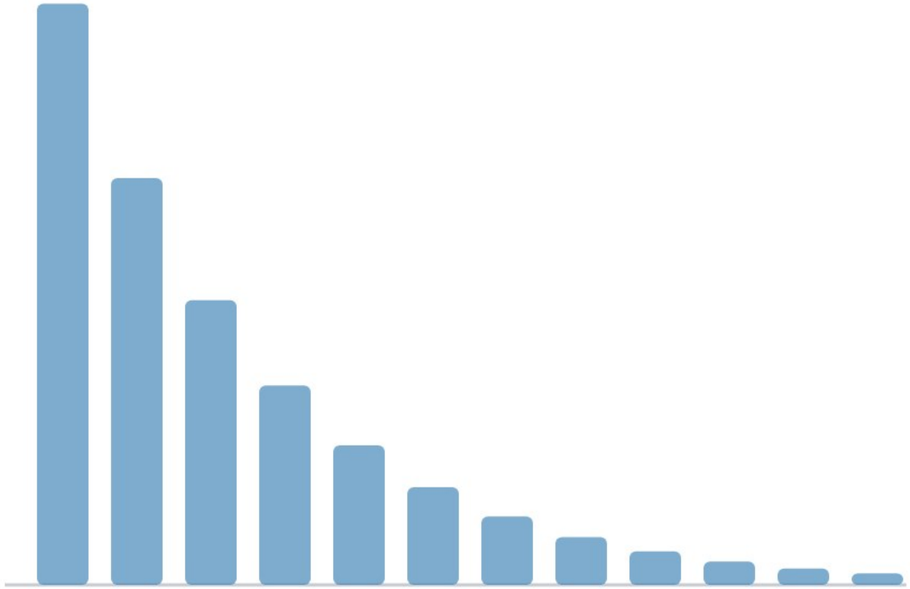
ECON 104 · DISCUSSION · FALL 2026

Autocorrelation, Autoregressive Models, and Distributed Lags



Jake Anderson
UCLA

Plan for Today

1

Autocorrelation

Is a series correlated with its own earlier values?

Test each sample autocorrelation with $Z = \sqrt{T}\, r_k$.

Spring 2024 Final V2, Question 7 (modified)

2

Autoregressive (AR) models

GDP growth this quarter depends on growth in the two quarters before.

Forecast next quarter's growth.

Winter 2026 Midterm 1 (Rojas), Question 21

3

Distributed lags (DL)

Income this quarter changes consumption this quarter and in the next two quarters.

What does the coefficient on Y_{t-1} measure?

Spring 2026 Final (Boswell), Question 36

Is GPA Last Quarter Related to GPA This Quarter?

Why might a student's GPA this quarter depend on their GPA last quarter?

Missed material

A weak quarter leaves gaps in material that the next courses build on, so the next quarter is harder too.

Study habits persist

A student who studies steadily this quarter usually studies the same way next quarter.

Health problems last

An illness or injury that lowers grades this quarter often lasts into the next one.

Money problems last

A student working long hours to pay rent this quarter is likely still working them next quarter.

Confidence

A good quarter raises confidence and motivation; a bad quarter lowers them.

Every story is momentum

A high GPA tends to be followed by a high GPA, and a low one by a low one: y_t is correlated with y_{t-1}. When a variable has momentum, measure it with the autocorrelation, then model it with an autoregression.

Each Lag Is a Copy of y Moved Down One Row

t	y_t	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-4}	y_{t-5}	y_{t-6}	y_{t-7}	y_{t-8}	y_{t-9}	y_{t-10}
1	0.8	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
2	0.5	0.8	NA	NA	NA	NA	NA	NA	NA	NA	NA
3	0.8	0.5	0.8	NA	NA	NA	NA	NA	NA	NA	NA
4	0.5	0.8	0.5	0.8	NA	NA	NA	NA	NA	NA	NA
5	0.5	0.5	0.8	0.5	0.8	NA	NA	NA	NA	NA	NA
6	0.0	0.5	0.5	0.8	0.5	0.8	NA	NA	NA	NA	NA
7	0.8	0.0	0.5	0.5	0.8	0.5	0.8	NA	NA	NA	NA
8	0.7	0.8	0.0	0.5	0.5	0.8	0.5	0.8	NA	NA	NA
9	1.3	0.7	0.8	0.0	0.5	0.5	0.8	0.5	0.8	NA	NA
10	0.6	1.3	0.7	0.8	0.0	0.5	0.5	0.8	0.5	0.8	NA
11	0.4	0.6	1.3	0.7	0.8	0.0	0.5	0.5	0.8	0.5	0.8
12	0.5	0.4	0.6	1.3	0.7	0.8	0.0	0.5	0.5	0.8	0.5
13	1.0	0.5	0.4	0.6	1.3	0.7	0.8	0.0	0.5	0.5	0.8
14	1.0	1.0	0.5	0.4	0.6	1.3	0.7	0.8	0.0	0.5	0.5
15	0.9	1.0	1.0	0.5	0.4	0.6	1.3	0.7	0.8	0.0	0.5
16	1.2	0.9	1.0	1.0	0.5	0.4	0.6	1.3	0.7	0.8	0.0
17	1.0	1.2	0.9	1.0	1.0	0.5	0.4	0.6	1.3	0.7	0.8
18	0.3	1.0	1.2	0.9	1.0	1.0	0.5	0.4	0.6	1.3	0.7
19	1.1	0.3	1.0	1.2	0.9	1.0	1.0	0.5	0.4	0.6	1.3
20	0.5	1.1	0.3	1.0	1.2	0.9	1.0	1.0	0.5	0.4	0.6

Illustration: GDP growth values are made up for this table.

THE DATA

20 quarters of GDP growth, one row per quarter. t is the quarter and y_t is growth in that quarter.

