

# LPM vs. Logit/Probit

Jake Anderson

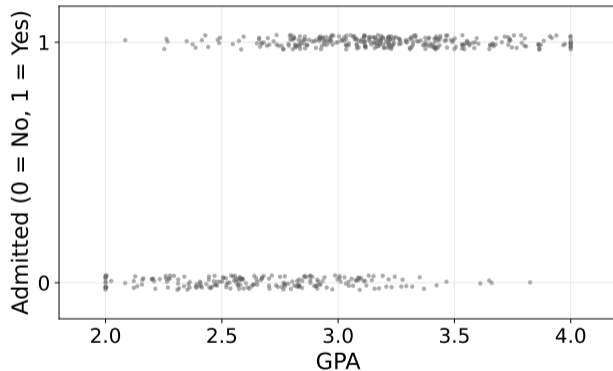
May 16, 2026

# Outline

- 1 The Linear Probability Model
- 2 The S-Curve Solution
- 3 Interpreting Logit Coefficients
- 4 Logit vs. Probit
- 5 When Is the LPM Acceptable?
- 6 Maximum Likelihood Estimation

# The Data

A university admissions office records **GPA** and **admission outcome** (admit/reject) for 500 applicants.



The outcome is binary: 0 (rejected) or 1 (admitted). How do we model the probability of admission?

## Natural Instinct: Run OLS

The simplest approach: regress the 0/1 outcome on GPA, just like any other regression.

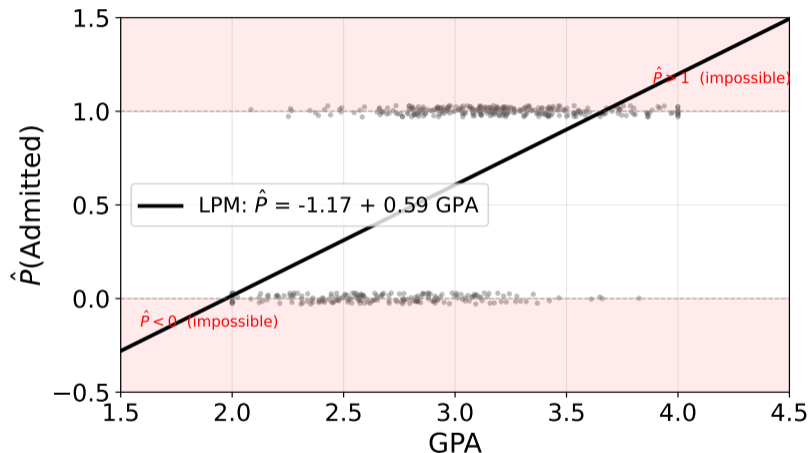
This is the **Linear Probability Model** (LPM):

$$P(\text{Admit}_i = 1 \mid \text{GPA}_i) = \beta_0 + \beta_1 \text{GPA}_i$$

The coefficients have a direct interpretation:

- $\beta_1$  = change in the *probability of admission* for a one-unit increase in GPA

Sounds reasonable. Let's see what happens.



$\hat{P}(\text{Admit}) = -1.17 + 0.59 \cdot \text{GPA}$ . At GPA = 4.0:  $\hat{P} = 1.20$ . At GPA = 2.0:  $\hat{P} = 0.02$ .

$\implies$  Probabilities **must** lie in  $[0, 1]$ . A straight line cannot respect this constraint.

## Problem 1: Impossible Predictions

A probability model should produce  $\hat{P} \in [0, 1]$  for all observations. The LPM violates this.

For any linear function  $\hat{P} = \beta_0 + \beta_1 x$ :

- If  $x$  is large enough  $\implies \hat{P} > 1$
- If  $x$  is small enough  $\implies \hat{P} < 0$

In our data: applicants with GPA above  $\approx 3.7$  get predicted probabilities exceeding 1.

$\implies$  The LPM is a line forced through inherently nonlinear data. It works in the middle but fails in the tails.

