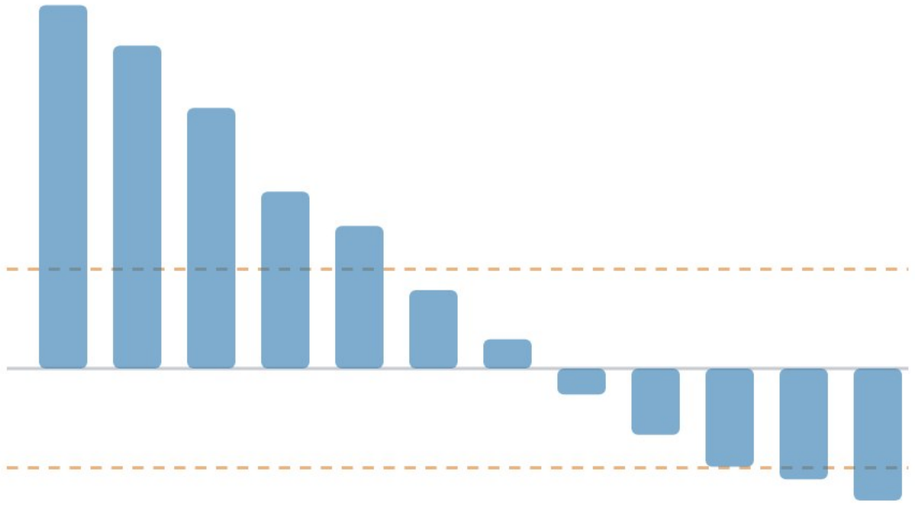
ECON 104 · DISCUSSION · FALL 2026

Lag Length, Serial Correlation Tests, and HAC Standard Errors



Jake Anderson
UCLA

Plan for Today

1

Lag length

How many past values of y should an AR model use?

Compare AIC and BIC across lag lengths and pick the lowest.

Spring 2025 Midterm V4 (Boswell), Question 12 (modified)

2

Breusch-Godfrey test

Are the errors of the chosen model autocorrelated?

Write the auxiliary regression, then compare $LM = T \times R^2$ to a χ^2 critical value.

Winter 2026 Midterm 1 (Rojas), Questions 23 and 24

3

HAC standard errors

The test says the errors are autocorrelated. What now?

Build a confidence interval with standard errors that allow for the autocorrelation.

Spring 2023 Midterm V1, Question 12

The Question: How Many Lags?

You fit AR(1) to AR(5) models to a stationary time series y_t and get the following output. Based on AIC and BIC, what lag length would you choose?

Criterion	Lag 1	Lag 2	Lag 3	Lag 4	Lag 5
AIC	-5.43	-5.88	-5.51	-5.50	-5.30
BIC	-5.20	-5.45	-5.02	-4.89	-4.67

a 1

b 2

c 3

d AIC 2; BIC 3

e None of the given answer options are correct.

Source: Spring 2025 Midterm V4 (Boswell), Question 12, modified: the original asks about VARSelect output for two series; here the five models are AR models for one series.

Each Extra Lag Improves the Fit a Little and Adds a Penalty

SCHWARZ CRITERION (SC), ALSO CALLED BIC

$$SC = \underbrace{\ln(SSE/T)}_{\text{fit}} + \underbrace{k\ln(T)/T}_{\text{penalty}}$$

SSE: sum of squared residuals. T: number of periods used. k: number of coefficients, p lags $+1$ intercept.

FIT

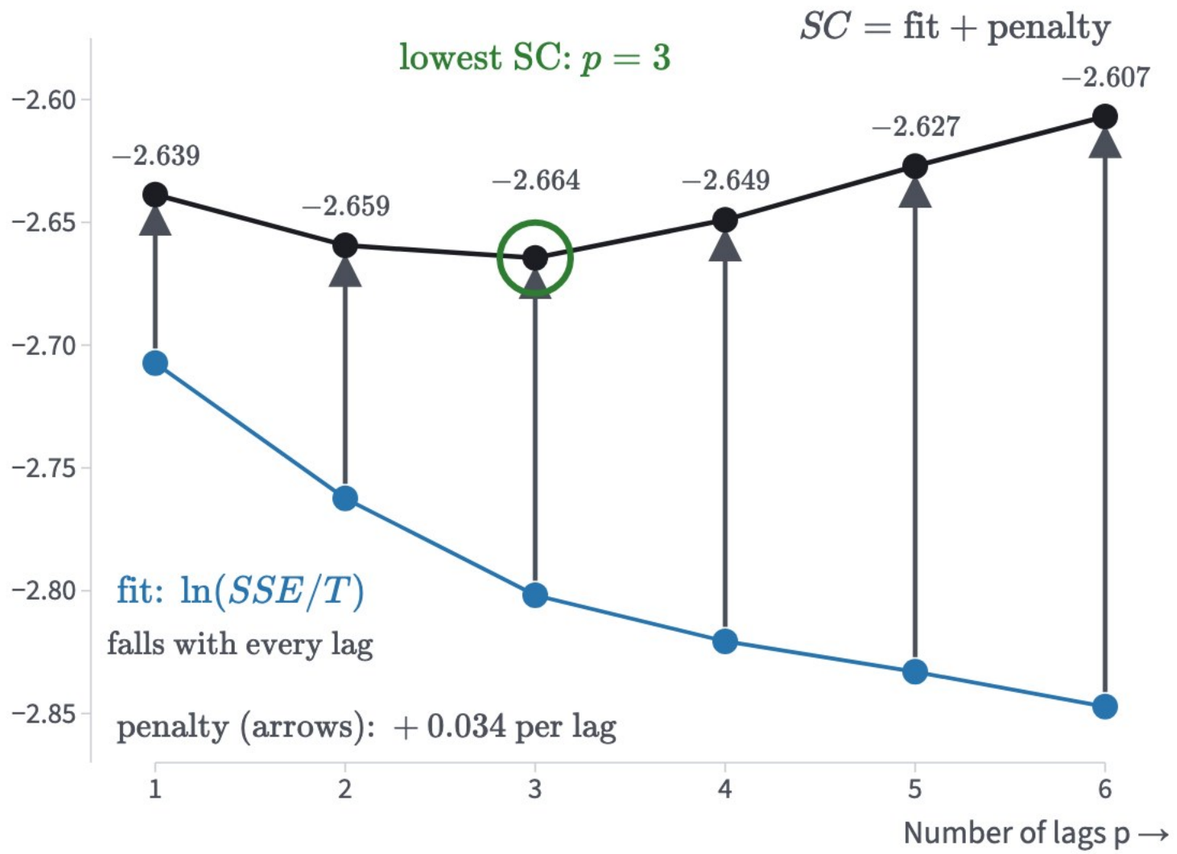
On the same 145 quarters, adding a lag cannot raise SSE (OLS could set its coefficient to 0), so $\ln(SSE/T)$ falls with every lag.

