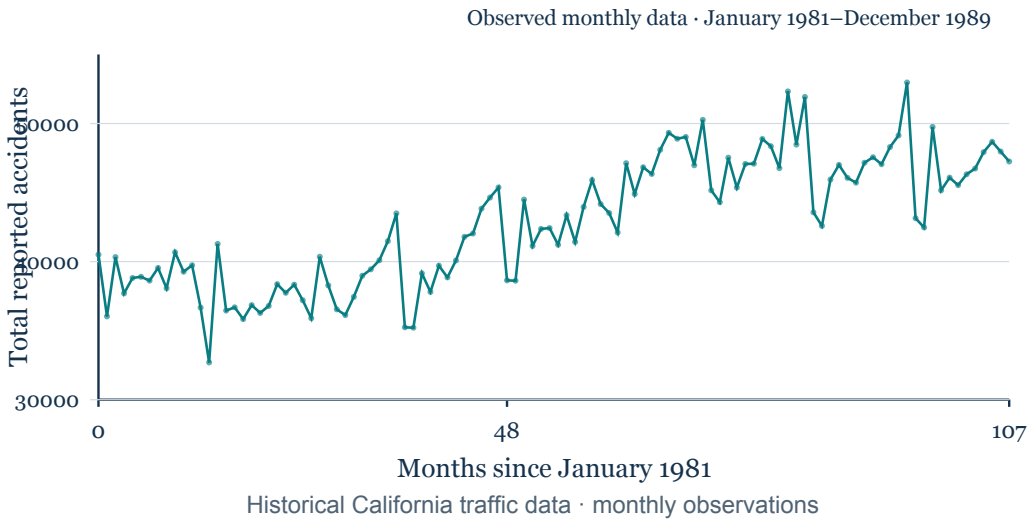
Why can the same estimate have different uncertainty?

Monthly observations may share information; uncertainty calculations must account for it.



Practice

Name a factor that might affect accidents for several consecutive months but is not fully measured in the dataset.

Explanation: A lasting change in road conditions, enforcement or reporting could affect several months. Such a shared influence can make the unexplained parts of their accident totals related.

Define the monthly accident model

State the outcome, law indicators, calendar controls and error term.

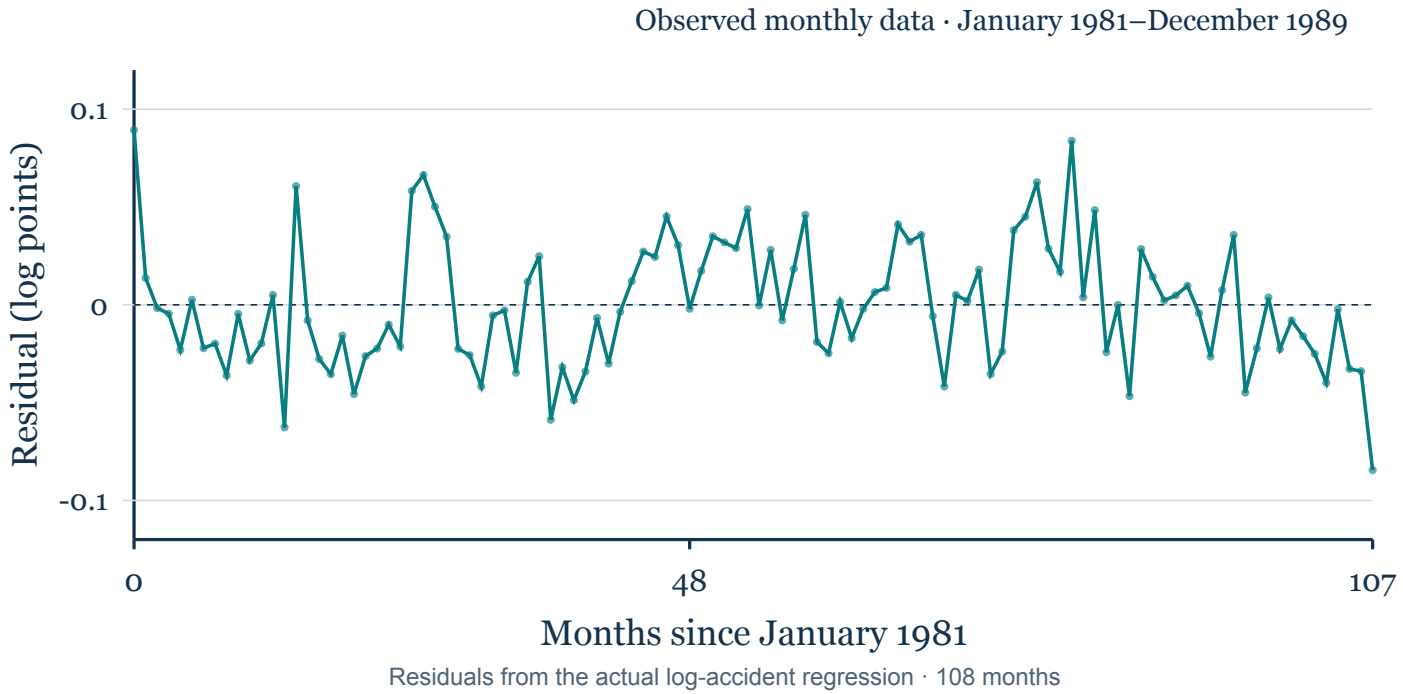
> $\log(A_t) = a + bB_t + cS_t + dU_t + fW_t + gt + \sum_{m=2}^{12} \gamma_m D_{mt} + u_t$