## Problem 2: Constant Marginal Effects

The LPM says: each additional GPA point increases admission probability by **0.59**, regardless of where you start.

- Going from GPA 2.0 to 3.0: +0.59 probability
- Going from GPA 3.0 to 4.0: +0.59 probability

Is that realistic?

No. Consider the S-shaped relationship we expect:

- Near the middle of the GPA range, the probability is changing rapidly (steep part of the curve)
- At the extremes, the probability is near 0 or near 1, so an extra GPA point makes little difference (flat parts of the curve)

⇒ Marginal effects should be **largest near the midpoint** and diminish in the tails, not constant everywhere.

## Problem 3: Heteroskedastic Errors

When  $y$  can only be 0 or 1, its variance is the variance of a Bernoulli random variable:

$$\text{Var}(y_i | x_i) = P(x_i)(1 - P(x_i))$$

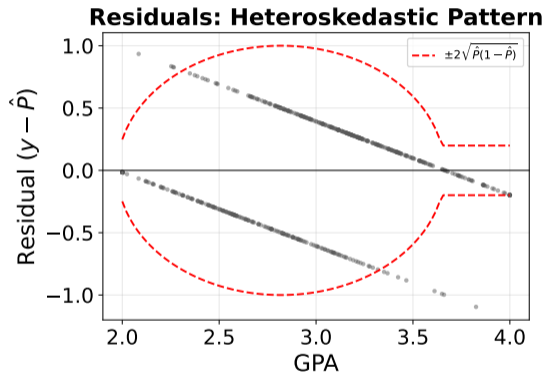
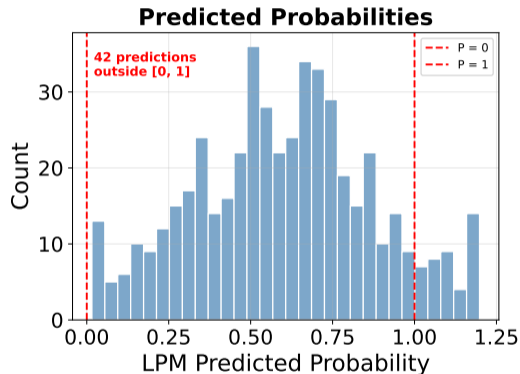
This varies with  $x$  by construction  $\implies$  **heteroskedasticity is guaranteed.**

Consequences:

- OLS coefficients are still **unbiased**
- But OLS standard errors are **wrong** (too small or too large)
- Hypothesis tests and confidence intervals are unreliable

This problem is fixable: robust standard errors correct the SEs. But the impossible predictions and constant marginal effects remain.

# LPM Problems: Visualized



Left: some predictions fall outside [0, 1]. Right: residuals fan out, confirming heteroskedasticity.

All three LPM problems share one structural mismatch:

A straight line has no bounds, but a probability does.

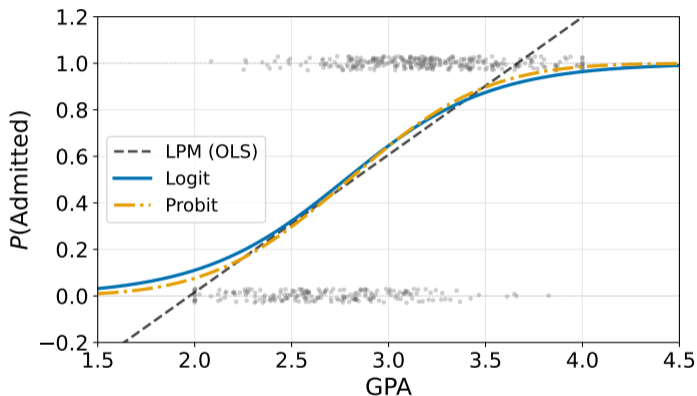
- Problems 1 and 2 are two faces of this mismatch:
  - **Out of bounds**  $\implies$  the line overshoots  $[0, 1]$
  - **Constant slope**  $\implies$  the line cannot flatten as it approaches 0 or 1
- Problem 3 (heteroskedasticity) is a byproduct, and can be patched with robust SEs

$\implies$  We need a **curve**, not a line: something that starts near 0, rises steeply through the middle, and flattens near 1.