LAG 1

Each value moves down one row: row t of y_{t-1} holds y from row $t-1$. Row 1 has no earlier quarter, so it is NA (missing).

```
library(dplyr)  # its lag() moves values
df <- df %>%
  mutate(y_lag1 = lag(y, 1))
```

LAG 2

Move y_{t-1} down one more row. Rows 1 and 2 are NA.

```
  mutate(y_lag2 = lag(y, 2))
```

UP TO LAG 10

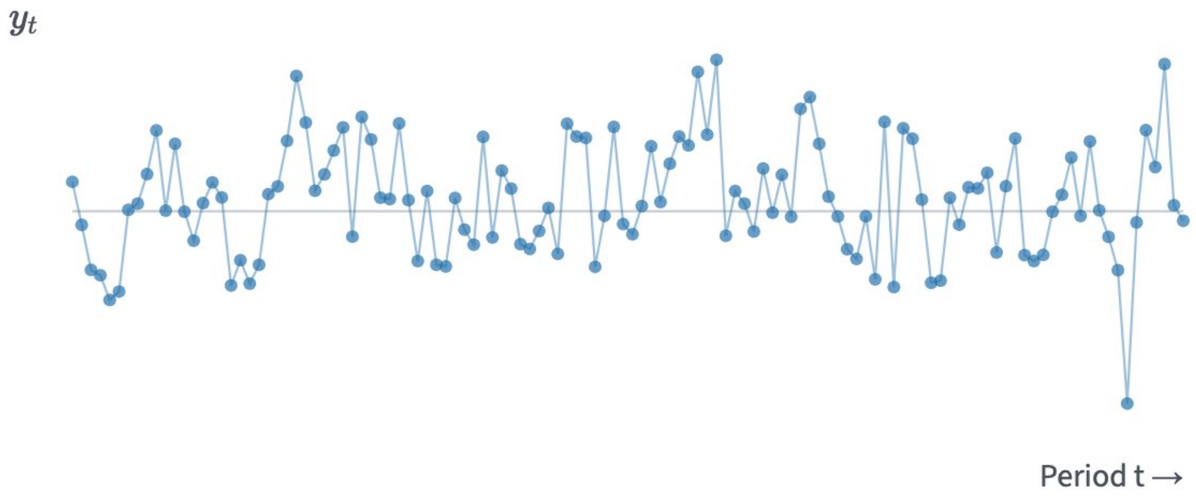
y_{t-k} starts with k NAs.

Rows with an NA cannot be used. Pairing y_t with y_{t-k} leaves $T-k$ complete rows: 19 for lag 1, 18 for lag 2, and only 10 of the 20 quarters for lag 10.

Correlation Between Values One Period Apart

THE SERIES

One value per period: $y_1, y_2, \ldots, y_T$, with mean $\bar{y}$.



PAIR EACH VALUE WITH THE ONE BEFORE

Plot y_t against y_{t-1}, the first two columns of the lag table.

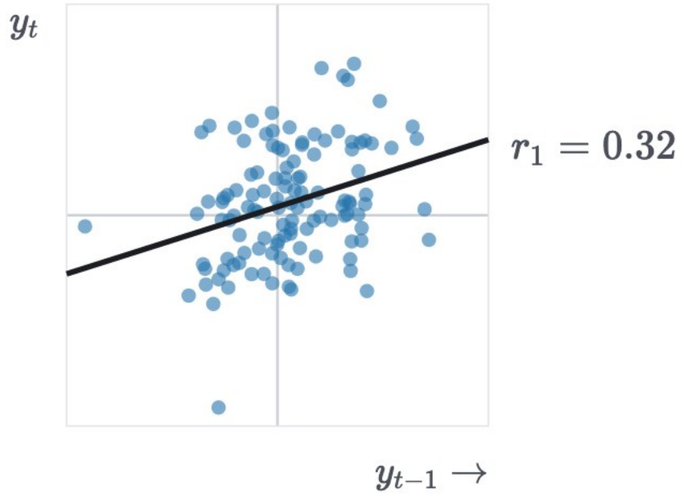


Illustration: 120 periods of a series whose lag-1 autocorrelation is about 0.32.

$$r_1 = \mathrm{Corr}(y_t, y_{t-1})$$

The correlation of the pairs in the scatter. Here $r_1 = 0.32$: a weak positive relationship.

$$= \frac{\mathrm{Cov}(y_t, y_{t-1})}{\mathrm{Var}(y_t)}$$

A correlation is the covariance divided by both standard deviations. Both columns are the same series, so that product is $\mathrm{Var}(y_t)$.

$$= \frac{\sum_{t=2}^{T}(y_t - \bar{y})(y_{t-1} - \bar{y})}{\sum_{t=1}^{T}(y_t - \bar{y})^2}$$

The covariance and variance written as sums (the $1/T$ in each cancels). A product on top is positive when y_t and y_{t-1} are on the same side of $\bar{y}$.

r_k uses y_{t-k} in place of y_{t-1}. Problem 1 gives r_1 to r_4 for $T = 680$ periods: which are different from 0?

The Question: Which Autocorrelations Are Significant?

Say we are given the following sample autocorrelations:

Lag	r_k
1	0.32
2	-0.21
3	0.08
4	-0.01

Which lags are statistically significant at the 5% level, if $Z_{0.975} = 1.96$ and there are 680 periods in the study?

a. Lag 1 only

b. Lags 1 and 2 only

c. Lags 1 and 3 only

d. Lags 1, 2, and 3 only

e. Lags 1, 2, 3, and 4

Source: Spring 2024 Final V2, Question 7, modified: r_2 changed from -0.91 to -0.21 so that the four values could come from real data.

Compute $Z = \sqrt{T}\, r_k$ for Each Lag

HYPOTHESES

For each lag k: $H_0\colon \rho_k = 0$ vs $H_1\colon \rho_k \neq 0$

ρ_k is the population autocorrelation at lag k; r_k is its estimate from the sample.

