

Who Heard the Alarm?

Module 8 · ARDL, information, and exogeneity

Gary Cornwall

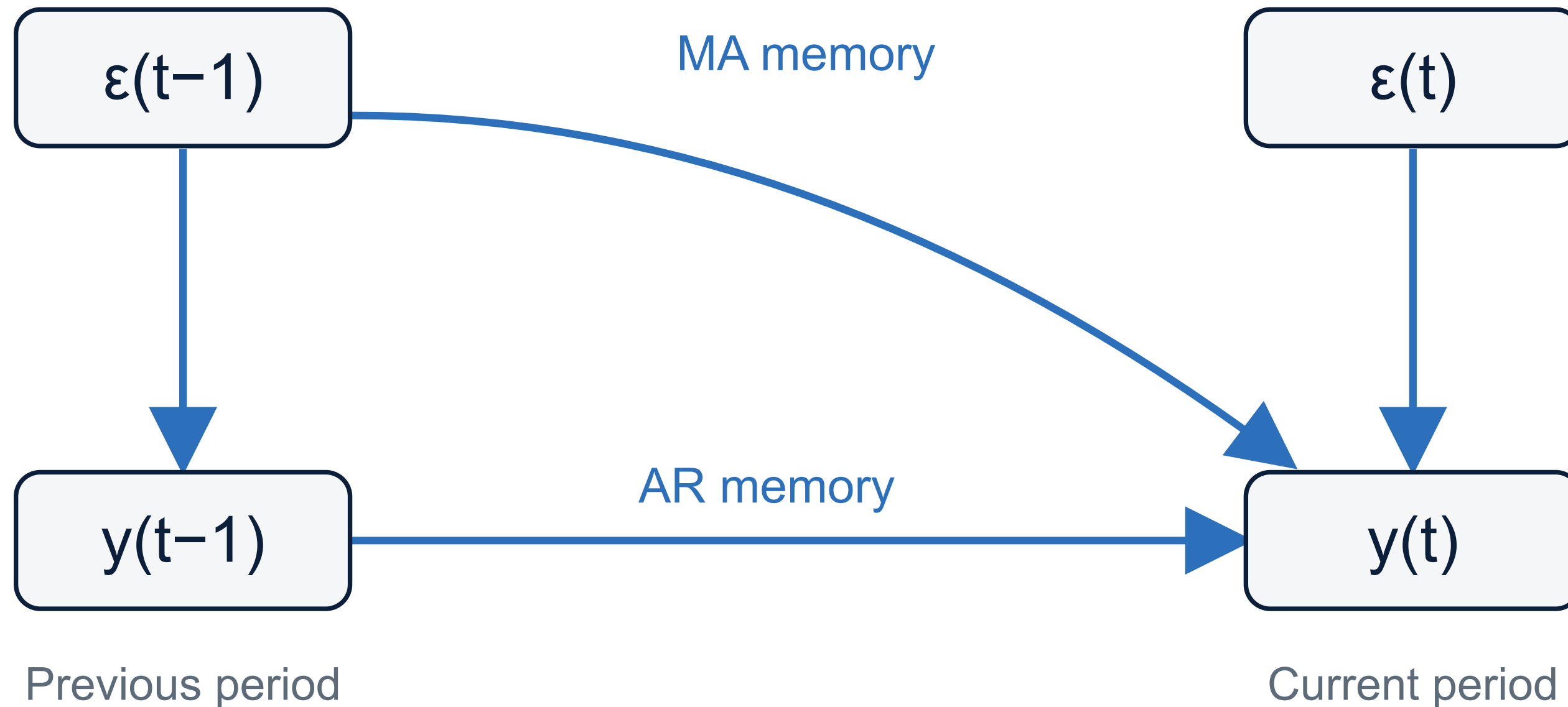
Econ 6376 · The George Washington University

Act I • Letting another variable into the room

One series' memory becomes a relationship across series.

ARMA: one series remembers outcomes and surprises

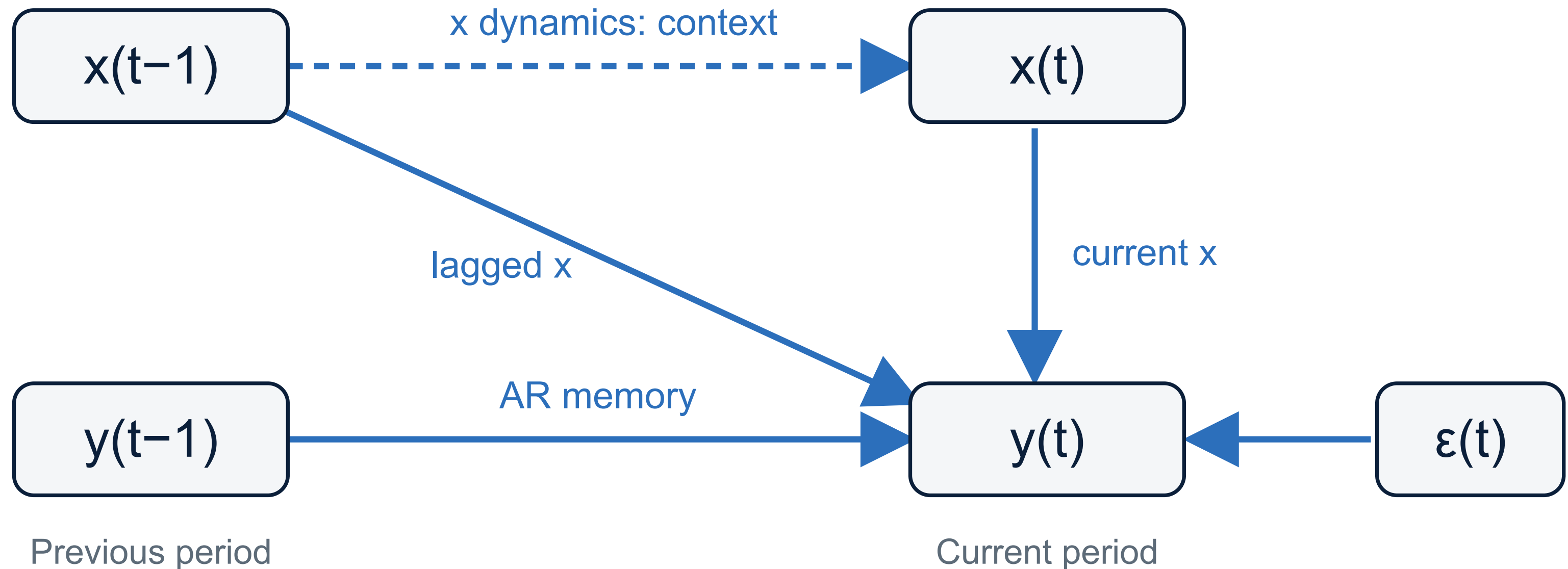
$$y_t = \alpha + \phi_1 y_{t-1} + \theta_1 \epsilon_{t-1} + \epsilon_t$$



A time-unrolled DAG: each date is a different node. Innovations ϵ are unobserved DGP quantities.