PENALTY

Each coefficient adds $\ln(145)/145 = 0.034$. AR(1) already has $k = 2$: $2 \times 0.034 = 0.069$.

Up to $p = 3$ the fit falls faster than the penalty rises; after that, more slowly. SC is lowest at $p = 3$: pick the lag length with the **lowest** value.

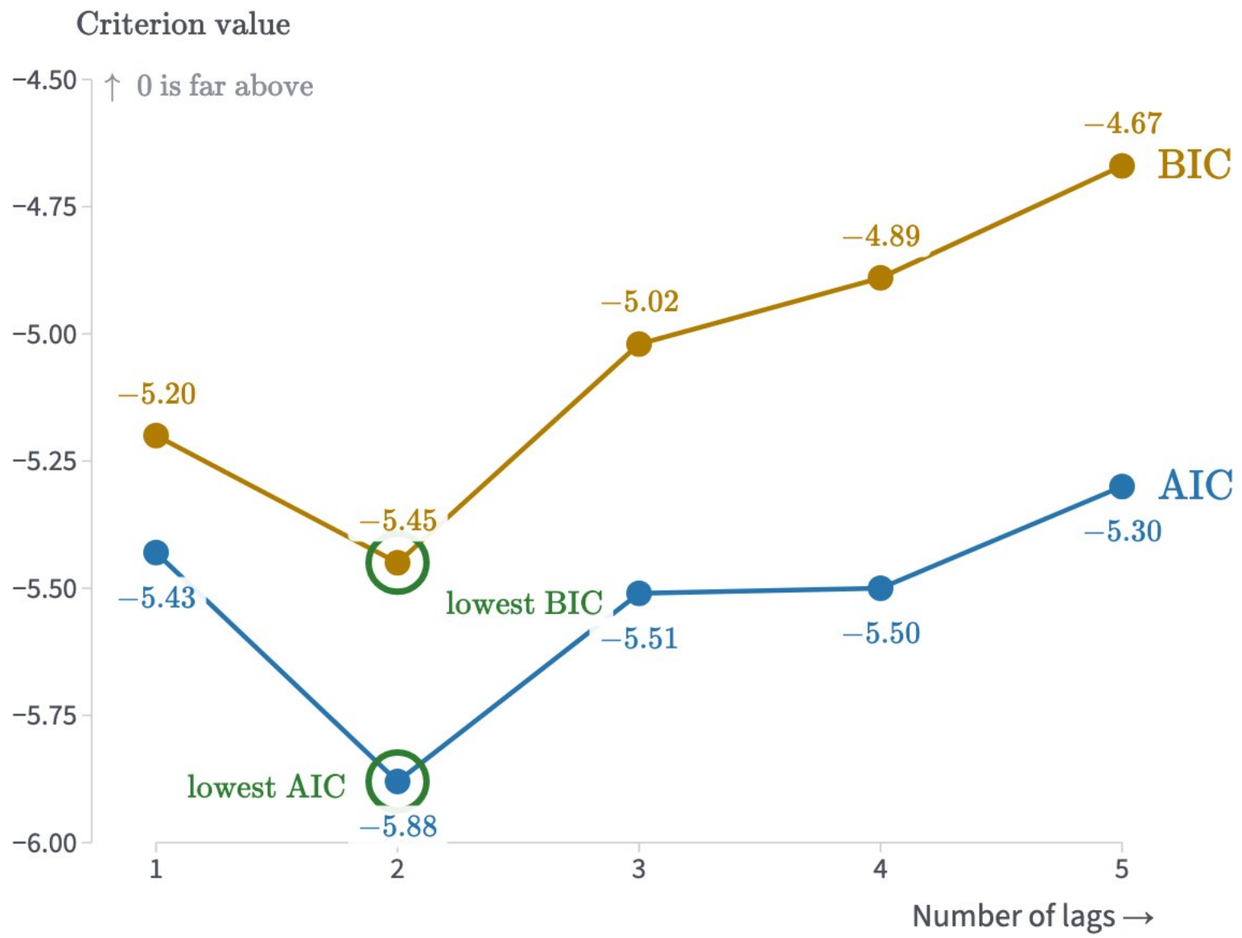
> AIC uses a smaller penalty, $2k/T$ (0.014 per coefficient here), and is lowest at $p = 4$. This series is not the one in the exam question.



A different series from the exam question, computed from `okun5_aus` (POE5): AR(p) models for the quarterly change in Australian unemployment, the same 145 quarters for every p. The lecture's formulas, $n\ln(SSE/n) + k\ln(n)$ and $2k - 2\ln(L)$, are T times these (plus a constant for AIC), so they pick the same p.

Find the Lowest AIC and the Lowest BIC

Ask: Which is lower, -5.43 or -5.88?



AIC

Lowest at lag 2: -5.88 is further below 0 than every other AIC value.

BIC

Lowest at lag 2: -5.45. At every lag BIC is higher than AIC because its penalty is larger ($\ln T > 2$ once $T \geq 8$).

LAG 2 Both criteria pick lag 2 $\Longrightarrow$ AR(2).

WHEN AIC AND BIC DISAGREE

Choose BIC. Its penalty per lag is larger, so it picks fewer lags, and extra lags that only fit noise are left out.

Checking Each Answer Choice

a 1

AIC at lag 1 (-5.43) is higher than at lag 2 (-5.88): it is closer to 0, not lower.

c 3

Neither criterion is lowest at lag 3.

d AIC 2; BIC 3

BIC at lag 3 (-5.02) is higher than at lag 2 (-5.45).

e None of the given answer options are correct.

Choice (b) is correct.

b 2

Answer: (b)

INTERPRETATION

Use an AR(2): y_t depends on the two most recent periods. Next, check the residuals of the chosen model for autocorrelation (Problem 2).

