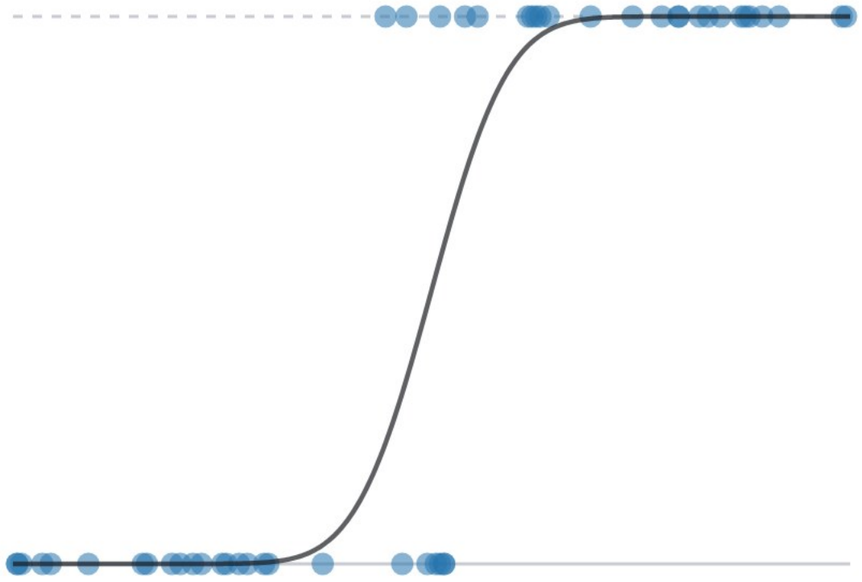
ECON 104 · DISCUSSION · FALL 2026

Granger Causality, the Linear Probability Model, and Probit



Jake Anderson
UCLA

Plan for Today

1

Granger causality

Do past values of one series improve the forecast of another?

Read two F-tests, one in each direction, and choose the model they support.

Spring 2026 Midterm (Boswell), Question 11

2

Linear probability model

A predicted probability of 1.1 and errors whose variance changes with x.

Two problems with the LPM and the correction for each.

Fall 2024 Midterm V1 (Boswell), Questions 15 and 17

3

Probit predictions

The same user, now with a probit model.

Read the probability $\Phi(z)$ and compare to the LPM.

Fall 2024 Midterm V1 (Boswell), Question 16

The Question: Which Way Does Granger Causality Run?

A macroeconomist is studying whether consumer confidence (Conf) helps predict retail sales growth (Sales) in the US, using quarterly data from 2000–2023 ($T = 96$), where both series have been confirmed stationary. For the following Granger causality tests:

```
Granger causality test
Model 1: Sales ~ L(Sales, 1:2) + L(Conf, 1:2)
Model 2: Sales ~ L(Sales, 1:2)
  Res.Df Df      F  Pr(>F)
1     91
2     93 -2 6.8741  0.0016 **

Granger causality test
Model 1: Conf ~ L(Conf, 1:2) + L(Sales, 1:2)
Model 2: Conf ~ L(Conf, 1:2)
  Res.Df Df      F  Pr(>F)
1     91
2     93 -2 1.2033  0.3047
```

which of the following correctly interprets the test results and their implication for model specification?

a Consumer confidence Granger causes retail sales and retail sales Granger causes consumer confidence; a VAR model is required to capture the bidirectional relationship.

b Retail sales Granger causes consumer confidence, but consumer confidence does not Granger cause retail sales; lagged Sales should be used to forecast Conf instead.

c Consumer confidence Granger causes retail sales, but retail sales does not Granger cause consumer confidence; the ARDL model with lagged Conf as a predictor of Sales is appropriate.

d Neither series Granger causes the other; the ARDL specification is unsupported and both distributed lag terms should be dropped.

e The test is inconclusive because the residual degrees of freedom row for Model 1 is missing from the output.

Source: Spring 2026 Midterm (Boswell), Question 11

Police and Crime: Which Variable Is X?

SO FAR: ARDL

One outcome Y, and a variable X that affects Y but is not affected by Y.

Consider the size of a city's police force and the city's crime rate.

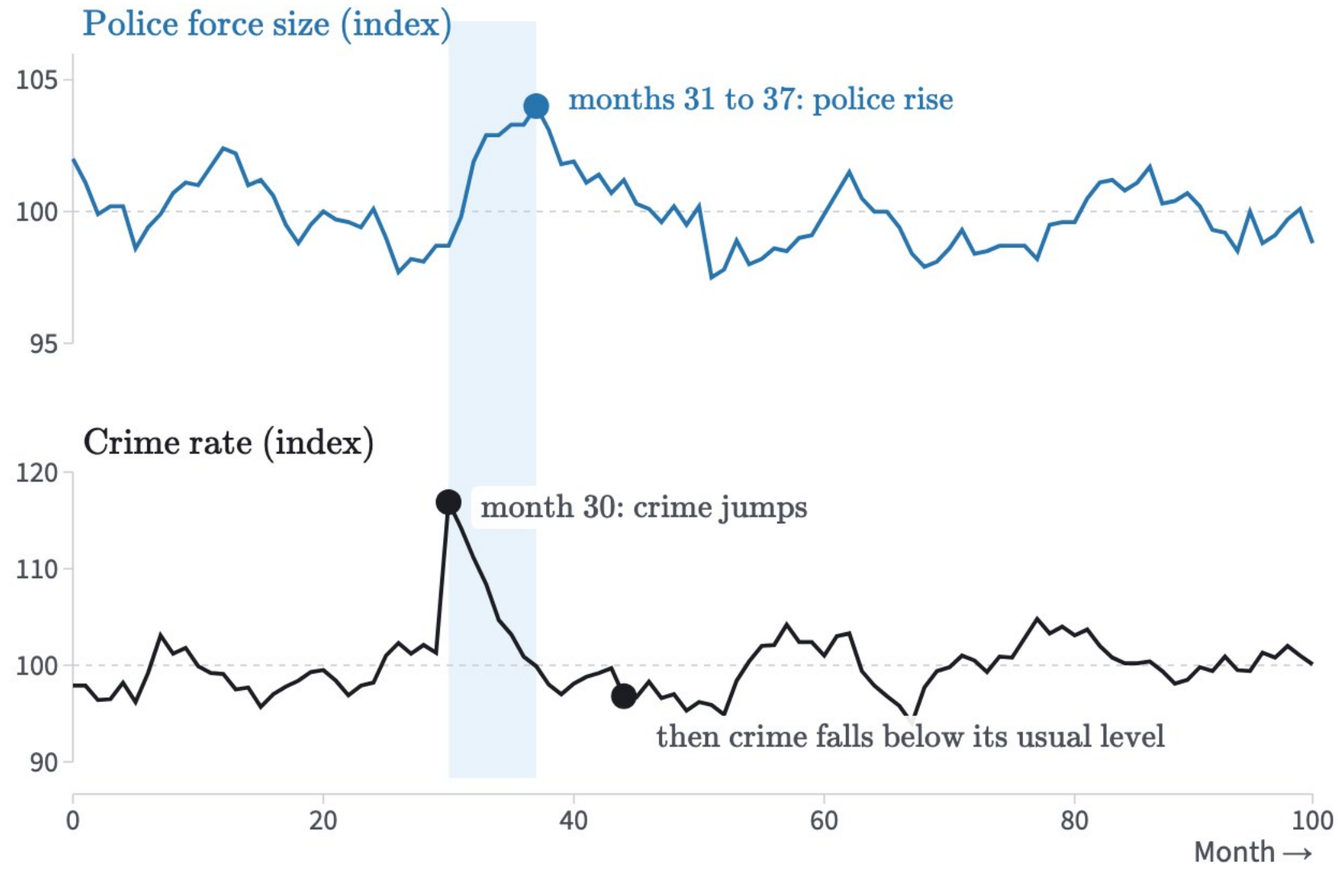
POLICE DEPEND ON PAST CRIME

High crime $\Longrightarrow$ the city hires more officers. Hiring takes months.

CRIME DEPENDS ON PAST POLICE

More officers on the street $\Longrightarrow$ less crime.

$\Longrightarrow$ There is no clean Y and X. Each variable belongs on both sides.



Made-up monthly data for one city, built so that each variable depends on the other's past.

The Granger Test: Does Past Police Improve the Forecast of Crime?