ARDL: another series brings its own history

$$y_t = \alpha + \phi_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \epsilon_t$$



Observed x replaces the explicit MA block. Drawing an arrow does not identify a causal effect.

Where we are: expand the mixing board

$$y_t = \alpha + \delta t + \sum_{j=1}^p \phi_j y_{t-j} + \sum_{\ell=1}^q \theta_{\ell} \epsilon_{t-\ell} + \epsilon_t$$

Channel	Module 8 baseline
Intercept and outcome memory	On: $\alpha, \phi_j y_{t-j}$
Trend and explicit MA memory	Off
Observed predictor history	New: $\sum_{\ell=0}^q \beta_{\ell} x_{t-\ell}$
Current disturbance	Still present; its relation to x now matters

Module 7 combined forecasts. Today we add information absent from that forecast pool.

One conditional equation, many possible lags

$$y_t = \alpha + \sum_{j=1}^p \phi_j y_{t-j} + \sum_{\ell=0}^q \beta_{\ell} x_{t-\ell} + \epsilon_t$$

Restriction	Model left on the board
All $\beta_{\ell} = 0$	AR (p)
All $\phi_j = 0$	Distributed-lag regression
$q = 0$	Outcome lags and current x
$\beta_0 = 0$	Lag-only predictor block

What does ARDL(1,2) include that ARDL(1,1) does not?

Here q means the maximum x lag. It is not the MA order.

The DGP and our fitted equation remain different

DGP: parameters and disturbances we know only in a simulation

$$y_t = \alpha + \phi_1 y_{t-1} + \beta_0 x_t + \beta_1 x_{t-1} + \epsilon_t$$

Estimated equation: hats and an observed residual

$$y_t = \hat{\alpha} + \hat{\phi}_1 y_{t-1} + \hat{\beta}_0 x_t + \hat{\beta}_1 x_{t-1} + e_t$$

$$e_t = y_t - \hat{y}_t$$

Fitted values: $\hat{y}$. Forecasts from Module 7 onward: $\tilde{y}_{t+h|t}$.

Keep the stationary baseline visible

$$\Phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$$

Stationary inputs + stable dynamics + suitable disturbances.

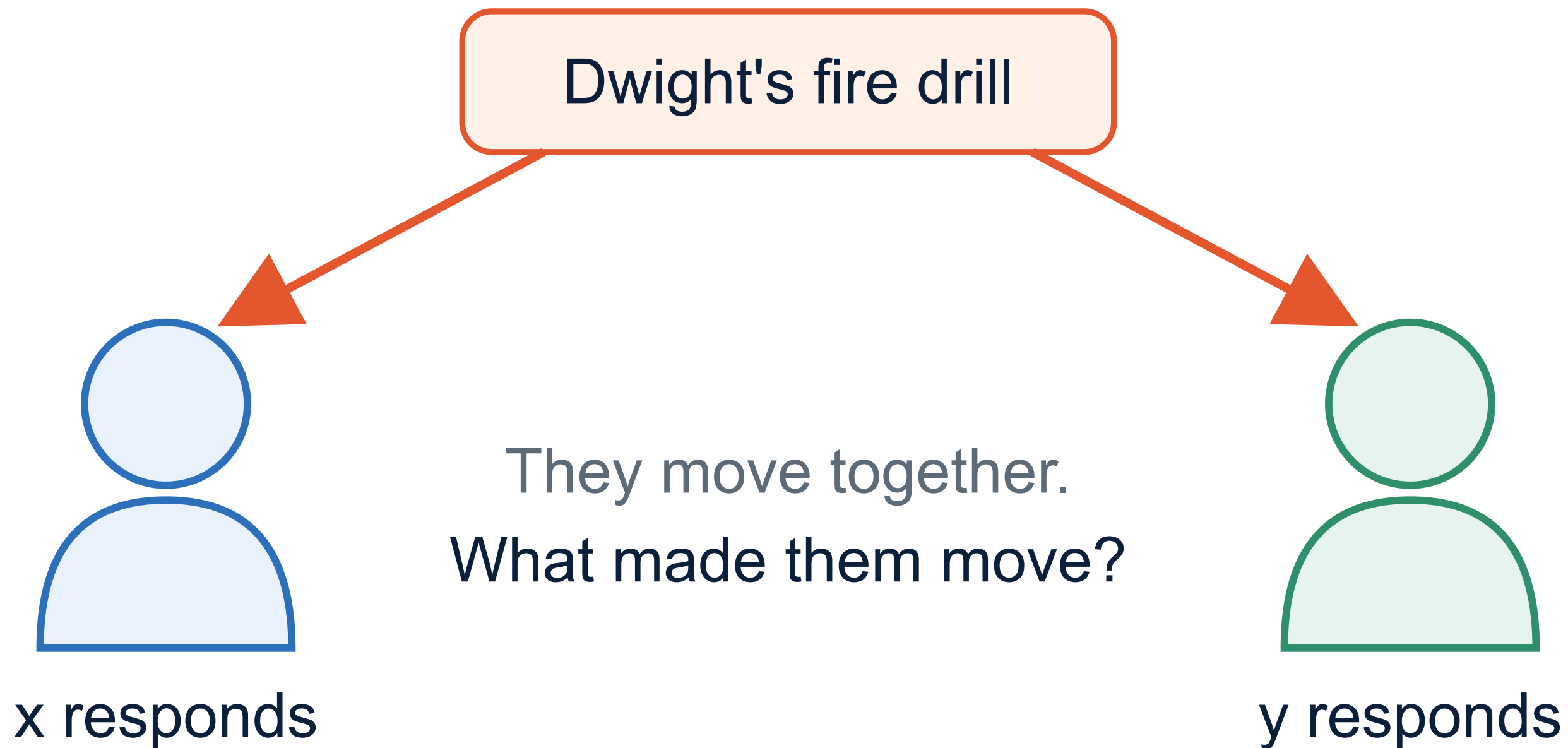
- AR lag-polynomial roots lie **outside** the unit circle.
- For one outcome lag: $|\phi_1| < 1$.
- If x responds to past y , check stability of the **joint system**.