# Outline

- 1 The Linear Probability Model
- 2 The S-Curve Solution**
- 3 Interpreting Logit Coefficients
- 4 Logit vs. Probit
- 5 When Is the LPM Acceptable?
- 6 Maximum Likelihood Estimation

# LPM vs. Logit vs. Probit



The LPM (dashed) overshoots at both ends. Logit and probit replace the line with an S-shaped curve that stays in  $[0, 1]$ .

# Where Does the S-Curve Come From?

Imagine each applicant has a **latent** (unobserved) “admissibility” score:

$$y_i^* = \beta_0 + \beta_1 \text{GPA}_i + \varepsilon_i$$

*Latent* just means we never see  $y_i^*$  directly. We only observe the binary decision:

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

What is the probability of admission?

$$\begin{aligned} P(\text{Admit} = 1 \mid \text{GPA}) &= P(y^* > 0) \\ &= P(\varepsilon > -\beta_0 - \beta_1 \text{GPA}) \end{aligned}$$

⇒ The distribution we assume for  $\varepsilon$  determines the shape of the curve.

# The Latent Variable and Choice Probability

Starting from:

$$P(\text{Admit} = 1) = P(\varepsilon > -\beta_0 - \beta_1 \text{ GPA})$$

If  $\varepsilon$  has a *symmetric* distribution (by symmetry of the CDF):

$$P(\varepsilon > -z) = P(\varepsilon < z) = F(z)$$

So the probability is just the CDF of  $\varepsilon$  evaluated at  $\beta_0 + \beta_1 \text{ GPA}$ :

$$P(\text{Admit} = 1 \mid \text{GPA}) = F(\beta_0 + \beta_1 \text{ GPA})$$

$\implies$  Any CDF maps  $(-\infty, +\infty) \rightarrow [0, 1]$ , which is exactly what we need. Two standard choices give us two models.

## Two Distributions, Two Models

**Logistic distribution** for  $\varepsilon$ :

$$P(\text{Admit} = 1 \mid \text{GPA}) = \Lambda(\beta_0 + \beta_1 \text{GPA}) = \frac{e^{\beta_0 + \beta_1 \text{GPA}}}{1 + e^{\beta_0 + \beta_1 \text{GPA}}}$$

This is the **logit** model.

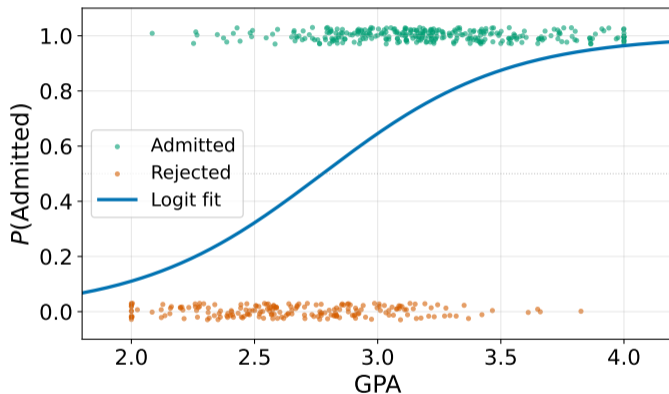
**Standard normal distribution** for  $\varepsilon$ :

$$P(\text{Admit} = 1 \mid \text{GPA}) = \Phi(\beta_0 + \beta_1 \text{GPA})$$

This is the **probit** model.

Both produce S-shaped curves bounded in  $[0, 1]$ . The logistic CDF ( $\Lambda$ ) has slightly heavier tails than the normal CDF ( $\Phi$ ), but in practice the two are nearly indistinguishable.