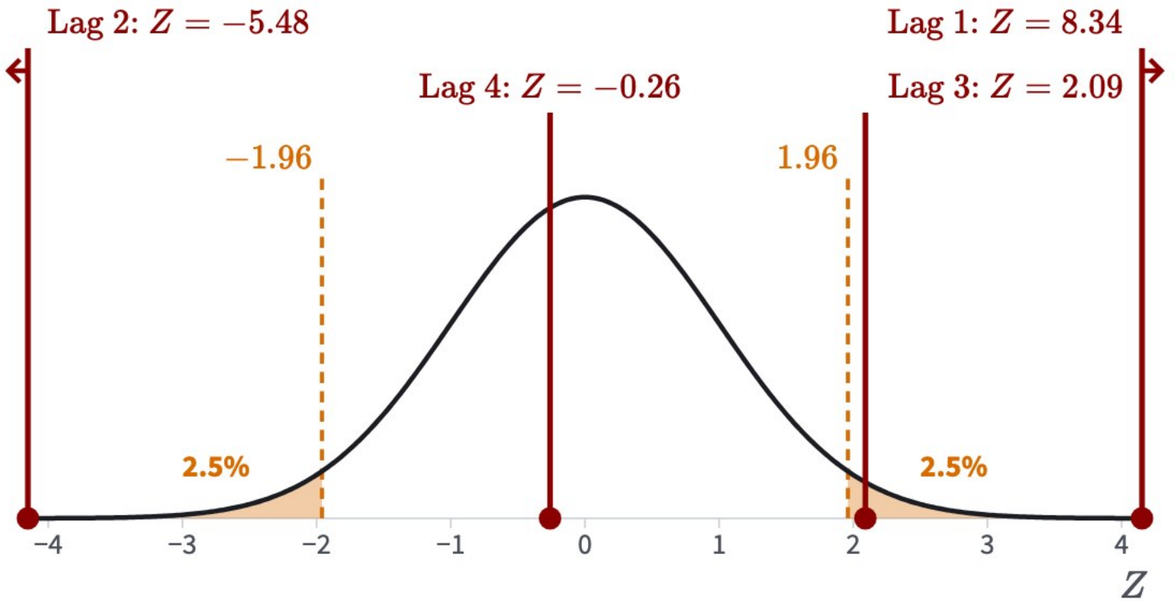
STATISTIC

Under H_0, r_k is approximately normal with mean 0 and variance $1/T$:

$$Z = \frac{r_k - 0}{\sqrt{1/T}} = \sqrt{T}\, r_k \sim N(0,1)$$

with $\sqrt{680} = 26.08$.

Lag	r_k	$Z = 26.08\, r_k$	$\lvert Z\rvert > 1.96$?
1	0.32	8.34	Yes
2	−0.21	−5.48	Yes
3	0.08	2.09	Yes
4	−0.01	−0.26	No

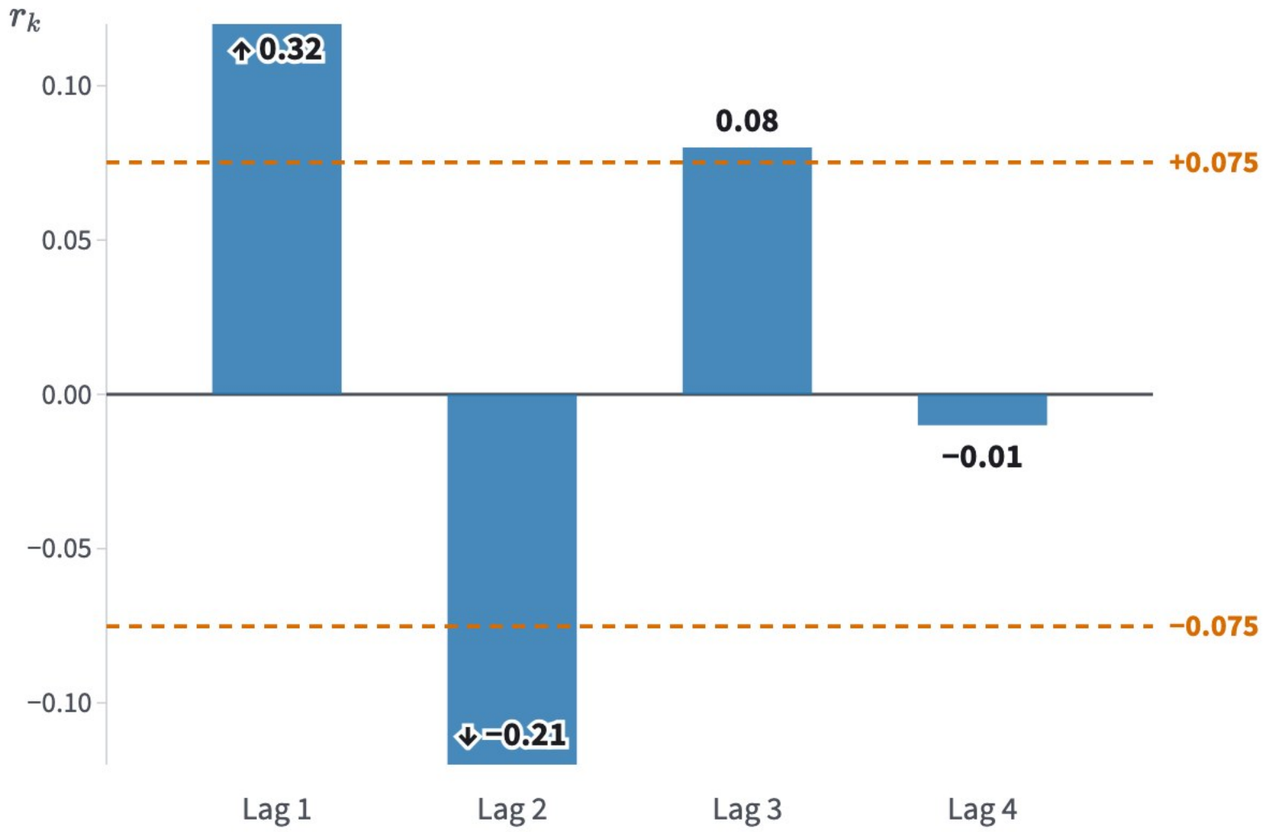


REJECT H_0 for lags 1, 2, and 3. Lag 4: $\lvert -0.26\rvert < 1.96 \implies$ fail to reject.

INTERPRETATION

At the 5% level, y_t is correlated with y_{t-1}, y_{t-2}, and y_{t-3}: positively at lags 1 and 3, negatively at lag 2. The data give no evidence of a correlation with y_{t-4}.

The Correlogram Gives the Same Decision



CORRELOGRAM

One bar for each sample autocorrelation r_k.

BOUNDS

Lines at $\pm 1.96/\sqrt{T} = \pm 1.96/26.08 = \pm 0.075$. A bar outside the lines $\iff |Z| > 1.96$.

ZOOM IN

Lag 3: $0.08 > 0.075$, just outside. Lag 4: -0.01, inside.

> The lecture slides draw the lines at $\pm 2/\sqrt{T} = \pm 0.077$. Lag 3 is still outside.

