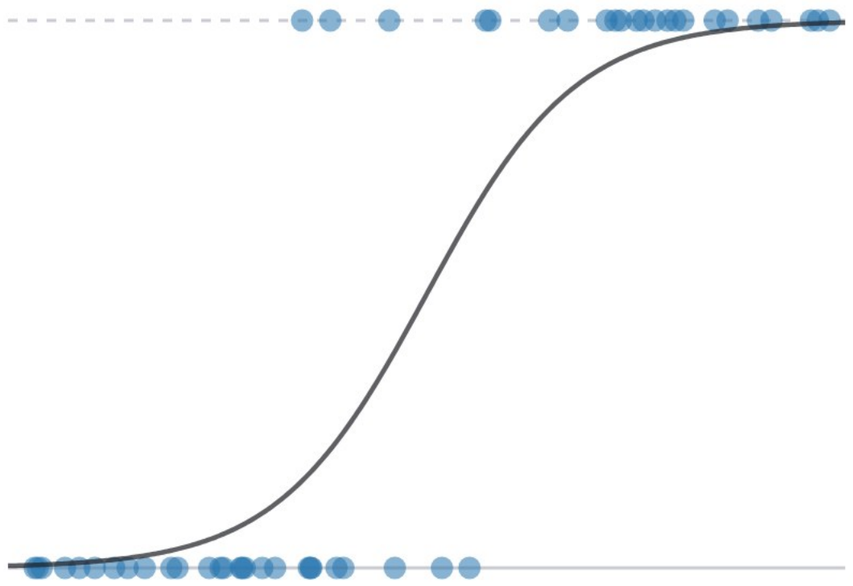
ECON 104 · MIDTERM REVIEW · FALL 2026

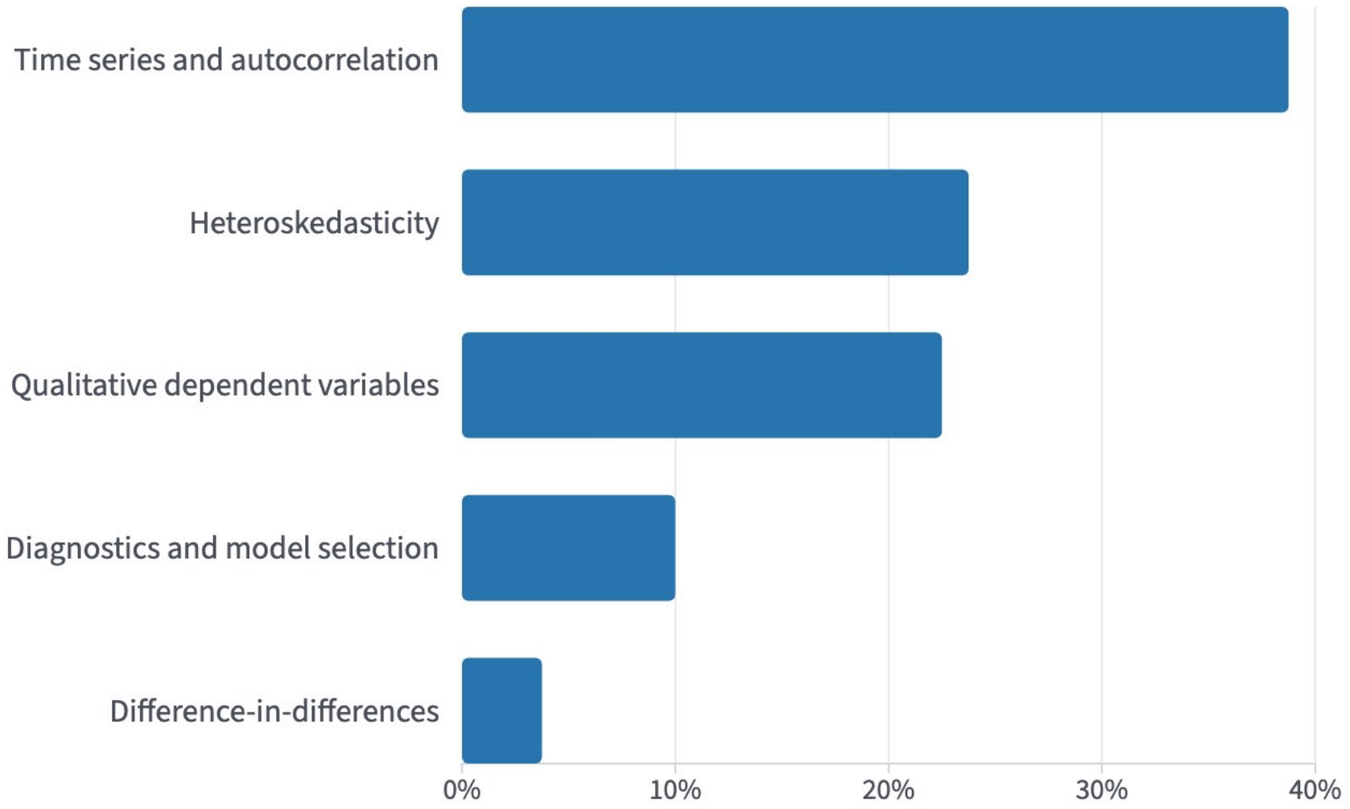
Midterm Review

Jake Anderson
UCLA



PAST EXAMS · 4 MIDTERMS, 80 QUESTIONS

Which Topics Appear on the Midterm



Share of questions by topic on the Fall 2024, Spring 2025, Fall 2025, and Spring 2026 midterms (20 questions each, all by Professor Boswell).

ON EVERY ONE OF THE 4 EXAMS

A heteroskedasticity test (Goldfeld-Quandt or Breusch-Pagan)
Stationarity, an ACF, or a PACF
Granger causality
A logit or probit probability or marginal effect
Comparing binary models: confusion matrix, ROC, or AIC

Four Kinds of Questions and an Approach for Each

The question asks you to	Count	Share	Approach
Pick the correct statement	42	53%	Ask what the statement is about: the coefficient (bias) or its standard error (efficiency, inference)? Most wrong choices mix up the two.
Compute a number	24	30%	Write the formula first, then plug in. Check units ($1000s, 10-minute units) and which value goes in the numerator.
Read R output or a figure	12	15%	Write H_0 before looking at the numbers. Then compare the p-value to 0.05 and say the conclusion in words.
Match a scenario to a model	2	3%	Name the type of the dependent variable first: 0/1, a count, ranked categories, unranked choices, or censored.

Each of the 80 questions was tagged by topic by hand.

The Ten Problems

1
VIF and overfitting
Which variable to drop; train vs test error

2
Goldfeld-Quandt and FGLS
The F ratio, the FGLS weight, robust SEs

3
Breusch-Pagan and WLS
NR^2, robust intervals, OLS vs WLS

4
Stationarity and ACF
ADF, $r_k\sqrt{T}$, choosing AR(p)

5
Forecasting
AR(2) interval, BG test, corrections

6
Lags and Granger
Lag operator, IDL multiplier, ARDL or VAR

7
Logit
Probability, marginal effect, odds ratio

8
Probit and LPM
Φ, symmetry, predictions above 1

9
Choosing a model
Precision, recall, F1, ROC, model types

10
Difference-in-differences
Reading the coefficients, parallel trends

Each problem is built from 3 to 6 linked questions, mostly from past midterms. Parts marked **new** cover Lecture 9 and 10 material that no past midterm has tested yet.

PROBLEM 1(A) · MULTICOLLINEARITY

The Question: Which Variable Should Be Dropped?

You examine a regression model for TikTok engagement using 150 creators: Engagement is likes + shares in one week; Followers is in 1,000s; Posts is the number of posts that week; Hashtags is the average number of hashtags per post.

$$\ln(\text{Engagement}_i) = \beta_0 + \beta_1 \text{Followers}_i + \beta_2 \text{Posts}_i + \beta_3 \text{Hashtags}_i + \beta_4 \text{AvgCaptionLength}_i + u_i$$

	Followers	Posts	Hashtags	AvgCaptionLength
Followers	1.00	0.22	0.35	0.88
Posts	0.22	1.00	0.41	0.20
Hashtags	0.35	0.41	1.00	0.29
AvgCaptionLength	0.88	0.20	0.29	1.00

	Followers	Posts	Hashtags	AvgCaptionLength
VIF	6.1	1.5	1.9	5.8

Given the correlation matrix and the variance inflation factors (VIFs), what would be the best guidance regarding variable selection?

a. No variables should be removed; all VIFs are under 10

b. Posts and Hashtags should be removed because their VIFs are low

c. All variables should be retained; correlation does not imply multicollinearity

d. Hashtags should be removed because it is strongly correlated with Followers

e. Followers and AvgCaptionLength are collinear and one of them should be dropped

Source: Fall 2025 Midterm V1 (Boswell), Question 1

VIF: The Other Regressors' Fit to x_j

DEFINITION

Regress x_j on all the other regressors and keep its R_j^2. Then

$$\text{VIF}_j = \frac{1}{1 - R_j^2}$$

FOLLOWERS: VIF = 6.1

$R_j^2 = 1 - 1/6.1 = 0.84$. The other three regressors explain 84% of the variation in Followers.

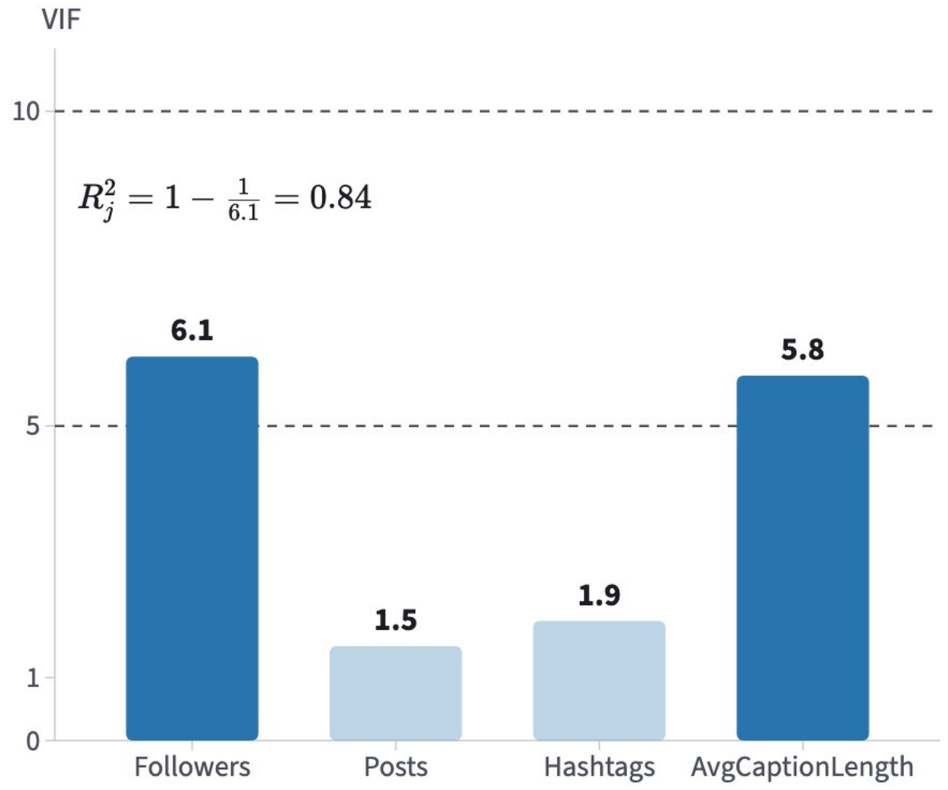
EFFECT ON THE STANDARD ERROR

The variance of $\hat{\beta}_1$ is 6.1 times what it would be if Followers were uncorrelated with the others, so its standard error is $\sqrt{6.1} = 2.5$ times larger.

WHICH PAIR

The correlation matrix shows where the overlap comes from: Followers and AvgCaptionLength have correlation 0.88. Every other pair is 0.41 or lower.

$\Longrightarrow$ Drop one of Followers and AvgCaptionLength. The one you keep has a VIF near 1 once its partner is gone.



Dashed lines: the cutoffs used in lecture. Above 5 is a warning; above 10 is serious.

Checking Each Answer Choice

a No variables should be removed; all VIFs are under 10.

10 is one cutoff, but 6.1 and 5.8 are above 5, and the 0.88 correlation shows the two variables carry nearly the same information.

b Remove Posts and Hashtags because their VIFs are low.

A low VIF is good: the variable adds information the other regressors do not have.

c Keep all; correlation does not imply multicollinearity.

High correlation between regressors is what multicollinearity means. The phrase being borrowed is "correlation does not imply causation."

d Remove Hashtags because it is strongly correlated with Followers.

Their correlation is 0.35, which is not strong, and the VIF of Hashtags is 1.9.

e Followers and AvgCaptionLength are collinear and one of them should be dropped.

Answer: (e)

> The printed VIFs do not match this correlation matrix, which gives 4.6, 1.2, 1.3, and 4.4. On the exam, use the printed numbers. The answer is (e) either way, because of the 0.88 correlation.

The Question: The Analyst Drops Both Variables

A health insurer models whether a household enrolls in a supplemental dental plan (Enrolled = 1) using Income (in $1,000s) and HasChildren (1 = yes) with an LPM, a probit, and a logit on 800 households. Age (in years) is added to the model, which gives these VIFs:

Variable	VIF
Income	10.6
Age	11.8
HasChildren	1.2

A junior analyst drops **both** Income and Age from all three models, keeping only HasChildren.

What are the consequences of this decision for the model's estimates?

- **a** The remaining HasChildren coefficient becomes unbiased, since removing correlated predictors always purges confounding.
- **b** Individual standard errors for HasChildren decrease, but the joint explanatory power of the model is unchanged.
- **c** The model gains efficiency because multicollinearity inflated all standard errors; dropping both variables corrects this for the remaining coefficient.
- **d** The HasChildren coefficient will absorb variation from both excluded variables, producing estimates that are biased and inconsistent; the efficiency gain from removing collinear regressors does not offset the omitted variable bias introduced.
- **e** The only consequence is a reduction in R^2; coefficient estimates remain unbiased because HasChildren is uncorrelated with Income and Age.

Source: Spring 2026 Midterm (Boswell), Question 5

PROBLEM 1(B) · MULTICOLLINEARITY

Multicollinearity vs Omitted Variables

THE MODEL WITH ALL THREE

$$\text{Enrolled} = \beta_0 + \beta_1 \text{Income} + \beta_2 \text{Age} + \beta_3 \text{HasChildren} + e$$

DROP INCOME AND AGE

They do not disappear. Their effects move into the error term: the new error is $\beta_1 \text{Income} + \beta_2 \text{Age} + e$.

OMITTED VARIABLE BIAS (LPM)

$$E[\tilde{\beta}_3] = \beta_3 + \beta_1 \delta_1 + \beta_2 \delta_2$$

δ_1, δ_2: slopes from regressing Income and Age on HasChildren. The bias is not zero when Income or Age affects enrollment and is related to HasChildren.

IS HASCHILDREN RELATED TO THEM?

VIF of HasChildren = 1.2 $\implies R^2 = 1 - 1/1.2 = 0.17$. Income and Age explain 17% of the variation in HasChildren: small, but not zero.

	Multicollinearity	Omitted variable
Coefficients	**Unbiased**	**Biased, inconsistent**
Standard errors	**Large**	Can be small
Fix	Drop *one* of Income and Age. The one you keep still controls for most of the shared information, so the HasChildren coefficient is barely affected. The kept variable's own coefficient now includes the dropped one's effect.	

$\implies$ A small standard error around a biased estimate is not an improvement.

Checking Each Answer Choice

a HasChildren becomes unbiased; removing correlated predictors purges confounding.

Removing a variable that affects y and is related to HasChildren creates confounding. It does not remove it.

b Standard errors decrease, but joint explanatory power is unchanged.

Income and Age explained part of enrollment. Without them, R^2 and the joint explanatory power fall.

c The model gains efficiency, which corrects the remaining coefficient.

HasChildren had VIF 1.2, so its standard error was not inflated much. Any gain in precision comes with bias.

e Only R^2 falls; HasChildren is uncorrelated with Income and Age.

VIF = 1.2 means $R^2 = 0.17$ with the others, so it is correlated with them.

d HasChildren absorbs the excluded variables' effects: biased and inconsistent. The efficiency gain does not offset the omitted variable bias.

Answer: (d)

PROBLEM 1(C) · OVERFITTING

The Question: Training Error vs Test Error

A student fits a final wage model on an 80% training set, achieving $R^2 = 0.87$ and MAE = \$1,820. Applied to the held-out 20% test set, MAE = \$4,350. Which of the following is the most complete and economically sound diagnosis of this gap?

a The gap is negligible; an R^2 of 0.87 is sufficiently high that out-of-sample performance is guaranteed to be similar.

b The gap proves the model is misspecified and must be re-estimated using the full sample before any conclusions can be drawn.

c A doubling of MAE on the test set is normal and expected; train/test gaps of this magnitude are standard in econometric models and require no further investigation.

d The gap most likely reflects overfitting: the model has been tuned too closely to features of the training data which do not generalise to new observations.

e The gap arises because MAE is not a valid out-of-sample metric; the student should report MSE instead, which would show no gap.

Source: Spring 2026 Midterm (Boswell), Question 20. A similar question: Spring 2025 Midterm V4, Question 19 (86% training accuracy, 41% test accuracy).