> B, S: law indicators; U: unemployment; W: weekend count.

> D_{mt}: month indicator; u_t: error.

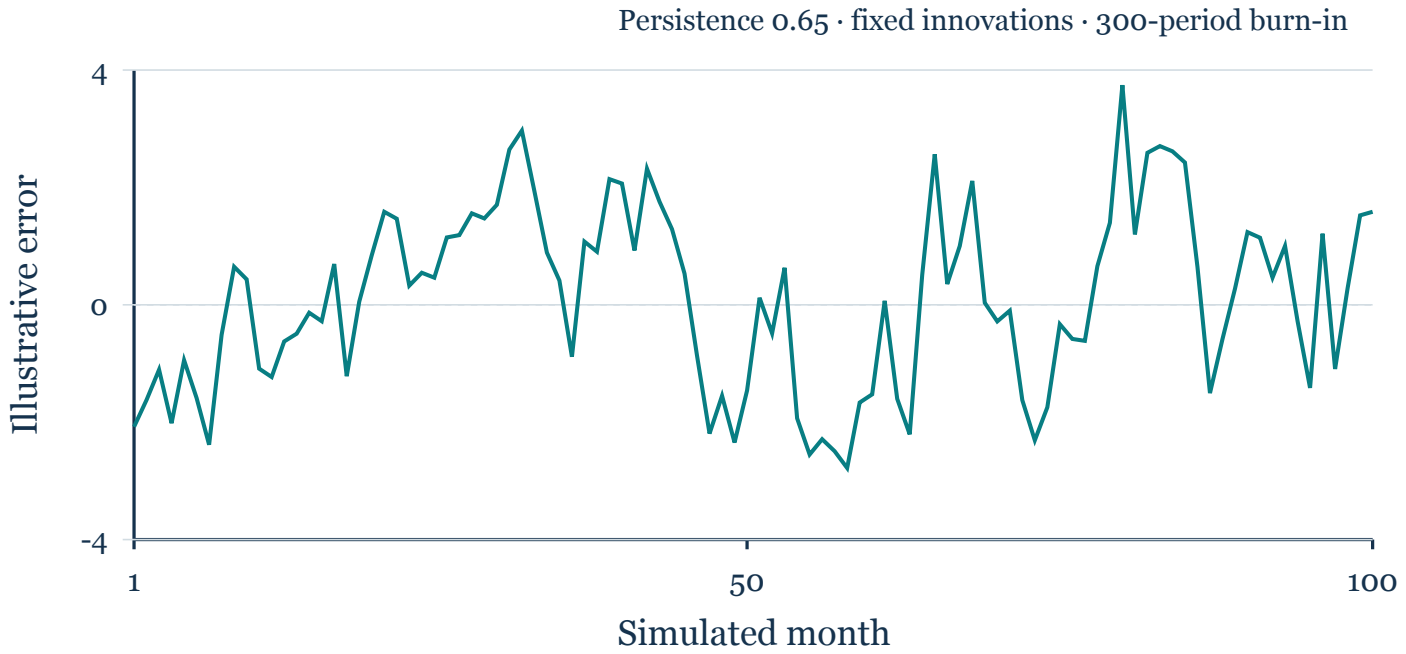
Follow the residual path

Look for sequences of positive or negative residuals.