DEFINITION

X Granger causes Y if past X improves the forecast of Y, beyond what past Y already gives.

$$\text{Crime}_t = \delta + \theta_1 \text{Crime}_{t-1} + \theta_2 \text{Crime}_{t-2} + \beta_1 \text{Police}_{t-1} + \beta_2 \text{Police}_{t-2} + e_t$$

H_0: $\beta_1 = \beta_2 = 0 \iff$ police does not Granger cause crime

RESTRICTED MODEL: LAGS OF CRIME ONLY

$SSE_R = 441.8$

UNRESTRICTED MODEL: ADD 2 LAGS OF POLICE

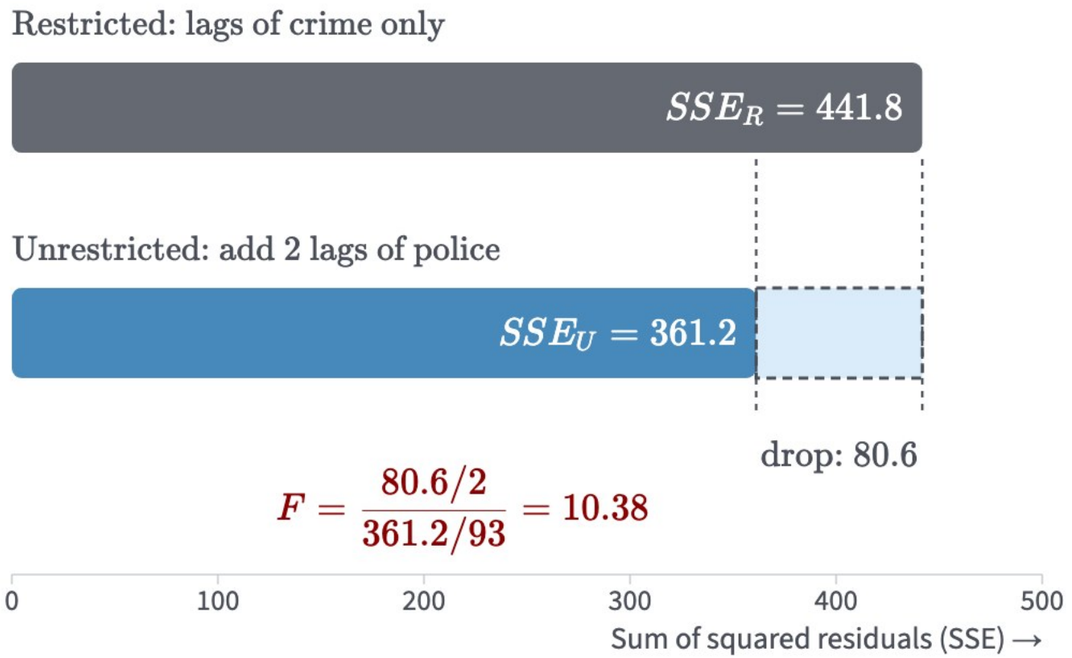
$SSE_U = 361.2$. The 2 lags lower the SSE by 80.6.

IS A DROP OF 80.6 LARGE? THE F-TEST

$$F = \frac{(SSE_R - SSE_U)/J}{SSE_U/(T-K)} = \frac{80.6/2}{361.2/(98-5)} = 10.38$$

$J = 2$ lags tested. $T = 98$ months, $K = 5$ coefficients.

REJECT H_0 $10.38 > F_{(0.95,\,2,\,93)} = 3.09 \implies$ police Granger causes crime.



SWAP THE ROLES

Police on 2 lags of police, then add 2 lags of crime: $F = 17.23 > 3.09 \implies$ crime Granger causes police. Both directions $\implies$ VAR: an equation for each variable, each with lags of both.

With real data on 56 U.S. cities and 49 states over 20+ years, Marvell and Moody (1996, *Criminology*) find Granger causation in both directions. The effect of police on crime is much larger.

Shuffle G and Rerun the Granger Test

LECTURE 7 TEST

Do 4 lags of GDP growth G improve the forecast of unemployment U, beyond 2 lags of U? Real data: $F = 5.98 > 2.41$, reject.

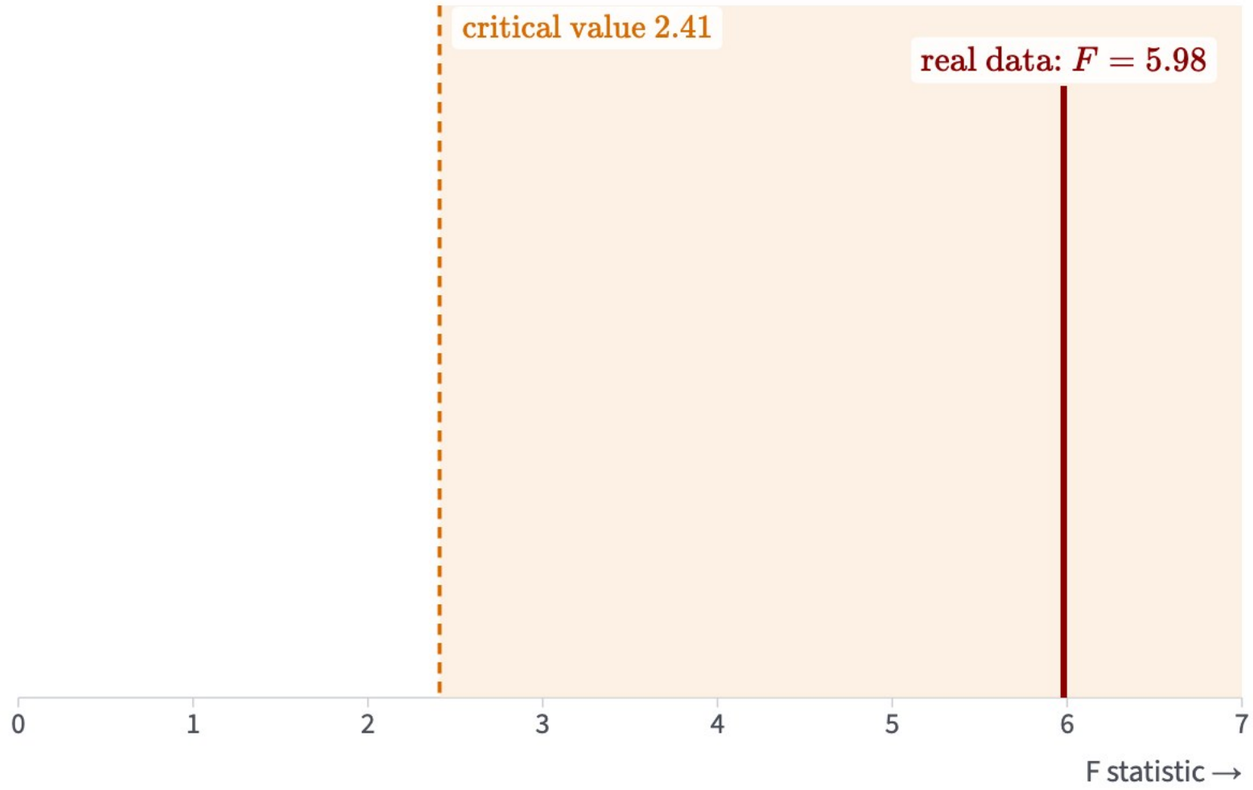
SHUFFLE

Put the 273 quarters of G in a random order and rerun the test. Shuffled G has no time order, so its lags carry no information about U.

Shuffle once | Shuffle 50 times | Reset

Press a button to shuffle.

Each dot is one shuffle. About 5% of shuffled F values land above the critical value 2.41. The real data give $F = 5.98$.



Computed in the browser from `usmacro` (POE5): the Lecture 7 models on the same 269 quarters.

Test 1: Do Lags of Conf Improve the Forecast of Sales?

```
Model 1: Sales ~ L(Sales, 1:2) + L(Conf, 1:2)
Model 2: Sales ~ L(Sales, 1:2)
  Res.Df Df      F  Pr(>F)
1     91
2     93 -2 6.8741 0.0016 **
```

TWO MODELS

`L(Conf, 1:2)` means lags 1 and 2 of Conf. Model 1 (unrestricted) includes them; Model 2 (restricted) drops them. H_0: both coefficients on lagged Conf are 0.

DF = −2

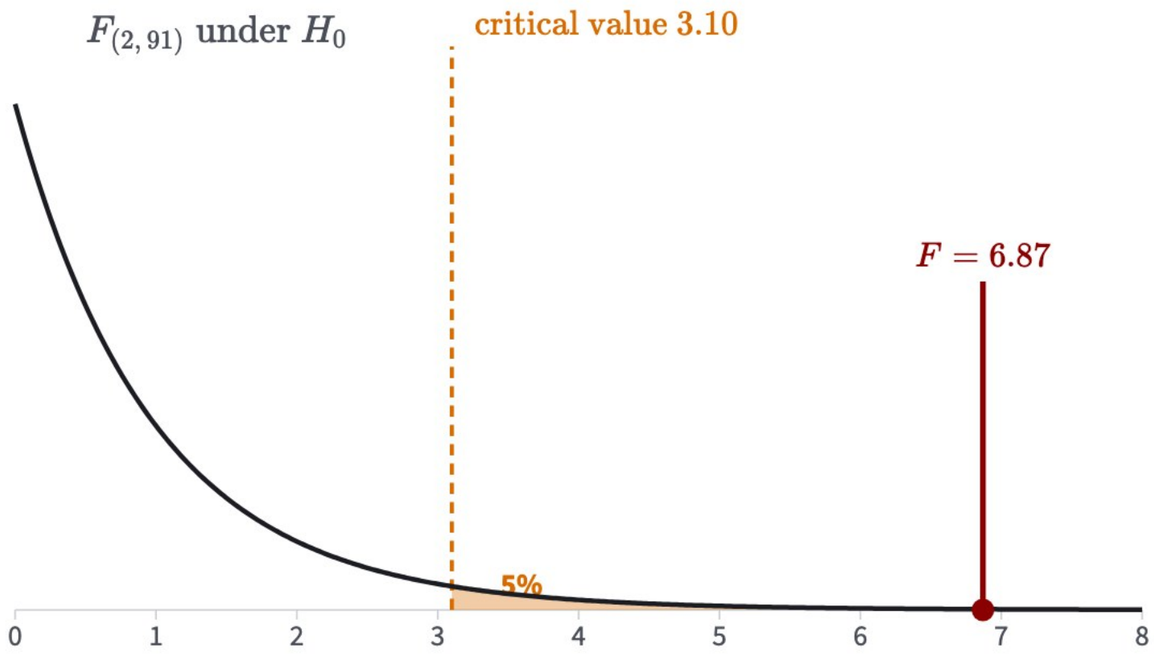
Res.Df is the residual degrees of freedom. Model 2 has 2 fewer coefficients, so its Res.Df is 2 higher; R prints the change as -2. That is $J = 2$ restrictions, and the F-test has $(2,\ 91)$ degrees of freedom, using Model 1's Res.Df as printed.