Checking Each Answer Choice

a Lag 1 only

Lags 2 and 3 also have $|Z| > 1.96$.

b Lags 1 and 2 only

Leaves out lag 3: $Z = 2.09 > 1.96$.

c Lags 1 and 3 only

Leaves out lag 2: $Z = -5.48$. The test is two-sided, so a large negative Z also rejects.

e Lags 1, 2, 3, and 4

Lag 4: $|Z| = 0.26 < 1.96$.

d Lags 1, 2, and 3 only

Answer: (d)

INTERPRETATION

The series is correlated with its own values 1, 2, and 3 periods earlier. It has momentum, so we could model it with its own past values: an autoregression (next section).

Today's Value Depends on Its Own Past

THE SAME PATTERN

GPA from quarter to quarter, mood from hour to hour, a city's crime rate from year to year, and GDP growth from quarter to quarter (Problem 2).

THE AR(1) MODEL

$$y_t = \delta + \theta_1 \, y_{t-1} + v_t$$

v_t is this period's new shock: mean 0, constant variance, and not correlated with earlier shocks. θ_1 measures how strongly this period's value depends on last period's.

THE AR(p) MODEL

Add the p most recent values of y: $y_t = \delta + \theta_1 y_{t-1} + \cdots + \theta_p y_{t-p} + v_t$. "Autoregressive" means y is regressed on its own past values. With p lags, **lm()** drops the NA rows and uses $T - p$ rows.

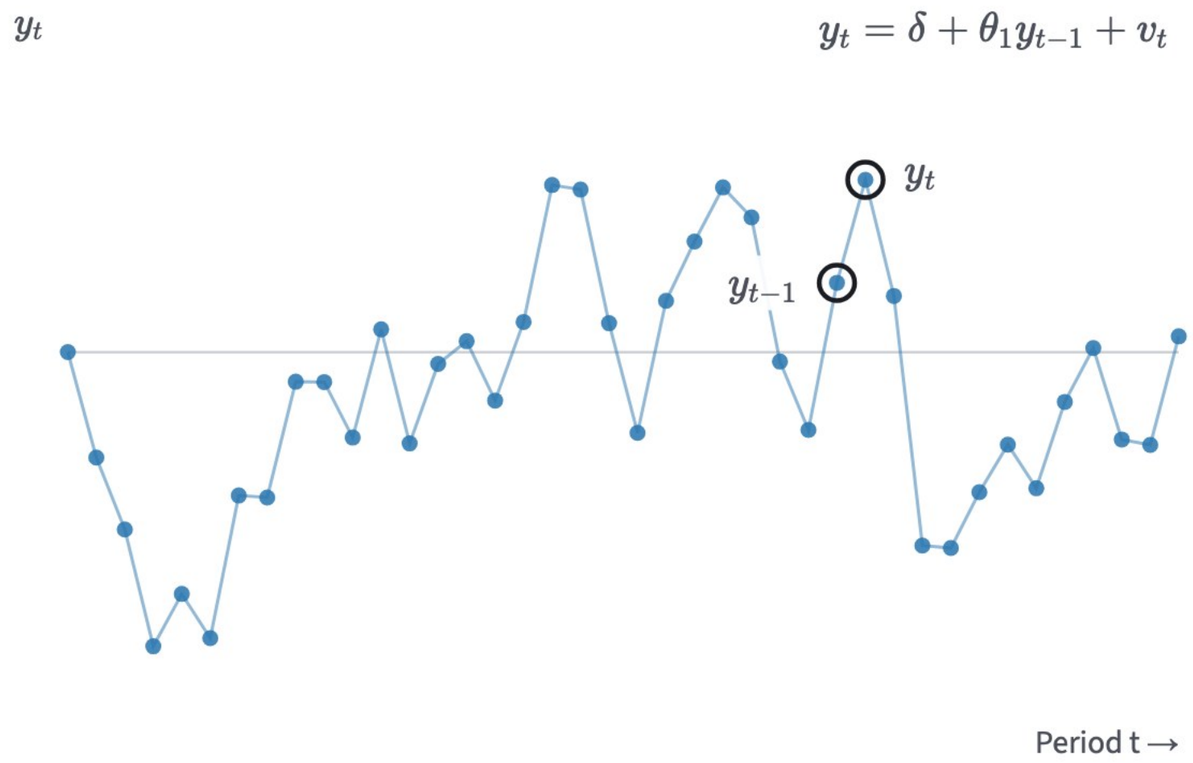


Illustration: 40 periods of an AR(1) series with $\theta_1 = 0.7$.

The Question: Forecasting GDP Growth

Consider an AR(2) model for GDP growth:

$$G_t = \delta + \theta_1 G_{t-1} + \theta_2 G_{t-2} + v_t$$

Estimated results give: $\hat{\delta} = 0.67, \hat{\theta}_1 = 0.12, \hat{\theta}_2 = -0.09$. If $G_T = 0.8$ and $G_{T-1} = -0.2$, what is the one-step-ahead forecast $\hat{G}_{T+1}$?

a 0.84

b 0.69

c 0.78

d 0.75

Source: Winter 2026 Midterm 1 (Rojas), Question 21

θ_1 Sets How Much of a Shock Is Left Each Period

ONE SHOCK

Take $y_t = \theta_1 y_{t-1} + v_t$ and start at $y = 0$. Set $v_0 = 1$ in period 0 and $v_t = 0$ in every later period. Then $y_k = \theta_1^k$.

$\theta_1 = 0.5$

$y_0 = 1, y_1 = 0.5, y_2 = 0.25, \ldots$ Each period keeps half of the last value. After 4 periods, $0.5^4 = 0.06$ of the shock is left: small, but not 0.