Trending levels need additional reasoning. Cointegration is a later route, not a blanket exemption.

Act II • Who else heard the news?

Exogeneity is the central question.

The Office: everyone reacts to the same emergency



If claims and unemployment rise in the same month, which story has the regression established?

Optional clip: [The Office · Fire Drill \(official\)](#). The diagram remains usable offline.

Exogeneity asks what remains in the disturbance

Let $\mathcal{F}_{t-1}$ contain the relevant histories of y and x .

$$E[\epsilon_t \mid \mathcal{F}_{t-1}, x_t] = 0$$

After using the modeled history and current x , no predictable mean component remains.

- This covers **all included regressors**, including lagged y .
- With stability, moments, and adequate regressor variation, it supports **large-sample OLS consistency**.
- It does not promise exact recovery in one finite sample.

Three claims that sound similar

x and y are related

Often why we included x. This is not the exogeneity restriction.

ϵ has no serial correlation

A statement about the disturbance across dates.

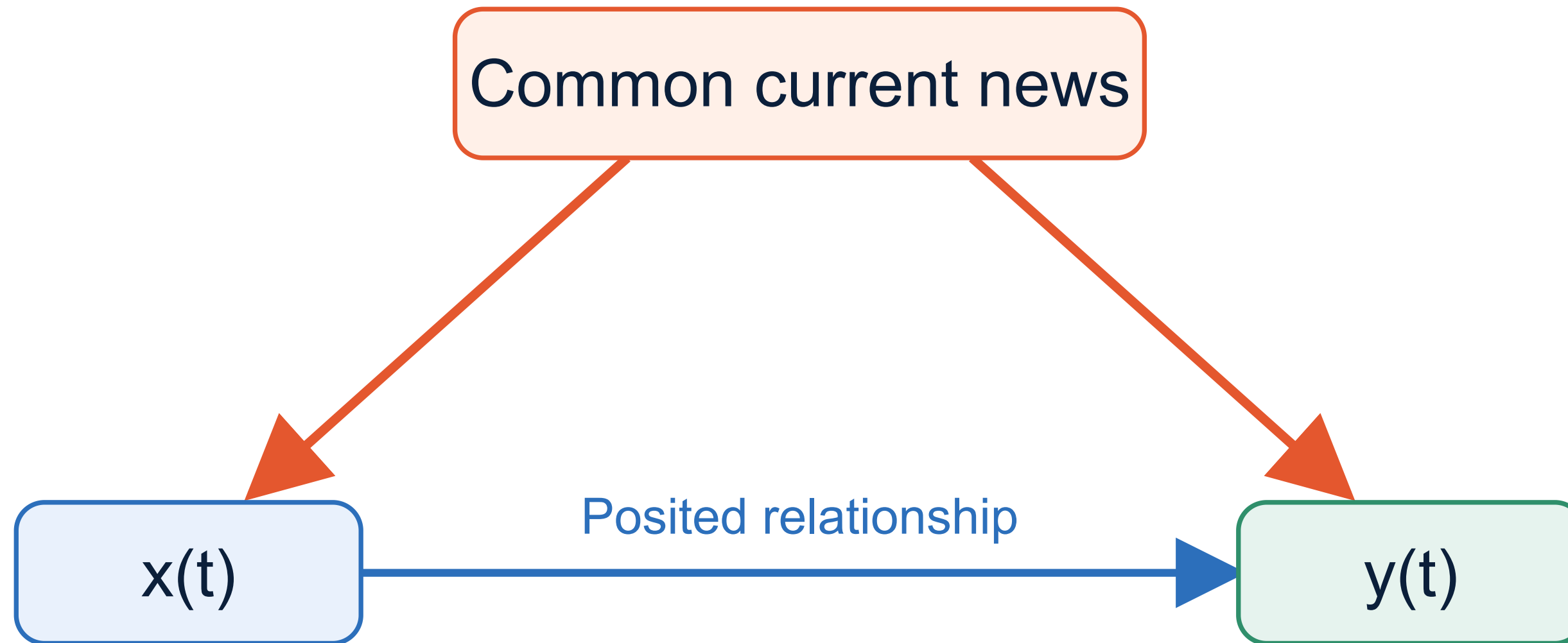
$$E[\epsilon_t \mid \mathcal{F}_{t-1}, x_t] = 0$$

A statement about the disturbance relative to the regressor information.

Could two white-noise shocks be correlated with each other today?

Yes. No memory is not independence from every other process.

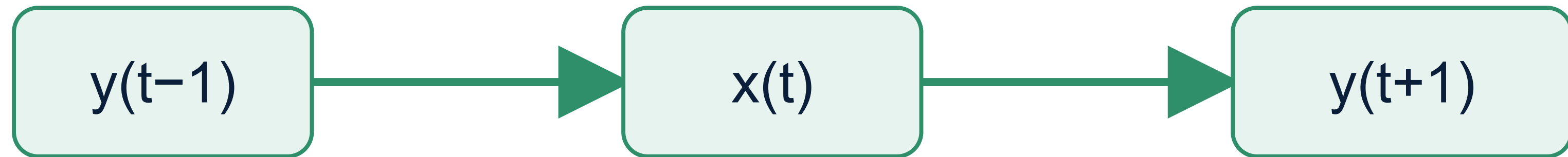
Common current news creates an omitted route



If the outcome disturbance contains shared news, current x may reveal part of that disturbance.