STATISTIC

$F = 6.87$. 5% critical value: $F_{(0.95,\,2,\,91)} = 3.10$.

REJECT H_0 $6.87 > 3.10$ $(p = 0.0016) \implies$ Conf Granger causes Sales.

> Both comparisons give the same decision: $F = 6.87 > 3.10$, or $p = 0.0016 < 0.05$. On the exam the p-value is enough.



Test 2: Do Lags of Sales Improve the Forecast of Conf?

```
Model 1: Conf ~ L(Conf, 1:2) + L(Sales, 1:2)
Model 2: Conf ~ L(Conf, 1:2)
  Res.Df Df      F  Pr(>F)
1     91
2     93 -2 1.2033 0.3047
```

THE OTHER DIRECTION

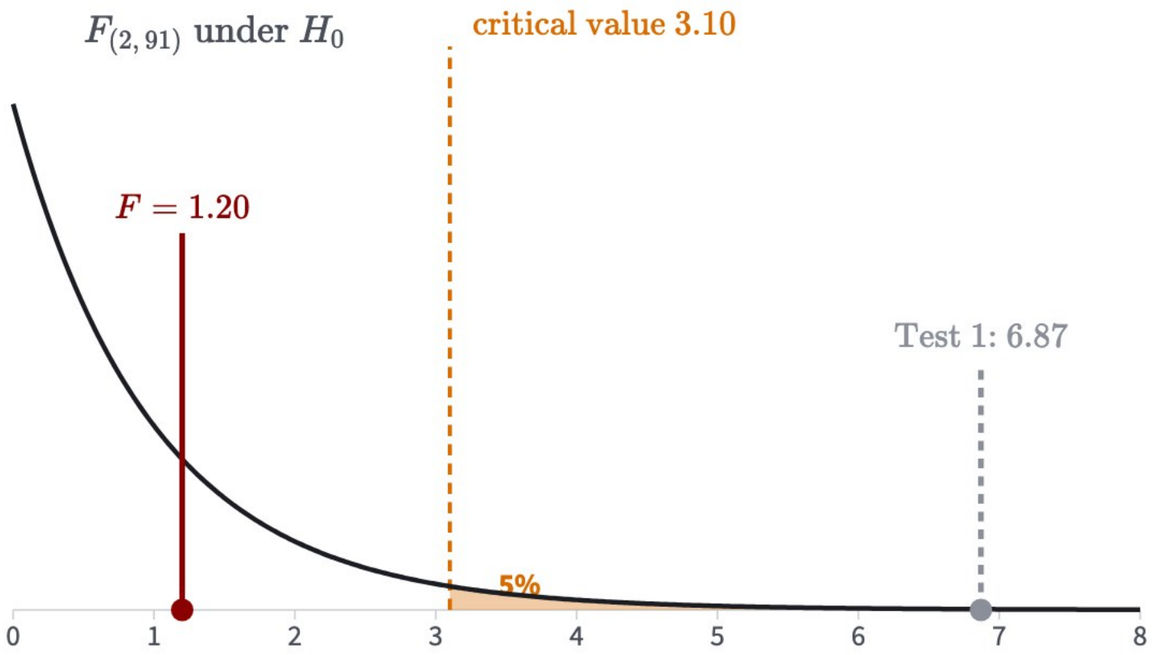
Now Conf is on the left. H_0: both coefficients on lagged Sales are 0. Same degrees of freedom, $(2,\ 91)$, so the same critical value, 3.10.

STATISTIC

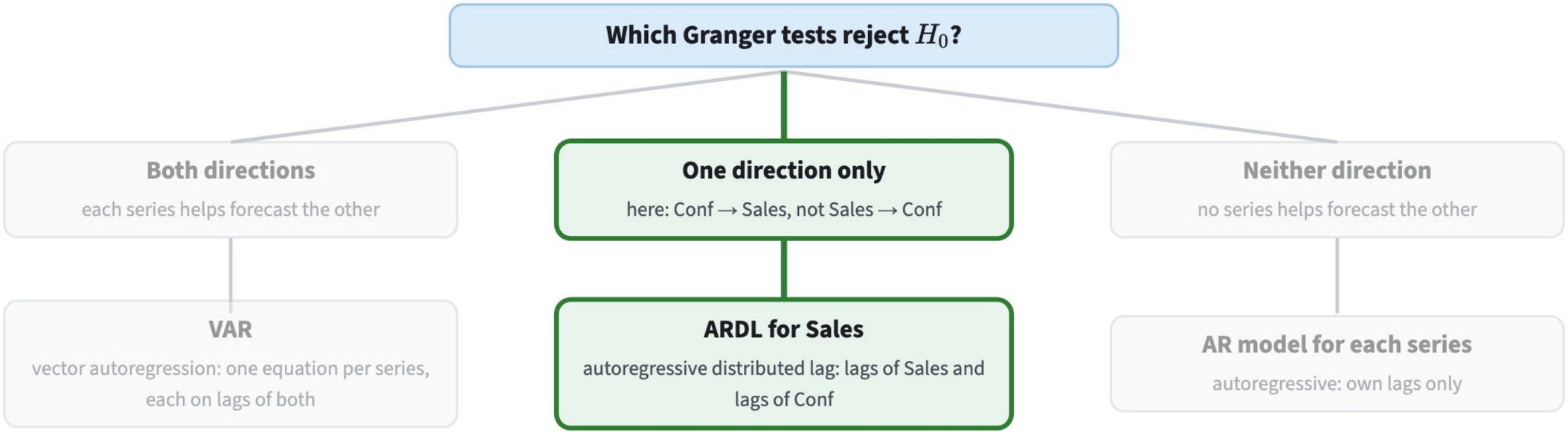
$F = 1.20$, below 3.10.

FAIL TO REJECT H_0 $1.20 < 3.10\ (p = 0.30) \implies$ no evidence that Sales Granger causes Conf.

Together: only the Conf → Sales test rejects at 5%. Conf Granger causes Sales, and we find no evidence that Sales Granger causes Conf.



Choose the Model from the Two Test Results



THE MODEL FOR SALES

$$\text{Sales}_t = \delta + \theta_1 \text{Sales}_{t-1} + \theta_2 \text{Sales}_{t-2} + \beta_1 \text{Conf}_{t-1} + \beta_2 \text{Conf}_{t-2} + e_t$$

Granger causality is about forecasting. It does not show that higher confidence causes higher sales.

Change the Two F Statistics and Watch the Model Choice

Critical value for both tests: $F_{(0.95,\,2,\,91)} = 3.10$. Start: the exam values.

Test 1: Conf → Sales **6.87**

6.87 > 3.10: reject H_0, Conf Granger causes Sales

Test 2: Sales → Conf **1.20**

1.20 ≤ 3.10: fail to reject H_0

	Test 2 rejects (Sales → Conf)	**Test 2 does not reject**
Test 1 rejects (Conf → Sales)	**VAR** each series on lags of both	**ARDL for Sales** lags of Sales and lags of Conf
Test 1 does not reject	**ARDL for Conf** lags of Conf and lags of Sales	**AR model for each** own lags only

Checking Each Answer Choice

a Both directions; a VAR model is required.

Test 2 does not reject ($p = 0.30$), so Sales does not Granger cause Conf.

b Sales Granger causes Conf, but not the reverse.

This reverses the direction. The test that rejects ($p = 0.0016$) has Sales on the left.

d Neither series Granger causes the other.

Test 1 rejects H_0: $6.87 > 3.10$.

e Inconclusive: the row for Model 1 is missing.

Nothing is missing. R prints Df and F once, on row 2, because the test compares the two rows.

c Conf Granger causes Sales, but Sales does not Granger cause Conf; the ARDL model with lagged Conf as a predictor of Sales is appropriate.

Answer: (c)

INTERPRETATION

Consumer confidence in the last two quarters improves the forecast of retail sales growth, beyond what past sales growth gives. We find no evidence that past sales growth improves the forecast of confidence.

> A note on the output: the printed Res.Df values (91 and 93) count all 96 quarters. The first 2 quarters have no lagged values, so only 94 quarters can be used, and the values should be $94 - 5 = 89$ and $94 - 3 = 91$. The 5% critical value is 3.10 either way, so the decision is the same.

Umbrellas Granger Cause Rain

THE DATA

Every day, count the umbrellas people carry across campus at 8 am, and record whether it rains at 3 pm.