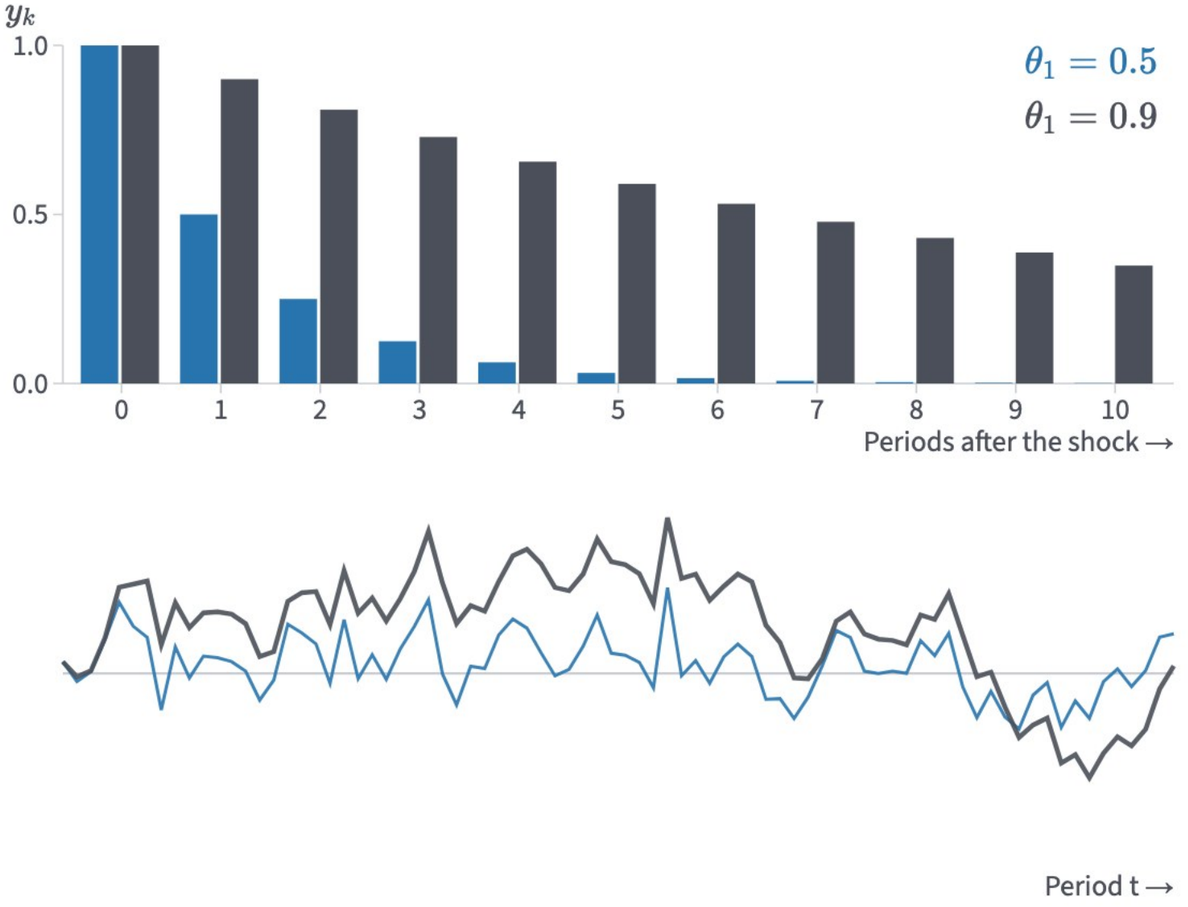
$\theta_1 = 0.9$

$y_1 = 0.9, y_2 = 0.81, \ldots$ Each period keeps 90% of the last value. After 4 periods, $0.9^4 = 0.66$ of the shock is left.

A NEW SHOCK EVERY PERIOD

Both series get the same shocks. With $\theta_1 = 0.9$ the series moves slowly and stays above or below 0 for long runs. With $\theta_1 = 0.5$ it moves back toward 0 faster.

If $\theta_1 \neq 0$, a shock affects every later period, because θ_1^k is never exactly 0. θ_1 sets how large that effect is: the effect in each period is θ_1 times the effect in the period before. If $|\theta_1| < 1$, the effect gets smaller every period. In the GDP model, $\hat{\theta}_1 = 0.12$ and $\hat{\theta}_2 = -0.09$ are small: past growth tells us little about next quarter.



Top: $y_k = \theta_1^k$ after one shock in period 0. For an AR(1), θ_1^k is also the autocorrelation at lag k. Bottom, illustration: 80 periods, the same random shocks for both values of θ_1.

Write the Model One Quarter Ahead

THE MODEL IN QUARTER $T + 1$

Replace t with $T + 1$:

$$G_{T+1} = \delta + \theta_1 G_T + \theta_2 G_{T-1} + v_{T+1}$$

G_T is the latest quarter and G_{T-1} the quarter before it. Both are already observed.

THE FORECAST

v_{T+1} is next quarter's shock. We do not know it, and its mean is 0, so the forecast sets it to 0 and uses the estimates:

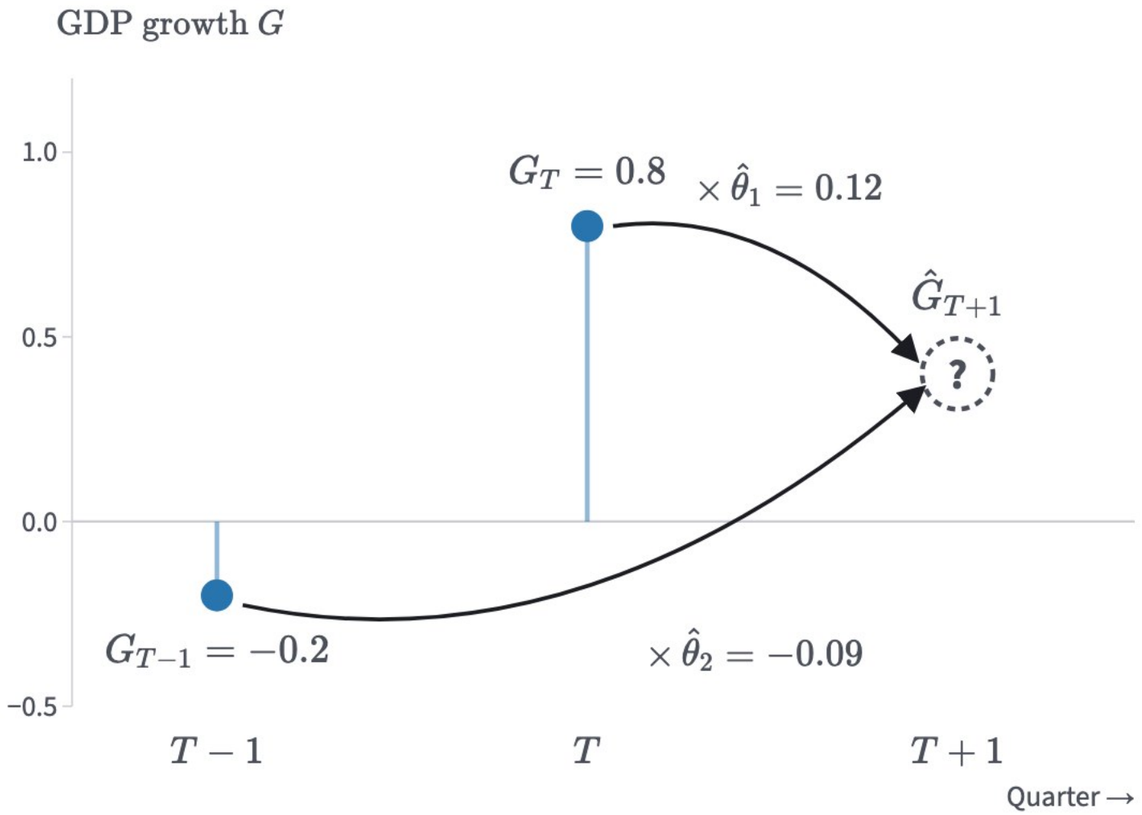
$$\hat{G}_{T+1} = \hat{\delta} + \hat{\theta}_1 G_T + \hat{\theta}_2 G_{T-1}$$