> When AIC and BIC disagree, choose BIC: its penalty per lag is larger, so it picks fewer lags.

The Question: Which Auxiliary Regression?

Question 23. The R code below estimates an ARDL model (answer: ARDL(1,1)).

```
model <- dynlm(du_ts ~ L(du_ts, 1) +
                       L(g_ts, 0:1))
```

`L(du_ts, 1)` — du_{t-1}: last quarter's change in unemployment

`L(g_ts, 0:1)` — g_t and g_{t-1}: GDP growth this quarter and last quarter

Question 24. Based on question 23, below we test for serial correlation in an ARDL model's residuals using the Breusch-Godfrey test of order up to 4 (`bgtest(ardl model, order = 4)`). Which auxiliary regression is used for the BG test? Let $\hat{e}_t$ be the OLS residual from the estimated ARDL model. du_t is the change in unemployment and g_t is GDP growth.

a $\hat{e}_t = \gamma_0 + \gamma_1 du_{t-1} + \gamma_2 g_t + \gamma_3 g_{t-1} + \rho \hat{e}_{t-1} + v_t$

b $\hat{e}_t = \gamma_0 + \rho_1 \hat{e}_{t-1} + \rho_2 \hat{e}_{t-2} + \rho_3 \hat{e}_{t-3} + \rho_4 \hat{e}_{t-4} + v_t$

c $\hat{e}_t = \gamma_0 + \gamma_1 du_{t-1} + \gamma_2 g_t + \gamma_3 g_{t-1} + \rho_1 \hat{e}_{t-1} + \rho_2 \hat{e}_{t-2} + \rho_3 \hat{e}_{t-3} + \rho_4 \hat{e}_{t-4} + v_t$

d $\Delta\hat{e}_t = \gamma_0 + \gamma_1 \Delta du_{t-1} + \gamma_2 \Delta g_t + \gamma_3 \Delta g_{t-1} + \rho_1 \Delta\hat{e}_{t-1} + \rho_2 \Delta\hat{e}_{t-2} + \rho_3 \Delta\hat{e}_{t-3} + \rho_4 \Delta\hat{e}_{t-4} + v_t$

Source: Winter 2026 Midterm 1 (Rojas), Questions 23 and 24

Why Check the Errors of This Model?

THE MODEL

$$du_t = \delta + \theta\, du_{t-1} + \beta_0 g_t + \beta_1 g_{t-1} + e_t$$

θ: how much of last quarter's change in unemployment carries into this quarter. OLS gives $\hat{\theta} = 0.379$.

IT COULD BE THE CASE THAT

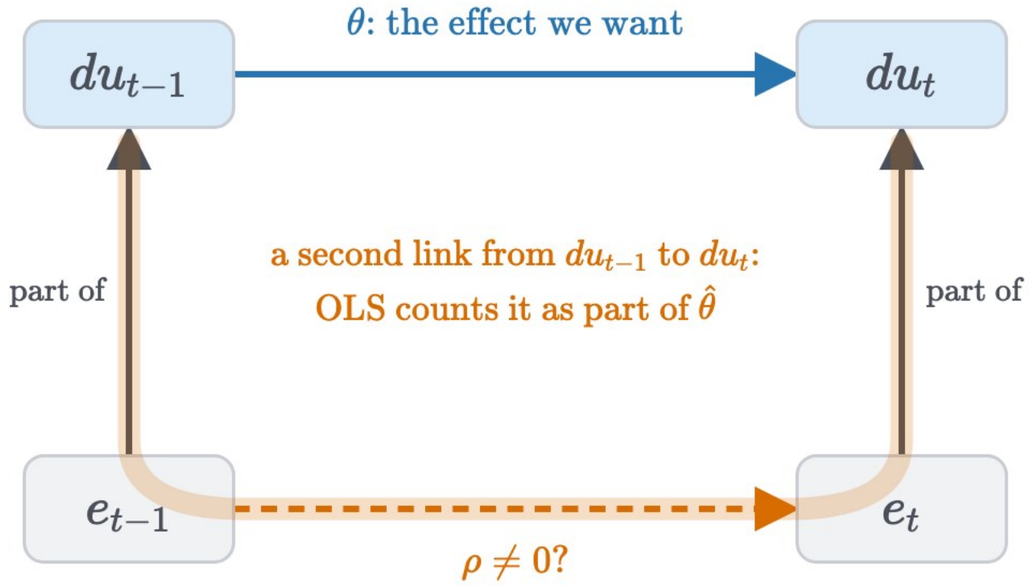
this quarter's error is related to last quarter's error: $e_t = \rho\, e_{t-1} + v_t$ with $\rho \neq 0$. For example, something the model leaves out, such as a slowdown in one industry, raises unemployment for several quarters.

THEN THE REGRESSOR CONTAINS THE ERROR

One quarter earlier, $du_{t-1} = \delta + \theta\, du_{t-2} + \beta_0 g_{t-1} + \beta_1 g_{t-2} + e_{t-1}$. So du_{t-1} contains e_{t-1}, which is related to e_t.

IF SO $\hat{\theta} = 0.379$ is estimating θ **plus the part of e_t that moves with e_{t-1}** instead of θ**, the effect we want**. More data does not fix this: OLS is inconsistent.

$\Longrightarrow$ Before trusting $\hat{\theta}$, test whether the errors are related over time: the Breusch-Godfrey test.



Are This Quarter's Residuals Related to Earlier Ones?

THE MODEL IN R, ESTIMATED ON AUSTRALIAN DATA (151 QUARTERS)

```
model <- dynlm(du_ts ~ L(du_ts, 1) + L(g_ts, 0:1))
```

$$\widehat{du}_t = 0.174 + 0.379\, du_{t-1} - 0.086\, g_t - 0.137\, g_{t-1}$$

R adds the intercept (0.174) without being asked.

ERROR AND RESIDUAL

e_t: the true error, never observed. $\hat{e}_t = du_t - \widehat{du}_t$: the residual, our estimate of e_t. The test uses the residuals in place of the errors.

RESIDUAL CORRELOGRAM

$r_1 = -0.162$ and $r_2 = 0.163$, just outside the bounds $\pm 1.96/\sqrt{151} = \pm 0.160$. By eye, it is hard to decide.