THE TEST

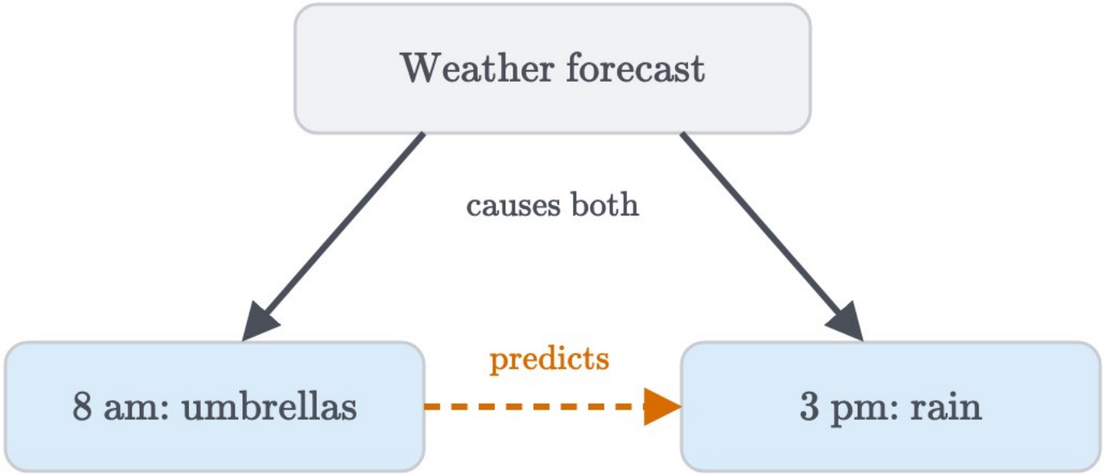
Days with many umbrellas in the morning are more often rainy in the afternoon. Morning umbrellas improve the forecast of rain: umbrellas Granger cause rain.

THE COMMON CAUSE

People check the weather forecast before they leave. The forecast causes both. Taking everyone's umbrella away would not stop the rain.

BACK TO REGRESSION

Conf may predict Sales because news about the economy appears in survey answers before it appears in sales. A causal effect needs an identification strategy: an instrument, a natural experiment, or a quasi-experiment (Lecture 7).



hiding the umbrellas does not stop the rain

The Question: Predicted Probability from the LPM

Suppose we're modeling whether someone will subscribe to an influencer's YouTube channel (Y: 1 = subscribe, 0 = doesn't subscribe) based on the number of Youtube video views at the time of watching the video and the number of ads that pop up during the video. Use the following table of coefficients from a sample of 5,000 users:

Model	Intercept	Ads	Video Views
Probit	-1.0	0.3	0.05
LPM	0.2	0.1	0.02

Question 15. What is the estimated predicted probability of subscribing to the YouTube channel for a user who sees 5 ads and where the video has 20 views, according to the LPM (censor your results where necessary if it falls outside the bounds of probability assumptions.)?

a 0.0

b 0.5

c 0.6

d 1.0

e None of the above

Source: Fall 2024 Midterm V1 (Boswell), Question 15

A Straight Line Leaves the Range 0 to 1

THE LPM FROM QUESTION 15

$$\overline{\text{Probability of subscribing}} = \hat{p} = 0.2 + 0.1\,\text{Ads} + 0.02\,\text{Video Views}$$

SAME SLOPE EVERYWHERE

Each ad adds 0.1 to $\hat{p}$, and each video view adds 0.02, however many the user has already seen.

HOLD VIDEO VIEWS AT 20

$\hat{p} = 0.2 + 0.02(20) + 0.1\,\text{Ads} = 0.6 + 0.1\,\text{Ads}$

ABOVE 1 AFTER 4 ADS

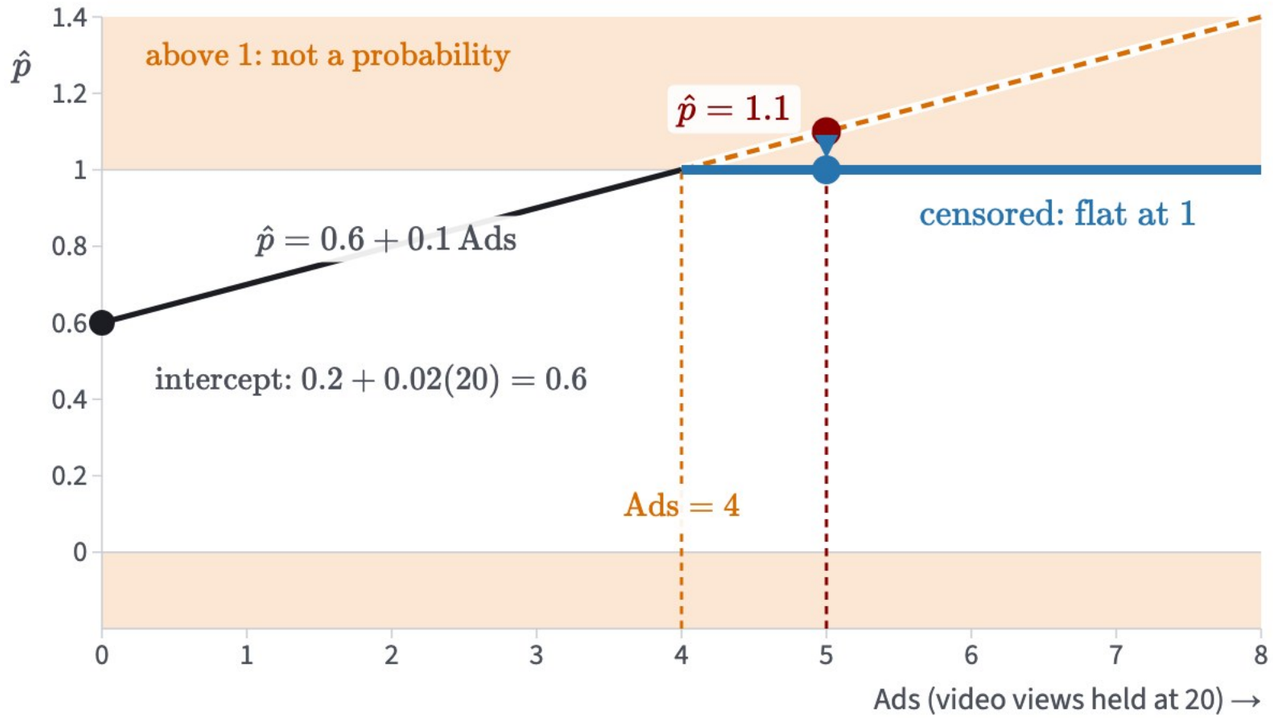
$0.6 + 0.1\,\text{Ads} > 1 \iff \text{Ads} > 4$. A straight line keeps rising, so it passes 1.

The exam user, 5 ads and 20 views:

$$\hat{p} = 0.2 + 0.1(5) + 0.02(20) = 1.1$$

COMMON PRACTICE: CENSOR

Report 1 for any $\hat{p}$ above 1 (and 0 for any below 0). The line is flat at 1 after 4 ads, so the exam user gets 1.0.



Checking Each Answer Choice: Question 15

a 0.0

Censoring moves 1.1 down to 1, not to 0.

b 0.5

Only the ads term, 0.1×5.

c 0.6

Leaves out the ads term: $0.2 + 0.4$.

e None of the above

Choice (d) holds after censoring.

d 1.0

Answer: (d)

INTERPRETATION

The LPM predicts 1.1 for this user. The question says to censor: replace a value above 1 with 1 (and a value below 0 with 0). The reported prediction is 1.0, the largest possible probability.

Censoring changes the reported number, not the model: the slope is still 0.1 per ad at every number of ads. The correction is a model whose predictions stay between 0 and 1 for every x: probit or logit (Problem 3).

The Question: Confidence Interval for the LPM Marginal Effect

Same setup: subscribing to a YouTube channel, 5,000 users.

Model	Intercept	Ads	Video Views
Probit	-1.0	0.3	0.05
LPM	0.2	0.1	0.02

Question 17. If the marginal effect of ads on the probability of subscribing has a standard error of 0.04 for the LPM, calculate the 95% confidence interval for the marginal effect, rounded to two decimal places. (The critical value is 1.96). Is this confidence estimate precise, and why or why not?

a $[0.02, 0.18]$; it is precise because the interval is narrow.

b $[0.1, 0.5]$; it is not precise because the interval is wide due to small sample size.

c $[0.1, 0.5]$; it is not precise because the interval is wide due to multicollinearity.

d $[0.02, 0.18]$; it is not precise because the interval is wide due to heteroscedasticity.

e None of the above answers are correct and we need more information to speak about the precision.

Source: Fall 2024 Midterm V1 (Boswell), Question 17

The LPM Error Variance Changes with x

y IS 0 OR 1 (BERNOULLI)

$$\mathrm{Var}(y_i \mid x_i) = p_i(1 - p_i), \;\; p_i = \beta_1 + \beta_2 x_i$$

THE VARIANCE DEPENDS ON p

$p = 0.5$: variance 0.25. $p = 0.1$ or 0.9: variance 0.09.