PROBLEM 1(C) · TRY IT

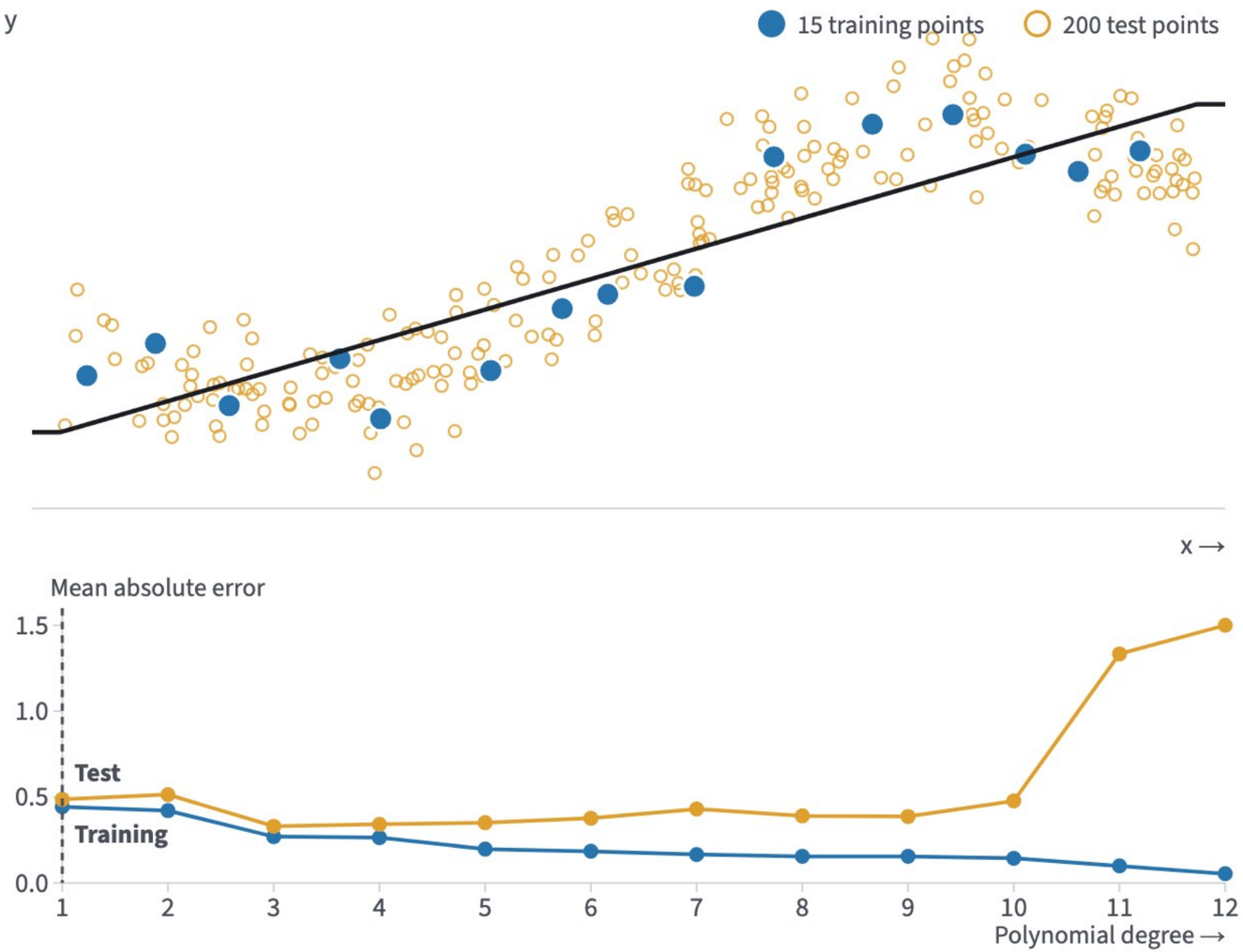
Add Terms: Training Error Falls, Test Error Does Not

Polynomial degree 1

Training MAE: **0.444**

Test MAE: **0.487**

Lowest test MAE: degree 3 (0.330)



TRAINING ERROR

OLS minimizes the sum of squared training errors, so an added term can never raise that sum. Training MAE also falls here, from 0.44 (degree 1) to 0.05 (degree 12).

TEST ERROR

Lowest at degree 3 (0.33). After that the added terms fit the noise in the 15 training points, and test MAE rises to 1.50 at degree 12.

$\Longrightarrow$ The student's test MAE is 2.4 times the training MAE $(4{,}350/1{,}820)$: the sign of a model fit too closely to its training data. Choose the model by its test or cross-validation error, not by training R^2.

Checking Each Answer Choice

a The gap is negligible; $R^2 = 0.87$ guarantees similar out-of-sample performance.

R^2 is measured on the training data. An overfit model has a high training R^2 by construction.

b The gap proves misspecification; re-estimate on the full sample.

Using the full sample removes the test set, so the gap can no longer be measured. A gap shows overfitting; it does not prove misspecification.

c Doubling of MAE is normal and needs no investigation.

Test MAE 2.4 times training MAE is a warning sign. Some gap is normal; this one is large.

e MAE is not a valid out-of-sample metric; MSE would show no gap.

MAE is a valid out-of-sample measure. MSE squares the same test errors, so it would show a gap too.

d Overfitting: the model is tuned too closely to features of the training data that do not generalize.

Answer: (d)

The Question: Does the Variance Grow with Followers?

Same TikTok data as Problem 1(a): 150 creators. You suspect the variance of the error term increases with the number of followers. You sort the data by Followers, omit the middle 50 observations, and run this regression separately on the bottom 50 and top 50 creators:

$$\ln(\text{Engagement}_i) = \beta_0 + \beta_1 \text{Followers}_i + \beta_2 \text{Posts}_i + \beta_3 \text{Hashtags}_i + u_i$$

The sums of squared errors from the small and large follower groups are 460 and 670. Compute the Goldfeld-Quandt test statistic and interpret it, using a critical value of 1.96.

a $F_{GQ} = 1.46$, and we reject the null and conclude heteroskedasticity exists with respect to followers

b $F_{GQ} = 0.69$, and we reject the null and conclude heteroskedasticity exists with respect to followers

c $F_{GQ} = 1.46$, and we fail to reject the null and conclude heteroskedasticity does not exist with respect to followers

d $F_{GQ} = 0.69$, and we fail to reject the null and conclude heteroskedasticity does not exist with respect to followers

e We need more information to answer this question

Source: Fall 2025 Midterm V1 (Boswell), Question 2

PROBLEM 2(A) · GOLDFELD-QUANDT

Goldfeld-Quandt: Compare the Variance in Two Groups

HYPOTHESES

H_0: $\sigma^2_{\text{large}} = \sigma^2_{\text{small}}$ (homoskedastic). H_1: $\sigma^2_{\text{large}} > \sigma^2_{\text{small}}$.

STATISTIC

Each group's variance estimate is $\hat{\sigma}^2 = SSE/(n-k)$. Both groups have $n = 50$ and $k = 4$ coefficients, so $n - k = 46$ cancels:

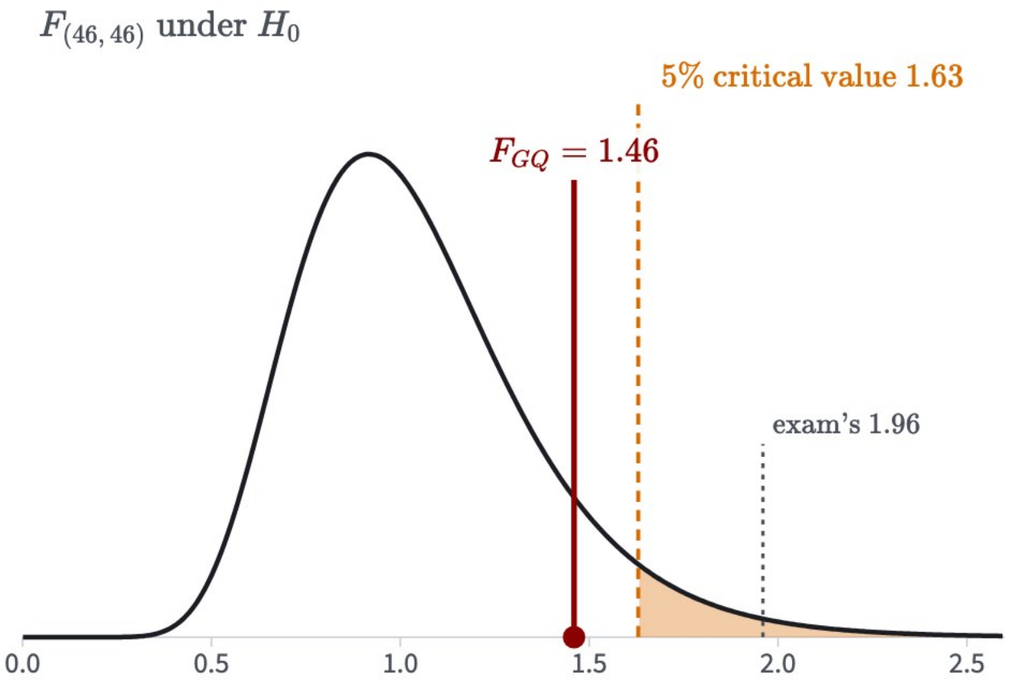
$$F_{GQ} = \frac{670/46}{460/46} = \frac{670}{460} = 1.46$$

The group that H_1 says has the larger variance (high Followers) goes on top. Here that is 670.

FAIL TO REJECT H_0 $1.46 < 1.96$

> The exam gives 1.96 as the critical value. The exact 5% value from the $F_{(46,\,46)}$ distribution is 1.63; 1.46 is below it too ($p = 0.10$), so the decision is the same.

$\Longrightarrow$ No evidence that the variance rises with Followers. This does not prove the errors are homoskedastic: the test only looked at Followers.



PROBLEM 2(A) · TRY IT

Run the Goldfeld-Quandt Test on Simulated Data

Spread grows with x **0.00**

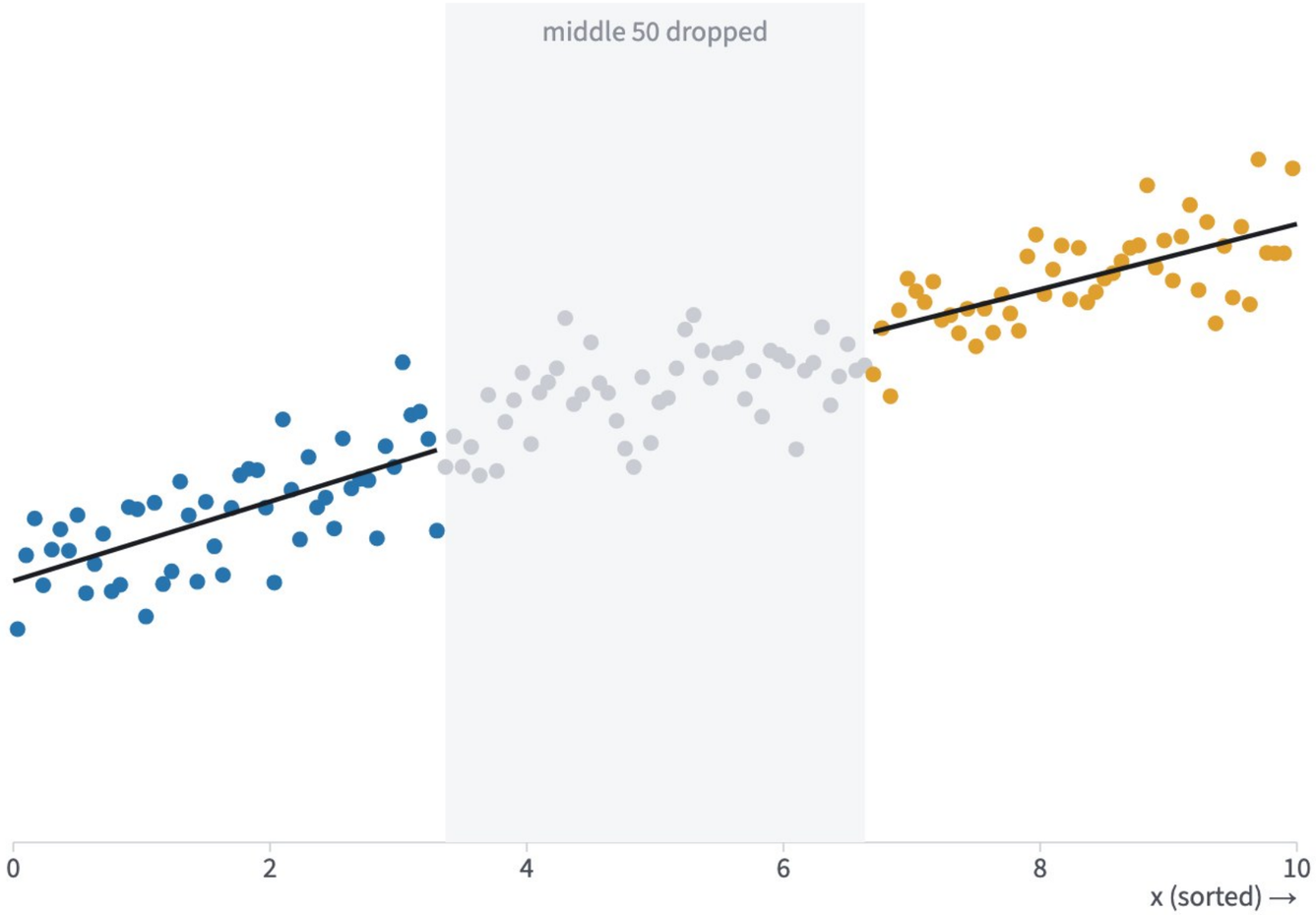
New sample

SSE, low 50: **114** SSE, high 50: **79**

F_{GQ} = 0.69: **fail to reject H_0**

150 points, sorted by x. The middle 50 are dropped. Each group of 50 gets its own regression line; F_{GQ} divides the high-x group's SSE by the low-x group's SSE. One regressor here, so $n - k = 48$ and the 5% critical value is $F_{(48,\,48)} = 1.62$.

At 0 the errors are homoskedastic, and about 1 sample in 20 still rejects.



PROBLEM 2(A) · GOLDFELD-QUANDT

Checking Each Answer Choice

a $F_{GQ} = 1.46$; reject.

The statistic is right, but $1.46 < 1.96$, so we fail to reject.

b $F_{GQ} = 0.69$; reject.

$460/670 = 0.69$ puts the smaller variance on top. And 0.69 is not above any critical value.

d $F_{GQ} = 0.69$; fail to reject.

The decision matches, but the ratio is upside down: the high-Followers group's SSE goes in the numerator.

e We need more information.

Both groups have the same $n - k = 46$, so the SSEs are enough.

c $F_{GQ} = 1.46$; fail to reject: no evidence of heteroskedasticity with respect to followers.

Answer: (c)

> Choice (c) says heteroskedasticity "does not exist." The test can only say there is not enough evidence of it, and only with respect to Followers. This is why Question 4 on the same exam recommends robust standard errors anyway.

PROBLEM 2(B) · FGLS WEIGHT

The Question: What Is the FGLS Weight?

Your classmate proposes an FGLS correction to the previous question in which the error variance is proportional to:

$$\mathrm{Var}(e_i) = \hat{\sigma}_i^2 = e^{0.00042\,\mathrm{Followers}_i}$$

For a TikTok creator with 12,000 followers, 4 posts, and an average of 5 hashtags, calculate the weight that should multiply all variables in this observation in the FGLS transformation. Also, determine if this transformation is valid based on the prior GQ test result. Round to four decimal places and check the variable scaling before Question 1.

a. 1.0051; yes this transformation is valid

b. 1.0025; no this transformation is invalid and worsens our estimates

c. 0.0805; no, this transformation is invalid and worsens precision

d. 0.9975; no, this transformation is invalid and worsens precision

e. None of the answers above provide the appropriate weight; this transformation is invalid and worsens our estimates.

Source: Fall 2025 Midterm V1 (Boswell), Question 3

PROBLEM 2(B) · FGLS WEIGHT

The FGLS Weight: Check the Units First

FGLS

Divide every variable in observation i (including the intercept column) by $\hat{\sigma}_i$. The transformed error has variance 1 for every observation. So the weight is

$$w_i = \frac{1}{\hat{\sigma}_i} = \frac{1}{\sqrt{e^{\,0.00042\,\text{Followers}_i}}}$$

UNITS

Followers is measured in 1,000s: 12,000 followers is $\text{Followers} = 12$. Posts and Hashtags do not enter the variance function.

PLUG IN

$\hat{\sigma}^2 = e^{0.00504} = 1.00505,\ \hat{\sigma} = 1.00252,\ w = 1/1.00252 = 0.9975$

IS FGLS WARRANTED?

The GQ test in 2(a) found no evidence that the variance changes with Followers. Weighting by a variance function the data do not support can make the estimates less precise than OLS.

> The official solution also says the bias increases. Weights that depend only on the regressors keep the estimator unbiased, so the cost is precision, which is what choice (d) says.

PROBLEM 2(B) · FGLS WEIGHT

Checking Each Answer Choice

a 1.0051; valid.

$1.0051 = \hat{\sigma}^2$, the variance itself, not the weight. And the GQ test gave no reason to use FGLS.

b 1.0025; invalid.

$1.0025 = \hat{\sigma}$. The weight is $1/\hat{\sigma}$.

c 0.0805; invalid.

Uses 12,000 instead of 12. Followers is in 1,000s.

e None of the weights are right.

Choice (d) has the correct weight.

d 0.9975; no, this transformation is invalid and worsens precision.

Answer: (d)

The Question: Larger Robust SEs and Their Effect on t and p

For the original log-engagement relationship you estimate the OLS model and then compute robust (heteroskedasticity-consistent) standard errors. You observe that the robust standard errors are larger than the original standard errors. What is the most likely effect on the reported t-statistics and p-values of the regression coefficients?

a the absolute value of the t-statistics will increase; p-values will decrease

b the absolute value of the t-statistics will increase; p-values will remain the same

c Both the absolute value of the t-statistics and p-values will remain unchanged

d the absolute value of the t-statistics will decrease; p-values will decrease

e the absolute value of the t-statistics will decrease; p-values will increase

Source: Fall 2025 Midterm V1 (Boswell), Question 5

PROBLEM 2(C) · ROBUST STANDARD ERRORS

Checking Each Answer Choice

a $|t|$ up; p down.

A larger denominator makes $|t|$ smaller, not larger.

b $|t|$ up; p the same.

$|t|$ falls, and p moves with $|t|$.

c Both unchanged.

The coefficient is unchanged; its standard error is not, so t changes.

d $|t|$ down; p down.

A smaller $|t|$ is closer to 0, so the p-value rises.

e $|t|$ down; p up.

Answer: (e)

ONLY THE SE CHANGES

Robust SEs change only the standard errors. $\hat{\beta}$ is the same number.

$$t = \frac{\hat{\beta}}{SE(\hat{\beta})}$$

A larger SE $\Longrightarrow$ smaller $|t|$ $\Longrightarrow$ larger p-value $\Longrightarrow$ coefficients look less significant.

QUESTION 4 ON THE SAME EXAM

In large samples, robust SEs are valid whether or not the errors are heteroskedastic, so they can be used as a safeguard even when a test fails to reject.

PROBLEM 3(A) · BREUSCH-PAGAN

The Question: Is the Variance Related to Truck Weight?

A logistics company analyzes fuel consumption (Y_i) based on truck weight (X_i). The Breusch-Pagan (BP) test for heteroskedasticity is performed, yielding an auxiliary regression of:

```
Call:
lm(formula = I(eh^2) ~ weight)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.77682    0.05336   14.56  < 2e-16 ***
weight       0.22286    0.05422    4.11 4.63e-05 ***

Residual standard error: 1.193 on 498 degrees of freedom
Multiple R-squared:  0.03281,  Adjusted R-squared:  0.03086
F-statistic: 16.89 on 1 and 498 DF,  p-value: 4.627e-05
```

If $N = 500$, and a 5% critical value of $\chi^2_{(1)} = 3.84$, what is your conclusion?

a. Reject H_0; LM = 16.4, thus heteroskedasticity is present with respect to weight.

b. Fail to reject H_0: LM = 0.035, not enough evidence to conclude heteroskedasticity is present with respect to weight

c. Reject H_0; LM = 3.84, thus heteroskedasticity is present with respect to weight.

d. Fail to reject H_0: LM = 16.4, not enough evidence to conclude heteroskedasticity is present with respect to weight

e. Cannot conclude; need the F-statistic instead

Source: Spring 2026 Midterm (Boswell), Question 17

Breusch-Pagan: $N \times R^2$ From the Auxiliary Regression

AUXILIARY REGRESSION

Regress the squared OLS residuals on the variables that might drive the variance: $\hat{e}_i^2 = \alpha_1 + \alpha_2 \,\text{weight}_i + v_i$. H_0: $\alpha_2 = 0$, the variance does not depend on weight.