THE BG TEST CHECKS LAGS 1 TO 4 TOGETHER

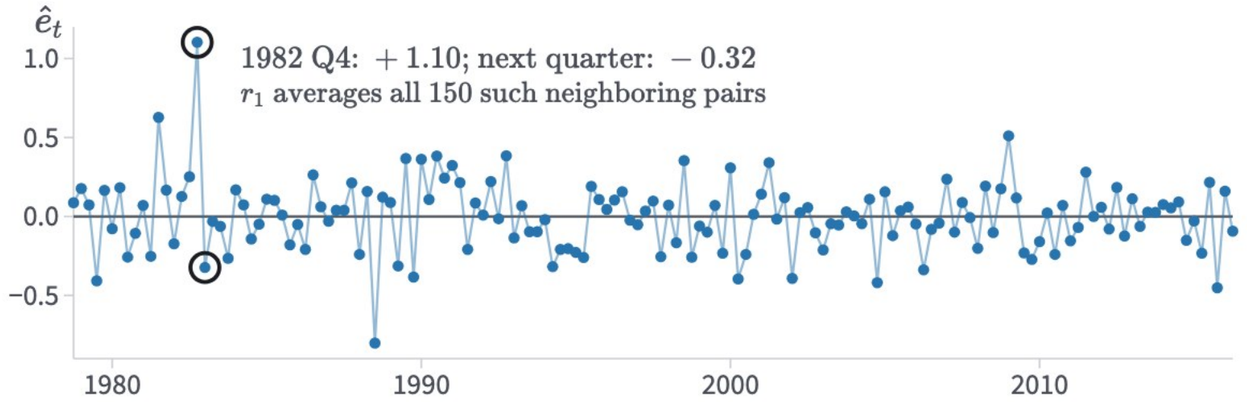
$$e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + \rho_3 e_{t-3} + \rho_4 e_{t-4} + v_t$$

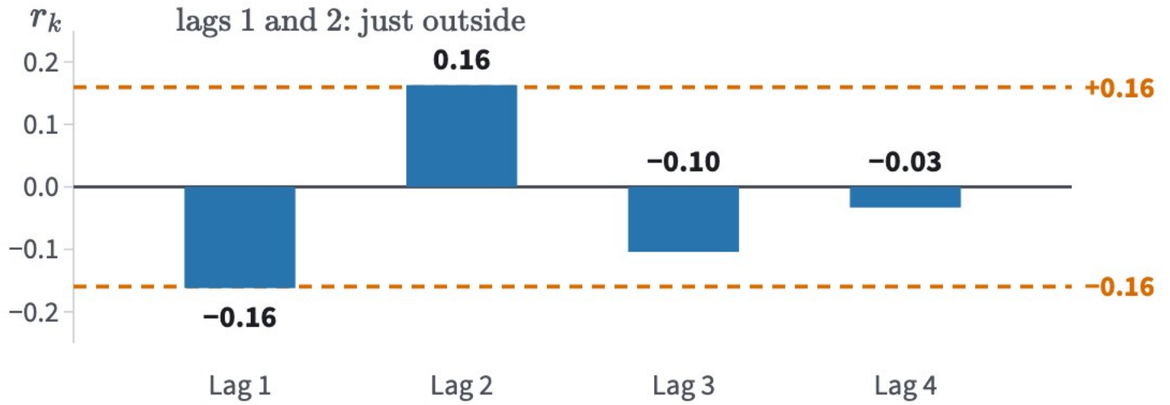
H_0: $\rho_1 = \rho_2 = \rho_3 = \rho_4 = 0$ $\iff$ no autocorrelation up to lag 4

H_1: at least one $\rho_j \neq 0$

The auxiliary regression in Step 2 also includes the model's regressors.

> Breusch-Pagan (week 2) regresses $\hat{e}^2$ and tests heteroskedasticity.
> Breusch-Godfrey regresses $\hat{e}$ on lagged $\hat{e}$ and tests autocorrelation.





Computed from okun5_aus (POE5), the data behind this model: 1978 Q4 to 2016 Q2.

Save the Residuals and Make Four Lagged Copies

Quarter	$\hat{e}_t$	$\hat{e}_{t-1}$	$\hat{e}_{t-2}$	$\hat{e}_{t-3}$	$\hat{e}_{t-4}$
1978 Q4	0.087	NA	NA	NA	NA
1979 Q1	0.177	0.087	NA	NA	NA
1979 Q2	0.073	0.177	0.087	NA	NA
1979 Q3	−0.407	0.073	0.177	0.087	NA
1979 Q4	0.165	−0.407	0.073	0.177	0.087
1980 Q1	−0.078	0.165	−0.407	0.073	0.177
1980 Q2	0.183	−0.078	0.165	−0.407	0.073
1980 Q3	−0.257	0.183	−0.078	0.165	−0.407
1980 Q4	−0.107	−0.257	0.183	−0.078	0.165
1981 Q1	0.070	−0.107	−0.257	0.183	−0.078
⋮ 151 quarters in all, to 2016 Q2					

ORDER = 4

Four lagged copies of the residuals, $\hat{e}_{t-1}$ to $\hat{e}_{t-4}$, each moved down one more row (as in last week's lag table). The first rows have no earlier quarter: NA.

FILL = 0 (THE DEFAULT)

`bgtest()` replaces each NA with 0 and keeps all $T = 151$ rows. 0 is the mean of the residuals, so the filled rows add no pattern of their own.

FILL = NA

Drops the first 4 rows instead: $T - p = 151 - 4 = 147$ rows. This is the lecture's $n = T - p$. Both versions give the same decision here (Step 3).

Regress $\hat{e}_t$ on the Original Regressors and the Lagged Residuals

AUXILIARY REGRESSION

A second regression that we run only to compute the test statistic. (The lecture writes α where this question writes γ.)

Ask: Do the ARDL model's regressors go in?

$$\hat{e}_t = \gamma_0 + \gamma_1 du_{t-1} + \gamma_2 g_t + \gamma_3 g_{t-1} + \rho_1 \hat{e}_{t-1} + \rho_2 \hat{e}_{t-2} + \rho_3 \hat{e}_{t-3} + \rho_4 \hat{e}_{t-4} + v_t$$

Every regressor of the ARDL model, copied as it is: du_{t-1}, g_t, g_{t-1} in levels, not differences.

order = 4 $\Longrightarrow$ the residual at lags 1, 2, 3, and 4.