See how an error can carry into the next month

An AR(1) process adds a new shock to a fraction of the previous error.



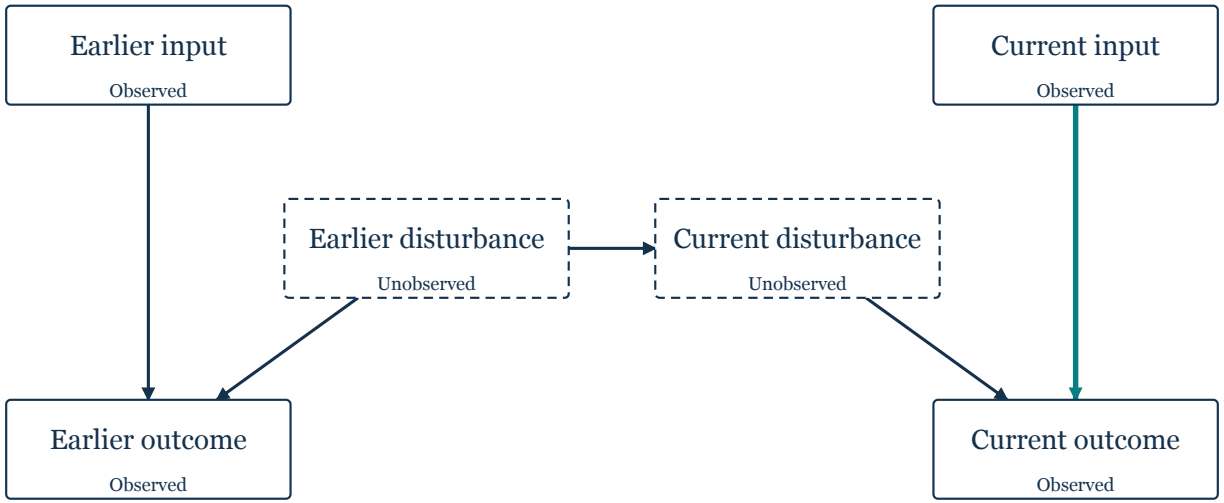
Illustrative simulation · fixed innovations · vertical range adapts to the selected persistence

$u_t = \rho u_{t-1} + \varepsilon_t$

ρ: persistence; ε_t: new unpredictable error

Separate the comparison, model and uncertainty

Related errors do not, by themselves, tell us whether the coefficient is valid.



1 open directed route · 0 open alternatives · gold: conditioned on

Assumed causal structure · dashed boxes: unobserved · gold boxes: conditioned on

- Exogeneity: required lack of relation between regressors and error
- Mean model: which variables predict the outcome?
- SE calculation: how much uncertainty remains?

Why dependence changes uncertainty

Related errors add covariance terms to the variance calculation.

> $$\operatorname{Var}(\overline{u}) = \frac{\sum_t \operatorname{Var}(u_t) + 2\sum_{s<t} \operatorname{Cov}(u_s, u_t)}{n^2}$$

> Variance measures spread; covariance measures how two errors move together.

> Two unit-variance errors: covariance $0 \Rightarrow 0.5$; covariance $0.5 \Rightarrow 0.75$

Serial covariance enters the variance through cross-date terms

The direction of the SE change depends on both regressor and error patterns.

> Let $a_t = X_t - \overline{X}$, $S = \sum_t a_t^2$.

> $\operatorname{Var}(b_1 \mid X) = S^{-2}[\sum_t a_t^2 \operatorname{Var}(u_t \mid X) + 2\sum_{t<s} a_t a_s \operatorname{Cov}(u_t, u_s \mid X)]$.