## The Logit Model: Fitted Curve



The logit curve passes through the middle of the data, stays in  $[0, 1]$ , and has the steepest slope near  $P = 0.5$ .

# Outline

- 1 The Linear Probability Model
- 2 The S-Curve Solution
- 3 Interpreting Logit Coefficients**
- 4 Logit vs. Probit
- 5 When Is the LPM Acceptable?
- 6 Maximum Likelihood Estimation

# The Interpretation Problem

Our logit estimates:  $\hat{\beta}_0 = -7.46$ ,  $\hat{\beta}_1 = 2.69$ .

How would you interpret  $\hat{\beta}_1 = 2.69$ ?

**Tempting but wrong:** “A one-unit increase in GPA raises the probability of admission by 2.69.”

Why wrong? Because the logit is **nonlinear**. The coefficient 2.69 operates on the **log-odds** scale, not the probability scale. A probability change of 2.69 is not even possible.

⇒ To interpret logit coefficients, we need to understand what they actually measure.

## Log-Odds: What the Coefficient Measures

Define the **odds** of admission:

$$\text{Odds} = \frac{P(\text{Admit} = 1)}{P(\text{Admit} = 0)} = \frac{P}{1 - P}$$

The logit model is **linear in log-odds**:

$$\underbrace{\ln\left(\frac{P}{1 - P}\right)}_{\text{log-odds}} = \beta_0 + \beta_1 \text{ GPA}$$

$\implies \beta_1 = 2.69$  means: a one-unit increase in GPA raises the **log-odds** of admission by 2.69.

Equivalently, the **odds ratio**:

$$e^{\beta_1} = e^{2.69} \approx 14.7$$

A one-unit increase in GPA **multiplies** the odds of admission by  $\approx 14.7$ . For example, going from GPA 2.5 to 3.5 multiplies the odds by this factor.

## Marginal Effects: What We Actually Want

The effect on *probability* depends on where you start:

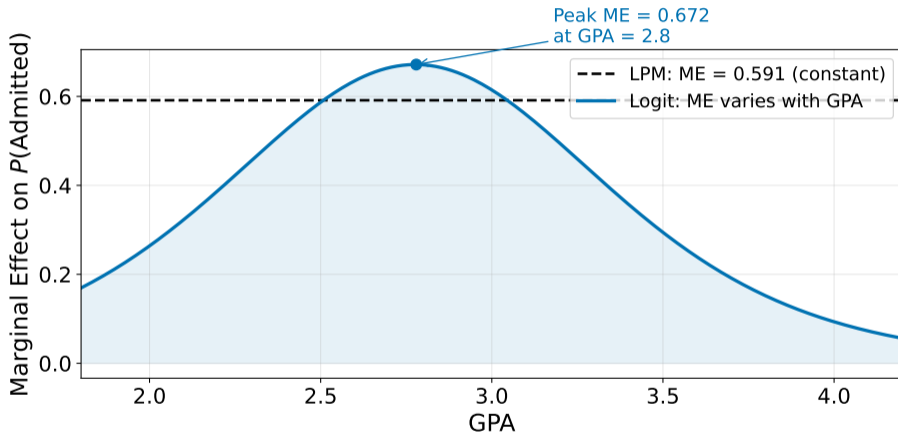
$$\underbrace{\frac{\partial P}{\partial \text{GPA}}}_{\text{marginal effect}} = \beta_1 \cdot \Lambda(\beta_0 + \beta_1 \text{GPA}) \cdot (1 - \Lambda(\beta_0 + \beta_1 \text{GPA}))$$

This equals  $\beta_1 \cdot P \cdot (1 - P)$ , which is largest when  $P = 0.5$ .

GPA	$\hat{P}(\text{Admit})$	Marginal Effect
2.0	0.11	0.26
2.5	0.32	0.59
3.0	0.65	<b>0.61</b>
3.5	0.87	0.29
4.0	0.96	0.09

⇒ The same one-unit GPA increase has roughly 7x more impact near the middle than at the top.