WHY KEEP THE REGRESSORS?

du_{t-1} contains last quarter's error, so it is correlated with $\hat{e}_{t-1}$ (correlation 0.72 in these data). Without du_{t-1}, the test rejects too rarely. With these data it gives $LM = 8.42$ and does not reject (Step 3).

Compute $LM = T \times R^2$ and Compare to $\chi^2_{(4)}$

```
Auxiliary regression (c), 151 rows:
Multiple R-squared: 0.11049
```

STATISTIC

$LM = T \times R^2_{\text{aux}} = 151 \times 0.11049 = \mathbf{16.68}$

Ask: Degrees of freedom: 4 or 8?

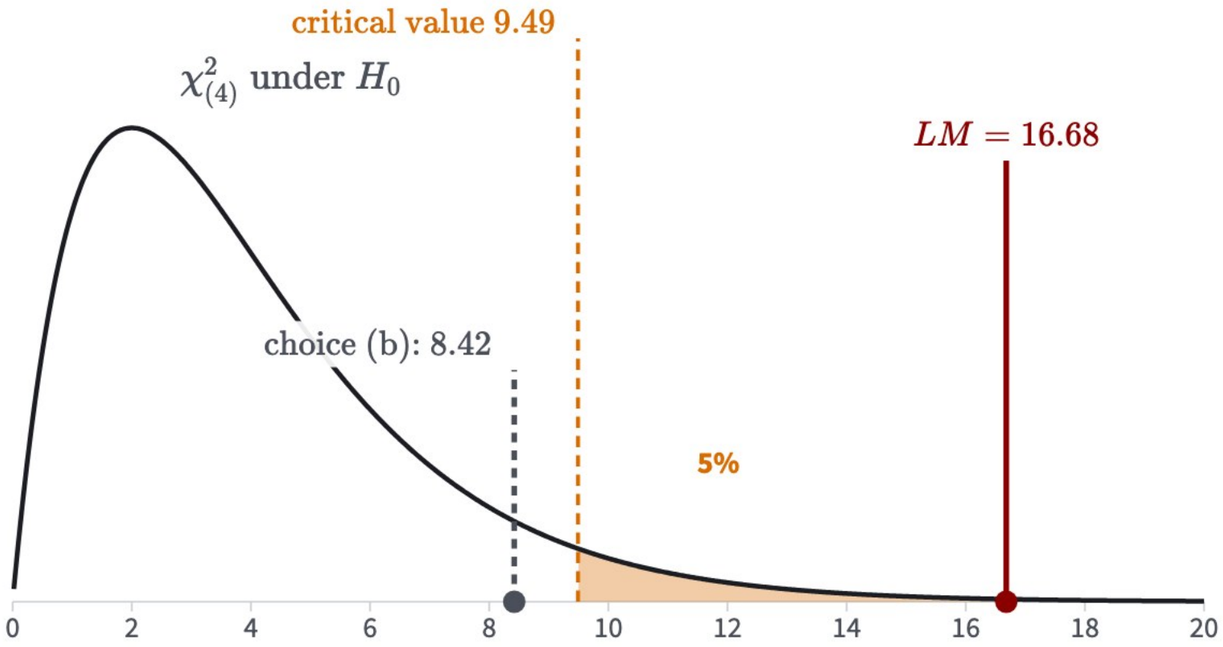
DEGREES OF FREEDOM

Count the lagged residuals: 4. Not 8: the 3 regressors and the intercept are not counted. 5% critical value: $\chi^2_{(4)} = 9.49$.

REJECT H_0 $16.68 > 9.49$ $(p = 0.002) \implies$ the errors are autocorrelated.

Lecture version, $n = T - p$ (fill = NA): $R^2 = 0.10340$, $LM = 147 \times 0.10340 = 15.20 > 9.49$. Same decision.

Choice (b), without the regressors: $R^2 = 0.05577$, $LM = 151 \times 0.05577 = 8.42 < 9.49$. That test does not reject H_0.



Checking Each Answer Choice

a $\hat{e}_t = \gamma_0 + \gamma_1 du_{t-1} + \gamma_2 g_t + \gamma_3 g_{t-1} + \rho \hat{e}_{t-1} + v_t$

One lagged residual: this tests lag 1 only, but order = 4.

b $\hat{e}_t = \gamma_0 + \rho_1 \hat{e}_{t-1} + \rho_2 \hat{e}_{t-2} + \rho_3 \hat{e}_{t-3} + \rho_4 \hat{e}_{t-4} + v_t$

Leaves out du_{t-1}, g_t, g_{t-1}. With these data it gives $LM = 8.42$, which does not reject H_0.

d $\Delta\hat{e}_t = \gamma_0 + \gamma_1 \Delta du_{t-1} + \gamma_2 \Delta g_t + \gamma_3 \Delta g_{t-1} + \rho_1 \Delta\hat{e}_{t-1} + \cdots + \rho_4 \Delta\hat{e}_{t-4} + v_t$

The test uses the residuals and regressors as they are in the model. Nothing is differenced.

c $\hat{e}_t = \gamma_0 + \gamma_1 du_{t-1} + \gamma_2 g_t + \gamma_3 g_{t-1} + \rho_1 \hat{e}_{t-1} + \cdots + \rho_4 \hat{e}_{t-4} + v_t$

Answer: (c)

Ask: The test rejects. Can we just switch to HAC standard errors?

INTERPRETATION

The errors of this ARDL(1,1) are autocorrelated. Write the model one quarter earlier: $du_{t-1} = \delta + \theta\, du_{t-2} + \beta_0 g_{t-1} + \beta_1 g_{t-2} + e_{t-1}$. So the regressor du_{t-1} contains e_{t-1}, and with autocorrelated errors e_{t-1} is correlated with e_t. OLS is then **inconsistent**: its estimates do not get close to the true coefficients even with a very large sample. HAC standard errors do not change that.

NEXT STEP

Add lags of du and g, choose how many with SC (Problem 1), and run the BG test again.

The posted solutions for this exam list (d) but describe the regression in (c) (“the original regressors plus lagged residuals up to order 4”). Answer letters can differ across exam versions; in the version shown here, the answer is (c).