STATISTIC

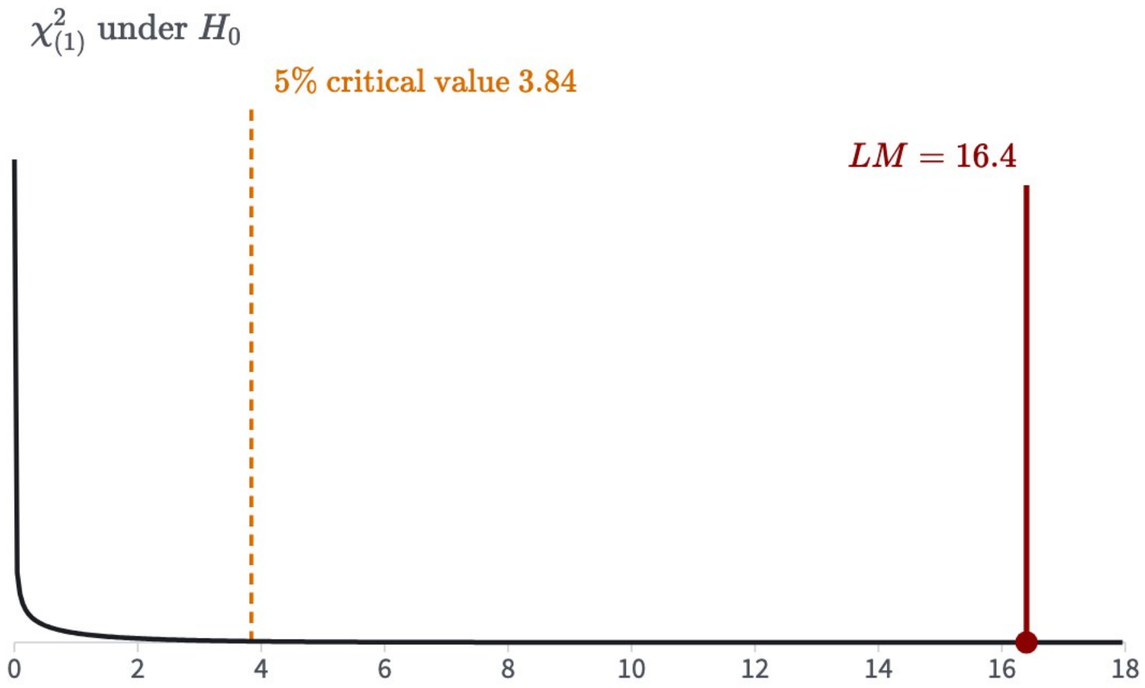
$LM = N \times R^2 = 500 \times 0.03281 = 16.4$. Use the auxiliary regression's R^2, not its adjusted R^2.

DEGREES OF FREEDOM

The number of variables in the auxiliary regression, not counting the intercept: 1.

REJECT H_0 $16.4 > 3.84 \implies$ the variance changes with weight.

> With one variable, the t-test on weight ($p = 4.6 \times 10^{-5}$) and the F-test ($F = 16.89$) give the same decision. With several variables, use NR^2 or the F-test, not one t-test.



Checking Each Answer Choice

b Fail to reject; LM = 0.035.

0.035 is about the size of R^2 itself (0.033). The statistic is $N \times R^2 = 16.4$.

c Reject; LM = 3.84.

3.84 is the critical value, not the statistic.

d Fail to reject; LM = 16.4.

$16.4 > 3.84$, so we reject.

e Need the F-statistic.

NR^2 is enough. The F-statistic is printed too and gives the same answer.

a Reject H_0; LM = 16.4: heteroskedasticity is present with respect to weight.

Answer: (a)

The Question: Is There Heteroskedasticity at the 1% Level?

1,080 houses sold in Baton Rouge in 2005. $\log(\text{PRICE}) = \beta_1 + \beta_2\,\text{SQFT100} + \beta_3\,\text{AGE} + \beta_4\,\text{AGE}^2 + e$, with SQFT100 = square feet / 100. Standard errors in parentheses.

	b_1	b_2	b_3	b_4
(i) Least squares	11.120	0.03876	−0.01756	0.0001734
	(0.027)	(0.00087)	(0.00136)	(0.0000227)
(ii) HC standard errors	11.120	0.03876	−0.01756	0.0001734
	(0.033)	(0.00123)	(0.00175)	(0.0000372)

```
lm(formula = res2 ~ age + sqft100)
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.0115081  0.0153786  -0.748   0.4544
age          0.0035429  0.0003100  11.429   <2e-16 ***
sqft100      0.0009758  0.0005287   1.846   0.0652 .
Residual standard error: 0.1734 on 1077 degrees of freedom
Multiple R-squared: 0.1082, Adjusted R-squared: 0.1066
F-statistic: 65.35 on 2 and 1077 DF,  p-value: < 2.2e-16
```

Test for heteroskedasticity using a Breusch-Pagan test and the variables AGE and SQFT100. Given that $\chi^2_{(2)} = 9.21$ at the 1% level, assume $\text{Var}(e) = h(\alpha_1 + \alpha_2\,\text{AGE} + \alpha_3\,\text{SQFT100})$.

a We fail to reject H_0 and conclude that conditional heteroskedasticity is present

b We reject H_0 and conclude that conditional heteroskedasticity is present

c We fail to reject H_0 and conclude that conditional heteroskedasticity is not present

d We reject H_0 and conclude that conditional heteroskedasticity is not present

e We cannot come to a conclusion; we need more information.

Source: Spring 2025 Midterm V4 (Boswell), Question 1

PROBLEM 3(B) · BREUSCH-PAGAN

Checking Each Answer Choice

a Fail to reject; heteroskedasticity present.

These two halves contradict each other. Failing to reject cannot show heteroskedasticity.

c Fail to reject; not present.

$NR^2 = 1080 \times 0.1082 = 116.9$**, far above 9.21.**

d Reject; not present.

Rejecting H_0 (homoskedasticity) means the evidence says heteroskedasticity is present.

e Need more information.

N and R^2 are given.

b Reject H_0: conditional heteroskedasticity is present.

Answer: (b)

STATISTIC

$LM = NR^2 = 1080 \times 0.1082 = 116.9$

df = 2 (age and sqft100) $\implies 116.9 > 9.21$: reject at 1%.

> sqft100 alone is not significant at 5% ($p = 0.065$). The test is about the two variables together; age is the significant one ($p < 0.001$).

The Question: Which Standard Error Goes in the Interval?

Given the conclusion from the previous question, using the appropriate table, construct a 95% confidence interval for the effect of SQFT100 on log(PRICE), with a critical value of 1.96.

	b_1	b_2	b_3	b_4
(i) Least squares	11.120	0.03876	−0.01756	0.0001734
	(0.027)	(0.00087)	(0.00136)	(0.0000227)
(ii) HC standard errors	11.120	0.03876	−0.01756	0.0001734
	(0.033)	(0.00123)	(0.00175)	(0.0000372)

a [0.03753, 0.03999]

b [0.03705, 0.04047]

c [−0.02998, 0.1075]

d [0.03635, 0.04117]

e None of the given answer options are correct.

Source: Spring 2025 Midterm V4 (Boswell), Question 2

PROBLEM 3(C) · ROBUST CONFIDENCE INTERVAL

Checking Each Answer Choice

a [0.03753, 0.03999]

0.03876 ± 0.00123: the robust SE without the 1.96.

b [0.03705, 0.04047]

$0.03876 \pm 1.96 \times 0.00087$: uses the least squares SE, which is wrong under heteroskedasticity.

c [−0.02998, 0.1075]

The width does not come from either standard error of b_2.

e None are correct.

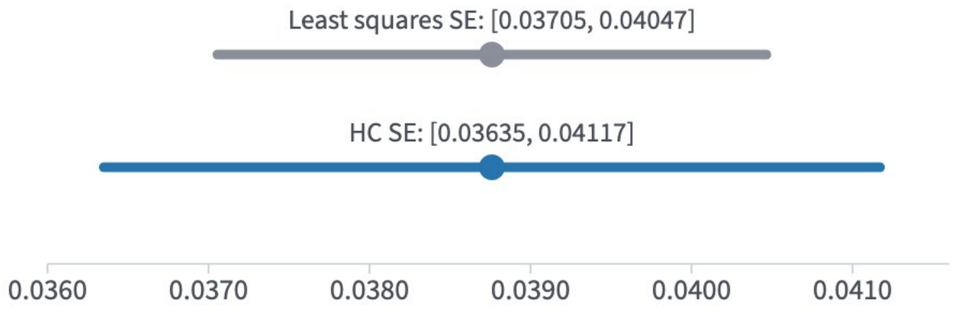
Choice (d) is correct.

d $[0.03635, 0.04117] = 0.03876 \pm 1.96 \times 0.00123$

Answer: (d)

USE THE HC ROW

3(b) found heteroskedasticity, so the least squares SEs are wrong. The coefficient is the same in both rows; only the SE changes.



PROBLEM 3(D) · OLS VS WLS

The Question: Same Estimates, So Is OLS Fine?

An economist estimates a cross-sectional wage model and suspects heteroskedasticity. Comparing OLS and WLS (weighted by firm size), the slope on education is $\hat{\beta}_{OLS} = 1.42$ with standard error 0.31, while $\hat{\beta}_{WLS} = 1.39$ with standard error 0.180. A colleague argues that because the two point estimates are nearly identical, heteroskedasticity is not a concern and OLS should be preferred for simplicity. Which of the following best identifies the flaw in this reasoning?

a Similarity in point estimates does not imply OLS standard errors are valid; heteroskedasticity leaves OLS coefficients unbiased but renders standard errors inefficient, and the narrower WLS standard error reflects a genuine efficiency gain.

b The colleague is correct; when slope estimates are similar across OLS and WLS, heteroskedasticity is confirmed absent and OLS is BLUE.

c WLS always produces smaller standard errors than OLS by construction, so the difference in standard errors carries no informational content.

d WLS estimates are always biased when the variance function is misspecified, making OLS preferable under any uncertainty about the form of heteroskedasticity.

e OLS is preferred here because WLS is only appropriate when the sample size exceeds 1,000 observations.

Source: Spring 2026 Midterm (Boswell), Question 19

PROBLEM 3(D) · TRY IT

Draw Many Samples: Compare the Spread of OLS and WLS

True slope $\beta = 1.4$. In each sample of 100 workers, the error's standard deviation grows with x. Estimate the slope by OLS and by WLS (weights $1/x^2$) in every sample.

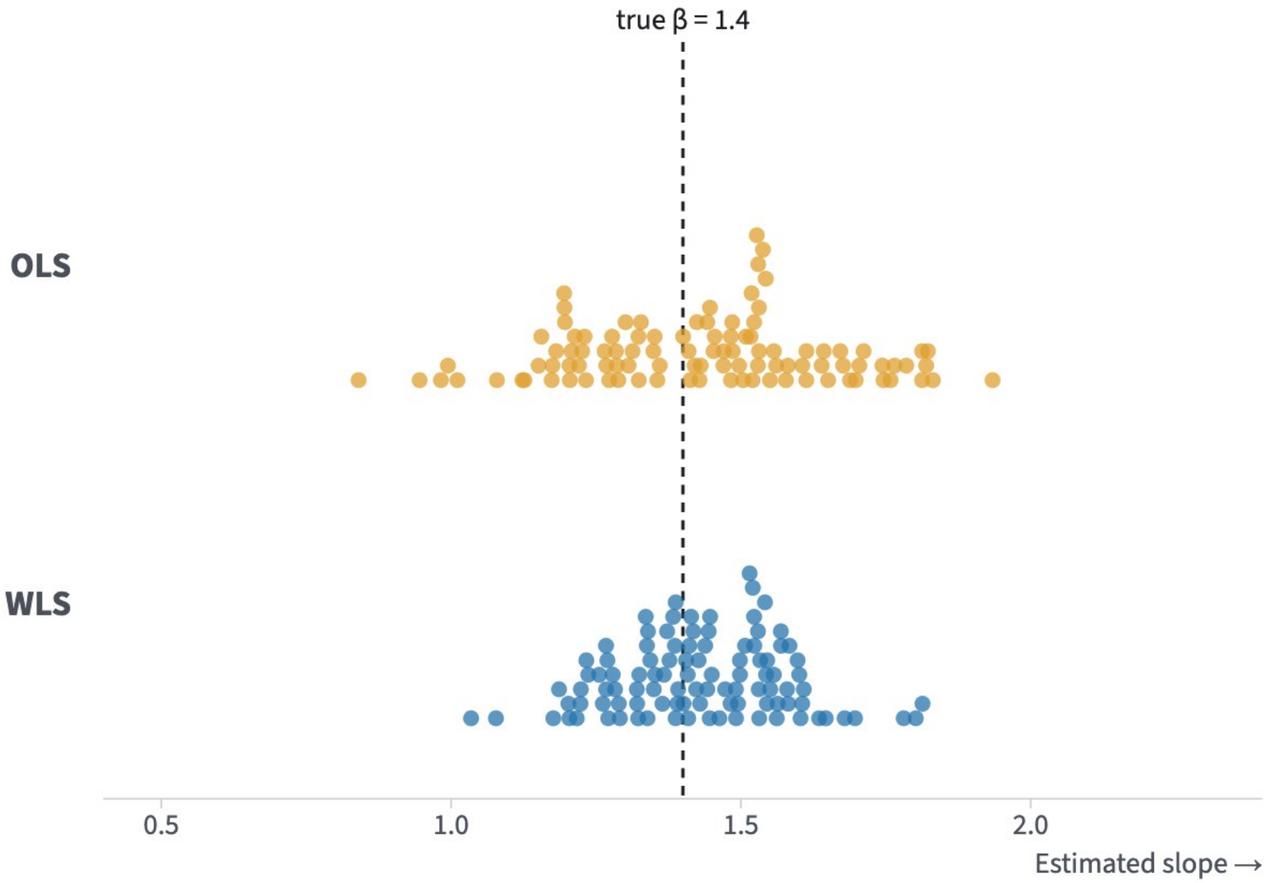
Draw 1 sample | Draw 100 | Reset

100 samples

OLS: mean **1.43**, SD **0.224**

WLS: mean **1.43**, SD **0.148**

Both sets of estimates center on 1.4: **both are unbiased**. The WLS estimates are less spread out: **WLS is more efficient**. One sample's two estimates are usually close, as 1.42 and 1.39 are.



Checking Each Answer Choice

b Similar estimates confirm no heteroskedasticity; OLS is BLUE.

Under heteroskedasticity OLS is still unbiased, so similar estimates are expected either way. They say nothing about the variance.

c WLS SEs are always smaller by construction.

WLS SEs are not smaller by construction: with badly wrong weights WLS can be less precise than OLS. The smaller WLS SE fits weights that are about right, but that SE assumes the weights are correct.

d WLS is always biased if the variance function is misspecified.

Wrong weights cost efficiency, not unbiasedness, as long as the weights depend only on the regressors.

e WLS needs more than 1,000 observations.

There is no such rule.

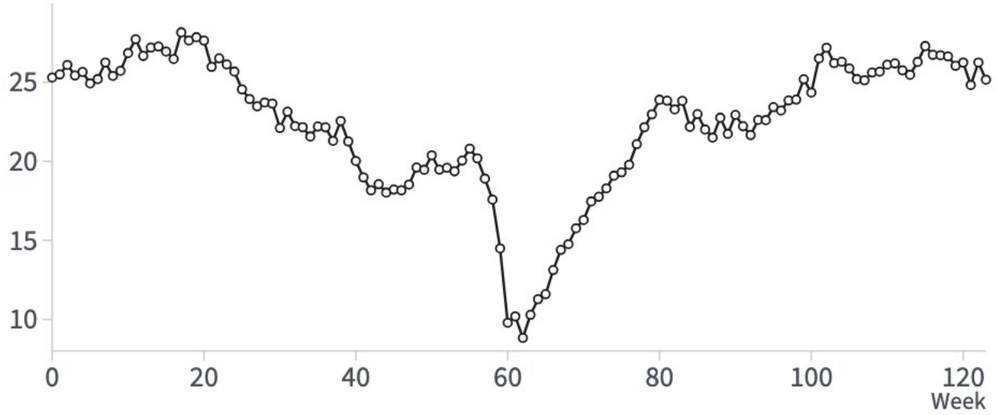
a Similar estimates do not make OLS SEs valid; OLS is unbiased but inefficient, and the narrower WLS SE is a real efficiency gain.

Answer: (a)

PROBLEM 4(A) · STATIONARITY

The Question: What Does the ADF Test Say?

The diagram shows weekly paper towel sales (in 10,000s of rolls) for the most recent 124 weeks, paired with the results of the ADF test:




```
Augmented Dickey-Fuller Test
Dickey-Fuller = -1.6311, Lag order = 2, p-value = 0.7298
alternative hypothesis: stationary
```

What can be concluded about this time series?

a. We fail to reject the null of the ADF test and conclude that our time series is stationary.

b. We reject the null of the ADF test and conclude that our time series is stationary.

c. We fail to reject the null of the ADF test and conclude that our time series is non-stationary.

d. We reject the null of the ADF test and conclude that our time series is non-stationary.

e. None of the given answers are correct.

Source: Fall 2024 Midterm V1 (Boswell), Question 10. The plot is traced by hand from the exam figure; values are approximate.

PROBLEM 4(A) · STATIONARITY

Checking Each Answer Choice

a Fail to reject; stationary.

Failing to reject keeps the null, and the null of the ADF test is non-stationarity.

b Reject; stationary.

$p = 0.73 > 0.05$: we do not reject.

d Reject; non-stationary.

We do not reject, and rejecting would mean stationary.

e None are correct.

Choice (c) is correct.

c Fail to reject the null: the series is non-stationary.

Answer: (c)

READ THE HYPOTHESES FIRST

R prints "alternative hypothesis: stationary." So H_0: unit root (non-stationary). H_1: stationary.

DECISION

$p = 0.73 > 0.05 \implies$ fail to reject $\implies$ treat the series as non-stationary. The plot agrees: the level wanders and does not return to a fixed mean.

> Fix before using it in a regression: first-difference the series, then test again.

The Question: Which Autocorrelations Are Significant?