## Marginal Effects: Visualized



The LPM assumes a constant effect (dashed). The logit captures the realistic bell shape: largest effect near  $P = 0.5$ , vanishing in the tails.

# Average Marginal Effect (AME)

Reporting a marginal effect at a single GPA is incomplete. Researchers typically report the **Average Marginal Effect**:

$$\text{AME} = \frac{1}{n} \sum_{i=1}^n \hat{\beta}_1 \cdot \hat{P}_i \cdot (1 - \hat{P}_i)$$

Compute the marginal effect *at each observation's actual GPA*, then average.

In our data:  $\text{AME} \approx 0.49$ .

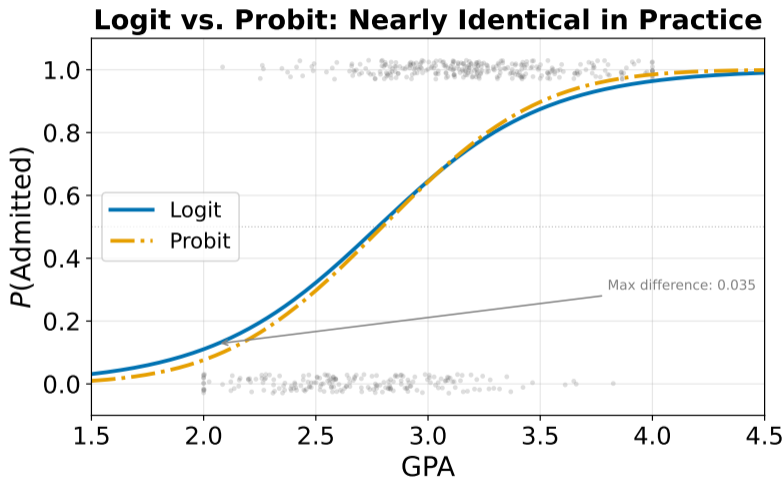
“On average, a one-unit increase in GPA is associated with a 49 percentage point increase in the probability of admission.”

A full GPA point is a large change (e.g., 2.5 to 3.5), so this large AME makes sense in context.

⇒ AME gives a single summary number comparable to the LPM coefficient (0.59). The LPM overstates the average effect because it ignores diminishing returns.

# Outline

- 1 The Linear Probability Model
- 2 The S-Curve Solution
- 3 Interpreting Logit Coefficients
- 4 Logit vs. Probit**
- 5 When Is the LPM Acceptable?
- 6 Maximum Likelihood Estimation



The two curves are almost indistinguishable. The largest difference is in the tails, where both curves are near 0 or 1.

## Logit vs. Probit: Coefficients

The logit and probit coefficients are on different scales:

	$\hat{\beta}_0$	$\hat{\beta}_1$	Scale
<b>Logit</b>	-7.46	2.69	Log-odds
<b>Probit</b>	-5.03	1.80	z-score (std. normal)

Three numbers you may see for the logit/probit coefficient ratio:

- $\sqrt{\pi^2/3} \approx 1.81$ : the *theoretical* ratio, from the fact that the logistic distribution has variance  $\pi^2/3$  while the standard normal has variance 1
- $\approx 1.6$ : a coarser textbook approximation (Amemiya)
- Here:  $2.69/1.80 = 1.49$ : the actual ratio in this finite sample

⇒ Marginal effects and predicted probabilities are nearly identical regardless. The choice between logit and probit rarely changes conclusions. Logit is more common in economics because of the odds-ratio interpretation.

# Outline

- 1 The Linear Probability Model
- 2 The S-Curve Solution
- 3 Interpreting Logit Coefficients
- 4 Logit vs. Probit
- 5 When Is the LPM Acceptable?**
- 6 Maximum Likelihood Estimation

# In Defense of the LPM

Despite its problems, the LPM is widely used in applied research. When is it acceptable?