Lagged feedback is a different issue

$$x_t = 0.5x_{t-1} + 0.2y_{t-1} + \epsilon_{x,t}$$



Every arrow moves forward in time.

With independent current shocks, independent of the past, this feedback can coexist with our conditional-mean restriction.

Stronger assumptions and narrower diagnostics

Strict exogeneity asks today’s disturbance to be unrelated to the entire x path—including future x .

Lagged feedback may violate that stronger requirement while the sequential conditional mean remains valid.

Proposed fix	What it cannot establish
“Use a lag”	Earlier does not guarantee exogenous
“The residual ACF looks clean”	No proof of structural exogeneity
“Use HAC standard errors”	No repair for invalid coefficient moments

Are we making a prediction claim, a conditional-path claim, or a causal claim?

Act III • Build it, then break it

Same structural coefficients. Different information.

Two equations we control

$$x_t = 0.5x_{t-1} + c y_{t-1} + \epsilon_{x,t},$$

$$y_t = 0.2 + 0.6y_{t-1} + 0.4x_t + 0.2x_{t-1} + \epsilon_t.$$

Current shocks have variance one and correlation r . Shock pairs are independent across dates.

Case	Feedback c	Shock correlation r
Baseline	0	0
Shared current shocks	0	0.7
Lagged feedback only	0.2	0

Live code: the for-loop is the model

LIVE CODE · build, fit, then change one assumption

```
1 # Excerpt from ardl_simulator(); initialization is in the full helper.
2 epsilon_x <- rnorm(total)
3 epsilon <- shock_corr * epsilon_x +
4   sqrt(1 - shock_corr^2) * rnorm(total)
5 for (t in 2:total) {
6   x[t] <- 0.5*x[t-1] + feedback*y[t-1] + epsilon_x[t]
7   y[t] <- 0.2 + 0.6*y[t-1] + 0.4*x[t] +
8     0.2*x[t-1] + epsilon[t]
9 }
```

Which change should move OLS away from the structural coefficient 0.4?

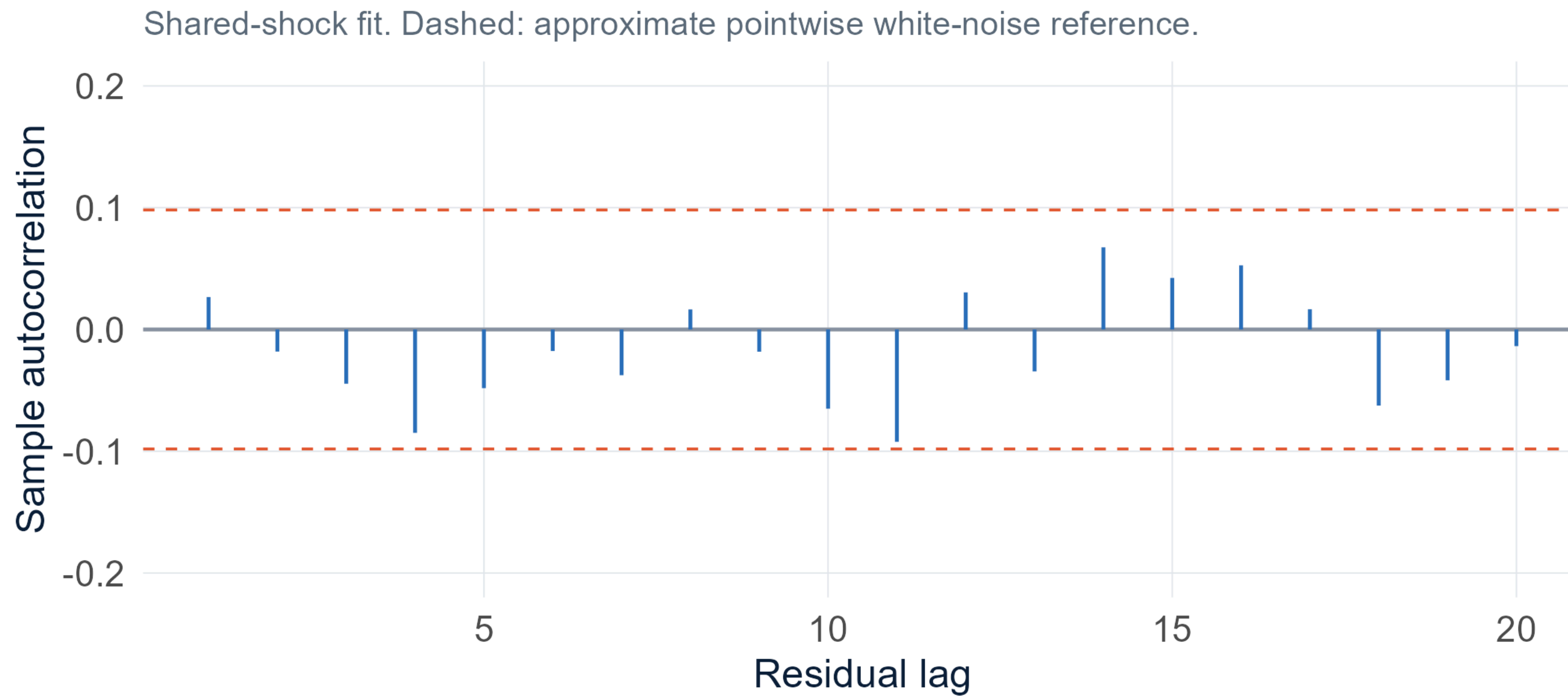
One sample is not the distribution

200 samples, $n = 400$. Dashed: structural 0.4. Dotted: conditional 1.1.



More observations reduce sampling noise. They cannot identify the structural coefficient from an invalid restriction.

The wrong structural answer can have clean residuals



What did the ACF check—and what did it leave unanswered?

OLS is answering a different conditional question

For the Gaussian shared-shock case,

$$E[\epsilon_t \mid \epsilon_{x,t}] = 0.7\epsilon_{x,t}, \quad \epsilon_{x,t} = x_t - 0.5x_{t-1}.$$

Therefore,

$$E[y_t \mid \mathcal{F}_{t-1}, x_t] = 0.2 + 0.6y_{t-1} + 1.1x_t - 0.15x_{t-1}.$$

Coefficient	Structural DGP	Conditional regression
Current x	0.4	1.1
Lagged x	0.2	-0.15

A useful conditional predictor need not recover the causal mechanism we wanted.