> Both regressor weights and cross-date error covariance affect precision.

Check whether past information is still missing

Dynamic completeness concerns whether the model uses the relevant past information.

> Dynamic completeness: additional relevant past information does not improve the model's conditional mean

> Adding a lag changes the fitted equation and may change the sample

> HAC addresses uncertainty from dependence that remains

Choose the response that matches the problem

Adding lags, Newey–West and Prais–Winsten change different parts of the analysis.

> Missing mean dynamics → reconsider the equation.

> Valid mean model with suitable remaining dependence → justify covariance estimation.

> A structured error transformation is developed in the next meeting.

Durbin–Watson is a residual-difference statistic

Its familiar approximation is a descriptive guide with conditions.

$$DW = \frac{\sum_{t=2}^{T}(\widehat{u}_t - \widehat{u}_{t-1})^2}{\sum_{t=1}^{T}\widehat{u}_t^2} \approx 2(1-\hat{\rho}_1).$$

Define the model and testing assumptions before turning a descriptive statistic into a decision.

A Breusch–Godfrey test keeps the original regressors in its auxiliary model

State the residual lags and the joint null.

> $\widehat{u}_t = X_t\prime\gamma + \sum_{j=1}^{q} \rho_j \widehat{u}_{t-j} + v_t.$

> $H_0 : \rho_1 = \cdots = \rho_q = 0$; a conventional LM reference is χ^2_q.

> State the sample, lag orders, implementation and covariance assumptions.

Read the serial-correlation diagnostic

Breusch–Godfrey tests a stated no-serial-correlation null.

```
. estimates restore conventional
(results conventional are active now)

. estat bgodfrey, lags(1 4)

Breusch–Godfrey LM test for autocorrelation
------------------------------------------------------------------------------
    lags(p) |          chi2               df                 Prob > chi2
------------+-----------------------------------------------------------------
       1    |         13.010               1                   0.0003
       4    |         16.907               4                   0.0020
------------------------------------------------------------------------------
                       H0: no serial correlation
```

- Supporting regression: current residual on original regressors and earlier residuals.
- Null: added residual-lag coefficients are jointly zero.
- Chi-square reference degrees of freedom: number of tested lag restrictions.

Reproduce and interpret dated residual diagnostics

Try the comparison before opening the explanation.

Practice

Reproduce the diagnostics block in overleaf_technical.do using traffic.csv. Report the DW statistic and BG tests at lag orders one and four. Explain the nulls, why four months are not four quarters, and why a diagnostic result does not by itself identify a policy effect.

Reproduce and interpret dated residual diagnostics: explanation

Follow the calculation and interpret its units.

> Use the same 108-month regression, dates and controls. DW is about 1.200475. The BG statistics are about 13.010 for one lag (p about 0.0003) and 16.907 for four lags (p about 0.0020), rejecting their stated nulls under the reference conditions. DW describes adjacent fitted-residual differences. The BG tests jointly set the specified residual-lag coefficients to zero under their reference assumptions. Four lags mean four months. Record the actual statistics and p-values from your successful run, then distinguish a conditional-mean revision from a covariance correction and from a causal identification argument.

Choose an action that addresses the actual problem

Connect the observed concern to the model, comparison or uncertainty calculation.