MATCH EACH NUMBER TO ITS SYMBOL

$\hat{\theta}_1 = 0.12$ multiplies the most recent value, $G_T = 0.8$.
$\hat{\theta}_2 = -0.09$ multiplies the older value, $G_{T-1} = -0.2$.

PLUG IN EVERY NUMBER

$$\begin{aligned}\hat{G}_{T+1} &= 0.67 + (0.12)(0.8) + (-0.09)(-0.2) \\ &= 0.67 + 0.096 + 0.018 = \mathbf{0.784}\end{aligned}$$



Plug In and Watch the Signs

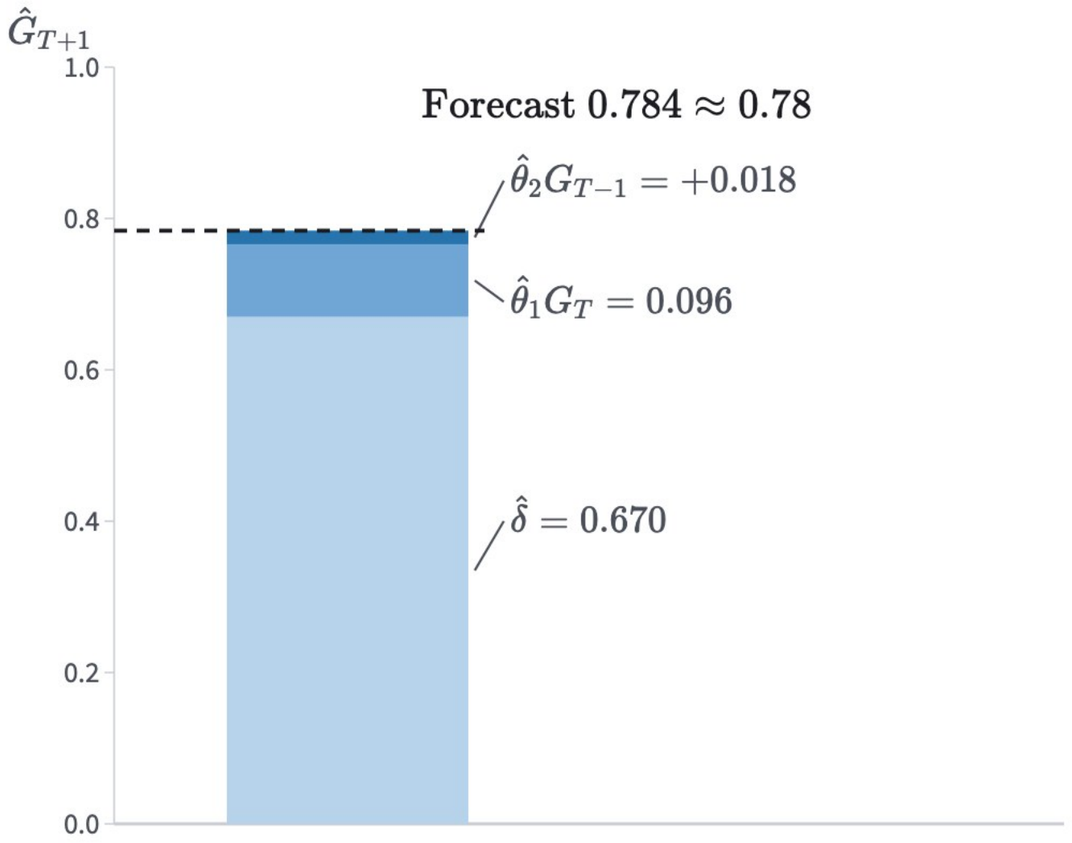
$$\hat{G}_{T+1} = 0.67 + (0.12)(0.8) + (-0.09)(-0.2)$$

Term	Estimate × value	Product
$\hat{\delta}$	0.67	0.670
$\hat{\theta}_1 G_T$	0.12×0.8	0.096
$\hat{\theta}_2 G_{T-1}$	$(-0.09) \times (-0.2)$	$+0.018$
Forecast $\hat{G}_{T+1}$		0.784

A negative coefficient times a negative value is positive: $(-0.09)(-0.2) = +0.018$.

INTERPRETATION

GDP growth next quarter is forecast at 0.78. The average growth rate of this model is $\mu = \hat{\delta}/(1 - \hat{\theta}_1 - \hat{\theta}_2) = 0.67/0.97 = 0.69$: the θ's are small, so the forecast is close to that average.



Checking Each Answer Choice

b 0.69

$0.67 + (-0.09)(-0.2) = 0.688$: **leaves out $\hat{\theta}_1 G_T = 0.096$.**

d 0.75

$0.67 + 0.096 - 0.018 = 0.748$: **treats $(-0.09)(-0.2)$ as negative.**

a 0.84

$0.67 + 0.096 + 0.09(0.8) = 0.838$: **uses G_T in both terms and drops the minus sign on $\hat{\theta}_2$.**

c 0.78

Answer: (c) $0.67 + 0.096 + 0.018 = 0.784$

THE RECIPE

1. Write the model one period ahead.
2. Set the unknown shock v_{T+1} to 0.
3. Multiply each $\hat{\theta}$ by the value it goes with: $\hat{\theta}_1$ with G_T, $\hat{\theta}_2$ with G_{T-1}.
4. Keep the signs.