Act IV • Follow every path

A coefficient is one route. A response follows all routes.

First specify the input experiment

$$y_t = 0.6y_{t-1} + 0.4x_t + 0.2x_{t-1} + \epsilon_t$$

Imposed change in x	Today	Next period	Later
Pulse	+1	0	0
Permanent step	+1	+1	+1

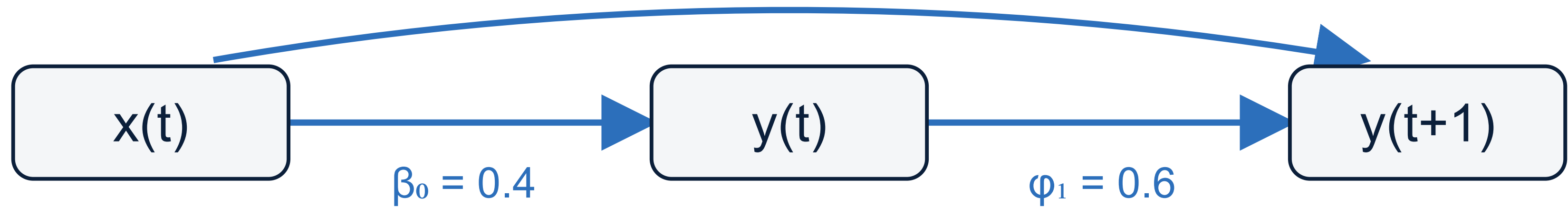
Hold initial outcome history and outcome shocks fixed; compare the two y paths.

Should a one-period pulse and a permanent step have the same long-run response?

Tomorrow has two routes from today's x

$$m_0 = \beta_0 = 0.4$$

$$\beta_1 = 0.2$$

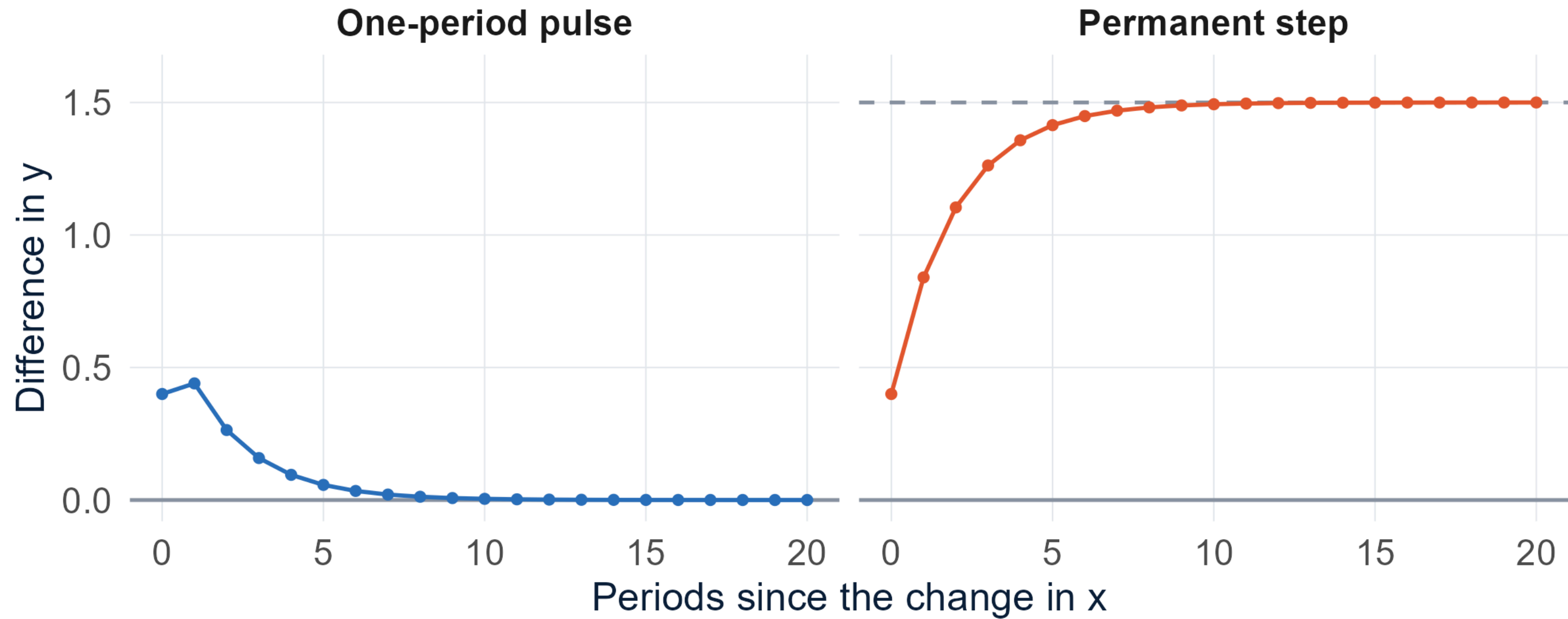


$$m_1 = \underbrace{0.6 (0.4)}_{\text{inherited}} + \underbrace{0.2}_{\text{direct}} = 0.44$$

β_1 alone is not the total next-period response.