AND p DEPENDS ON x

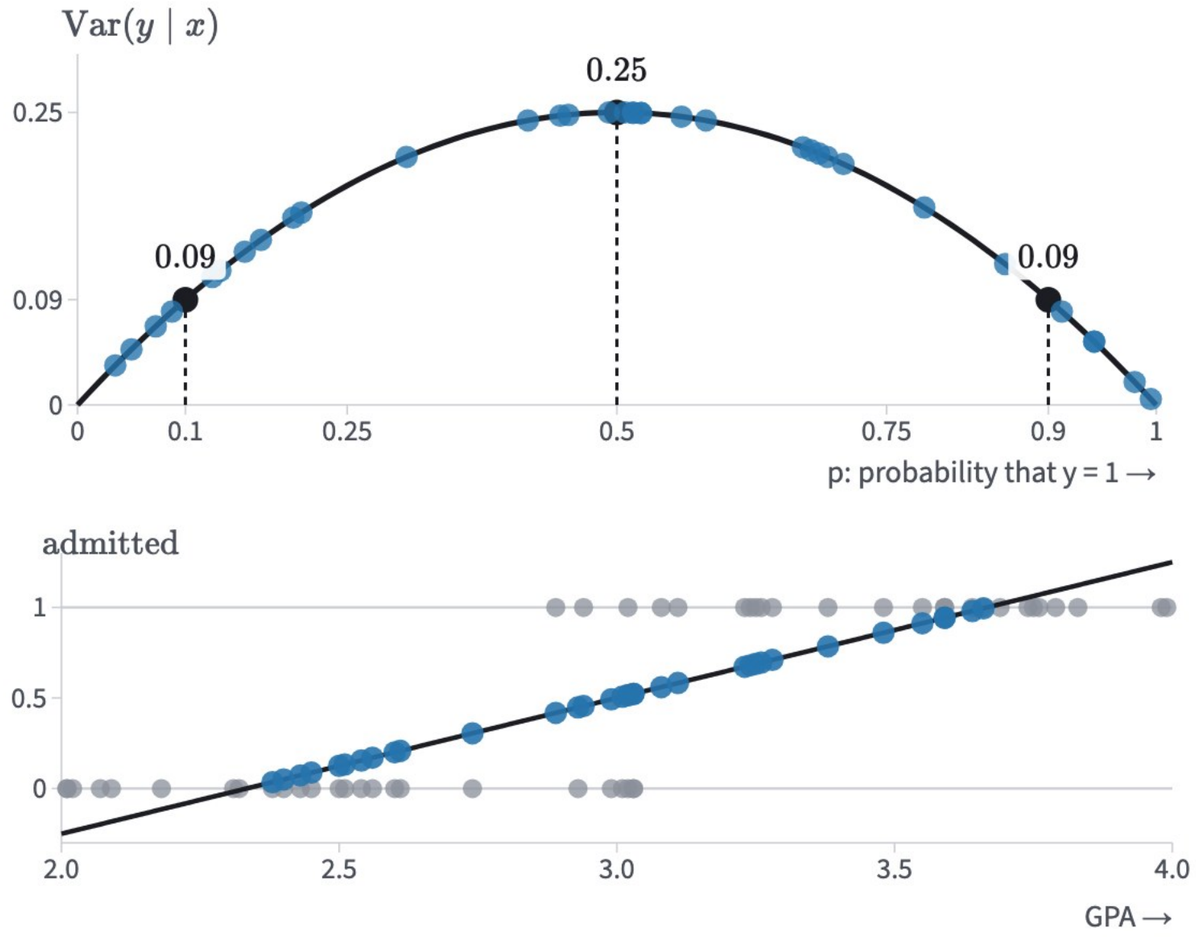
The error is $e_i = y_i - p_i$, so $\mathrm{Var}(e_i \mid x_i) = \mathrm{Var}(y_i \mid x_i)$. Because p_i changes with x_i, the LPM errors are heteroskedastic unless $\beta_2 = 0$.

CONSEQUENCE

OLS coefficients are still unbiased (Lecture 8) but not efficient, and the usual OLS standard errors are not valid.

CORRECTION

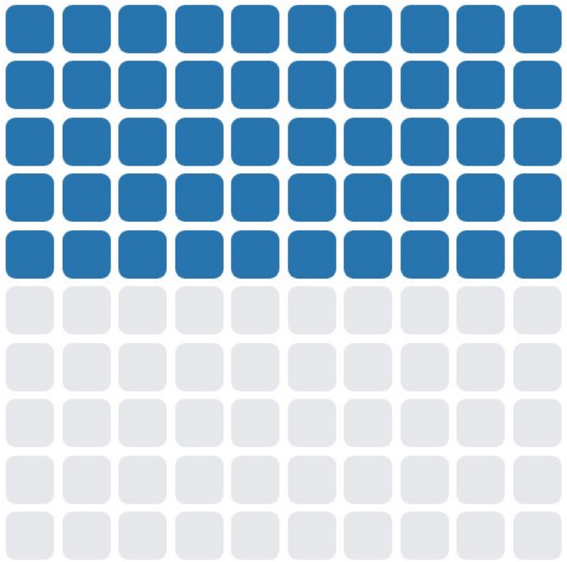
Use heteroskedasticity-robust standard errors (week 2 discussion), or feasible generalized least squares (FGLS) with weights $1/\big(\hat{p}_i(1 - \hat{p}_i)\big)$, which needs every $\hat{p}_i$ strictly between 0 and 1.



Bottom: made-up admissions data (50 applicants) and its LPM line, $\hat{p} = -1.75 + 0.75\,\mathrm{GPA}$. Blue: the fitted values between 0 and 1, on both charts.

Move p: The Variance of a 0/1 Outcome

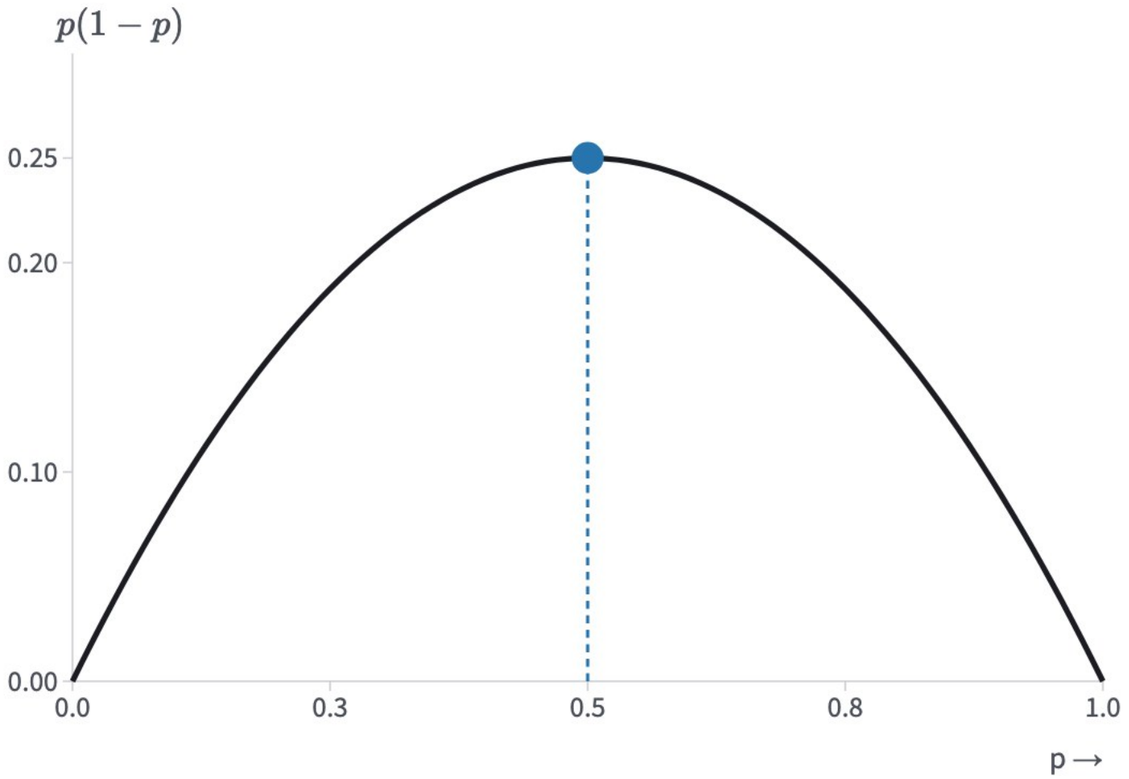
$p = \Pr(y = 1)$ **0.50**



100 people with the same x. Blue: $y = 1$. Gray: $y = 0$.

50 of 100 have y = 1

Variance: 0.50 × 0.50 = **0.2500**



All 1s or all 0s: no variation. Half and half: the most variation. In the LPM, p moves with x, so the variance moves with x.

Build the Interval for the Ads Coefficient

FORMULA

$$\hat{\beta}_{\text{Ads}} \pm 1.96 \times se(\hat{\beta}_{\text{Ads}})$$

In the LPM the marginal effect of ads is the coefficient itself: 0.1.