Does Air Pollution This Week Affect Hospital Visits Later?

Why might more pollution this week raise hospital visits in later weeks?

Symptoms build up

Lung inflammation from polluted air can take days to become serious enough for a hospital visit.

People wait

Many people wait to see whether a cough goes away before they see a doctor.

Infections follow

Irritated airways make it easier to catch a respiratory infection a week or two later.

Serious cases take longer

Heart and lung damage builds slowly, so hospital admissions for it come weeks after the pollution starts.

Every story

Pollution in week t raises visits in week t and in weeks $t+1, t+2, \ldots \implies y_t$ depends on $x_t, x_{t-1}, x_{t-2}, \ldots$

Today’s Value Depends on Past Values of Another Variable

THE SAME PATTERN

Pollution and hospital visits over several weeks, ad spending and sales over the next few quarters, fertilizer and crop yield months later, and income and consumer spending over three quarters (Problem 3).

THE DISTRIBUTED LAG MODEL

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \cdots + \beta_q x_{t-q} + e_t$$

β_s is the effect of x on y s periods later. The effect of a change in x is spread over $q + 1$ periods.

AR model: lags of y itself. Distributed lag model: lags of another variable x. A model with both is an autoregressive distributed lag (ARDL) model.

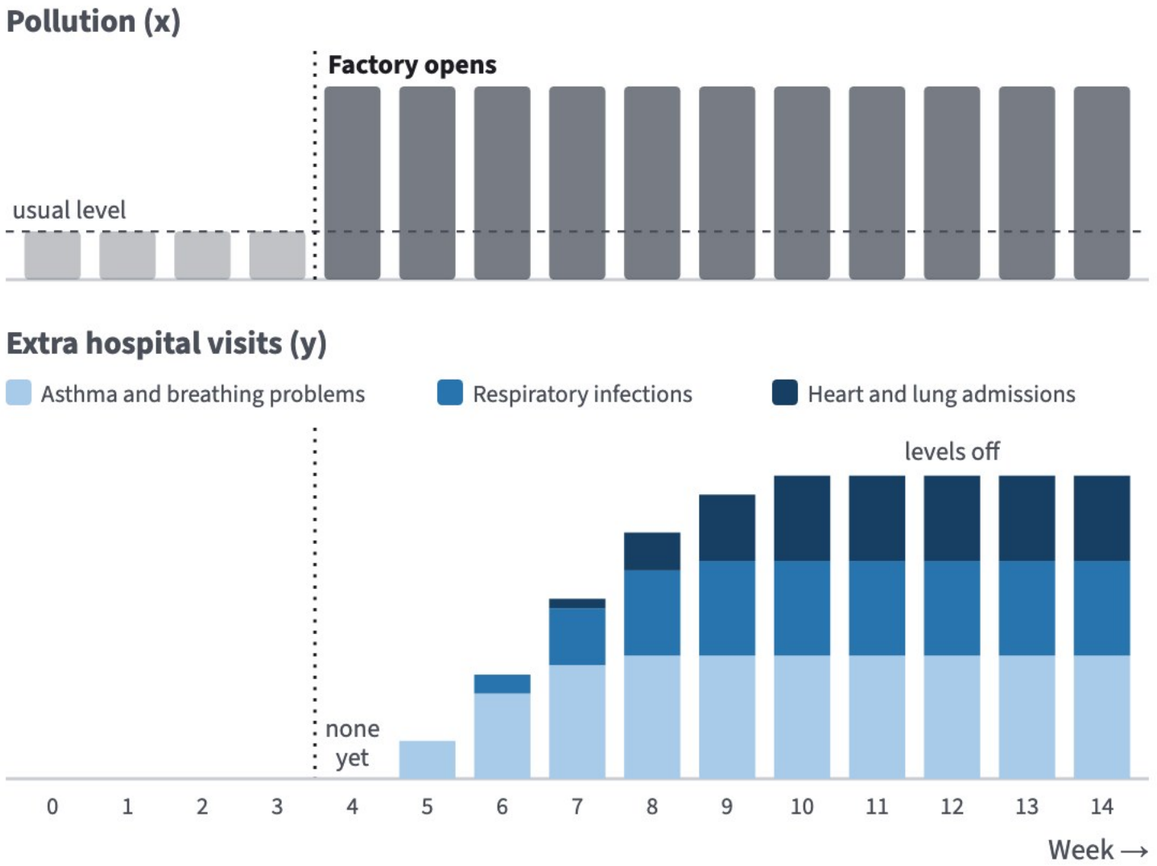


Illustration: a factory opens in week 4 and pollution rises to 4 times its usual level and stays there. Extra visits start one week later, rise for 6 weeks, then level off. Heart and lung admissions come last.

PROBLEM 3 · DISTRIBUTED LAGS

The Question: Income and Consumer Spending

Consider the following finite distributed lag model (FDL) of order 2 using quarterly data ($T = 100$):

$$\widehat{C}_t = 1.20 + 0.45\,Y_t + 0.28\,Y_{t-1} + 0.16\,Y_{t-2}$$

where C_t is consumer spending and Y_t is disposable income (both in \$billion). The coefficient on Y_{t-1} is best interpreted as:

a The total long-run effect of a \$1 increase in income on consumption.

b The immediate effect of a \$1 increase in income on consumption in the same period.

c The additional effect on consumption in period $t+1$ from a \$1 increase in income in period t, holding all else constant.

d The proportion of the long-run multiplier realised after one period.

e The effect of income growth on consumption growth one quarter ahead.

Source: Spring 2026 Final (Boswell), Question 36

Follow a One-Time Increase in Income

SETUP

Raise income by \$1 billion in quarter t only. Income in every other quarter is unchanged.