Pulse decays; repeated pulses accumulate

Same coefficients, different input paths. Stable step limit: 1.5.



$$m_2 = 0.264, \quad \text{step at } h = 2 = 0.4 + 0.44 + 0.264 = 1.104.$$

The long-run multiplier needs stability

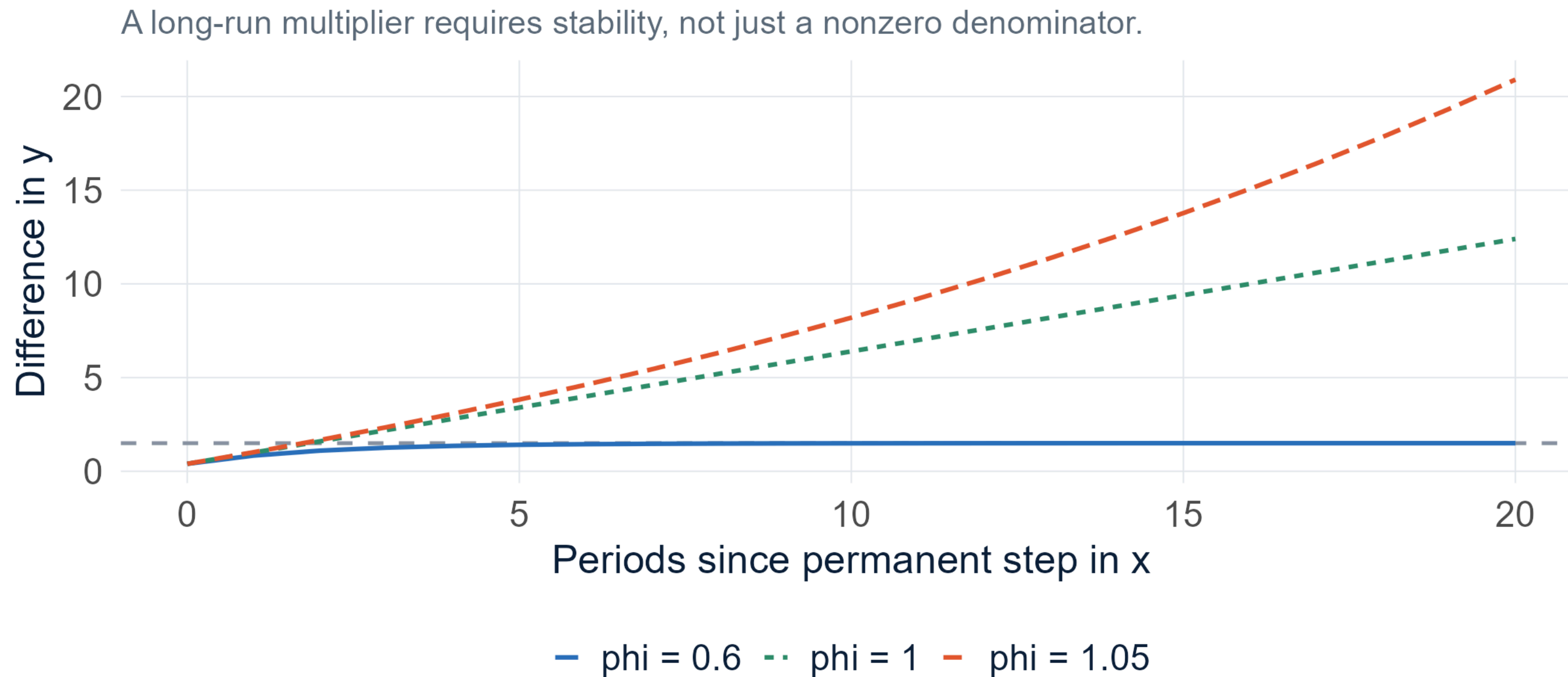
$$\text{Long-run step multiplier} = \frac{\sum_{\ell=0}^q \beta_{\ell}}{1 - \sum_{j=1}^p \phi_j}$$

$$\frac{0.4 + 0.2}{1 - 0.6} = 1.5$$

- A permanent +1 in x eventually changes y by **+1.5**.
- The one-period pulse itself converges to **zero**.
- Estimated multipliers carry hats; a denominator near zero makes estimates sensitive.

If $\hat{\phi}_1 = 1.05$, does the nonzero denominator make -12 an eventual response?

Break it: a finite ratio can describe no limit



At $\phi_1 = 1.05$ the path grows. The algebraic ratio -12 is not its limit.

An x innovation can create a whole input path

In the independent-shock, no-feedback baseline,

$$x_t = 0.5x_{t-1} + \epsilon_{x,t}.$$

A unit innovation creates an **x-path deviation** of $(1, 0.5, 0.25, \dots)$, not an imposed pulse.

Outcome response at $h = 1 = 0.44 + 0.4(0.5) = 0.64$.

Conditional path response, response to an innovation, and causal intervention are distinct objects.

What else must be followed if x responds to past y?

Act V • A predictor needs a calendar

Initial claims and unemployment, with an honest forecast origin.

The familiar workflow gets two new questions

Inspect → specify

Units, transformations, lag lengths, release dates.

Estimate → diagnose

Common estimation sample; residual plots and regression diagnostics.

Revise → evaluate

Training-only choices; identical holdout dates and benchmarks.

When was x available?

What interpretation do the assumptions support?

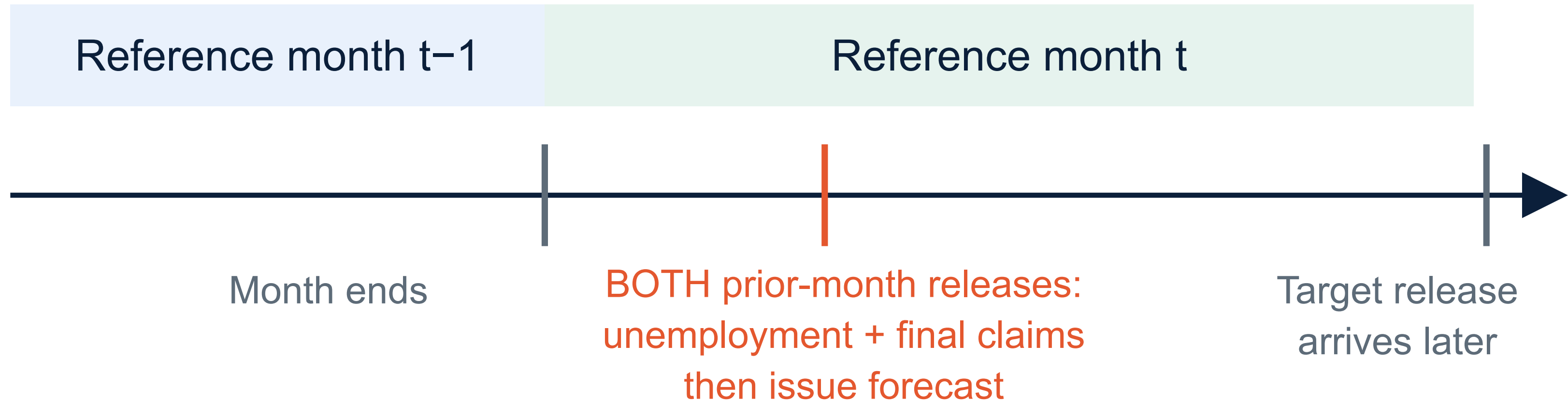
A future outcome may need a future predictor

$$\tilde{y}_{t+1|t} = \widehat{\alpha} + \widehat{\phi}_1 y_t + \widehat{\beta}_0 \tilde{x}_{t+1|t} + \widehat{\beta}_1 x_t$$

Information about future x	Claim we can make
Forecast from its own model	Forecast with predictor uncertainty
Stipulated path	Conditional / scenario forecast
Released before the target outcome	Nowcast using that information

If an evaluation substitutes the future realized x , what advantage has it given the model?