Say we are given the following sample autocorrelations:

Lag	1	2	3	4
r_k	0.58	−0.12	0.04	−0.36

Which, of the first 4 lags, are statistically significant at the 5% level, if $Z_{0.975} = 1.96$ and there are 1,125 effective periods in the study? If we are told that all lags thereafter are statistically insignificant, what conclusion can we draw about an appropriate model?

a Lag 1 only, suggesting an autoregressive (AR) model with a small coefficient

b Lags 1 and 2 only, suggesting a moving average (MA) model

c Lags 1 and 3 only, suggesting a moving average (MA), autoregressive (AR) or autoregressive moving average (ARMA) model

d Lags 1, 2, and 4 only, suggesting a moving average (MA), autoregressive (AR) or autoregressive moving average (ARMA) model

e Lags 1, 2, 3, and 4, suggesting an autoregressive (AR) model

Source: Fall 2024 Midterm V1 (Boswell), Question 14

Testing One Autocorrelation: $r_k\sqrt{T}$

TEST

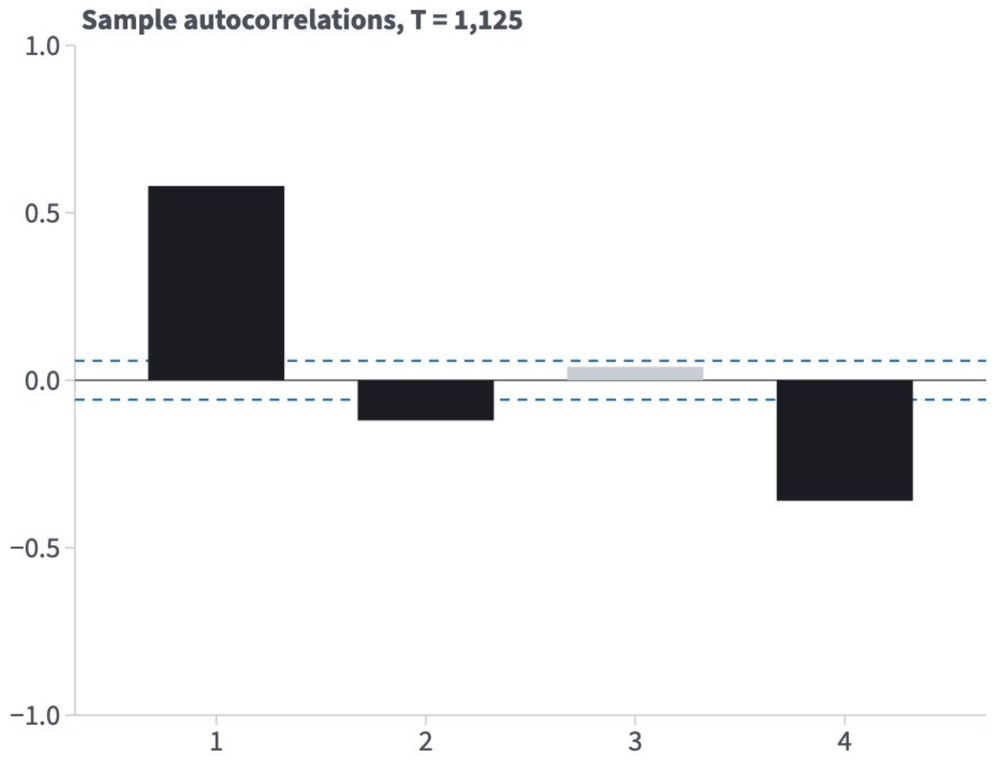
H_0: $\rho_k = 0$. Under H_0, $Z = r_k\sqrt{T}$ is approximately $N(0,1)$. Reject if $|Z| > 1.96$.

SAME TEST AS A BAND

Reject if $|r_k| > 1.96/\sqrt{T} = 1.96/\sqrt{1125} = 0.058$. This is the dashed band on an R correlogram.

Lag	r_k	$r_k\sqrt{1125}$	Significant?
1	0.58	19.45	Yes
2	−0.12	−4.02	Yes
3	0.04	1.34	No
4	−0.36	−12.07	Yes

> Fall 2025, Question 9 uses $T = 200$: the band is $1.96/\sqrt{200} = 0.139$, so $r_1 = 0.58$ is significant and $r_2 = 0.12$ is not. A larger T makes the band narrower.



Checking Each Answer Choice

a Lag 1 only.

Lags 2 and 4 also pass: $|-4.02| > 1.96$ **and** $|-12.07| > 1.96$.

b Lags 1 and 2 only.

Misses lag 4 ($r_4\sqrt{T} = -12.07$**).**

c Lags 1 and 3 only.

Lag 3 is the one that is not significant: $1.34 < 1.96$.

e All four lags.

Lag 3 is not significant.

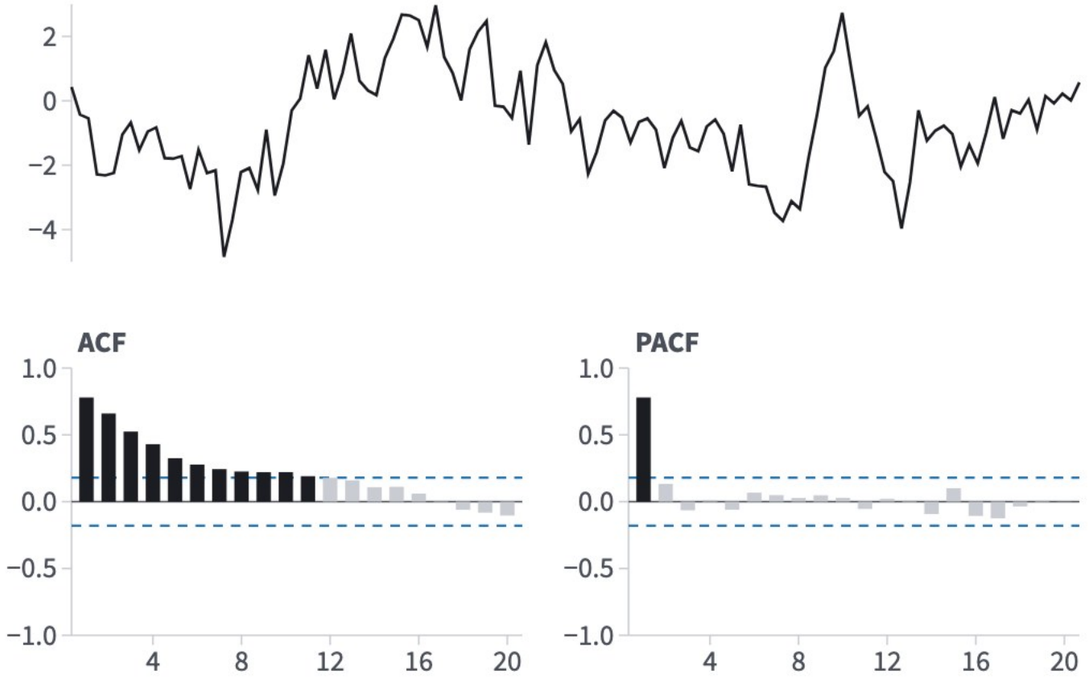
d Lags 1, 2, and 4 only, suggesting an MA, AR, or ARMA model

Answer: (d)

The ACF alone does not pick the model. It cuts off after lag 4, which fits an MA(4); an AR or ARMA model could also produce it. The PACF is needed to tell them apart (next question).

The Question: Which Model Produced This Series?

The output below shows a stationary quarterly series measuring the output gap (% deviation from potential GDP), 1994 Q1 to 2023 Q4. Based on the time series, ACF and PACF panels, which statement about the data-generating process is most accurate?



a The ACF cuts off sharply after lag 1 and the PACF decays slowly, identifying this as an MA(1) process; an ARIMA(0,0,1) is the most appropriate specification.

b Several ACF lags around lags 11–13 cross the significance bounds, so an ARIMA(12,0,0) is necessary to fully capture the autocorrelation structure.

c Both the ACF and PACF decay gradually from lag 1, which is the identifying signature of an ARIMA(1,0,1); a pure AR or MA specification should be ruled out.

d The ACF and PACF plots are insufficient for model identification; AIC and BIC must always be consulted before any model order can be proposed.

e The PACF has a single dominant spike at lag 1 with all subsequent lags inside the significance bounds, and the ACF decays gradually, consistent with an ARIMA(1,0,0). An ARIMA(1,0,1) could also be considered if residual diagnostics reveal remaining structure.

Source: Spring 2026 Midterm (Boswell), Question 14. The panels are redrawn from a simulated AR(1) series of the same length (120 quarters) with the same pattern as the exam figure.

PROBLEM 4(C) · TRY IT

Set ϕ_1 and ϕ_2: Watch the ACF and PACF

$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t$, 400 periods.

ϕ_1 **0.75**

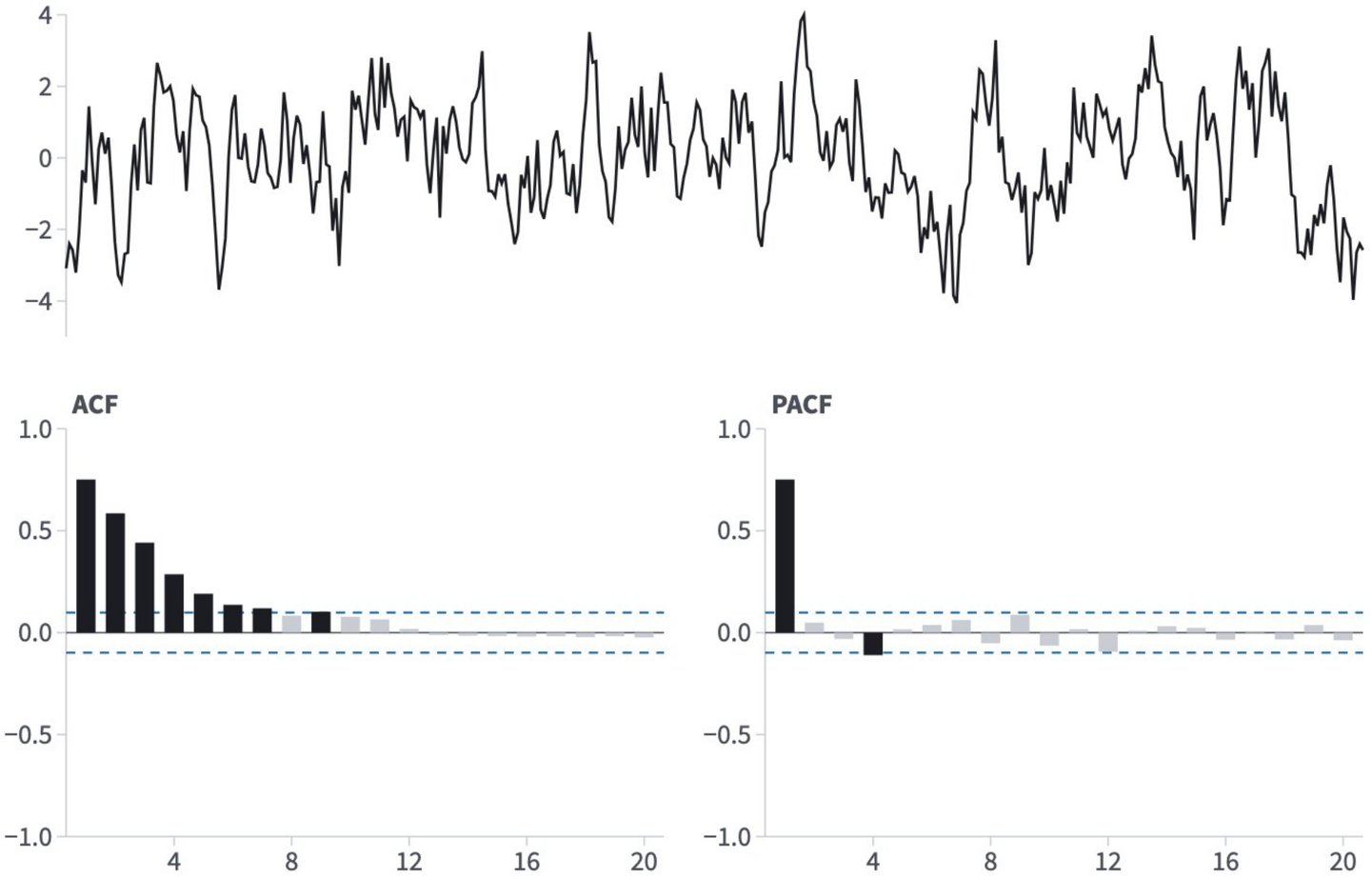
ϕ_2 **0.00**

New draw

AR(1): one PACF spike, ACF decays.

	ACF	PACF
AR(p)	Decays gradually	Cuts off after lag p
MA(q)	Cuts off after lag q	Decays gradually

ARIMA(p, 0, q) is ARMA(p, q): ARIMA(1,0,0) is AR(1), and ARIMA(0,0,1) is MA(1).



Checking Each Answer Choice

a ACF cuts off after lag 1; MA(1).

The ACF decays over several lags; it is the PACF that cuts off after lag 1.

b Lags 11–13 cross the bounds; ARIMA(12,0,0).

The PACF has one spike, so one AR term is enough. ACF bars far out belong to the gradual decay that an AR(1) produces (or to chance: with a 5% band, about 1 lag in 20 crosses), so they do not justify 12 AR terms.

c Both decay; ARIMA(1,0,1).

The PACF does not decay: it has one spike and then stays inside the bounds.

d ACF and PACF are insufficient; always use AIC and BIC.

The plots are the standard first step for choosing the order; AIC and BIC are a check, not a requirement.

e One PACF spike at lag 1 and a gradually decaying ACF: ARIMA(1,0,0), with ARIMA(1,0,1) if residual checks show remaining structure.

Answer: (e)

PROBLEM 5(A) · FORECAST INTERVAL

The Question: A 95% Prediction Interval for $\hat{Y}_{t+1}$

An economist forecasts inflation expectations using a simple AR(2) model:

$$\hat{Y}_t = 0.5 + 0.6\,Y_{t-1} + 0.2\,Y_{t-2}$$

If the current expectation is $Y_t = 3.0$, the previous expectation is $Y_{t-1} = 2.5$, and the regression variance is $\sigma_\varepsilon^2 = 0.25$, calculate the 95% prediction interval for the one-period-ahead forecast $\hat{Y}_{t+1}$. Use $z_{0.025} = 1.96$. (Round to 2 d.p.)

a. [1.82, 3.78]

b. [2.32, 3.28]

c. [1.32, 3.28]

d. [2.02, 3.98]

e. None of the given answer options represent the prediction interval

Source: Spring 2026 Midterm (Boswell), Question 12

PROBLEM 5(A) · FORECAST INTERVAL

Forecast, Then Add $\pm 1.96\,\sigma$

SHIFT THE EQUATION FORWARD ONE PERIOD

$\hat{Y}_{t+1} = 0.5 + 0.6\,Y_t + 0.2\,Y_{t-1} = 0.5 + 0.6(3.0) + 0.2(2.5) = 2.8.$
The lag-1 coefficient multiplies the most recent value, Y_t.

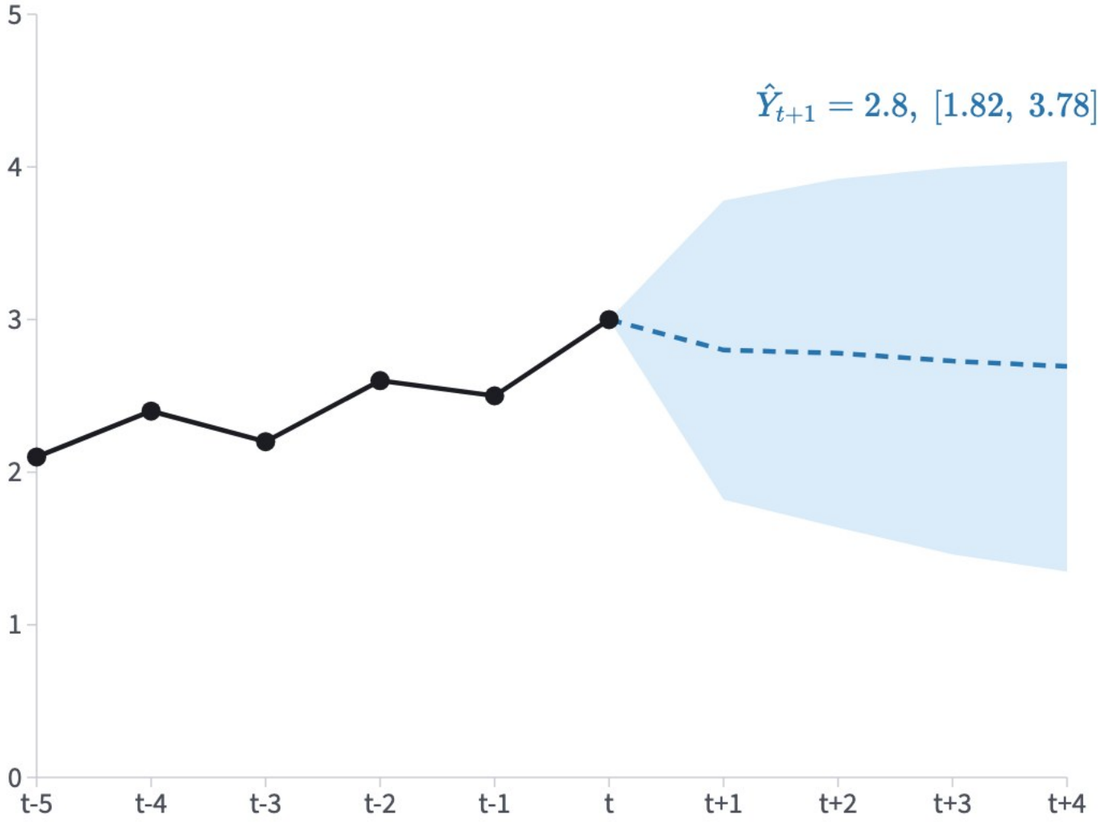
STANDARD DEVIATION, NOT VARIANCE

$\sigma = \sqrt{0.25} = 0.5$. Interval: $2.8 \pm 1.96 \times 0.5 = [1.82,\ 3.78]$.

FURTHER AHEAD

Two steps ahead, Y_{t+1} is itself a forecast, so its error adds in: the standard deviation becomes $\sigma\sqrt{1 + 0.6^2} = 0.58$. The slider changes σ.

σ_ε **0.50**



PROBLEM 5(A) · FORECAST INTERVAL

Checking Each Answer Choice

b [2.32, 3.28]

Uses the variance 0.25 in place of $\sigma = 0.5$.

c [1.32, 3.28]

Centered at 2.30, not at the forecast 2.8.

d [2.02, 3.98]

Centered at 3.0, the current value Y_t, instead of the forecast.

e None are correct.

Choice (a) is correct.

a $[1.82, 3.78] = 2.8 \pm 1.96 \times 0.5$

Answer: (a)

PROBLEM 5(B) · BREUSCH-GODFREY

The Question: Effect of Autocorrelated Errors on the Forecast

Given the outcome of the BG test of the model in 5(a):

```
Breusch-Godfrey test for serial correlation of order up to 2
data:  ar1_model
LM test = 15.152, df = 2, p-value = 0.000514
```

which of the following statements are correct about the forecast and confidence interval?
(i) The point forecast $\hat{Y}_{t+1}$ is biased because OLS estimates are inconsistent when lagged dependent variables are regressed on autocorrelated errors.
(ii) The forecast confidence interval is too narrow because the true variance of the forecast error is understated by OLS standard errors under autocorrelation.
(iii) The point forecast remains valid and unaffected; only the width of the confidence interval is unreliable.
(iv) The forecast confidence interval may have incorrect coverage, meaning the stated 95% interval does not reliably contain the true value 95% of the time.
(v) Autocorrelation has no consequence for forecasting as long as the model's R^2 is sufficiently high.

a (ii) and (iv) only

b (iii) and (v) only

c (i) and (iii) only

d (i), (ii), and (iv) only

e (i), (ii), (iv), and (v) only

Source: Spring 2026 Midterm (Boswell), Question 13

PROBLEM 5(B) · BREUSCH-GODFREY

Breusch-Godfrey Test: A Lagged Y Makes OLS Inconsistent

AUXILIARY REGRESSION (ORDER 2)

Regress the residual on *all* the original regressors plus 2 lagged residuals:

$$\hat{e}_t = \gamma_0 + \gamma_1 Y_{t-1} + \gamma_2 Y_{t-2} + \rho_1 \hat{e}_{t-1} + \rho_2 \hat{e}_{t-2} + v_t$$

H_0: $\rho_1 = \rho_2 = 0$. $LM = T \times R^2 \sim \chi^2_{(2)}$.