QUARTER t

$$C_t = 1.20 + 0.45\,Y_t + 0.28\,Y_{t-1} + 0.16\,Y_{t-2} \implies +0.45$$

QUARTER $t+1$

$$C_{t+1} = 1.20 + 0.45\,Y_{t+1} + 0.28\,Y_t + 0.16\,Y_{t-1} \implies +0.28$$

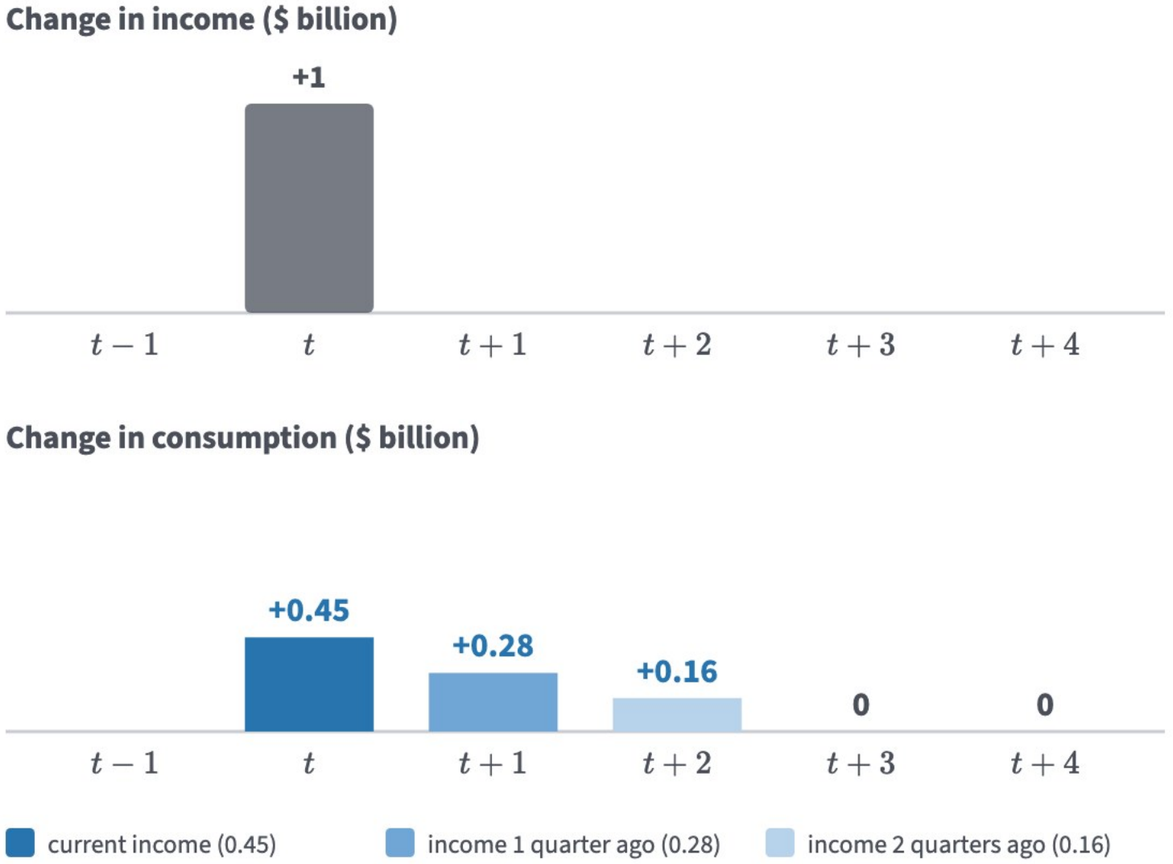
QUARTER $t+2$

$$C_{t+2} = 1.20 + 0.45\,Y_{t+2} + 0.28\,Y_{t+1} + 0.16\,Y_t \implies +0.16$$

QUARTER $t+3$

Y_t is no longer in the equation $\implies$ no effect.

The coefficient on Y_{t-1}, 0.28, is the change in consumption one quarter after the change in income.



A Permanent Increase: Add the Weights

SETUP

Raise income by $\$1$ billion in quarter t and keep it $\$1$ billion higher in every later quarter.

QUARTER t

Only Y_t is higher: 0.45. This is the impact multiplier.

QUARTER $t+1$

Y_{t+1} and Y_t are both higher: $0.45 + 0.28 = 0.73$. This is an interim multiplier.

QUARTER $t+2$

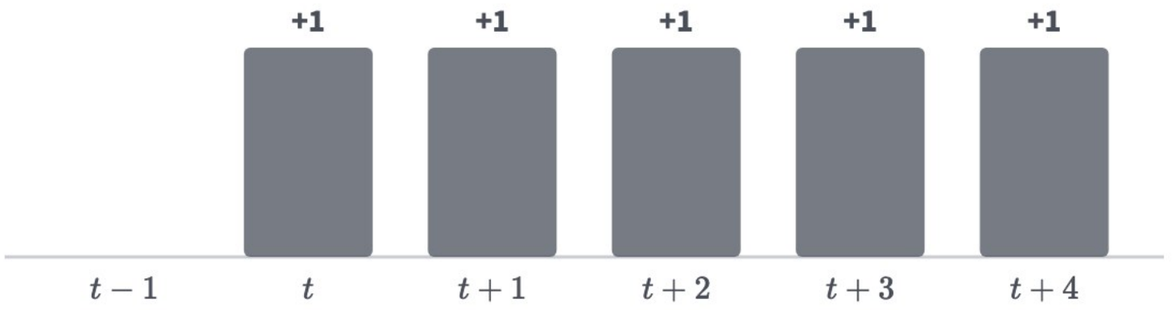
$0.45 + 0.28 + 0.16 = 0.89$.

QUARTER $t+3$ AND LATER

Still 0.89: only two lags enter the model. This is the total (long-run) multiplier.

> One coefficient is the effect of income at one lag. The sum of the coefficients is the effect of a permanent change in income.

Change in income ($ billion)



Change in consumption ($ billion)



Checking Each Answer Choice

a The total long-run effect of a $\$1$ increase in income on consumption.

The total effect is the sum: $0.45 + 0.28 + 0.16 = 0.89$.

b The immediate effect of a $\$1$ increase in income on consumption in the same period.

The immediate (impact) effect is the coefficient on Y_t, 0.45.

d The proportion of the long-run multiplier realised after one period.

That share is $(0.45 + 0.28)/0.89 = 0.82$.

e The effect of income growth on consumption growth one quarter ahead.

The model is in levels of C and Y, not growth rates.

c The additional effect on consumption in period $t + 1$ from a $\$1$ increase in income in period t, holding all else constant.

Answer: (c)

INTERPRETATION

A $\$1$ billion increase in income this quarter raises consumption next quarter by $\$0.28$ billion, holding income in all other quarters fixed. Read the same coefficient backward: last quarter's income raises this quarter's consumption by 0.28 per dollar.

Thank you!

jakeanderson@g.ucla.edu