The Question: Taxes and the Top 1% Income Share

We are interested in the following model to determine the effect of the tax rate on inequality, where SHARE is the percentage income share held by those with the top 1% of incomes: $SHARE_t = \alpha_1 + \alpha_2 TAX_t + \alpha_3 YEAR_t + \alpha_4 YEAR_t^2 + e_t$

We run the following in R:

```
round(coeftest(reg.mod, vcov.=NeweyWest(reg.mod)), 3)
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   17.808      2.616   6.807   <2e-16 ***
ca_tax        -0.107      0.064  -1.675    0.098 .
year          -0.174      0.132  -1.319    0.191
year2          0.002      0.001   1.588    0.116
```

We then perform the following test:

```
bgtest(reg.mod, order=3, fill=0)
Breusch-Godfrey test for serial correlation of order up to 3
LM test = 56.568, df = 3, p-value = 3.177e-12
```

What is the confidence interval for the marginal effect of taxes on inequality, given the critical value, $t_{0.975,76} = 1.99$? Is the use of HAC errors justified?

a The confidence interval is $[-0.23, 0.02]$ and the use of HAC errors is justified

b The confidence interval is $[0.23, 23.02]$ and the use of HAC errors is justified

c The confidence interval is $[-0.23, 0.02]$ and the use of HAC errors is not justified

d The confidence interval is $[-.44, 0.02]$ and the use of HAC errors is not justified

e None of the above

Source: Spring 2023 Midterm V1, Question 12

Ten Friends, One Study Guide

THE QUESTION

Is the midterm curved? You ask 10 friends, and 9 say yes.

TEN SEPARATE ANSWERS

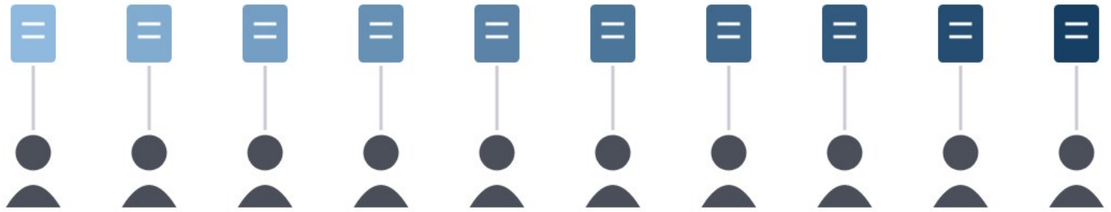
Each friend found out on their own. That is 10 independent answers, and 9 out of 10 is strong evidence.

ONE SHARED SOURCE

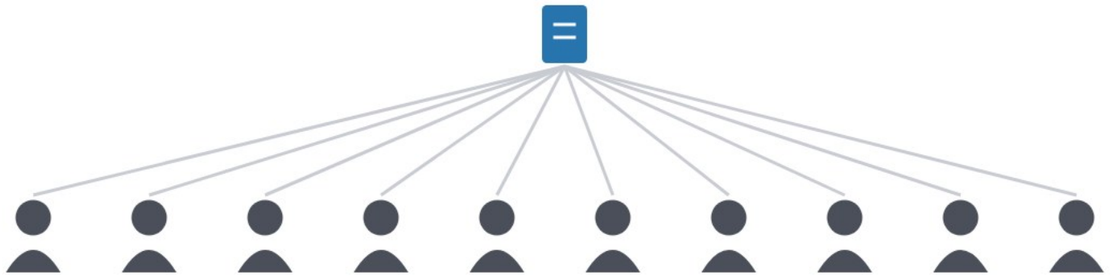
All 10 read the same study guide, written by one student. That is about one answer repeated 10 times. Counting it as 10 makes you too confident.

BACK TO REGRESSION

The usual OLS standard error treats the 80 yearly errors as 80 independent pieces of information. If each year's error is close to last year's, the data contain less information than that, and the usual standard error is too small.



10 sources: 10 independent answers



1 source: about 1 answer, heard 10 times

This Year's Error Is Close to Last Year's

THE DATA

Canada, 1921 to 2000: $T = 80$ years. $SHARE$: income share of the top 1% (%). TAX: median marginal tax rate of the top 1% (%). The plot shows the OLS residuals.

LONG RUNS

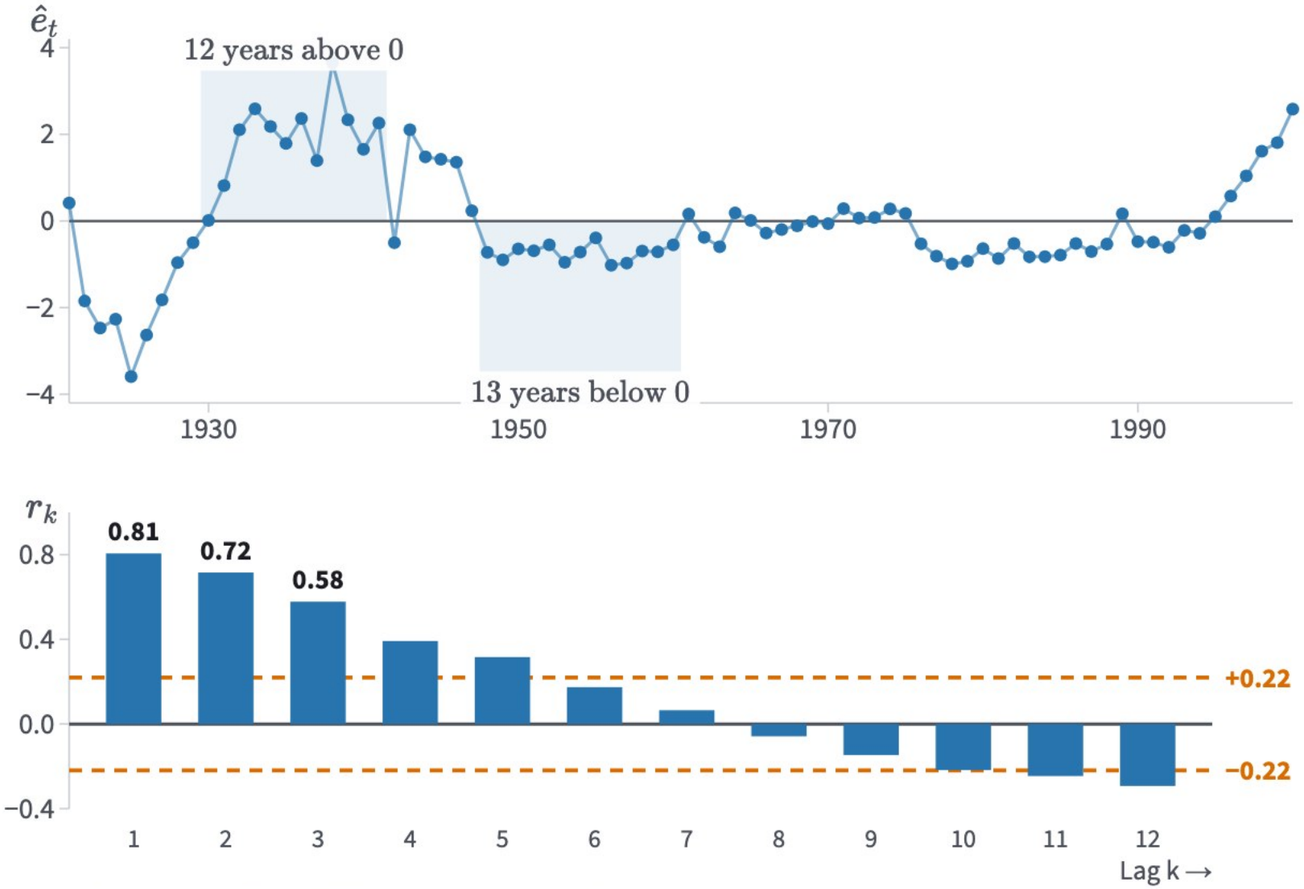
12 years in a row above 0, then later 13 years in a row below 0. The sign changes only 14 times in 80 years; independent errors would change sign about 40 times.

