Autocorrelation in Time-Series Regression

Detecting and Correcting Serially Correlated Errors

Jake Anderson

May 16, 2026

Outline

Time-Series Data and Persistent Shocks

In cross-sectional data, observations $i = 1, \ldots, n$ have no natural order. The standard OLS assumption $\text{Cov}(e_i, e_j) = 0$ for $i \neq j$ is plausible: two random households are unlikely to share an unobserved shock.

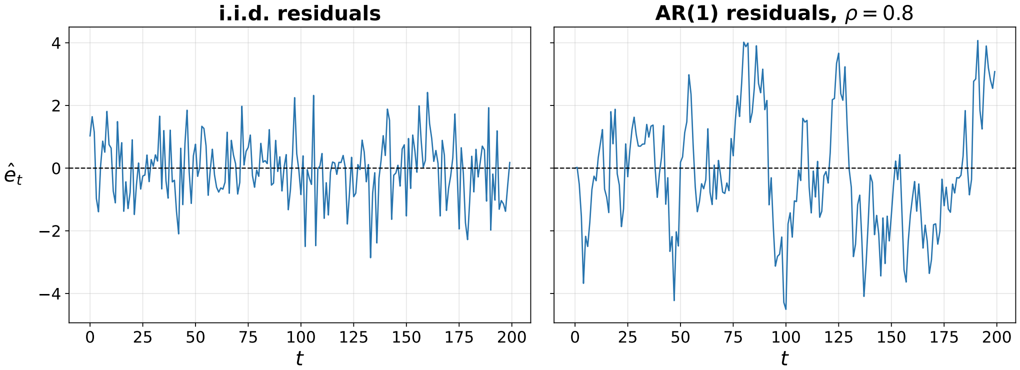
In time-series data, observations $t = 1, \ldots, T$ are ordered. The same disturbance can persist across periods:

- A demand shock today still moves prices tomorrow.
- A monetary policy surprise propagates through GDP for several quarters.
- A weather shock affects agricultural output for a full growing season.

$\implies$ Errors are likely **serially correlated**: $\text{Cov}(e_t, e_{t-k}) \neq 0$ for some lag $k \geq 1$.

What Serial Correlation Looks Like

Compare two residual series from the same regression fit on simulated data:



Left: i.i.d. errors. Sign flips frequently; no visible pattern.
Right: AR(1) errors with $\rho = 0.8$. Long positive and negative runs; the series *wanders*.

$\Longrightarrow$ Today's residual carries information about tomorrow's.

Outline

Two Sources of Dynamics

With time-indexed data, today's outcome typically depends on:

- **Its own past:** $y_{t-1}, y_{t-2}, \ldots$ (momentum, habit, inertia).
- **The past of explanatory variables:** $x_{t-1}, x_{t-2}, \ldots$ (delayed effects, accumulation).

Three workhorse specifications encode these:

- **AR**: lags of y only.
- **DL**: lags of x only.
- **ARDL**: both.

$\Longrightarrow$ Specifying dynamics in y and x comes first; thinking about dynamics in the errors comes after.

Autoregressive (AR) Models

An AR(p) model regresses y_t on its own lags:

$$y_t = \alpha + \phi_1 \, y_{t-1} + \phi_2 \, y_{t-2} + \cdots + \phi_p \, y_{t-p} + u_t, \qquad u_t \sim \text{i.i.d.}(0, \sigma_u^2).$$

Intuition: the system carries momentum forward $\implies$ a shock today still moves y tomorrow because tomorrow's y depends on today's.

Examples:

- GPA this quarter depends on GPA last quarter (study habits, cumulative mastery).
- Mood this hour depends on mood last hour (emotional inertia).
- Crime rate this year depends on crime rate last year (retaliation, network effects).

Stationarity requires $|\phi_1 + \cdots + \phi_p| < 1$ (and stricter conditions for $p \geq 2$). Without it, shocks accumulate without bound.

Distributed Lag (DL) Models

A DL(q) model regresses y_t on contemporaneous and lagged values of x:

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

Intuition: the effect of x on y builds up over time rather than landing all at once.

Examples:

- Air pollution today $\rightarrow$ respiratory illness over the next several weeks.
- Ad spend this month $\rightarrow$ sales over the next few quarters.
- Fertilizer application $\rightarrow$ crop yield months later.