REJECT H_0 $LM = 15.15 > 5.99$ (5% $\chi^2_{(2)}$ critical value; $p = 0.0005$) $\Longrightarrow$ the errors are autocorrelated.

LAGGED Y PLUS AUTOCORRELATION

Y_{t-1} contains e_{t-1}. If e_t is correlated with e_{t-1}, then the regressor Y_{t-1} is correlated with the error e_t $\Longrightarrow$ OLS is **inconsistent**, so the forecast is biased: (i).

THE INTERVAL

The usual standard errors understate the forecast error variance, so the interval is too narrow (ii) and its coverage is not 95% (iv).

	Regressors exogenous	Lagged Y as a regressor
Coefficients	**Unbiased, consistent**	**Inconsistent**
Standard errors	**Wrong**	**Wrong**

$\Longrightarrow$ (iii) is false because of (i), and (v) is false: a high R^2 does not fix either problem.

PROBLEM 5(B) · BREUSCH-GODFREY

Checking Each Answer Choice

a (ii) and (iv) only

Leaves out (i): with a lagged dependent variable, autocorrelation makes OLS inconsistent.

b (iii) and (v) only

Both are false: the forecast is affected, and R^2 does not help.

c (i) and (iii) only

(i) and (iii) contradict each other.

e (i), (ii), (iv), and (v) only

(v) is false.

d (i), (ii), and (iv) only

Answer: (d)

The Question: Which Is Not a Correction for Autocorrelation?

Which of the following is NOT a valid correction for autocorrelation?

a. Applying HAC (Newey-West) robust standard errors to account for the serial correlation in inference.

b. Adding lagged dependent variables to the model to directly capture the remaining autocorrelation structure.

c. Applying Cochrane-Orcutt feasible GLS, which estimates the autocorrelation parameter in the error term and re-estimates the transformed model.

d. First-differencing the series

e. Applying Weighted Least Squares (WLS) using the inverse of the squared residuals as weights

Source: Spring 2026 Midterm (Boswell), Question 10

PROBLEM 5(C) · CORRECTIONS

Checking Each Answer Choice

a HAC (Newey-West) SEs

Valid: keeps OLS coefficients and fixes the standard errors.

b Add lagged dependent variables

Valid: models the dynamics so the remaining error is closer to white noise. Then test the residuals again.

c Cochrane-Orcutt FGLS

Valid: estimates ρ and transforms $y_t - \hat{\rho} y_{t-1}$.

d First-differencing

Valid: removes a unit root and much of the persistence.

e WLS with the inverse of the squared residuals as weights

Answer: (e)

WLS reweights observations to fix a variance that changes across observations: heteroskedasticity. It does nothing about the correlation between e_t and e_{t-1}.

PROBLEM 6(A) · LAG OPERATOR

The Question: Write the ARDL(1,3) With the Lag Operator

Consider the following ARDL(1,3) model:

$$y_t = \alpha + \phi y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \beta_3 x_{t-3} + u_t$$

Which of the following correctly rewrites this model using lag operator notation, with y_t as the subject?

a $y_t = \alpha + \phi y_{t-1} + (\beta_0 + \beta_1 L + \beta_2 L^2 + \beta_3 L^3) x_t + u_t$

b $y_t = (1 - \phi L)^{-1}[\alpha + (\beta_0 + \beta_1 L + \beta_2 L^2 + \beta_3 L^3) x_t + u_t]$

c $y_t = \alpha + (\beta_0 + \beta_1 L + \beta_2 L^2 + \beta_3 L^3) x_t + u_t$

d $y_t = (1 - \phi L)[\alpha + (\beta_0 + \beta_1 L + \beta_2 L^2 + \beta_3 L^3) x_t + u_t]$

e None of the above options correctly depict the lag notation.

Source: Fall 2025 Midterm V1 (Boswell), Question 11

PROBLEM 6(A) · LAG OPERATOR

Checking Each Answer Choice

a Keeps ϕy_{t-1} on the right.

Uses L for x but not for y, and y_t appears on both sides, so y_t is not the subject.

c Drops ϕy_{t-1}.

That is a distributed lag model, not the ARDL.

d Multiplies by $(1 - \phi L)$.

Moving $\phi L y_t$ to the left gives $(1 - \phi L)y_t$; to isolate y_t you divide by it, not multiply.

e None are correct.

Choice (b) is correct.

b $y_t = (1 - \phi L)^{-1}[\alpha + (\beta_0 + \beta_1 L + \beta_2 L^2 + \beta_3 L^3)x_t + u_t]$

Answer: (b)

STEPS

$Lx_t = x_{t-1},\ L^k x_t = x_{t-k},\ Ly_t = y_{t-1}.$

$y_t = \alpha + \phi L y_t + (\beta_0 + \beta_1 L + \beta_2 L^2 + \beta_3 L^3)x_t + u_t$

$(1 - \phi L)\, y_t = \alpha + (\beta_0 + \cdots + \beta_3 L^3)x_t + u_t$

Divide both sides by $(1 - \phi L)$.

PROBLEM 6(B) · INFINITE DISTRIBUTED LAG

The Question: The Rate of Decline and Stability

We estimate the effect of permanent changes in oil prices (X_t) on long-term inflation expectations (Y_t) using an infinite distributed lag (IDL) model with geometrically declining lags:

$$Y_t = 1.2 + 0.8X_t + 0.4X_{t-1} + 0.2X_{t-2} + \cdots$$

For the IDL model to be stable, what is the geometric decline parameter (λ), and what condition must it satisfy for the long-run multiplier to be finite?

a. $\lambda = 0.4$; requires $|\lambda| > 1$

b. $\lambda = 0.5$; requires $|\lambda| < 1$

c. $\lambda = 0.8$; requires $|\lambda| < 1$

d. $\lambda = 0.2$; requires $|\lambda| \neq 0$

e. $\lambda = 0.5$; requires $|\lambda| > 0$

Source: Spring 2026 Midterm (Boswell), Question 15

PROBLEM 6(B) · INFINITE DISTRIBUTED LAG

Checking Each Answer Choice

a $\lambda = 0.4; |\lambda| > 1$

0.4 is δ_1, the second lag weight, not the ratio. With $|\lambda| > 1$ the lag weights grow and the sum is infinite.

c $\lambda = 0.8$

0.8 is δ_0, the first lag weight (the impact effect).

d $\lambda = 0.2$

0.2 is δ_2, the third lag weight. And $\lambda \neq 0$ is not the stability condition.

e $\lambda = 0.5; |\lambda| > 0$

Right λ, wrong condition: $\lambda = 1.5$ satisfies $|\lambda| > 0$ but the weights explode.

b $\lambda = 0.5$; requires $|\lambda| < 1$

Answer: (b)

GEOMETRIC LAGS

δ_k is the coefficient on X_{t-k}: $\delta_k = \delta_0 \lambda^k$, so λ is the ratio of consecutive weights: $0.4/0.8 = 0.2/0.4 = 0.5$.

PROBLEM 6(C) · INFINITE DISTRIBUTED LAG

The Question: The Total (Long-Run) Multiplier

We estimate the effect of permanent changes in oil prices (X_t) on long-term inflation expectations (Y_t) using an infinite distributed lag (IDL) model with geometrically declining lags:

$$Y_t = 1.2 + 0.8X_t + 0.4X_{t-1} + 0.2X_{t-2} + \cdots$$

Recall: $\sum_{i=0}^{\infty} ar^i = \frac{a}{1-r}$ for $|r| < 1$.

What is the estimated total multiplier: the full long-run change in Y due to a permanent one-unit change in X_t?

a. 0.5
b. 1.2
c. 1.6
d. 1.8
e. None of the given options are correct.

Source: Spring 2026 Midterm (Boswell), Question 16

PROBLEM 6(C) · TRY IT

Move λ: Add Up the Lag Weights

λ **0.50**

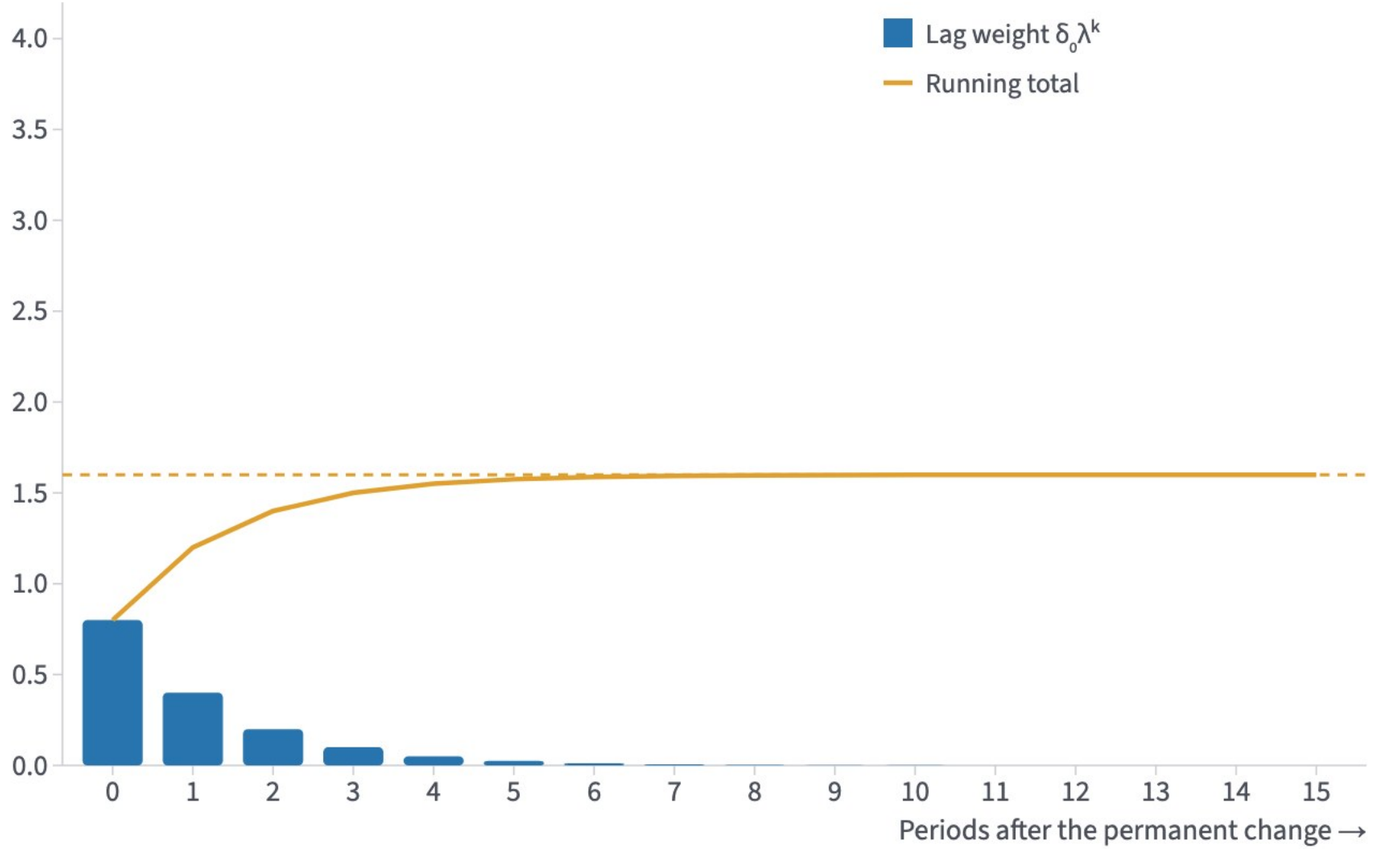
Weights: 0.80, 0.40, 0.20, 0.10, ...

Total multiplier: 0.8 / (1 − 0.50) = **1.60**

A permanent one-unit rise in X at time 0 raises Y by δ_0 at once, by $\delta_0 + \delta_1$ after one period, and so on. The running total approaches

$$\frac{\delta_0}{1-\lambda} = \frac{0.8}{1-0.5} = 1.6$$

> The intercept 1.2 is not part of the multiplier. As $\lambda \to 1$ the total grows without limit.



Checking Each Answer Choice

a 0.5

0.5 is λ, the ratio between weights.

b 1.2

1.2 is the intercept. It does not respond to X.

d 1.8

Not the sum of the lag weights: $0.8 + 0.4 + 0.2 + 0.1 + \cdots = 1.6$.

e None are correct.

Choice (c) is correct.

c $1.6 = 0.8/(1 - 0.5)$

Answer: (c)

The Question: Which Way Does Granger Causality Run?

A researcher studies advertising expenditure (AdExp) and sales revenue (Sales) for 10 companies over 10 years, and runs Granger causality tests in both directions:

Test 1, H_0: AdExp does not Granger cause Sales. p-value = 0.002
Test 2, H_0: Sales does not Granger cause AdExp. p-value = 0.23

a Both advertising expenditure and sales revenue Granger cause each other so we need a VAR model to represent this relationship.

b Advertising expenditure Granger causes sales revenue, but sales revenue does not Granger cause advertising expenditure.

c Sales revenue Granger causes advertising expenditure, but advertising expenditure does not Granger cause sales revenue.

d Neither advertising expenditure nor sales revenue Granger cause each other.

e There is no significant relationship between advertising expenditure and sales revenue.

Source: Fall 2024 Midterm V1 (Boswell), Question 13

PROBLEM 6(D) · GRANGER CAUSALITY

Checking Each Answer Choice

a Both directions; VAR.

Test 2 does not reject ($0.23 > 0.05$).

c Sales causes AdExp only.

Reversed: the test that rejects has AdExp's lags predicting Sales.

d Neither.

Test 1 rejects ($0.002 < 0.05$).

e No significant relationship.

Test 1 finds one.

b AdExp Granger causes Sales, but Sales does not Granger cause AdExp.

Answer: (b)

ONE TEST

Restricted: Sales on its own lags. Unrestricted: add lags of AdExp. An F-test asks whether the AdExp lags improve the forecast.

Result	Model
One direction (this question)	ARDL: Sales on lags of Sales and AdExp
Both directions (Spring 2025, Q13)	VAR: one equation for each

> Granger causality is about forecasting. It does not show that advertising causes sales.

PROBLEM 7(A) · LOGIT ODDS

The Question: What Does the Logit Coefficient 0.6 Mean?

A health insurer models whether an individual opts into a supplemental dental plan (Enrolled = 1). Income is in $1,000s; HasChildren = 1 if there are children in the household. 800 observations:

Model	Estimated equation
LPM	$\hat{P} = 0.12 + 0.008\,\text{Income} + 0.14\,\text{HasChildren}$
Probit	$\hat{P} = \Phi(-0.9 + 0.024\,\text{Income} + 0.36\,\text{HasChildren})$
Logit	$\hat{P} = \Lambda(-1.5 + 0.04\,\text{Income} + 0.6\,\text{HasChildren})$

The actuary argues: "A household with children has 0.6 times the enrollment odds of a household without children, all else equal, under a logit model." Which correctly identifies the error and restates the interpretation?
(i) There is no error; the coefficient 0.6 directly equals the odds ratio for HasChildren.
(ii) The error is that 0.6 measures the marginal effect on probability; having children raises the probability of enrollment by 0.6 percentage points.
(iii) The error is that 0.6 is the odds ratio only after exponentiating; since $e^{0.6} \approx 1.82$, having children is associated with 82% higher enrollment odds.
(iv) The error is that 0.6 is a log-odds change, not an odds ratio; having children is associated with the log odds of enrollment increasing by 0.6, holding income fixed.

a (i) only
b (ii) only
c (iii) only
d (iv) only
e (iii) and (iv) only

Source: Spring 2026 Midterm (Boswell), Question 1

PROBLEM 7(A) · LOGIT ODDS

Probability, Odds, and Log-Odds

ODDS

$\text{odds} = \dfrac{p}{1-p}$. A probability of 0.8 is odds of 4 (4 to 1). Lecture 9: rolling a 6 has odds $\frac{1/6}{5/6} = 0.2$.

$\Lambda(z) = 1/(1+e^{-z})$, the logistic CDF, turns the log-odds z into a probability.

THE LOGIT IS LINEAR IN THE LOG-ODDS

$$\ln\left(\frac{p}{1-p}\right) = -1.5 + 0.04\,\text{Income} + 0.6\,\text{HasChildren}$$

HasChildren from 0 to 1 **adds 0.6 to the log-odds**: statement (iv).

EXPONENTIATE TO GET THE ODDS RATIO

$\dfrac{\text{odds with children}}{\text{odds without}} = e^{0.6} = 1.82 \implies$ odds 82% higher: statement (iii).

NOT A PROBABILITY CHANGE

At Income = 50: $\Lambda(0.5) = 0.623$ without children, $\Lambda(1.1) = 0.750$ with. The probability rises by 0.128, not by 0.6 and not by 0.6 percentage points.

Scale	Effect of HasChildren
Log-odds	$+0.6$ (add)
Odds	$\times\, e^{0.6} = 1.82$ (multiply)
Probability	Depends on Income

> Lecture 9 example: $e^{0.156} = 1.17$, so one more math point multiplies the odds of honors placement by 1.17 (17% higher), at every math score.

PROBLEM 7(A) · LOGIT ODDS

Checking Each Answer Choice

a (i) only

0.6 is the change in log-odds. The odds ratio is $e^{0.6} = 1.82$.

b (ii) only

A logit coefficient is not a marginal effect on probability.

c (iii) only

(iv) is also correct: it states the same fact on the log-odds scale.

d (iv) only

(iii) is also correct: it states the same fact on the odds scale.

e (iii) and (iv) only

Answer: (e)

PROBLEM 7(B) · LOGIT PREDICTION

The Question: Probability and Marginal Effect at Income = 30

A health insurer models whether an individual opts into a supplemental dental plan (Enrolled = 1). Income is in $1,000s; HasChildren = 1 if there are children in the household. 800 observations:

Model	Estimated equation
LPM	$\hat{P} = 0.12 + 0.008\,\text{Income} + 0.14\,\text{HasChildren}$
Probit	$\hat{P} = \Phi(-0.9 + 0.024\,\text{Income} + 0.36\,\text{HasChildren})$
Logit	$\hat{P} = \Lambda(-1.5 + 0.04\,\text{Income} + 0.6\,\text{HasChildren})$

Using the logit model, what is the predicted probability of enrollment and the marginal effect of Income for a childless household earning $30,000?