CORRELOGRAM

$r_1 = 0.81, r_2 = 0.72, r_3 = 0.58$: far outside the bounds $\pm 1.96/\sqrt{80} = \pm 0.22$.

HAC STANDARD ERRORS

Heteroskedasticity and autocorrelation consistent (HAC) standard errors, computed by `NeweyWest()` in R. Same OLS estimates; standard errors that allow the errors to be correlated over time.



Computed from `inequality` (POE5): the variables `ca_share` and `ca_tax` in the exam's R output.

Is HAC Justified? Read the BG Test

```
bgtest(reg.mod, order=3, fill=0)
LM test = 56.568, df = 3, p-value = 3.177e-12
```

HYPOTHESES

H_0: no autocorrelation up to lag 3 vs H_1: autocorrelation at lag 1, 2, or 3. order = 3 $\implies \chi^2_{(3)}$, 5% critical value 7.81.

STATISTIC

$LM = 56.568$, far above 7.81.

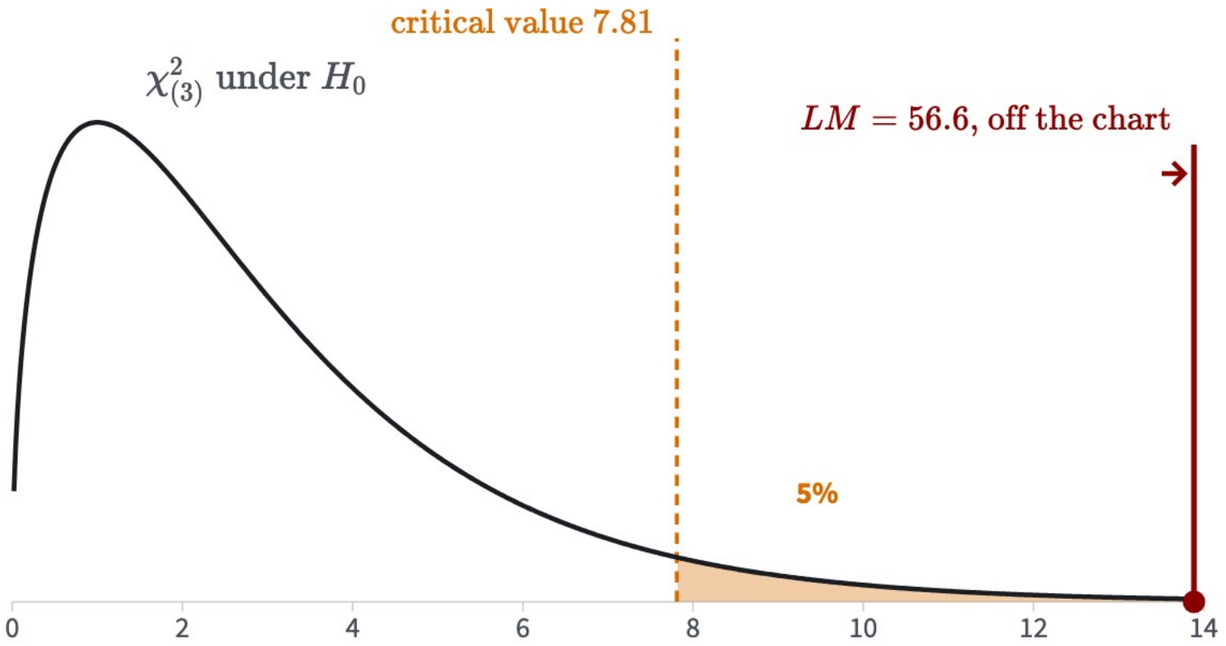
Ask: Is $p = 3.177\text{e-}12$ small or large?

P-VALUE

$3.177\text{e-}12 = 3.177 \times 10^{-12} = 0.000000000003177$. Not 3.177.

REJECT H_0 The errors are autocorrelated $\implies$ HAC standard errors are justified.

Unlike Problem 2, this model has no lagged $SHARE$ on the right side. If TAX_t is uncorrelated with e_t, the OLS estimates are consistent, and only the standard errors need correcting.



Build the Interval with the HAC Standard Error

FORMULA

$$\hat{\alpha}_2 \pm t_{0.975,\,76} \times se(\hat{\alpha}_2)$$

$df = T - 4 = 80 - 4 = 76$: 80 years, 4 coefficients.

Ask: Which row of the output, and which t value?

READ THE ROW FOR CA_TAX

Estimate -0.107, HAC standard error 0.064.