$\Longrightarrow$ The sequence $\beta_0, \beta_1, \ldots, \beta_q$ traces out the **lag distribution** of x's effect on y. The sum $\sum_{s=0}^{q} \beta_s$ is the long-run multiplier.

ARDL Models: The Combination

An ARDL(p, q) model has both kinds of dynamics:

$$y_t = \alpha + \sum_{j=1}^{p} \phi_j \, y_{t-j} + \sum_{s=0}^{q} \beta_s \, x_{t-s} + u_t.$$

Why combine them?

- AR alone captures persistence in y but ignores external drivers.
- DL alone captures effects of x but ignores y's own momentum.
- ARDL captures both $\implies$ richer dynamics with fewer lags of either.

ARDL is the dominant specification in applied time-series work because it nests both AR and DL as special cases.

$\implies$ The chapter's main interpretive payoff (lag weights, long-run multipliers, forecasting) lives inside the ARDL framework.

Errors That Remain Autocorrelated

Even after specifying an ARDL with as many lags of y and x as you think relevant, the errors e_t may *still* be autocorrelated. Reasons:

- Omitted lags of y or x leak into e_t and inherit persistence.
- Persistent unobserved shocks (weather, sentiment, policy uncertainty) move slowly.
- Misspecified functional form $\Longrightarrow$ time-correlated fit error.

Two parsimonious models for the error process:

$$\textbf{AR(1):}\quad e_t = \rho\, e_{t-1} + u_t, \qquad \textbf{MA(1):}\quad e_t = u_t + \theta\, u_{t-1},$$

with $u_t \sim \text{i.i.d.}(0, \sigma_u^2)$.

- **AR(1)**: today's error depends on yesterday's error $\Longrightarrow$ shocks persist geometrically, $\text{Cov}(e_t, e_{t-k}) = \rho^k\, \text{Var}(e_t)$.
- **MA(1)**: today's error is a weighted sum of today's and yesterday's white-noise shocks $\Longrightarrow$ a shock u_{t-1} affects only e_{t-1} and e_t, then disappears. Covariance is non-zero only at lag 1.

$\Longrightarrow$ AR gives long decaying correlation; MA gives short abrupt correlation. The rest of this deck targets AR(1)-type structure (the more common case) with tests that also pick up MA(1) and higher orders.

Outline

OLS Under Autocorrelation: The Good News

Suppose the regressors are strictly exogenous: $\mathrm{E}[e_t \mid x_1, \ldots, x_T] = 0$ for all t. Then even with serially correlated errors, OLS is still:

- **Unbiased:** $\mathrm{E}[\hat{\beta}] = \beta$.
- **Consistent:** $\hat{\beta} \xrightarrow{p} \beta$ as $T \to \infty$.

Why? The single-regressor formula

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_t (x_t - \bar{x})\, e_t}{\sum_t (x_t - \bar{x})^2}$$

has expectation β_1 as long as $\mathrm{E}[e_t \mid x] = 0$. Whether the e_t's correlate *with each other* does not enter.

$\implies$ The coefficient estimates themselves are fine (in the strictly exogenous case). The problem is the standard error.

Where the SE Formula Goes Wrong

For a single-regressor model, the true sampling variance of the slope is

$$\mathrm{Var}(\hat{\beta}_1) = \frac{1}{\left[\sum_t (x_t - \bar{x})^2\right]^2} \, \mathrm{Var}\left(\sum_t (x_t - \bar{x}) \, e_t \right).$$

Expand the inner variance:

$$\mathrm{Var}\left(\sum_t (x_t - \bar{x}) \, e_t \right) = \underbrace{\sum_t (x_t - \bar{x})^2 \, \mathrm{Var}(e_t)}_{\text{diagonal: usual term}} + \underbrace{2 \sum_{t<s} (x_t - \bar{x})(x_s - \bar{x}) \, \mathrm{Cov}(e_t, e_s)}_{\text{off-diagonal: missed by OLS}}.$$

The OLS formula $\widehat{\mathrm{Var}}(\hat{\beta}_1) = \dfrac{\hat{\sigma}^2}{\sum_t (x_t - \bar{x})^2}$ keeps only the diagonal term. Under autocorrelation, the cross terms $\mathrm{Cov}(e_t, e_s)$ are nonzero and the formula misses them entirely.

