wage}_i + \beta_2 \, \text{Hours}_i + \beta_3 \, \text{Support}_i$.
A higher wage increases $\text{XB}_i$, which shifts the entire density of $y_i^*$ to the right.




The cutpoints stay fixed; the density moves. Positive $\beta$ raises $P(\text{highest})$ and lowers $P(\text{lowest})$; middle categories are ambiguous.

# The Probability Formula: Derivation

The probability of observing rating $j$ depends on where $y_i^*$ falls:

$$P(\text{Rating}_i = j) = P(\mu_{j-1} < y_i^* \leq \mu_j)$$

Substitute $y_i^* = \text{XB}_i + \varepsilon_i$ and rearrange:

$$= P\big(\mu_{j-1} - \text{XB}_i < \varepsilon_i \leq \mu_j - \text{XB}_i\big)$$

Since $\varepsilon_i$ has CDF $F(\cdot)$:

$$= F(\mu_j - \text{XB}_i) - F(\mu_{j-1} - \text{XB}_i)$$

with boundary conditions $F(-\infty) = 0$ and $F(+\infty) = 1$.

# Probit vs. Logit: Choosing $F(\cdot)$

The general formula works with any CDF. The two standard choices:

- **Ordered probit:** $F = \Phi$ (standard normal CDF), so $\varepsilon_i \sim N(0, 1)$
- **Ordered logit:** $F = \Lambda$ (logistic CDF), so $\varepsilon_i \sim$ Logistic

$\Longrightarrow$ Same idea, different distributional assumption on $\varepsilon_i$. Results are usually similar in practice. Economists tend to use ordered probit; biostatisticians often prefer ordered logit.

# What Gets Estimated

The model estimates two sets of parameters:

1. **Coefficients** $(\beta_1, \beta_2, \beta_3)$, i.e. $(\beta_{\text{wage}}, \beta_{\text{hours}}, \beta_{\text{support}})$
   How each predictor shifts the latent variable $y_i^*$
2. **Cutpoints** $\mu_1, \mu_2, \mu_3, \mu_4$
   Where the boundaries between adjacent categories lie

There is **no intercept** in the equation. An intercept would be absorbed into the cutpoints (you cannot separately identify both), so we normalize by omitting it.

# The Parallel Regressions Assumption

The ordered model assumes that the coefficients $(\beta_1, \beta_2, \beta_3)$ are the **same for every cutpoint**. Visually: when a covariate changes, the density shifts horizontally, but all cutpoints stay fixed.

**Where the name comes from:** imagine running separate binary logits for each cumulative split:

- Rating $\leq 1$ vs. $> 1$
- Rating $\leq 2$ vs. $> 2$
- Rating $\leq 3$ vs. $> 3$
- Rating $\leq 4$ vs. $> 4$

The "parallel regressions" assumption says the slope on each predictor would be the **same across all four splits**. Only the intercept (cutpoint) differs.

$\Longrightarrow$ This is what makes the ordered model parsimonious: one set of slopes instead of four.

# Numeric Example: Computing Probabilities

Suppose a worker earns \$55k, works 42 hours, and has supervisor support (= 1). With the ordered probit coefficients $\beta_{\text{wage}} = 0.04$, $\beta_{\text{hours}} = -0.03$, $\beta_{\text{support}} = 0.80$:

$$\text{XB}_i = 0.04 \times 55 + (-0.03) \times 42 + 0.80 \times 1 = 2.20 - 1.26 + 0.80 = 1.74$$