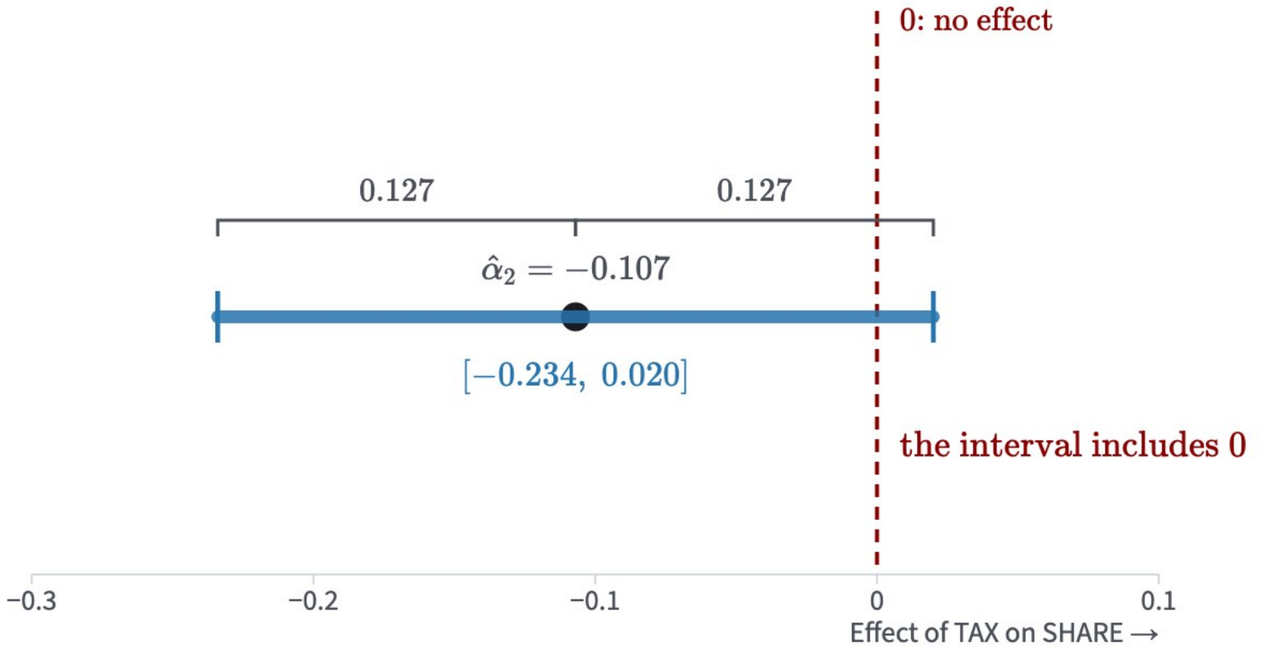
MARGIN

$1.99 \times 0.064 = 0.127$

INTERVAL

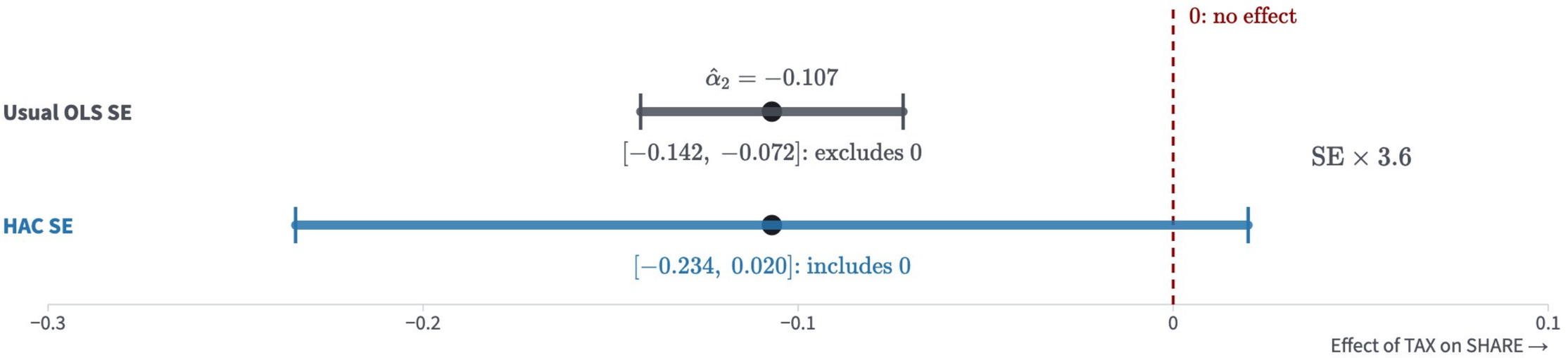
$$-0.107 \pm 0.127 = [-0.234,\ 0.020] \approx [-0.23,\ 0.02]$$

The interval includes 0. This agrees with the output: $p = 0.098 > 0.05$.



Usual Standard Error vs HAC Standard Error

Ask: If we drop HAC and use the usual standard errors, will -0.107 change?



USUAL OLS STANDARD ERROR

0.0177: interval $[-0.142,\ -0.072]$ does not include 0.

HAC STANDARD ERROR

0.064, about 3.6 times larger: interval $[-0.234,\ 0.020]$ includes 0.

With the usual SE the effect looks significant. That SE assumes no autocorrelation, and the BG test rejects that assumption.

Usual OLS standard error computed from `inequality` (POE5) with `coeftest(reg.mod)`; it is not in the exam output.

Checking Each Answer Choice

b $[0.23, 23.02]$, justified

The interval must contain the estimate -0.107; this one does not.

c $[-0.23, 0.02]$, not justified

Right interval. But the BG test rejects, so HAC is justified.

d $[-.44, 0.02]$, not justified

Its midpoint is -0.21, not -0.107. And HAC is justified.

e None of the above

Choice (a) holds.

a $[-0.23, 0.02]$, justified

Answer: (a)

INTERPRETATION

A tax rate 1 percentage point higher goes with a top 1% share 0.107 percentage points lower, holding $YEAR$ and $YEAR^2$ fixed. With HAC standard errors the interval includes 0: the effect is not statistically significant at the 5% level ($p = 0.098$).

Autocorrelated errors in…	OLS estimates	Correction
No lagged y (Problem 3)	consistent if TAX_t is uncorrelated with e_t	HAC
Lagged y (Problem 2: du_{t-1})	inconsistent	add lags; HAC is not enough

Thank you!

jakeanderson@g.ucla.edu