- **a** PP = 0.289; ME = 0.040
- **b** PP = 0.289; ME = 0.008
- **c** PP = 0.426; ME = 0.010
- **d** PP = 0.426; ME = 0.040
- **e** None of the given answer options are correct.

Source: Spring 2026 Midterm (Boswell), Question 2. PP: predicted probability. ME: marginal effect.

PROBLEM 7(B) · LOGIT PREDICTION

The Marginal Effect Is the Slope of the Curve

PROBABILITY

$$z = -1.5 + 0.04(30) + 0.6(0) = -0.3$$

$$\hat{P} = \Lambda(-0.3) = \frac{1}{1 + e^{0.3}} = 0.426$$

MARGINAL EFFECT: DERIVATIVE

$$\frac{\partial P}{\partial \text{Income}} = \Lambda(z)[1 - \Lambda(z)]\,\beta$$

$$= 0.426 \times 0.574 \times 0.04 = 0.0098$$

OR: ONE MORE UNIT

$\Lambda(-0.26) - \Lambda(-0.30) = 0.4354 - 0.4256 = 0.0098$. Both round to 0.010.

> 0.04 is the coefficient, and 0.008 is the LPM's slope. Only the LPM's coefficient is its marginal effect.

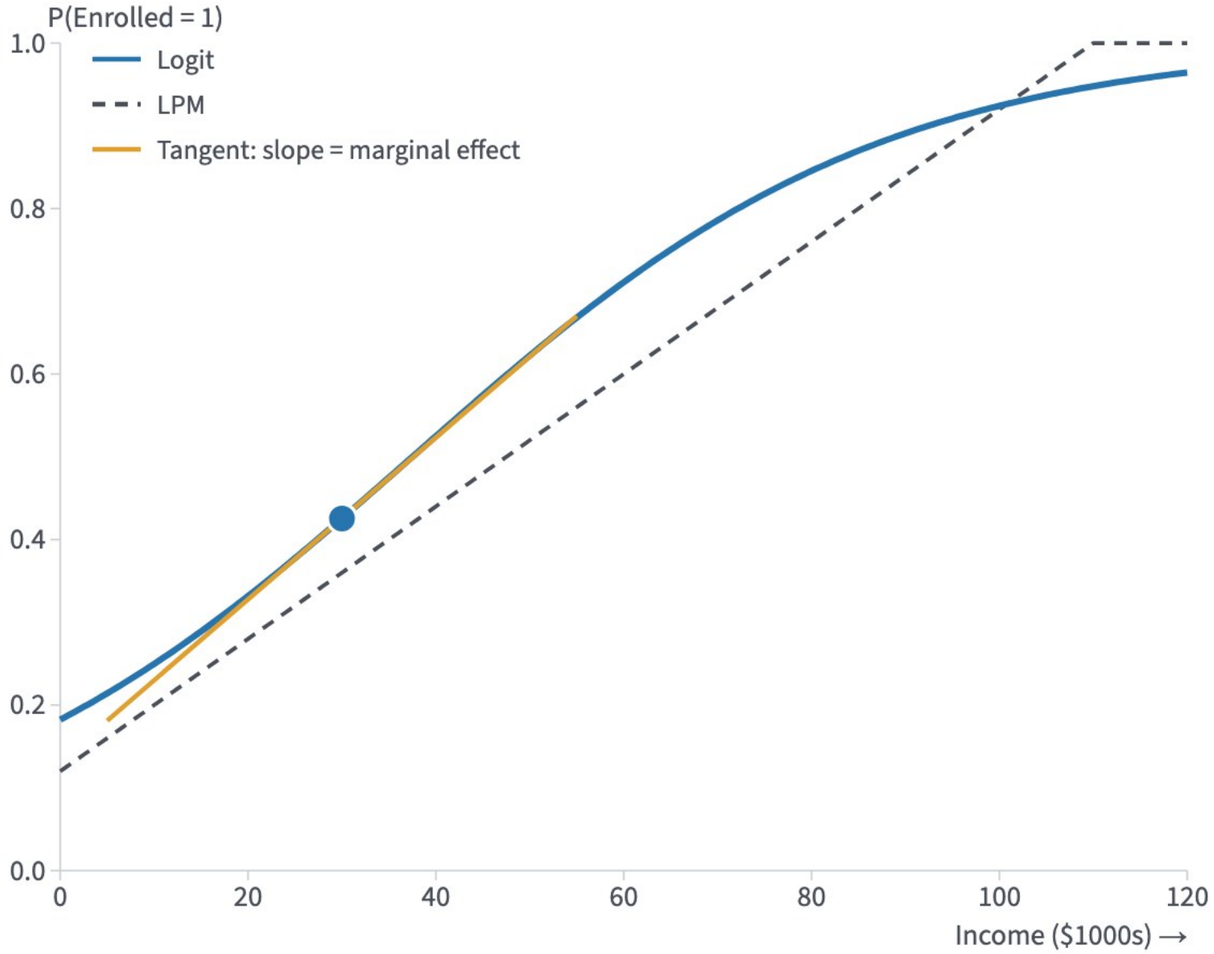
Income ($1000s) **30**

HasChildren: 0

z = −0.30, P = Λ(z) = **0.426**

Marginal effect: **0.0098** (derivative), **0.0098** (one more $1000)

LPM marginal effect: 0.008 at every income



PROBLEM 7(B) · LOGIT PREDICTION

Checking Each Answer Choice

a PP = 0.289; ME = 0.040

0.289 is $\Lambda(-0.9)$, not the probability at $z = -0.3$; 0.040 is the coefficient.

b PP = 0.289; ME = 0.008

Wrong probability, and 0.008 is the LPM's Income coefficient.

d PP = 0.426; ME = 0.040

Right probability, but the ME multiplies 0.04 by $\Lambda(1 - \Lambda) = 0.245$.

e None are correct.

Choice (c) is correct.

c PP = 0.426; ME = 0.010

Answer: (c)

PROBLEM 7(C) · LPM

The Question: Is an LPM Prediction of 0.22 Fine?

A health insurer models whether an individual opts into a supplemental dental plan (Enrolled = 1). Income is in $1,000s; HasChildren = 1 if there are children in the household. 800 observations:

Model	Estimated equation
LPM	$\hat{P} = 0.12 + 0.008\,\text{Income} + 0.14\,\text{HasChildren}$
Probit	$\hat{P} = \Phi(-0.9 + 0.024\,\text{Income} + 0.36\,\text{HasChildren})$
Logit	$\hat{P} = \Lambda(-1.5 + 0.04\,\text{Income} + 0.6\,\text{HasChildren})$

For a childless household earning $12,500, the LPM predicts exactly 0.22. A colleague argues this is fine because 0.22 is within [0, 1]. Which most precisely identifies why the LPM prediction might still be problematic?

a The LPM always produces biased coefficient estimates, so all predictions are unreliable regardless of their value.

b The LPM imposes a constant marginal effect of income across all income levels, which is inconsistent with the diminishing returns to income on probability typically observed in insurance take-up decisions.

c A prediction of 0.22 implies the model predicts negative probabilities for households with lower income, violating the probability axioms in other regions of the covariate space.

d The LPM cannot be estimated by OLS when the dependent variable is binary.

e The LPM standard errors are always smaller than probit standard errors, making inference invalid.

Source: Spring 2026 Midterm (Boswell), Question 3

PROBLEM 7(C) · LPM

Checking Each Answer Choice

a The LPM is always biased.

This is not the problem with this prediction, and the LPM is not always biased.

c Lower incomes give negative predictions.

At Income = 0 with no children, $0.12 > 0$. No household with income ≥ 0 gets a negative prediction.

d Cannot be estimated by OLS.

The LPM is OLS with a 0/1 dependent variable.

e LPM SEs are always smaller.

There is no such rule. The LPM's issue is heteroskedasticity: use robust SEs.

b The LPM forces the same marginal effect of income at every income level.

Answer: (b)

CONSTANT SLOPE

LPM: one more \$1,000 adds 0.008 at Income = 12.5 and at Income = 120.

Logit: the slope $\Lambda(1-\Lambda) \times 0.04$ is largest near $\hat{P} = 0.5$ and shrinks toward 0 and 1.

PROBLEM 7(D) · NEW: LECTURE 9

Three Ways to Report a Marginal Effect

Lecture 9's probit for driving to work: $\hat{P} = \Phi(-0.0644 + 0.3000\,\mathrm{DTIME})$, with DTIME = bus time minus car time, in 10-minute units. The marginal effect $\phi(z) \times 0.3$ is different for every person.

AT THE MEAN (MEM)

Plug in the sample mean DTIME $= -0.1224$: $\phi(-0.101) \times 0.3 = 0.1191$.

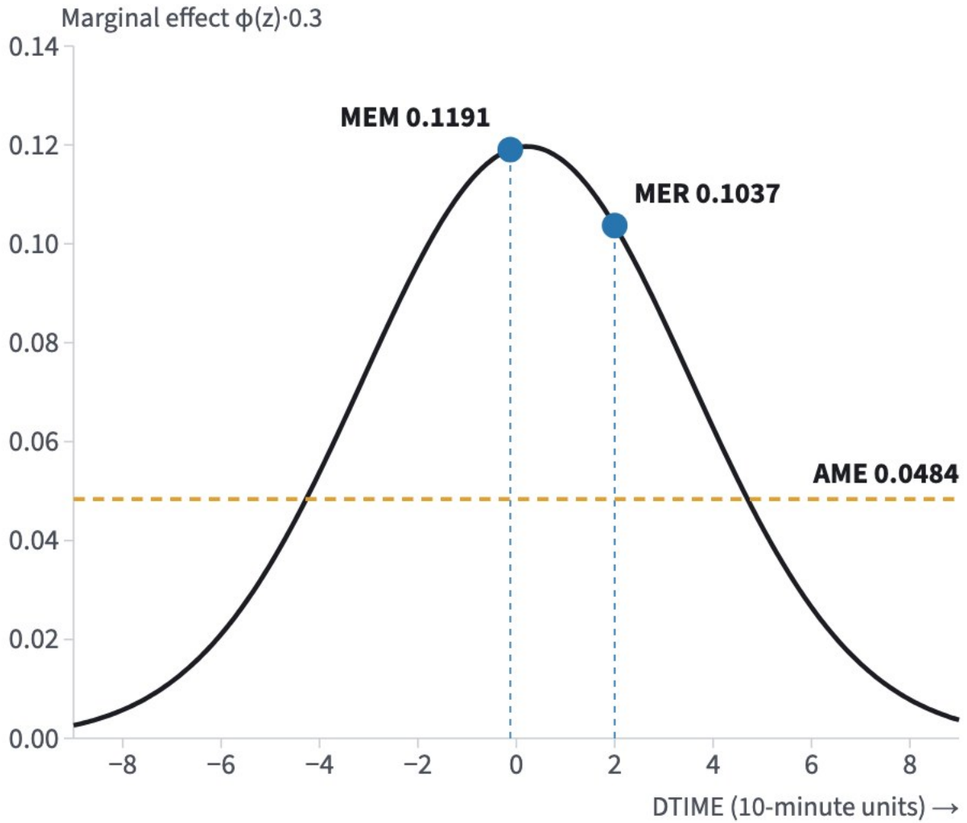
AT A REPRESENTATIVE VALUE (MER)

Pick a value of interest, e.g. the bus takes 20 minutes longer (DTIME = 2): $\phi(0.536) \times 0.3 = 0.1037$.

AVERAGE (AME)

Compute $\phi(z_i) \times 0.3$ for every person, then average:

$$\widehat{AME} = \frac{1}{N}\sum_{i=1}^{N} \phi(-0.0644 + 0.3\,\mathrm{DTIME}_i) \times 0.3 = 0.0484$$



PROBLEM 7(D) · NEW: LECTURE 9

The Question: Why Is the AME Smaller Than the MEM?

In Lecture 9's probit, the marginal effect of DTIME at the mean (MEM) is 0.1191, and the average marginal effect (AME) is 0.0484. Which statement explains the gap?

a The AME must be computed incorrectly: in any model, the average of the marginal effects equals the marginal effect at the average.

b At the mean DTIME, z is close to 0, where ϕ is largest. Many people have DTIME far from 0, where $\phi(z_i)$ is small, so their marginal effects pull the average down.

c The AME uses the logit density, which is flatter than the normal density.

d The MEM is larger because it ignores the sign of the coefficient.

e The two numbers measure different variables.

Source: New question, written from Lecture 9 (Reporting Marginal Effects)

PROBLEM 7(D) · NEW: LECTURE 9

Checking Each Answer Choice

a The average of MEs equals the ME at the average.

True only when the marginal effect is the same for everyone, as in the LPM. Here it depends on $\phi(z_i)$.

c The AME uses the logit density.

Both use the probit density ϕ.

d The MEM ignores the sign.

Both are positive because the coefficient 0.3 is positive.

e Different variables.

Both are the marginal effect of DTIME.

b ϕ is largest near $z = 0$; people far from 0 have small marginal effects and pull the average down.

Answer: (b)

PROBLEM 8(A) · LPM

The Question: The LPM Prediction for 5 Ads and 20 Views

We model whether someone subscribes to a YouTube channel ($Y = 1$) using the number of video views and the number of ads during the video. Coefficients from 5,000 users:

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

What is the estimated probability of subscribing for a user who sees 5 ads where the video has 20 views, according to the LPM? (Censor your result where necessary if it falls outside the bounds of probability.)

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

PROBLEM 8(A) · LPM

Checking Each Answer Choice

a 0.0

The raw prediction is above 1, not below 0.

b 0.5

0.5 is the Ads term alone: 0.1×5.

c 0.6

$0.2 + 0.02(20) = 0.6$: leaves out the Ads term 0.1×5.

e None of the above.

Choice (d) is correct.

d 1.0

Answer: (d)

$0.2 + 0.1(5) + 0.02(20) = 1.1$
A probability cannot exceed 1, so censor it to 1.0.

This is the LPM's first problem: a straight line leaves [0, 1]. Probit and logit stay inside by construction.

The Question: The Probit Prediction for the Same User

We model whether someone subscribes to a YouTube channel ($Y = 1$) using the number of video views and the number of ads during the video. Coefficients from 5,000 users:

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

What is the difference between the estimate in 8(a) and the probability of subscribing for the same user (5 ads, 20 views) using the probit model? ($\Phi(1.5) = 0.9332, \phi(1.5) = 0.1295$)

- **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

PROBLEM 8(B) · PROBIT

Checking Each Answer Choice

b 0.8705

$1 - \phi(1.5)$: uses the density, not the CDF.

c 0.1295

$\phi(1.5)$ is a height of the density, not a probability.

d 0.9332

This is the probit probability itself, not the difference.

e None are correct.

Choice (a) is correct.

a $1.0 - 0.9332 = 0.0668$

Answer: (a)

$z = -1.0 + 0.3(5) + 0.05(20) = 1.5$
$\hat{P} = \Phi(1.5) = 0.9332$
Difference: $1.0 - 0.9332 = 0.0668$

> Φ (CDF) gives probabilities. ϕ (density) appears only in marginal effects.

The Question: The Effect of One More Ad

We model whether someone subscribes to a YouTube channel ($Y = 1$) using the number of video views and the number of ads during the video. Coefficients from 5,000 users:

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

Using the probit model, what is the effect of one more ad on the probability of subscribing for a user who sees 5 ads where the video has 20 views? ($\Phi(1.5) = 0.9332, \phi(1.5) = 0.1295, \Phi(1.8) = 0.9641, \phi(1.8) = 0.0790$) Round to two decimal places.

a. 0.03

b. 0.04

c. 0.05

d. 0.06

e. None of the above are correct

Source: Fall 2024 Midterm V1 (Boswell), Question 18

PROBLEM 8(C) · PROBIT MARGINAL EFFECT

Two Ways to Get the Effect of One More Ad

DIFFERENCE

6 ads: $z = 1.8$. $\Phi(1.8) - \Phi(1.5) = 0.9641 - 0.9332 = 0.031 \approx 0.03$

DERIVATIVE

$\phi(1.5) \times 0.3 = 0.1295 \times 0.3 = 0.039 \approx 0.04$

SECANT VS TANGENT

The derivative is the slope of the tangent at 5 ads. The curve flattens as it nears 1, so the actual one-unit change (secant) is smaller.

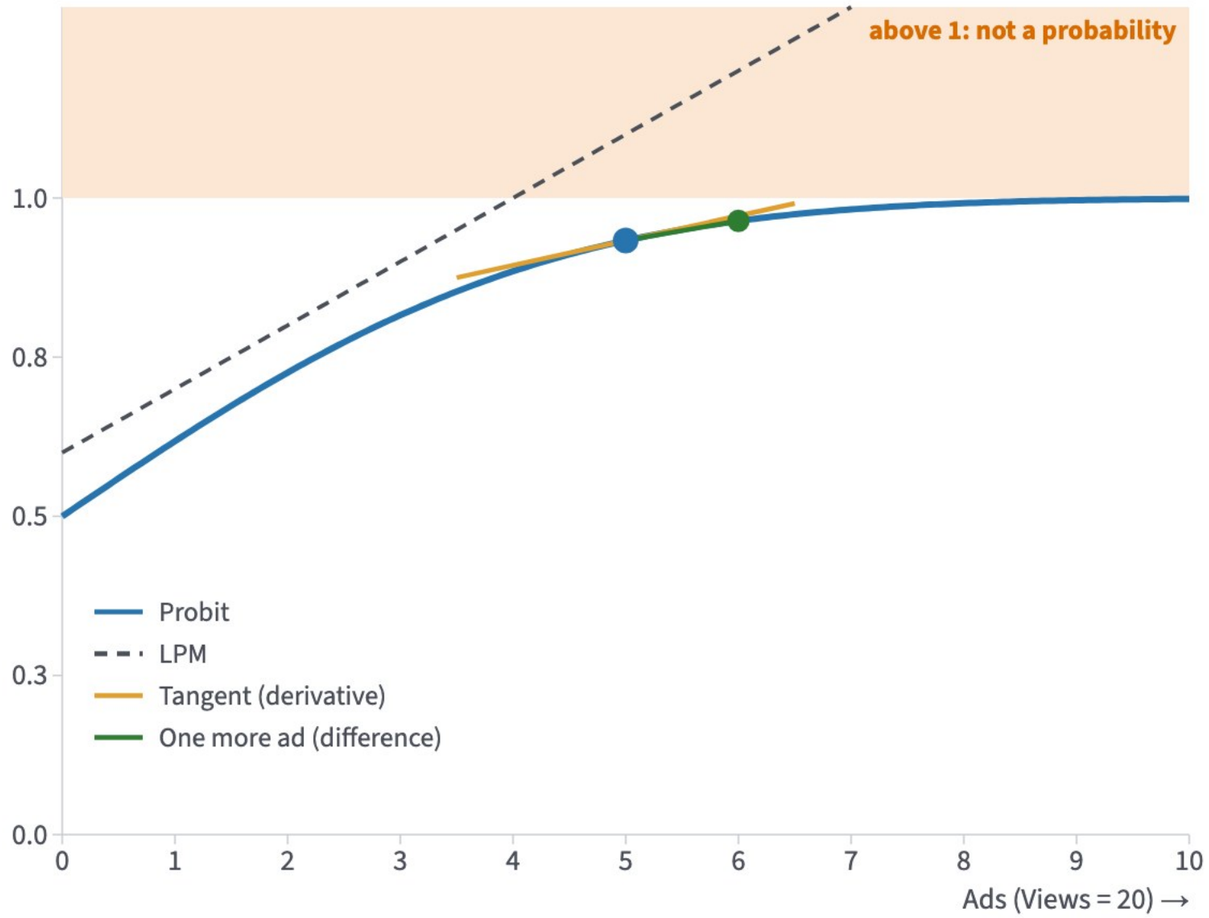
> Both (a) and (b) were accepted on the exam. If the question gives Φ at both points, it is pointing you to the difference.