$\Longrightarrow$ OLS standard errors are **inconsistent** under autocorrelation. *Software handles the multi-regressor case the same way (matrix sandwich form).*

Which Way Are SEs Wrong?

The sign of the error tells you whether OLS *under*-states or *over*-states uncertainty, so it is worth knowing the typical direction.

Positive autocorrelation ($\rho > 0$) and trending or persistent regressors (the typical macro case):

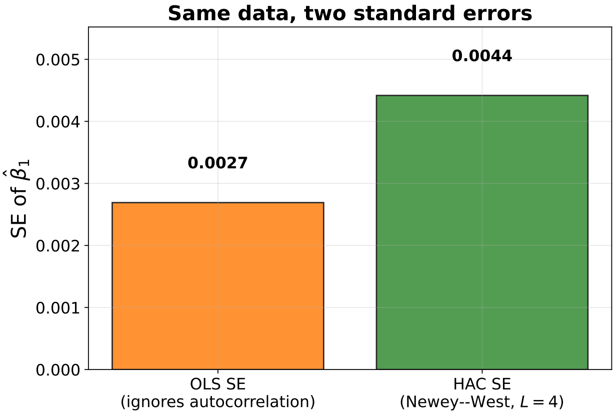
- The off-diagonal $\text{Cov}(e_t, e_s)$ terms are positive.
- They share sign with $(x_t - \bar{x})(x_s - \bar{x})$ when x is also persistent.
- True $\text{Var}(\hat{\beta}_1)$ is *larger* than what OLS reports.

$\Longrightarrow$ OLS SEs are **too small**; t-statistics are **too large**; we over-reject H_0 and find "significant" effects that are not really there.

Negative autocorrelation: the reverse can happen, but is rare in applied work.

Visualizing the Damage

The same data, the same point estimate, two different SEs:



The HAC bar (right) is consistent; the OLS bar (left) is what you would report if you ignored the autocorrelation.
$\Longrightarrow$ Using OLS SEs would shrink the CI by roughly half and inflate the t-statistic accordingly.

One Special Case: Lagged Dependent Variables

Suppose the model includes a lag of y as a regressor:

$$y_t = \beta_0 + \beta_1 y_{t-1} + e_t.$$

If e_t is autocorrelated, then e_t is correlated with e_{t-1}, which is part of y_{t-1}. Strict exogeneity fails:

$$\text{Cov}(y_{t-1}, e_t) \neq 0.$$

$\implies$ OLS is **biased and inconsistent** in this case, not just inefficient. Correcting SEs is not enough; we have to model the dynamics directly (covered in the dynamic models topic).

Implication for the DW test (coming up): DW assumes strict exogeneity and is *not* valid when a lagged dependent variable is a regressor.

Outline

Visual Detection: Three Pictures

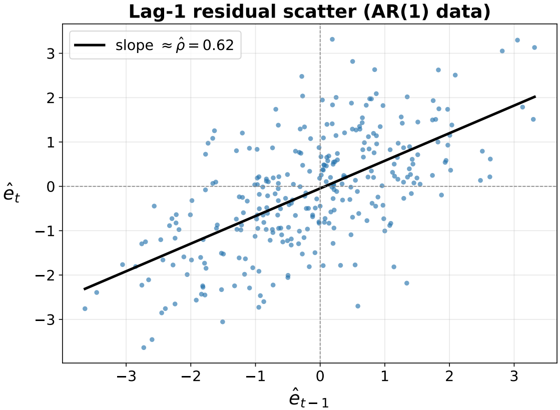
Before running a formal test, look at the residuals. Three plots cover most cases:

1. **Time plot of $\hat{e}_t$:** sign-flips frequently $\implies$ probably i.i.d.; long runs $\implies$ probably positively autocorrelated.
2. **Scatter of $\hat{e}_t$ vs $\hat{e}_{t-1}$:** cloud around the origin $\implies$ no autocorrelation; clear line $\implies$ AR(1) structure.
3. **Sample autocorrelogram (ACF):** bars at each lag k with $\pm 1.96/\sqrt{T}$ significance bounds.

None of these gives a p-value, but together they tell you what *kind* of autocorrelation you have, which guides which formal test to run.