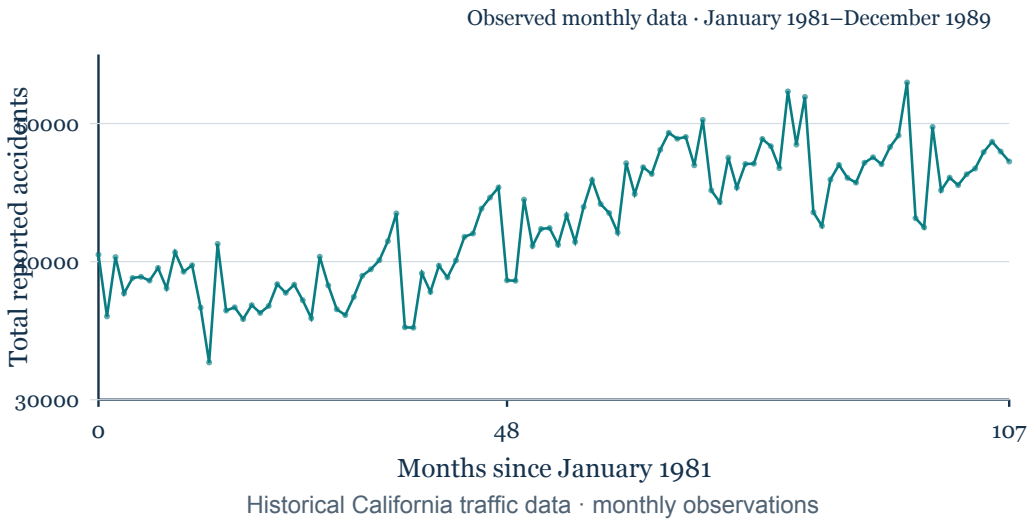
Practice

A policy indicator responds to unobserved accident risk. Is the first remedy a longer HAC window?

Explanation: No. The named problem is policy endogeneity. A longer HAC window changes uncertainty estimates but does not create an exogenous policy comparison. Address the design or identify a defensible source of policy variation.

Separate the outcome pattern from the error pattern

Serial correlation concerns errors across dates, after specifying the model.



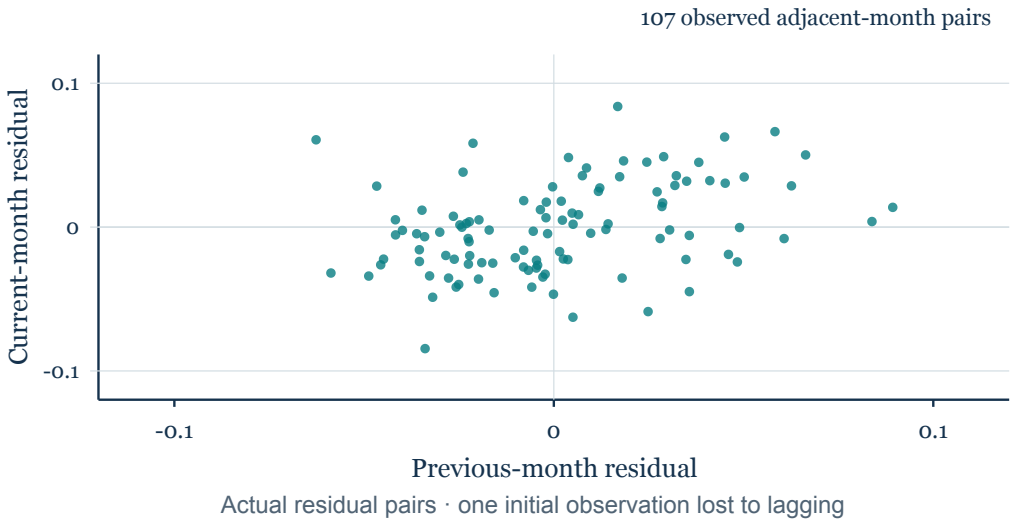
Practice

Does a seasonal accident plot prove that the errors in a seasonally adjusted regression are serially correlated?

Explanation: No. Outcome seasonality can be explained by the included month effects. Residual dependence must be assessed after fitting the specified mean model.

Compare one month with the previous month

Put the previous month's residual on the horizontal axis.



Practice

Why are there 107 residual pairs? What does a point in the upper-right region mean?

Explanation: The first residual has no observed predecessor. An upper-right point is a month with a positive residual preceded by another month with a positive residual.

What an SE adjustment cannot repair

HAC changes uncertainty; it does not remove a regressor's relation with the error.

Practice

A lagged outcome is correlated with the current error. What does an HAC option change, and what does it leave?

Explanation: It changes the estimated covariance around the fitted coefficients. It leaves the regressor–error correlation and potential coefficient inconsistency in place.

Apply the reasoning to quarterly inflation

Define the economic shock and time horizon before interpreting a future response.

> Inflation: rate of increase in the general price level

> Quarterly h = 4: the outcome one year ahead

> Local projection: a separate future-outcome regression for each horizon

> HAC lag 12 in quarterly data: a three-year covariance window