MARGIN

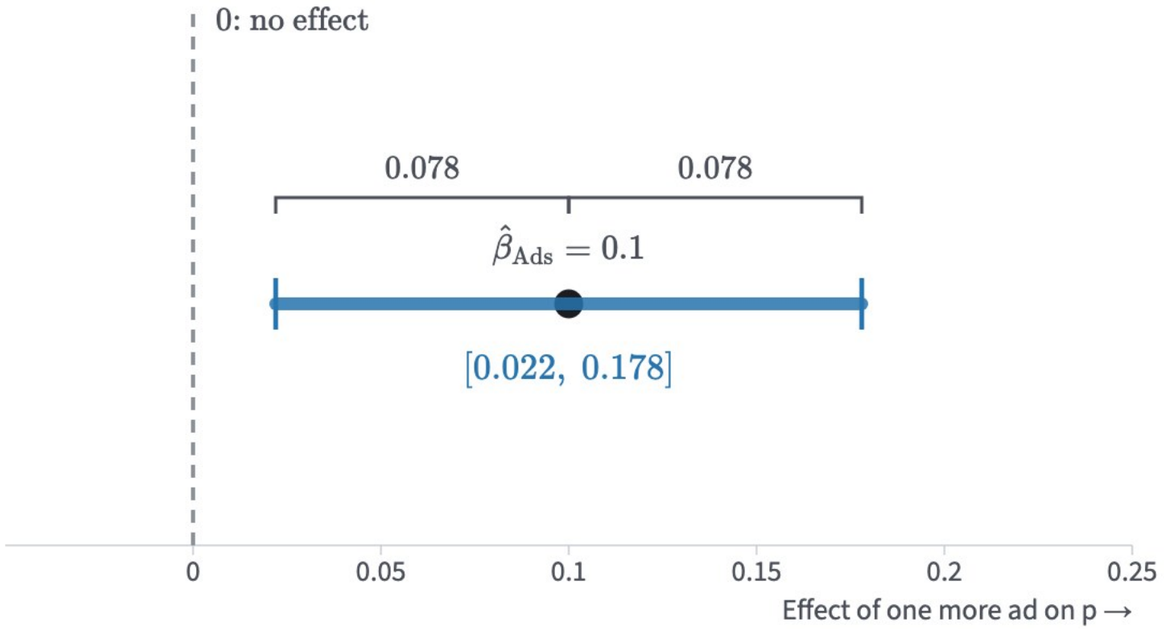
$1.96 \times 0.04 = 0.078$

INTERVAL

$$0.1 \pm 0.078 = [0.022,\ 0.178] \approx [0.02,\ 0.18]$$

The upper end is about 8 times the lower end: one more ad could raise the probability of subscribing by 2 percentage points or by 18.

> For the interval to be valid, 0.04 must be a heteroskedasticity-robust standard error (previous slide).



Checking Each Answer Choice: Question 17

b $[0.1, 0.5]$; wide due to small sample size.

0.1 ± 0.078 is not $[0.1, 0.5]$. And $n = 5{,}000$ is not small.

c $[0.1, 0.5]$; wide due to multicollinearity.

Wrong interval.

a $[0.02, 0.18]$; precise because the interval is narrow.

Right interval, but it runs from 0.02 to 0.18 around an estimate of 0.1.

e None of the above; need more information.

Can be argued, since the reason in (d) is weak. The posted answer is (d).

d $[0.02, 0.18]$; not precise, wide due to heteroscedasticity.

Answer: (d)

ON THE EXAM, CHOOSE (D)

Its reason is loosely worded. Heteroskedasticity does not by itself make an interval wider: it makes the usual OLS standard error invalid, and robust standard errors can be larger or smaller. The best case for (d): with heteroskedasticity OLS is not efficient, so FGLS would give a narrower interval.

LPM problem	Consequence	Correction
Straight line in x	$\hat{p}$ can be below 0 or above 1 (Q15: 1.1)	Probit or logit
Variance $p(1-p)$ changes with x	Usual standard errors not valid; OLS not efficient	Robust standard errors

PROBLEM 3 · PROBIT PREDICTIONS

The Question: The Same User, Now with Probit

Same setup: subscribing to a YouTube channel, 5,000 users.

Model	Intercept	Ads	Video Views
Probit	-1.0	0.3	0.05
LPM	0.2	0.1	0.02

Question 16. What is the difference between the estimate obtained in question 15, and the probability of subscribing to the YouTube channel for a user who currently sees 5 ads where the video has 20 views, using the probit model? (Note that $\Phi(1.5) = 0.9332$ (CDF) and $\phi(1.5) = 0.1295$ (PDF))

a 0.0668

b 0.8705

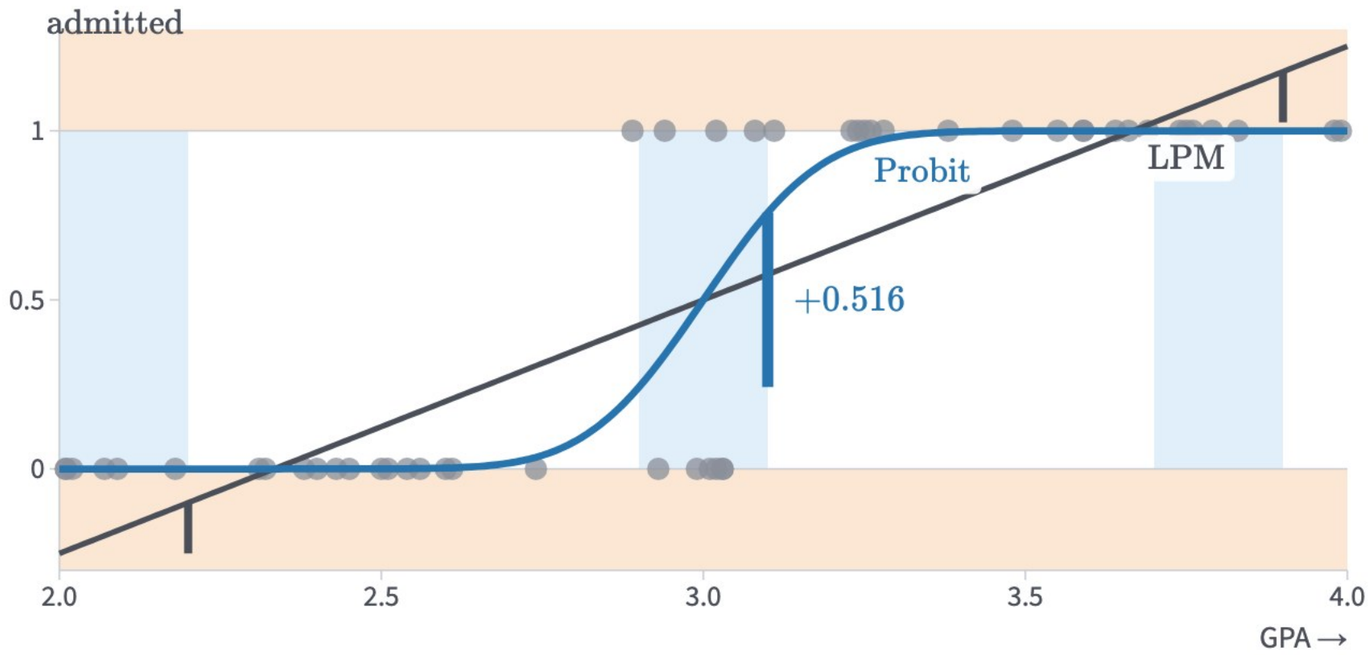
c 0.1295

d 0.9332

e None of the above answers are correct

Source: Fall 2024 Midterm V1 (Boswell), Question 16

Why Use Probit: GPA and the Probability of Admission



Made-up data for 50 applicants. $y = 1$ if admitted.

LPM

$$\widehat{\Pr}(\text{admit}) = -1.75 + 0.75\,\text{GPA}$$

PROBIT

$$\widehat{\Pr}(\text{admit}) = \Phi(-21 + 7\,\text{GPA})$$

Index $= 0$ at $\text{GPA} = 3.0$, so $\hat{p} = 0.5$ there.

Raise GPA by 0.2	LPM	Probit
2.0 → 2.2	+0.15	+0.000
2.9 → 3.1	+0.15	+0.516
3.7 → 3.9	+0.15	+0.000

THE LPM SAYS

"A little bit higher GPA makes a difference, no matter where you start!"

PROBIT SAYS

"GPA improvements don't really matter if you have a 2.0 or a 3.9, but it makes a big difference going from 2.9 to 3.1."

Probit Puts the Linear Index Inside the Normal CDF

PROBIT

$$p_i = \Phi(\beta_1 + \beta_2 x_i)$$

$\Phi(z) = \Pr(Z \leq z)$ for a standard normal Z: the normal CDF.