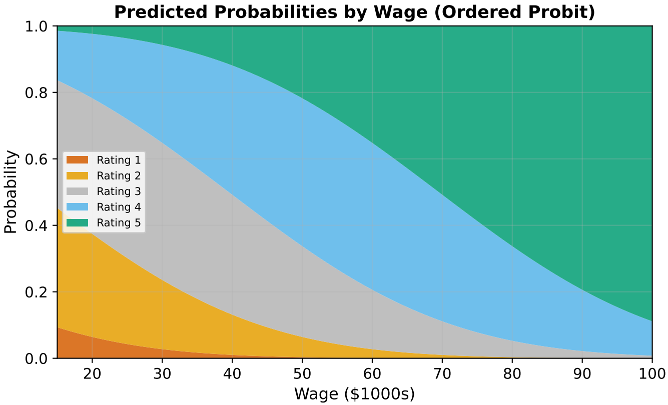
With cutpoints $\mu_1 = -1.5$, $\mu_2 = -0.3$, $\mu_3 = 0.8$, $\mu_4 = 2.0$:

$$\begin{aligned} P(\text{Rating} = 4) &= \Phi(\mu_4 - 1.74) - \Phi(\mu_3 - 1.74) \\ &= \Phi(2.0 - 1.74) - \Phi(0.8 - 1.74) \\ &= \Phi(0.26) - \Phi(-0.94) \\ &= 0.603 - 0.174 = 0.429 \end{aligned}$$

$\implies$ This worker has a 42.9% chance of reporting "Satisfied" (Rating 4). Software computes all five probabilities simultaneously.

# Outline

# Predicted Probabilities




As wage increases, probability shifts from lower ratings to higher ratings. At every wage level, the five probabilities sum to 1.

# Coefficient Interpretation: Sign Only

The estimated coefficients tell us the **direction** in which a predictor shifts $y_i^*$:

| **Variable** | $\hat{\beta}$ | **Interpretation** |
|---|---|---|
| Wage (\$1k) | $+$ | Higher wage $\Longrightarrow$ $y^*$ shifts right $\Longrightarrow$ higher satisfaction |
| Hours | $-$ | More hours $\Longrightarrow$ $y^*$ shifts left $\Longrightarrow$ lower satisfaction |
| Supervisor support | $+$ | Support $\Longrightarrow$ $y^*$ shifts right $\Longrightarrow$ higher satisfaction |

$\Longrightarrow$ You can interpret the **sign**, but not the magnitude directly. Saying "a \$1k raise increases satisfaction by 0.04" is wrong because 0.04 is in latent-variable units, which have no natural scale.

For magnitudes, compute **marginal effects** on probabilities.

# Marginal Effects: The Formula

The marginal effect on any single category depends on how much density sits near the two cutpoints that bracket that category.

The marginal effect of variable $x_k$ on $P(\text{Rating} = j)$ is:

$$\frac{\partial P(\text{Rating} = j)}{\partial x_k} = \left[ f(\mu_{j-1} - \text{XB}_i) - f(\mu_j - \text{XB}_i) \right] \cdot \beta_k$$

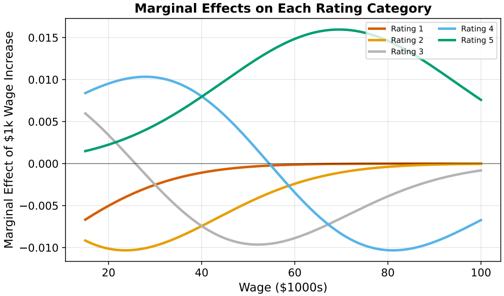
where $f(\cdot)$ is the density (derivative of $F$).

For the **extreme categories**, one of the boundary terms drops out:

- Rating 1: $\text{ME} = -f(\mu_1 - \text{XB}_i) \cdot \beta_k$ ($\beta_k > 0 \implies$ negative)
- Rating 5: $\text{ME} = f(\mu_4 - \text{XB}_i) \cdot \beta_k$ ($\beta_k > 0 \implies$ positive)

$\implies$ A positive coefficient always decreases $P(\text{lowest})$ and increases $P(\text{highest})$. But what about the middle categories?

# Middle Categories: Ambiguous Marginal Effects




The marginal effect on **middle categories changes sign** with the worker's starting point: at low wages a raise increases $P(\text{Rating} = 4)$; at high wages it decreases it (workers there are already shifting into Rating 5).

$\implies$ Report Average Marginal Effects (AMEs) for each category, not just the coefficient.

# Estimation: Maximum Likelihood

Each worker contributes one term to the likelihood: the probability of the rating they actually reported.