Ads **5.0**

z = 1.50, Φ(z) = **0.9332**; LPM: **1.10**

Derivative ϕ(z)·0.3 = **0.0389**

Difference Φ(z + 0.3) − Φ(z) = **0.0309**



PROBLEM 8(C) · PROBIT MARGINAL EFFECT

Checking Each Answer Choice

c 0.05

Not either method.

d 0.06

Not either method.

e None are correct.

0.03 and 0.04 are both correct, by the two methods.

a 0.03: the difference $\Phi(1.8) - \Phi(1.5)$

Accepted

b 0.04: the derivative $\phi(1.5) \times 0.3$

Accepted

PROBLEM 8(D) · PROBIT SYMMETRY

The Question: The Probability of Failing

You estimate a probit model for the probability of passing ECON 104 depending on hours of study, where $y_i = 1$ denotes passing. You get an intercept $\beta_0 = -0.5$ and a coefficient on hours of study $\beta_1 = 2$. For a student who studied half an hour, what is the probability of them failing the course?

a $\Phi(0.5)$

b $\phi(0.5)$

c $\Phi(-0.5)$

d $\phi(2)$

e None of the given answer options are correct.

Source: Past paper question shown in Lecture 9

PROBLEM 8(D) · PROBIT SYMMETRY

Checking Each Answer Choice

a $\Phi(0.5)$

This is the probability of passing.

b $\phi(0.5)$

A density height, not a probability.

d $\phi(2)$

Uses the coefficient instead of z, and the density instead of the CDF.

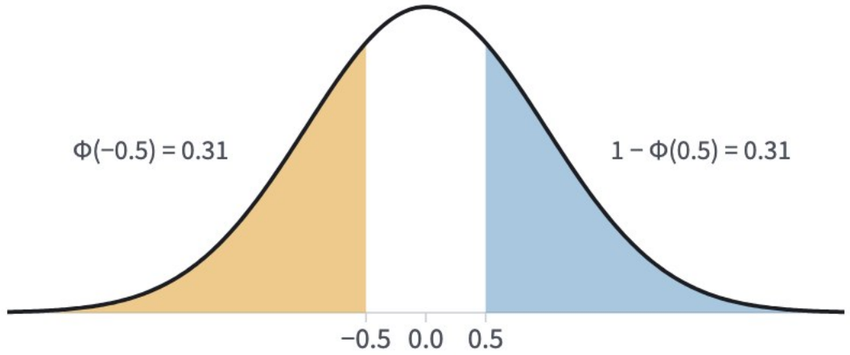
e None are correct.

Choice (c) is correct.

c $\Phi(-0.5)$

Answer: (c)

$z = -0.5 + 2(0.5) = 0.5$. Pass: $\Phi(0.5)$. Fail: $1 - \Phi(0.5)$.
The normal curve is symmetric: the area above 0.5 equals the area below -0.5, so $1 - \Phi(0.5) = \Phi(-0.5) = 0.31$.



PROBLEM 8(E) · LIKELIHOOD

The Question: Which Likelihood Is Maximized?

The model in 8(d) was estimated on a random sample of $n = 334$ students, where 15 failed. Which expression is the likelihood function maximized to estimate β_0 and β_1, where $y_i = 1$ denotes passing?

a $L = \sum_{i=1}^{334} [\Phi(\beta_0 + \beta_1 x_i)]^{y_i} + [1 - \Phi(\beta_0 + \beta_1 x_i)]^{1-y_i}$

b $L = \prod_{i=1}^{334} [\Phi(\beta_0 + \beta_1 x_i)]^{y_i} \, [1 - \Phi(\beta_0 + \beta_1 x_i)]^{1-y_i}$

c $L = \sum_{i=1}^{334} y_i \log[\Phi(\beta_0 + \beta_1 x_i)] + (1 - y_i) \log[1 - \Phi(\beta_0 + \beta_1 x_i)]$

d $L = \prod_{i=1}^{334} [\Phi(\beta_0 + \beta_1 x_i)]^{y_i}$

Source: Past paper question shown in Lecture 9

PROBLEM 8(E) · LIKELIHOOD

Checking Each Answer Choice

a A sum of the two pieces.

Independent observations multiply. A sum of probabilities is not a likelihood.

c The sum of logs.

This is the log-likelihood. It has the same maximizer, but the question asks for the likelihood L.

d Only the $y_i = 1$ factor.

Leaves out the 15 students who failed: each contributes $1 - \Phi$.

b $\prod_i \Phi^{y_i}(1-\Phi)^{1-y_i}$

Answer: (b)

Each student contributes $\Pr(y_i)$: $\Phi(z_i)$ if they passed, $1 - \Phi(z_i)$ if they failed. The exponents y_i and $1 - y_i$ keep Φ for students who passed and $1 - \Phi$ for students who failed. Maximum likelihood picks the βs that make the observed passes and fails most probable.

PROBLEM 9(A) · ACCURACY

The Question: Rank the Models by Accuracy

Using diabetes data (Type II diabetes = 1), you compare three binary classification models. Their confusion matrices for a test set of 100 observations:

Model	True pos	False pos	True neg	False neg
Probit	30	20	40	10
Logit	20	10	50	20
LPM	35	15	35	15

Which ranks the models from least to most accurate based on overall classification accuracy?

a. LPM < Logit < Probit
b. Logit < Probit < LPM
c. Logit < LPM < Probit
d. Probit < Logit < LPM
e. All have equal accuracy

Source: Fall 2025 Midterm V1 (Boswell), Question 18

PROBLEM 9(A) · ACCURACY

Checking Each Answer Choice

a LPM < Logit < Probit

All three are 0.70.

b Logit < Probit < LPM

All three are 0.70.

c Logit < LPM < Probit

All three are 0.70.

d Probit < Logit < LPM

All three are 0.70.

e All have equal accuracy

Answer: (e)

$$\text{Accuracy} = \frac{TP + TN}{\text{total}}$$

Probit: $(30 + 40)/100 = 0.70$

Logit: $(20 + 50)/100 = 0.70$

LPM: $(35 + 35)/100 = 0.70$

Same accuracy, different mistakes: Logit has the fewest false positives, Probit the fewest false negatives. The next question uses measures that separate them.

PROBLEM 9(B) · NEW: LECTURE 9

The Question: Which Model Catches the Most Cases?

Using diabetes data (Type II diabetes = 1), you compare three binary classification models. Their confusion matrices for a test set of 100 observations:

Model	True pos	False pos	True neg	False neg
Probit	30	20	40	10
Logit	20	10	50	20
LPM	35	15	35	15

A doctor screening for diabetes most wants to avoid missing true cases. Which model has the highest recall, and which has the highest F1?

a. Probit has the highest recall; LPM has the highest F1

b. Logit has the highest recall and the highest F1

c. LPM has the highest recall and the highest F1

d. Probit has the highest recall and the highest F1

e. All three are equal on both, because their accuracy is equal

Source: New question, written from Lecture 9 (F1) using the table from Fall 2025 Midterm V1, Question 18

PROBLEM 9(B) · NEW: LECTURE 9

Precision, Recall, Specificity, F1

PRECISION

$\frac{TP}{TP + FP}$: when the model says yes, how often is it right?

RECALL (SENSITIVITY, TRUE POSITIVE RATE)

$\frac{TP}{TP + FN}$: of the actual positives, how many does it catch?

SPECIFICITY (TRUE NEGATIVE RATE)

$\frac{TN}{TN + FP}$: of the actual negatives, how many does it correctly predict as 0?

F1

$\frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$: the harmonic mean of the two. Useful when one class is rare and both mistakes cost about the same.

From the table in 9(a)	Probit	Logit	LPM
Accuracy	0.70	0.70	0.70
Precision	0.60	0.67	0.70
Recall	0.75	0.50	0.70
Specificity	0.67	0.83	0.70
F1	0.67	0.57	0.70

> Lecture 9: with 1% fraud, predicting "not fraud" every time is 99% accurate and has recall 0. Accuracy alone can hide this.

> The LPM row has 50 actual positives (35 + 15); the other rows have 40. One test set would give the same count in every row. The table comes from the exam, so use the numbers as given.

PROBLEM 9(B) · NEW: LECTURE 9

Checking Each Answer Choice

b Logit: highest recall and F1.

Logit has the lowest recall (0.50): it misses 20 of 40 cases.

c LPM: highest recall and F1.

LPM has the highest F1 (0.70), but Probit's recall (0.75) is higher.

d Probit: highest recall and F1.

Probit's F1 is 0.67, below LPM's 0.70.

e All equal.

Equal accuracy does not mean equal recall or F1.

a Probit has the highest recall (0.75); LPM has the highest F1 (0.70).

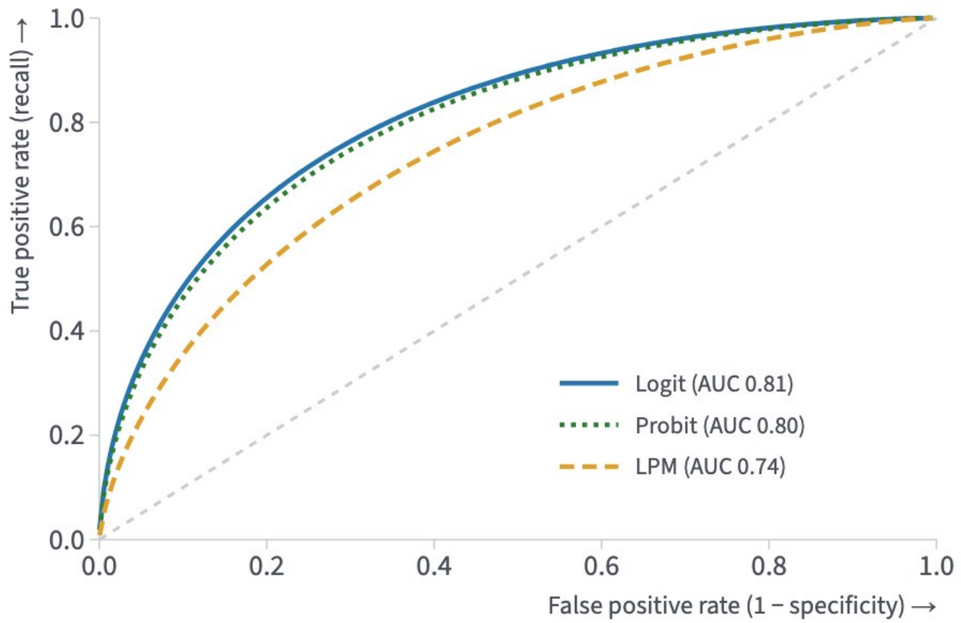
Answer: (a)

To avoid missing cases, choose the model with the highest recall: Probit. It also has more false positives (precision 0.60).

The Question: Which Model Performs Best?

The dental-plan insurer (Problem 7) evaluates all three models using ROC curves and fit statistics on the same 800 observations.

Model	Log-likelihood	AIC	BIC	AUC
Logit	−388.2	784.4	800.1	0.81
Probit	−389.7	787.4	803.1	0.80
LPM	—	—	—	0.74



Source: Spring 2026 Midterm (Boswell), Question 6. ROC curves redrawn with the reported AUCs.

- **a** Probit; although its AIC and BIC are slightly higher than the Logit's, the two models' ROC curves nearly overlap, so the Probit is preferred for its simpler latent-variable interpretation.
- **b** No model can be selected because the Logit and Probit AUCs differ by only 0.01, which is within sampling variability; a formal test of AUC equality is required before any recommendation.
- **c** LPM; AIC and BIC are unavailable for the LPM, so it cannot be penalized on those criteria, and its transparency makes it the safest default.
- **d** Logit; it has the lowest AIC, lowest BIC, and highest AUC, indicating the best in-sample fit and the strongest discriminatory power among the three models.
- **e** LPM; its ROC curve shows the steepest initial slope near the origin, indicating superior performance where it matters most.

PROBLEM 9(C) · TRY IT

Move the Threshold: One ROC Curve, Many Confusion Matrices

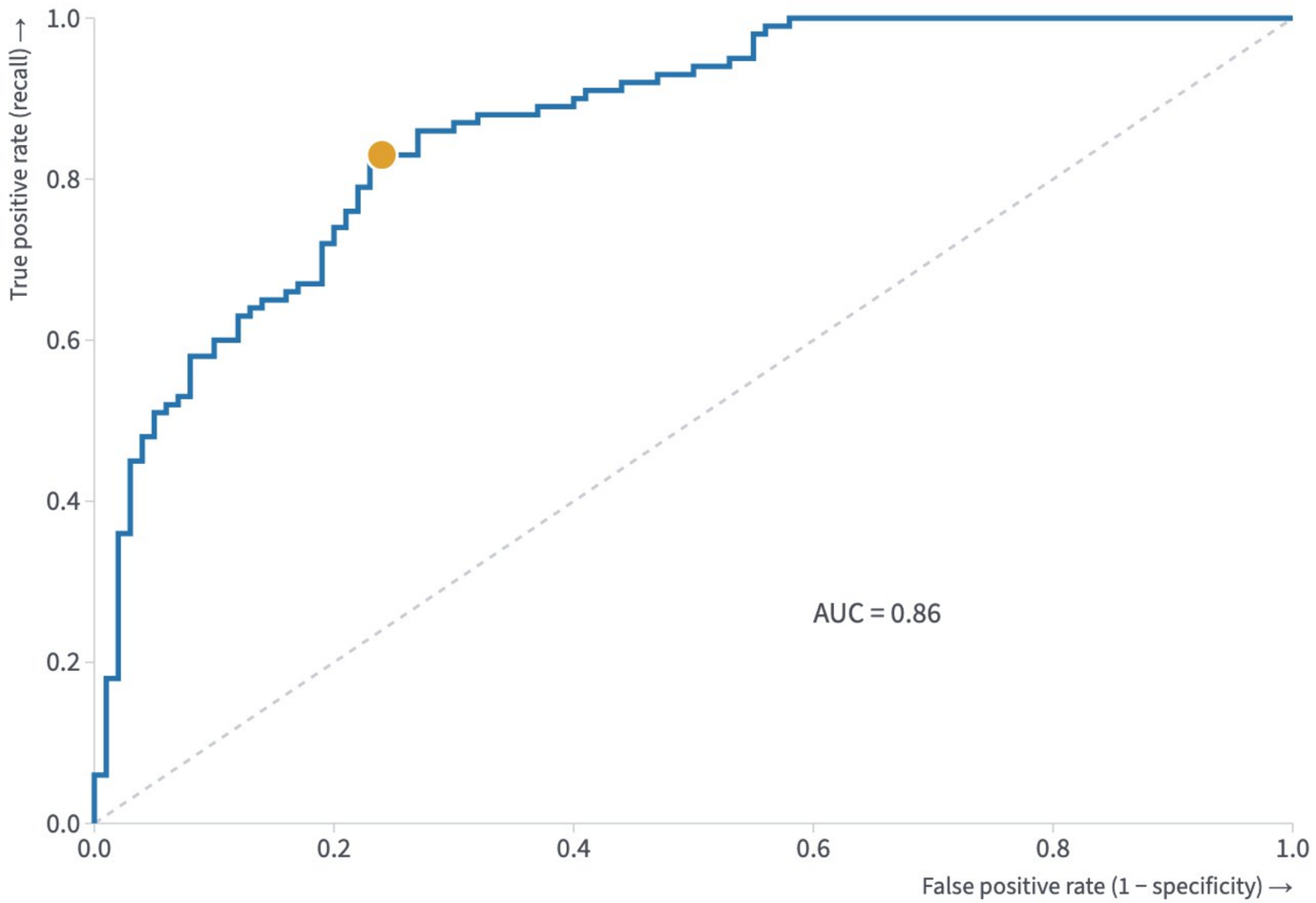
Threshold c **0.50**

Predict 1 when $\hat{P} > c$. 200 simulated people, 100 with $y = 1$.

	Predicted 1	Predicted 0
Actual 1	TP 83	FN 17
Actual 0	FP 24	TN 76

Accuracy **0.80** · Precision **0.78** · Recall **0.83**
Specificity **0.76** · F1 **0.80**

Lowering c catches more positives (recall up) and raises false alarms (specificity down). The ROC curve traces every threshold. AUC: 0.5 is guessing, 1 is perfect.



PROBLEM 9(C) · ROC AND AIC

Checking Each Answer Choice

a Probit, for its interpretation.

Probit is worse on AIC, BIC, and AUC; interpretation does not overturn that.

b No model can be selected.

Logit is ahead on all three measures; we are asked which is best, not whether the gap is significant.

c LPM, since AIC and BIC are unavailable.

Missing criteria are not a reason to choose it, and its AUC (0.74) is the lowest.

e LPM, steepest near the origin.

Its curve is the lowest, and its AUC is the lowest.

d Logit: lowest AIC, lowest BIC, highest AUC.

Answer: (d)

AIC and BIC: lower is better (fit, penalized for parameters). AUC: higher is better (how well the model ranks 1s above 0s, over every threshold).

> Lecture 9: McFadden's $R^2 = 1 - \ln L_{\text{full}} / \ln L_{\text{null}}$ cannot be compared with the LPM's OLS R^2; 0.2 to 0.4 is already a very good fit.

PROBLEM 9(D) · MODEL CHOICE

The Question: Match Each Scenario to a Model

(I) A researcher models the number of hospital visits individuals make in a year, which ranges from 0 to 20 and is skewed right.
(II) A company wants to model how applicants rank three job offers (A, B, and C) in order of preference.

Which pair correctly matches each scenario with the most appropriate model type?

a. Scenario 1: Ordered Logit; Scenario 2: Binary Logit

b. Scenario 1: Multinomial Logit; Scenario 2: Poisson Regression

c. Scenario 1: Binary Probit; Scenario 2: Multinomial Probit

d. Scenario 1: Rank-Ordered Logit; Scenario 2: Count Model

e. Scenario 1: Poisson Regression; Scenario 2: Ordered Logit

Source: Fall 2025 Midterm V1 (Boswell), Question 19

PROBLEM 9(D) · MODEL CHOICE

Match the Dependent Variable to the Model

The dependent variable is	Example	Model
0/1	Drives to work or not	LPM, probit, logit
One of 3+ unordered choices	Car, bus, or scooter	Multinomial logit (person's traits); conditional logit (choice's traits)
Ranked categories	Strongly disagree ... strongly agree	Ordered probit or ordered logit
A count: 0, 1, 2, ...	Hospital visits per year	Poisson: $E(y) = \exp(\beta_1 + \beta_2 x)$
Censored: all units kept, values capped	Speedometer readings stop at 85 mph	Tobit
Truncated: units beyond a limit are missing	Only cars slower than 85 mph are in the data	Truncated regression

> Multinomial and conditional logit assume IIA (independence of irrelevant alternatives): adding a new option does not change the relative odds of the existing two. Lecture 10: adding scooters should not change the car-to-bus odds.