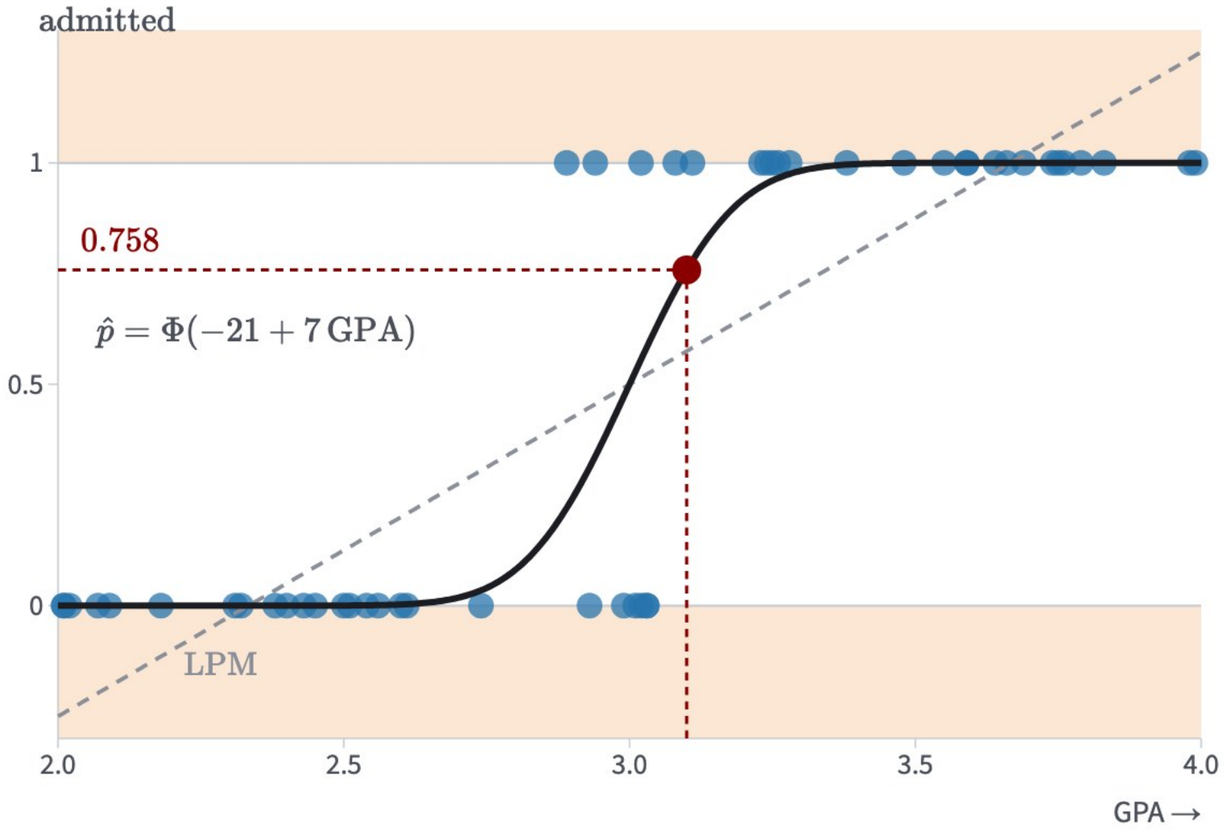
INDEX AND PROBABILITY

The index $z = \beta_1 + \beta_2 x$ can be any number. $\Phi(z)$ is always between 0 and 1.

ADMISSIONS DATA

$\hat{p} = \Phi(-21 + 7\,\text{GPA})$. With a GPA of 3.1, $z = -21 + 7(3.1) = 0.7$ and $\hat{p} = \Phi(0.7) = 0.758$.

> The marginal effect is no longer constant: the curve is steepest near $p = 0.5$ and nearly flat near 0 and 1.



Made-up admissions data for 50 applicants, the same data as slide 16. Dashed: the LPM line.

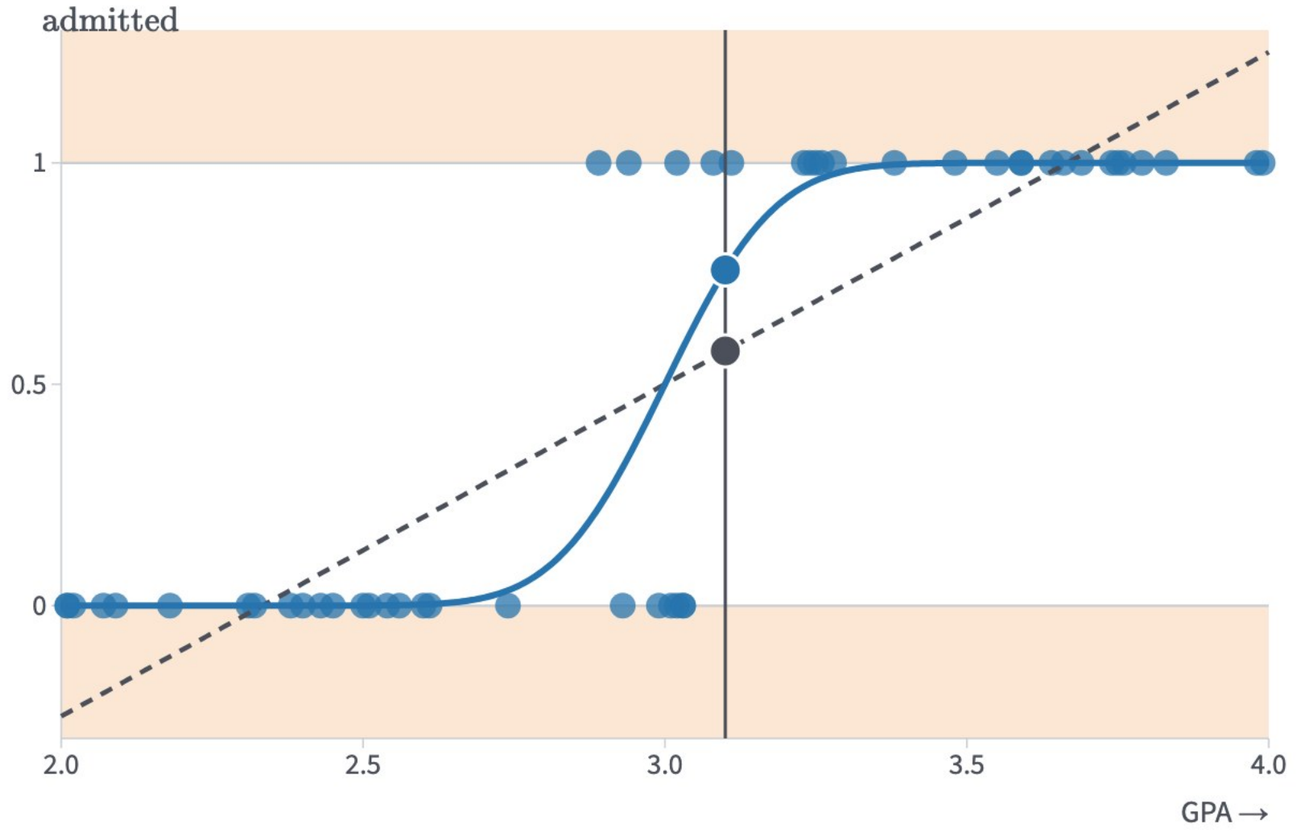
Move Along GPA: LPM and Probit Predictions

GPA **3.10**

LPM: $-1.75 + 0.75 \times 3.10 = \mathbf{0.575}$

Probit: $\Phi(0.70) = \mathbf{0.758}$

The LPM line goes below 0 for GPA below 2.33 and above 1 for GPA above 3.67. The probit curve stays between 0 and 1 for every GPA.



Made-up admissions data for 50 applicants.

The Probit Marginal Effect Changes Along the Curve

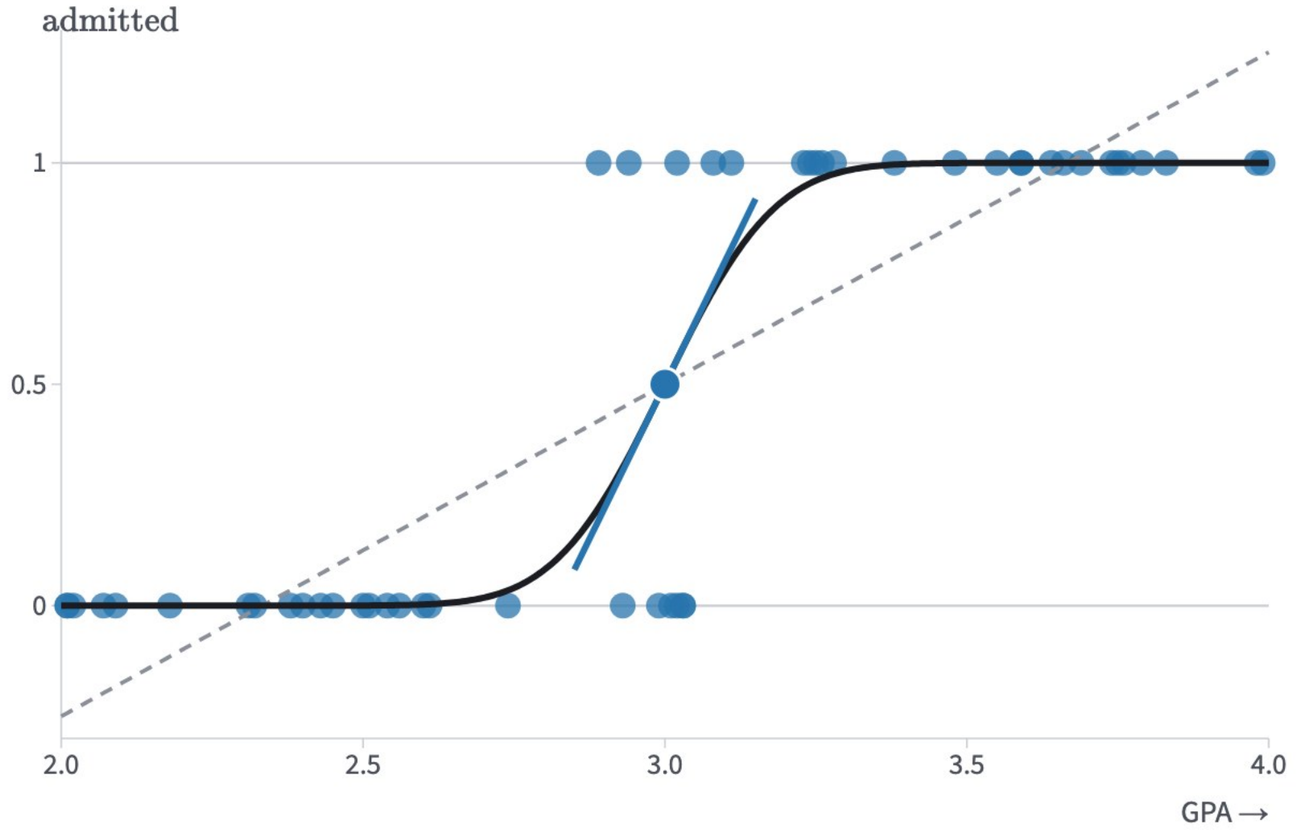
GPA **3.00**

$z = 0.000$, $\hat{p} = 0.500$

Probit slope: $\phi(0.00) \times 7 =$ **2.793**

LPM slope: 0.75

The slope of the tangent line is the marginal effect: $\phi(z) \times 7$, with $z = -21 + 7\,\mathrm{GPA}$. The LPM slope is 0.75 at every GPA.