Reference month is not release date



One step ahead in the monthly index; usually a nowcast of the current calendar month.

Schematic timing, not an exact historical release calendar. Use $\mathcal{J}_t^{\text{release}}$ for this origin.

What are we actually modeling?

$$y_t = \Delta U_t, \quad x_t = 100 \Delta \log C_t$$

U: unemployment rate

Monthly percent. y is its change in percentage points.

C: initial claims

Mean of weekly counts whose week-ending Saturdays fall in the month.

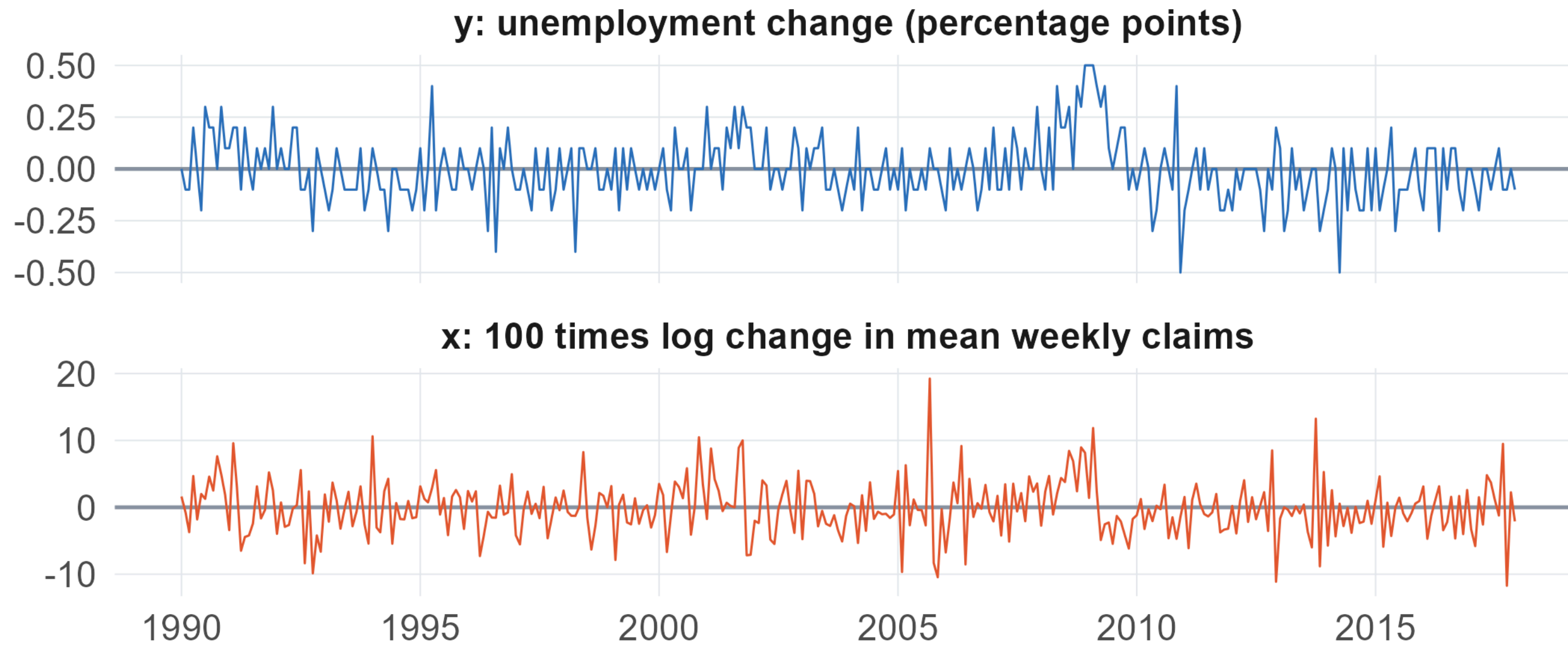
x: claims growth

$100 \times \log$ change, approximately percent growth in that monthly mean.

A monthly mean of weekly claims is not a monthly total.

Inspect the transformed training data

Training: January 1990-December 2017. Separate units and scales.



1990–2017 only. ADF rejects a unit root; KPSS does not reject level stationarity for either transformation at 5%.

Fit on a common sample; keep the holdout untouched

$$y_t = \alpha + \sum_{j=1}^p \phi_j y_{t-j} + \sum_{\ell=1}^q \beta_{\ell} x_{t-\ell} + \epsilon_t$$

TRAIN · Jan 1990–Dec 2017

336 common rows · choose by BIC

HOLDOUT

2018–2019 · 24 rows

Lag-only claims: $\beta_0 = 0$. No complete target-month claims average required.

First menu: $p = 1, 2, 3$; add one to three claims lags. Compare with AR benchmarks.