The log-likelihood:

$$\ell = \sum_{i=1}^{N} \sum_{j=1}^{5} d_{ij} \ln P(\text{Rating}_i = j)$$

where $d_{ij} = 1$ if worker $i$ reported rating $j$.

- $P(\text{Rating}_i = j) = F(\mu_j - \text{XB}_i) - F(\mu_{j-1} - \text{XB}_i)$
- Parameters: $(\beta_1, \beta_2, \beta_3)$ and $\mu_1, \ldots, \mu_4$
- No closed-form solution $\implies$ numerical optimization (same as binary logit)

$\implies$ Software estimates the coefficients and the cutpoints jointly.

# Outline

# This Is Not Multinomial Logit

Ordered choice and multinomial logit both handle categorical outcomes. But they are built for different situations.

**Multinomial logit:** categories have **no natural ordering** (Car, Bus, Bike, Walk). Each alternative gets its **own coefficient** ($\beta_j$ per mode).

**Ordered choice:** categories have a **natural ranking** ($5 > 4 > 3 > 2 > 1$). One set of coefficients ($\beta_1, \ldots, \beta_K$) shifts the entire latent distribution; the ordering is exploited rather than discarded.

$\Longrightarrow$ Using multinomial logit on ordered data wastes the ordering information and estimates far more parameters than necessary ($J - 1$ coefficient vectors instead of one).

# When Parallel Regressions Fails

Recall the parallel regressions assumption: the same slopes apply at every cumulative split. When does this break down?

**Example:** suppose wage has a strong effect on moving from "Dissatisfied" to "Neutral," but no effect on moving from "Satisfied" to "Very Satisfied." The ordered model cannot capture this because it forces a single $\beta_{\text{wage}}$.

**Test:** the Brant test checks whether the slopes are stable across cutpoints (note: applies specifically to ordered logit). If it rejects, consider:

- A **generalized ordered logit** (which allows slopes to vary by cutpoint)
- Multinomial logit as a fallback

$\Longrightarrow$ Always check parallel regressions before reporting ordered model results.

# Ordered Probit vs. Ordered Logit

The only difference is the assumed distribution of $\varepsilon_i$:

| | **Ordered Probit** | **Ordered Logit** |
|---|---|---|
| Error distribution | $\varepsilon_i \sim N(0,1)$ | $\varepsilon_i \sim$ Logistic |
| CDF used | $\Phi(\cdot)$ | $\Lambda(\cdot) = \frac{e^{(\cdot)}}{1+e^{(\cdot)}}$ |
| Tails | Thinner | Slightly heavier |
| Predicted probabilities | Nearly identical | Nearly identical |

$\Longrightarrow$ In practice, both give very similar results. Choose based on convention in your field.

# Decision Framework: Which Model to Use

1. **Binary** (yes/no, 0/1) $\implies$ Binary logit or probit
2. **Categorical, no natural order** (car, bus, bike, walk) $\implies$ Multinomial logit
3. **Categorical, natural ranking** (strongly disagree $\rightarrow$ strongly agree) $\implies$ Ordered probit or ordered logit
4. **Parallel regressions assumption fails** $\implies$ Generalized ordered logit, or fall back to multinomial logit

$\implies$ The ordered model is more efficient than multinomial logit when the ordering is genuine, because it estimates fewer parameters while exploiting the ranking structure.

# Summary: Back to Job Satisfaction

Recall the three OLS problems. How does ordered probit address each one?

1. **Predictions outside the valid range** $\Longrightarrow$ probabilities $\in (0, 1)$ that sum to 1. No predictions of 5.3.
2. **Equal-spacing assumption** $\Longrightarrow$ cutpoints $\mu_1, \ldots, \mu_4$ are estimated freely; no equal-distance assumption.
3. **Constant marginal effects** $\Longrightarrow$ marginal effects depend on the worker's current position on the satisfaction scale.

$\Longrightarrow$ Interpret the sign of $\hat{\beta}$ for direction; compute AMEs for magnitude. Check the parallel regressions assumption before reporting.

Thank you!

jakeanderson@g.ucla.edu