## The LPM works well when:

- 1 Predicted probabilities fall in  $[0.2, 0.8]$  for most observations  
⇒ The S-curve is approximately linear in this range
- 2 You only need the **average** effect, not predictions at extremes  
⇒ LPM coefficient  $\approx$  AME from logit
- 3 With robust standard errors to correct heteroskedasticity

## The LPM fails when:

- 1 You need predictions (e.g., credit scoring, medical diagnosis)
- 2 The outcome is rare or very common ( $P$  near 0 or 1)
- 3 You have covariates that push predictions far from 0.5

## LPM vs. Logit: Decision Framework

	<b>LPM</b>	<b>Logit / Probit</b>
Estimation	OLS	MLE
Predictions in $[0, 1]$ ?	No	Yes
Marginal effects	Constant	Vary with $x$
Coefficient = ME?	Yes	No (need AME)
Heteroskedasticity	Built in	Handled by MLE
Speed / simplicity	Fastest	Slightly more complex
With FE (many groups)	Easy	Bias risk (incidental parameters problem)

Incidental parameters: with many FE, logit MLE estimates one parameter per group  $\implies$  biased coefficients in short panels.

$\implies$  LPM + robust SEs is reasonable as a baseline; switch to logit/probit when predictions or nonlinear effects are central.

# Outline

- 1 The Linear Probability Model
- 2 The S-Curve Solution
- 3 Interpreting Logit Coefficients
- 4 Logit vs. Probit
- 5 When Is the LPM Acceptable?
- 6 Maximum Likelihood Estimation**

# Why Not OLS?

OLS minimizes the sum of squared residuals. With binary  $y$ , this creates the problems we saw: impossible predictions, heteroskedasticity.

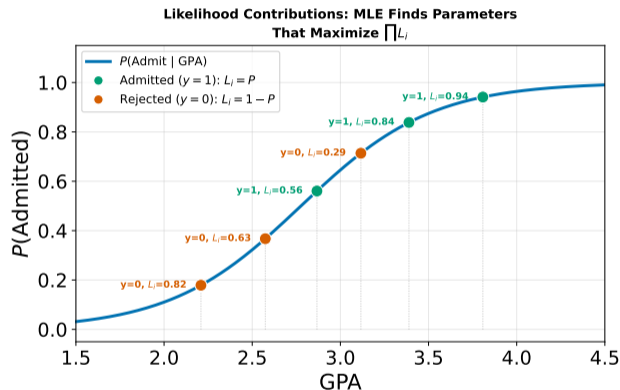
Logit and probit use a different estimation criterion: **Maximum Likelihood Estimation (MLE)**.

**MLE idea:** Find the parameters  $\beta_0, \beta_1$  that make the *observed data* most probable.

- For an admitted applicant ( $y_i = 1$ ): we want  $P_i$  to be **high**
- For a rejected applicant ( $y_i = 0$ ): we want  $P_i$  to be **low**

⇒ MLE finds the S-curve that best separates the admitted from the rejected.

# MLE: How It Works



Each observation contributes  $P_i$  (if admitted) or  $1 - P_i$  (if rejected) to the likelihood. MLE maximizes the product of these contributions.

# The Log-Likelihood

The **likelihood** for one observation:

$$L_i = P_i^{y_i} \cdot (1 - P_i)^{1-y_i}$$

The total **log-likelihood** (sum over all observations):

$$\ell(\beta_0, \beta_1) = \sum_{i=1}^n \left[ y_i \ln P_i + (1 - y_i) \ln(1 - P_i) \right]$$

where  $P_i = \Lambda(\beta_0 + \beta_1 \text{GPA}_i)$  for logit.

No closed-form solution  $\implies$  solved numerically (Newton-Raphson, gradient ascent). Software handles this automatically.

$\implies$  MLE is the standard estimation method for logit and probit. The resulting  $\hat{\beta}$  values are the ones that maximize this log-likelihood.

Thank you!  
jakeanderson@g.ucla.edu