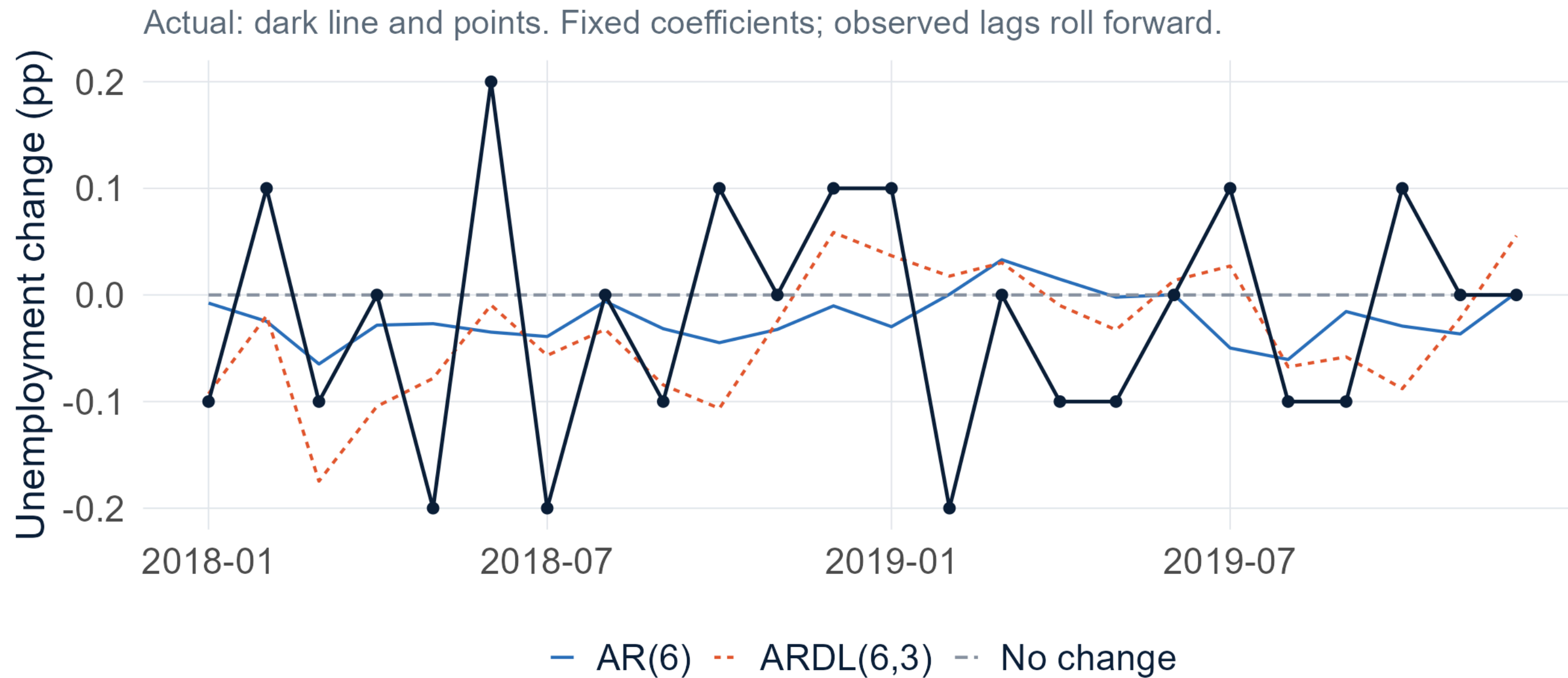
Does using yesterday's claims prove that omitted labor-market news is irrelevant?

Diagnose, revise once, and report what remains

Training decision	What happened
Small menu	AR(3) and lag-only ARDL(3,3) selected
Residual serial-correlation check	BG(12) $p \approx 0.0000085$
One justified expansion	Up to 12 outcome lags for both families
Revised selection	AR(6), lag-only ARDL(6,3)
Revised diagnostics	Stable AR roots; BG(12) $p \approx 0.0031$

Residual dependence remains. Stop this illustration and disclose it; do not search until the diagnostic passes.

Roll the information, keep the coefficients fixed



Observed lags update after release. Coefficients remain estimated on 1990–2017. No-change predicts $\Delta U = 0$.

The claims model improves this score—not every claim

Model	RMSE	MAE
AR(6)	0.1129	0.0929
Lag-only ARDL(6,3)	0.1057	0.0831
No change in unemployment	0.1099	0.0875

Unemployment percentage points · identical 24 dates · revised-data pseudo-out-of-sample evaluation.

Which statement is supported: “claims helped this forecast” or “raising claims causes unemployment to rise”?

A short-window ranking is descriptive evidence. Residual dependence and real-time-vintage limitations remain.

Act VI • Three doors out of ARDL

Brief orientation to the next modules.

Looking Ahead: does x 's past add predictive information?

MODULE 9 · GRANGER CAUSALITY · OPTIONAL PREVIEW

For a **lag-only** linear model and a stated conditioning set,

$$H_0: \beta_1 = \beta_2 = \cdots = \beta_q = 0.$$

A joint question: do these x lags add information beyond the history already included?

Predictive precedence can arise from an early signal of a common driver. It does not identify an intervention effect.

Would adding another predictor potentially change this answer?

Looking Ahead: can wandering levels share a stable gap?

MODULE 10 · COINTEGRATION · OPTIONAL PREVIEW

$$y_t \sim I(1), \quad x_t \sim I(1), \quad y_t - a - bx_t \sim I(0).$$

Harry and his dog, Mouse: both wander, but their separation can remain stationary.

- Cointegration concerns a **stationary combination** of nonstationary levels.
- An error-correction model links short-run changes to last period's deviation.
- Rewriting an ARDL as an ECM is **algebra, not a cointegration test**.

What would differ if they wandered independently with no stable separation?

Looking Ahead: give every variable an equation

MODULES 11–12 · VAR AND SHOCK INTERPRETATION

$$\begin{aligned}y_t &= a_y + b_{yy}y_{t-1} + b_{yx}x_{t-1} + \epsilon_{y,t}, \\x_t &= a_x + b_{xy}y_{t-1} + b_{xx}x_{t-1} + \epsilon_{x,t}.\end{aligned}$$

The ensemble cast: both histories can help explain both outcomes.

A system models joint evolution and forecast feedback. It still needs assumptions to identify causal shocks.

Key takeaways + next time

ARDL adds observed predictor history to outcome memory.

Exogeneity concerns the disturbance and the whole information set.

A response follows every path; state the input experiment first.

A forecast needs an honest calendar and a common benchmark.

Next: does one variable's past improve prediction beyond the information already being used?

Depth and runnable code: Module 8 notes §§8.1–8.7. Enders ch. 5 §2; Hamilton §8.2. Optional previews: Modules 9–10.