PROBLEM 9(D) · MODEL CHOICE

Checking Each Answer Choice

a Ordered logit; binary logit.

Visits are a count, and the job offers are not a yes/no outcome.

b Multinomial logit; Poisson.

Reversed: Poisson is for the count.

c Binary probit; multinomial probit.

Visits are not 0/1, and multinomial models ignore the ranking.

d Rank-ordered logit; count model.

Reversed.

e Scenario 1: Poisson regression; Scenario 2: ordered logit

Answer: (e)

PROBLEM 9(E) · IIA

The Question: A New Plan Changes the Old Choices

The insurer adds a vision plan, so individuals now choose between dental, vision, and no plan. This changes the relative probabilities of choosing dental vs. no plan substantially. What does this suggest about using the multinomial logit model?

a The model violated the assumption of linearity in parameters.

b The new choice introduces perfect multicollinearity.

c The independence of irrelevant alternatives (IIA) assumption may not hold.

d The logistic model overfits the original set of insurance options.

e The multinomial logit model is appropriate to use

Source: Spring 2026 Midterm (Boswell), Question 4 (wording shortened)

PROBLEM 9(E) · IIA

Checking Each Answer Choice

a Linearity in parameters.

Not the issue: the problem is how odds between existing options react to a new one.

b Perfect multicollinearity.

Adding an alternative does not make regressors collinear.

d Overfitting.

Overfitting is about training vs test fit, not about adding an option.

e MNL is appropriate.

The observed shift is what MNL rules out.

c The IIA assumption may not hold.

Answer: (c)

Under IIA the relative odds of two options do not depend on what else is offered. If adding vision moves the dental-to-none odds, IIA fails. Models without IIA: multinomial probit, nested logit.

PROBLEM 9(F) · NEW: LECTURE 10

The Question: Speedometers Stop at 85 mph

In the 1980s a federal law restricted speedometer readings to at most 85 mph. You want to predict a car's top speed from horsepower and engine size. Every car is in the data, but any car faster than 85 mph is recorded as 85. Which model fits, and why?

a. OLS, because top speed is a continuous variable

b. Tobit, because top speed is censored from above at 85

c. Truncated regression, because cars above 85 mph are missing from the data

d. Poisson, because speed takes whole-number values

e. Ordered probit, because faster cars rank above slower cars

Source: New question, written from Lecture 10 (Tobit example; truncation vs censoring)

PROBLEM 9(F) · NEW: LECTURE 10

Checking Each Answer Choice

a OLS.

Least squares on censored data is biased and inconsistent: the cars recorded at 85 were really faster.

c Truncated regression.

The fast cars are in the data with value 85, so the data are censored, not truncated.

d Poisson.

Speed is not a count of events.

e Ordered probit.

Speed is a continuous measurement, not a set of ranked categories.

b Tobit: top speed is censored from above at 85.

Answer: (b)

Censored: every observation is present, but some values are capped. **Truncated**: observations beyond the limit are not in the data at all.

The Question: The DiD Estimate from a 2×2 Table

Two similar cities: Los Angeles passed a law raising the minimum wage for fast food workers; Humboldt did not. Employment rates before and after the law:

City	Before	After
Humboldt	82%	81%
Los Angeles	80%	75%

Los Angeles is the treatment group. What is the difference-in-differences estimate of the effect of the minimum wage increase on employment?

- **a** −5 percentage points
- **b** −1 percentage point
- **c** −4 percentage points
- **d** −6 percentage points
- **e** +4 percentage points

Source: Numerical example in Lecture 10

PROBLEM 10(A) · DIFFERENCE-IN-DIFFERENCES

Two Differences: Over Time, Then Between Cities

GROUPS

Treatment: Los Angeles (passed the law). Control: Humboldt.

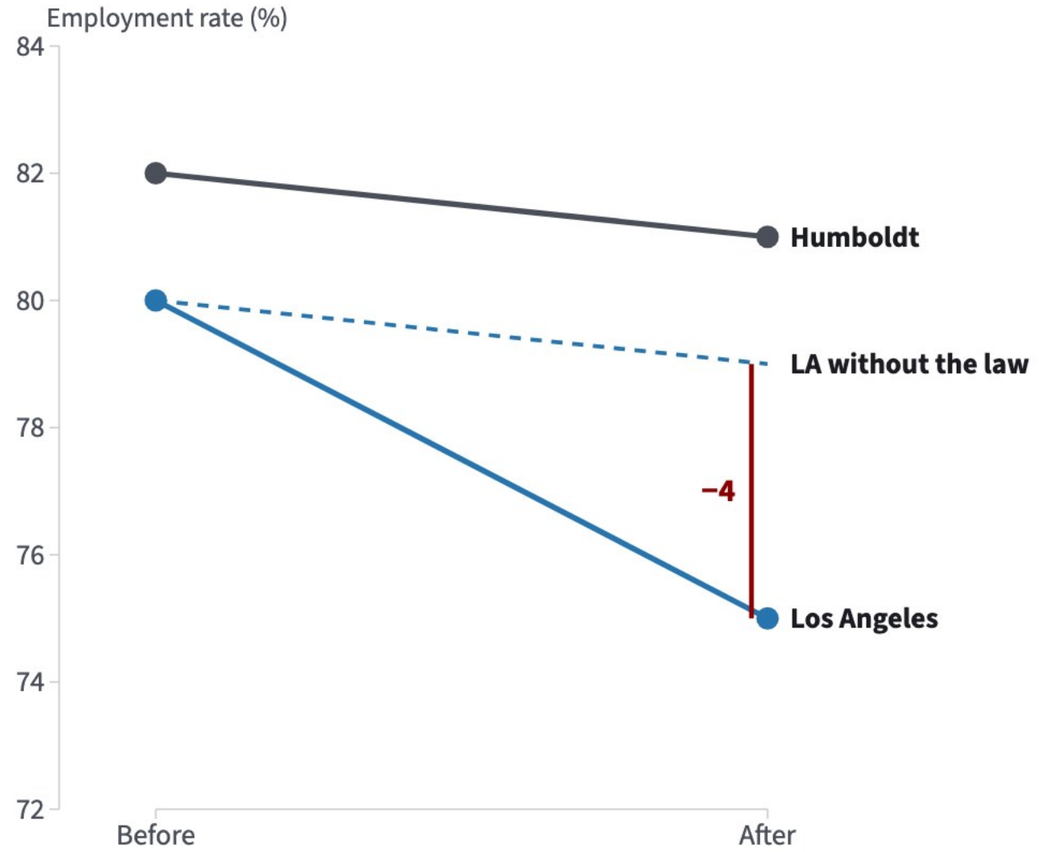
CHANGE IN THE TREATMENT GROUP

LA: $75 - 80 = -5$

CHANGE IN THE CONTROL GROUP

Humboldt: $81 - 82 = -1$. This is what would have happened to LA without the law, if the two cities share the same trend.

DiD $(-5) - (-1) = -4$ percentage points



Checking Each Answer Choice

a −5

LA's own change: ignores the common trend.

b −1

Humboldt's change: the control group's trend, not the effect.

d −6

$75 - 81$: the after-only gap, which includes the 2-point gap that existed before.

e +4

Wrong sign: LA fell by more than Humboldt.

c −4 percentage points

Answer: (c)

The Question: The Predicted Post-Bill Wage in a High-Funded State

A federal infrastructure bill passed in 2024 (Post). States receiving high funding (treatment) are compared with low-funding states (control), using 2023 as the pre-period. Y = local construction wages.

$$Y_{it} = \beta_0 + \beta_1 \text{Post}_t + \beta_2 \text{HighFund}_i + \delta(\text{Post}_t \times \text{HighFund}_i) + \varepsilon_{it}$$

$\hat{\beta}_0 = 22.0, \hat{\beta}_1 = 1.5, \hat{\beta}_2 = 3.0, \hat{\delta} = 2.8$; robust SEs (0.5, 0.6, 0.7, 0.9).

Using the estimated coefficients, what is the predicted post-bill wage in a high-funded state?

- **a** 23.5
- **b** 26.5
- **c** 27.8
- **d** 29.3
- **e** None of the given answers are correct

Source: Spring 2026 Midterm (Boswell), Question 7

Each Coefficient Is a Cell or a Difference

$$Y = \beta_0 + \beta_1 \text{Post} + \beta_2 \text{HighFund} + \delta(\text{Post} \times \text{HighFund})$$

	Pre (Post = 0)	Post (Post = 1)	Change
Low funding	β_0	$\beta_0 + \beta_1$	β_1
High funding	$\beta_0 + \beta_2$	$\beta_0 + \beta_1 + \beta_2 + \delta$	$\beta_1 + \delta$
Difference	β_2	$\beta_2 + \delta$	δ

COMMON TREND (β_1)

The change over time for the control group.

CHANGE IN THE TREATMENT GROUP ($\beta_1 + \delta$)

The common trend plus the treatment effect.

DID ESTIMATE (δ)

The treatment group's change minus the control group's change. β_2, the gap before treatment, cancels in the difference.

> Set Post and HighFund to 0 or 1 and add up the coefficients. Every question about this model can be answered from this table.

Checking Each Answer Choice

a 23.5

$\beta_0 + \beta_1$: the low-funding state after the bill.

b 26.5

$\beta_0 + \beta_1 + \beta_2$: leaves out δ.

c 27.8

$\beta_0 + \beta_2 + \delta$: leaves out the common trend β_1.

e None are correct.

Choice (d) is correct.

d $29.3 = 22.0 + 1.5 + 3.0 + 2.8$

Answer: (d)

The Question: The High-Funded State's Own Change

A federal infrastructure bill passed in 2024 (Post). States receiving high funding (treatment) are compared with low-funding states (control), using 2023 as the pre-period. Y = local construction wages.

$$Y_{it} = \beta_0 + \beta_1 \text{Post}_t + \beta_2 \text{HighFund}_i + \delta(\text{Post}_t \times \text{HighFund}_i) + \varepsilon_{it}$$

$\hat{\beta}_0 = 22.0, \hat{\beta}_1 = 1.5, \hat{\beta}_2 = 3.0, \hat{\delta} = 2.8$; robust SEs (0.5, 0.6, 0.7, 0.9).

What is the total post-bill change in predicted wages for a high-funded state relative to its own pre-period baseline?

- **a** 1.5
- **b** 2.8
- **c** 4.3
- **d** 4.5
- **e** 5.8

Source: Spring 2026 Midterm (Boswell), Question 8

Checking Each Answer Choice

a 1.5

β_1: the control group's change.

b 2.8

δ: the treatment effect, i.e. the change beyond the control group's change. The question asks for the total change.

d 4.5

$\beta_1 + \beta_2$: β_2 is a gap between groups, not a change over time.

e 5.8

$\beta_2 + \delta$: the gap between groups after the bill.

c $4.3 = \beta_1 + \delta = 1.5 + 2.8$

Answer: (c)

$$(\beta_0 + \beta_1 + \beta_2 + \delta) - (\beta_0 + \beta_2) = \beta_1 + \delta$$

PROBLEM 10(D) · PARALLEL TRENDS

The Question: High-Funded States Were Already Growing Faster

A critic argues that high-funding states were already experiencing faster wage growth before the bill passed, due to pre-existing labor market tightness. Which best describes the threat to the DiD estimate, and what empirical check would most directly address it?

a This challenges the parallel trends assumption; it can be assessed by plotting pre-treatment wage trends for both groups and, if sufficient pre-periods exist, testing for differential pre-trends.

b This describes omitted variable bias in the cross-section; it can be resolved by adding more control variables to the post-period regression only.

c This is a measurement error problem; wages in high-funded states are systematically mismeasured prior to treatment, biasing our difference-in-difference estimate downward.

d This describes reverse causality: high pre-treatment wages caused states to receive high funding, which is corrected by instrumenting for HighFund using geographic allocation formulas.

e The parallel trends assumption only applies to the post-period, so pre-existing differences in growth rates are irrelevant to the validity of the DiD estimate.

Source: Spring 2026 Midterm (Boswell), Question 9

PROBLEM 10(D) · TRY IT

Break Parallel Trends: The DiD Estimate Includes the Extra Growth

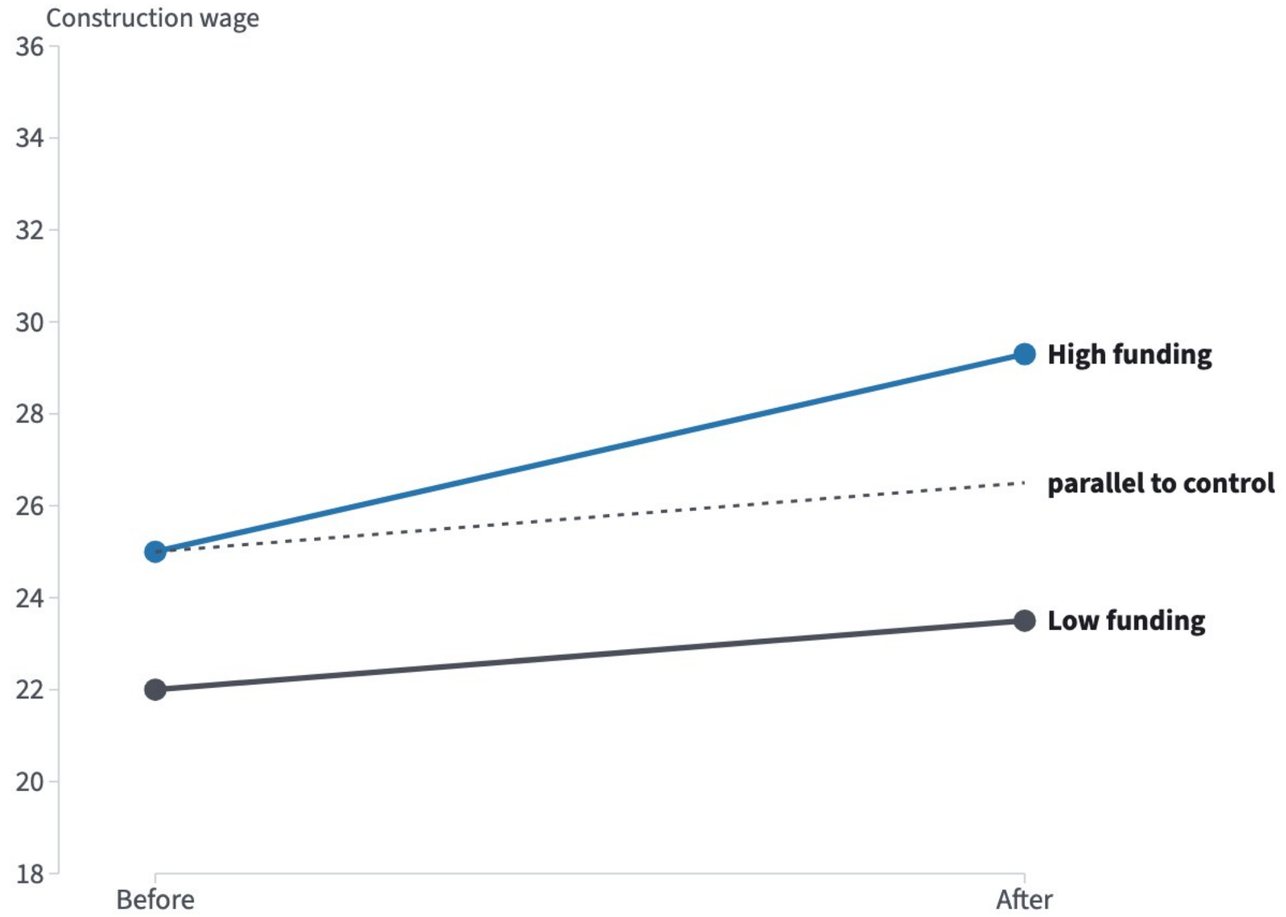
True effect **2.8**

Extra yearly growth, treated states **0.0**

True effect: **2.8**

DiD estimate: (4.3) − (1.5) = **2.8**

Parallel trends: without the bill, both groups would have changed by the same amount. If high-funded states were already growing faster, DiD adds that extra growth to the estimate.



CHECKS (LECTURE 10)

Plot several pre-treatment years: the lines should move in parallel. Placebo test: pretend the bill passed a year earlier; a large, significant "effect" means something is wrong.

Checking Each Answer Choice

b Omitted variable bias; add controls to the post period.

The problem is a difference in trends before treatment; controls in one period cannot reveal or remove it.

c Measurement error.

Nothing suggests wages were measured wrong.

d Reverse causality; use IV.

The critique is about growth rates before the bill, not about how funding was assigned.

e Parallel trends only applies after treatment.

It is about what would have happened without treatment; pre-period trends are the only evidence we can see.

a Parallel trends is in doubt: plot pre-treatment trends and test for different pre-trends.

Answer: (a)

Summary of the Ten Problems

#	Topic	Remember
1	VIF, overfitting	$\text{VIF} = 1/(1 - R_j^2)$; drop one of a collinear pair, not both; judge a model by its test error
2	Goldfeld-Quandt, FGLS	High-variance group's SSE on top; FGLS weight $= 1/\hat{\sigma}_i$; check units; robust SEs change t and p only
3	Breusch-Pagan, WLS	$LM = NR^2$, df = regressors in the auxiliary regression; heteroskedasticity: unbiased but inefficient
4	Stationarity, ACF, PACF	ADF H_0: unit root; band $\pm 1.96/\sqrt{T}$; AR(p): PACF cuts off after p
5	Forecasting	Interval uses σ, not σ^2; lagged Y + autocorrelation $\Longrightarrow$ inconsistent; WLS does not fix autocorrelation
6	Lags, Granger	$(1 - \phi L)^{-1}$; total multiplier $\delta_0/(1 - \lambda)$; Granger: one direction ARDL, both VAR
7	Logit	e^{β} = odds ratio; ME $= \Lambda(1 - \Lambda)\beta$; MEM, MER, AME
8	Probit, LPM	Probabilities from Φ; ME from ϕ; $1 - \Phi(z) = \Phi(-z)$; censor LPM predictions
9	Choosing a model	Precision, recall, F1; lower AIC/BIC, higher AUC; count: Poisson; ranked: ordered; censored: tobit
10	DiD	δ = treated change minus control change; parallel trends: check pre-period plots and placebo

On every question: write H_0 before reading a p-value; write the formula before plugging in; check units (thousands, variance vs SD); for a true/false statement, ask whether it is about the coefficient (bias) or the standard error.

Thank you!

jakeanderson@g.ucla.edu