Made-up admissions data for 50 applicants. Slopes are per 1 point of GPA; for 0.1 points, divide by 10.

The Probability Is the Area to the Left of 1.5

THE PROBIT PREDICTION: 5 ADS, 20 VIEWS

$$\hat{p} = \Phi(z), \quad z = \hat{\beta}_0 + \hat{\beta}_1 \,\text{Ads} + \hat{\beta}_2 \,\text{Views}$$
$$z = -1.0 + 0.3(5) + 0.05(20)$$
$$= -1.0 + 1.5 + 1.0 = 1.5$$
$$\hat{p} = \Phi(1.5)$$

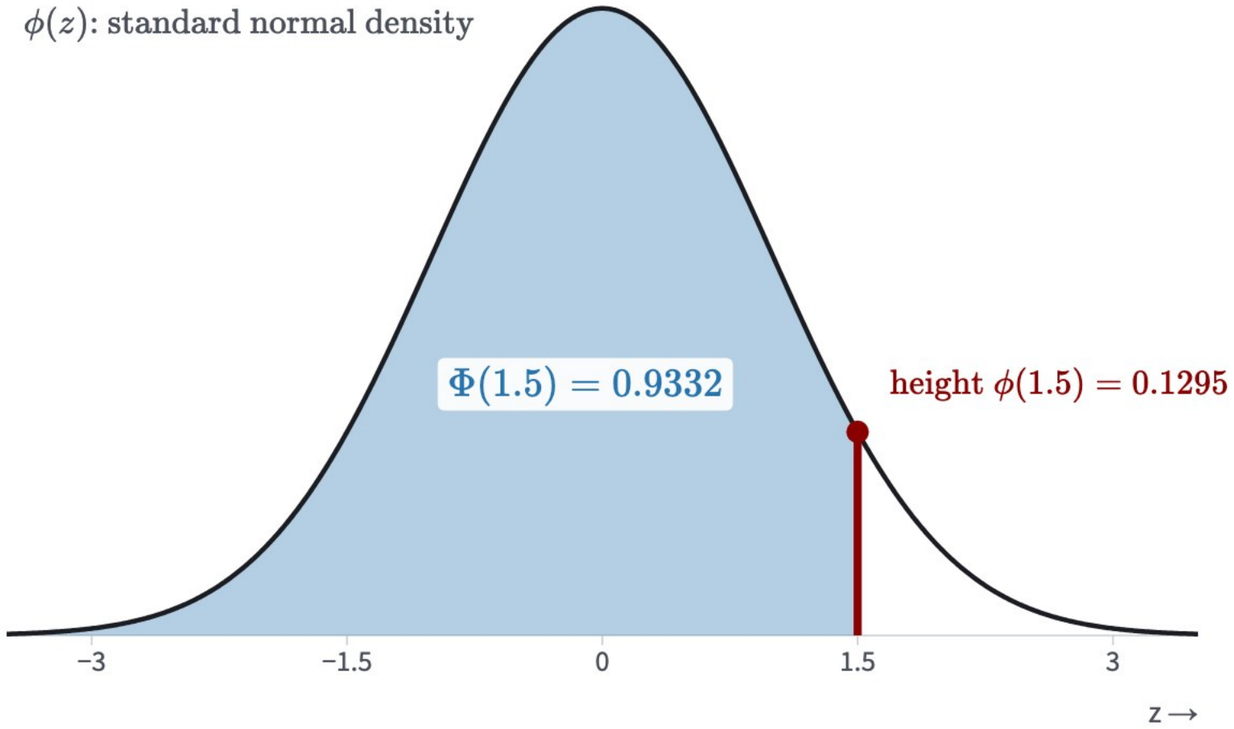
Ask: The question gives two numbers. Which one is the probability?

$\Phi(1.5) = 0.9332$

The area under the standard normal curve to the left of 1.5. This is $\hat{p}$, the probit probability of subscribing.

$\phi(1.5) = 0.1295$

The height of the curve at 1.5. It is used for marginal effects, $\phi(z)\,\beta$. It is not a probability.



Compare to the LPM Answer from Question 15

QUESTION 15 (LPM)

$\hat{p} = 1.1$, censored to 1.0.

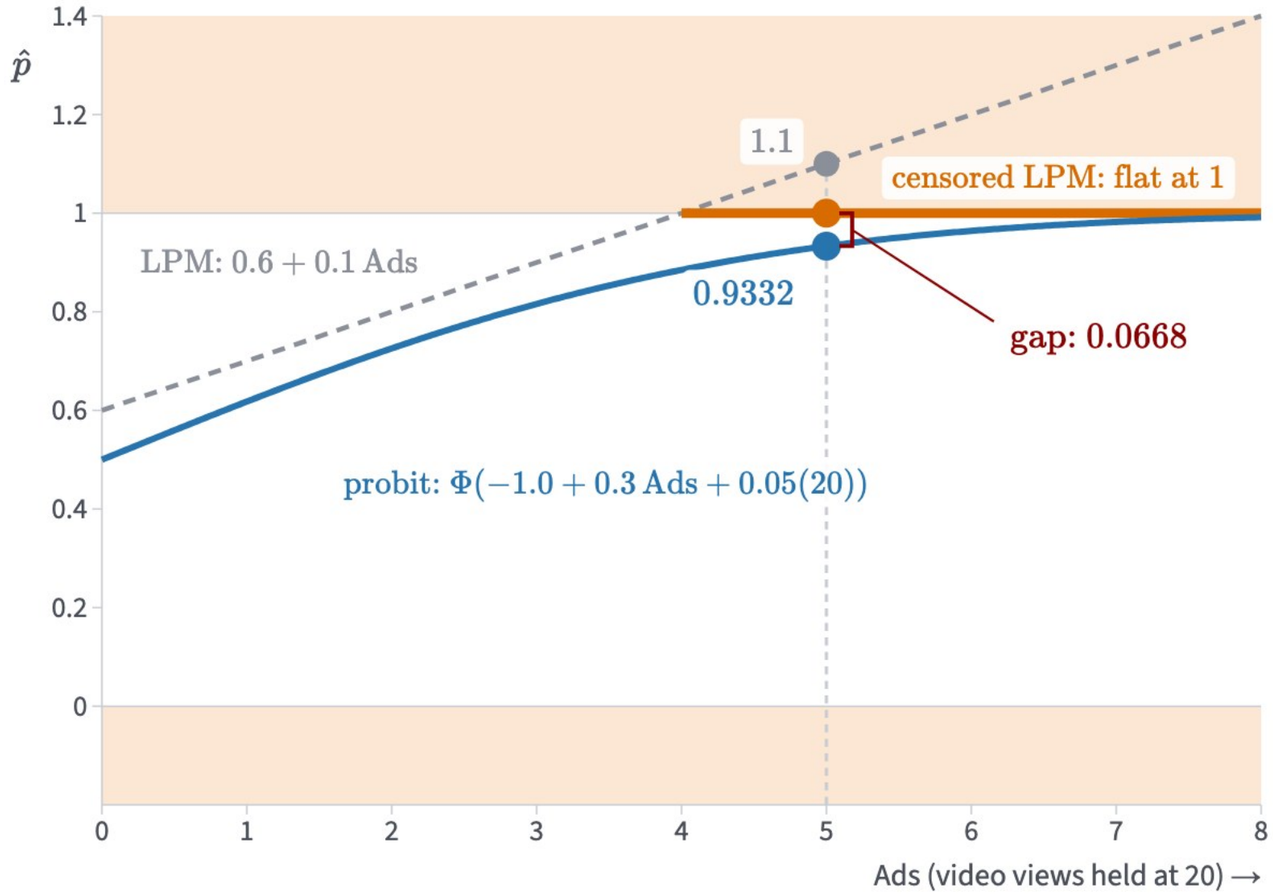
QUESTION 16 (PROBIT)

$\hat{p} = \Phi(1.5) = 0.9332$

DIFFERENCE

$1.0 - 0.9332 = 0.0668$

The difference is the area to the right of 1.5: $1 - \Phi(1.5) = \Phi(-1.5) = 0.0668$, the probit probability of not subscribing. This holds only because the censored LPM answer is exactly 1.



Checking Each Answer Choice: Question 16

c 0.1295

$\phi(1.5)$: the height of the curve, not a probability.

b 0.8705

$1 - \phi(1.5)$: uses the height in place of the area.

d 0.9332

The probit probability itself, not the difference from the LPM.

e None of the above answers are correct

Choice (a) holds.

a 0.0668

Answer: (a)

INTERPRETATION

For a user who sees 5 ads on a video with 20 views, the probit model predicts a 93.3% chance of subscribing. The censored LPM prediction, 100%, is 6.7 percentage points higher.

Thank you!

jakeanderson@g.ucla.edu