The Lag-1 Scatter

Plot today's residual against yesterday's:



A line through this cloud has slope approximately $\hat{\rho}$. Upward tilt $\implies$ positive autocorrelation at lag 1; the visual analogue of regressing $\hat{e}_t$ on $\hat{e}_{t-1}$.

The Sample Autocorrelation Function

Define the sample autocorrelation at lag k:

$$r_k = \frac{\sum_{t=k+1}^{T}(\hat{e}_t - \bar{\hat{e}})(\hat{e}_{t-k} - \bar{\hat{e}})}{\sum_{t=1}^{T}(\hat{e}_t - \bar{\hat{e}})^2}.$$

Under the null of i.i.d. residuals and for large T,

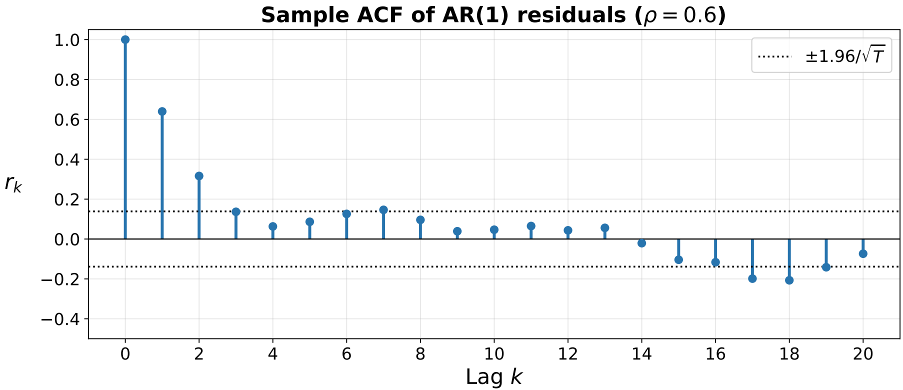
$$r_k \approx \mathcal{N}\left(0,\ \tfrac{1}{T}\right) \quad \Longrightarrow \quad \text{individual } r_k \text{ is ``significant'' if } |r_k| > \frac{1.96}{\sqrt{T}}.$$

The ACF plot shows r_k as vertical bars at $k = 0, 1, 2, \ldots$ with a horizontal band at $\pm 1.96/\sqrt{T}$.

Read it like this:

- Bar at lag 1 outside the band $\Longrightarrow$ probable AR(1) or MA(1).
- Bars at lags 1 and 2 outside the band $\Longrightarrow$ probable AR(2) or longer.
- Bar at lag 12 (monthly) or lag 4 (quarterly) outside the band $\Longrightarrow$ seasonal autocorrelation.

Example ACF: AR(1) Residuals



Bars at small lags poke above the band. The geometric decay pattern ($r_k \approx \rho^k$) is the visual signature of an AR(1) process.

$\implies$ A formal test with $p = 1$ (BG) or with the AR(1) alternative (DW) is the natural next step.

Formal Tests for Autocorrelation

Plots are suggestive, but “looks autocorrelated” is not a number to put in a paper. Two formal tests cover the standard cases:

- **Durbin–Watson:** tests $\rho = 0$ in the AR(1) model. Simple, classic, restrictive.
- **Breusch–Godfrey:** tests $\rho_1 = \cdots = \rho_p = 0$ for any chosen lag order p. Flexible, handles AR(p), works with lagged dependent variables.

$\Longrightarrow$ We work through each test in turn.

Outline

DW Test: The Idea

Premise: if errors follow AR(1) with coefficient ρ, then consecutive OLS residuals will tend to look alike when $\rho > 0$ and alternate sign when $\rho < 0$.

The Durbin–Watson statistic measures how different $\hat{e}_t$ is from $\hat{e}_{t-1}$ on average. Define

$$DW = \frac{\sum_{t=2}^{T}(\hat{e}_t - \hat{e}_{t-1})^2}{\sum_{t=1}^{T} \hat{e}_t^2}.$$

A bit of algebra (expand the square and use $\sum \hat{e}_t^2 \approx \sum \hat{e}_{t-1}^2$ for large T) gives the useful approximation

$$DW \approx 2(1 - \hat{\rho}), \qquad \text{where } \hat{\rho} = \frac{\sum_{t=2}^{T} \hat{e}_t \hat{e}_{t-1}}{\sum_{t=1}^{T} \hat{e}_t^2}.$$

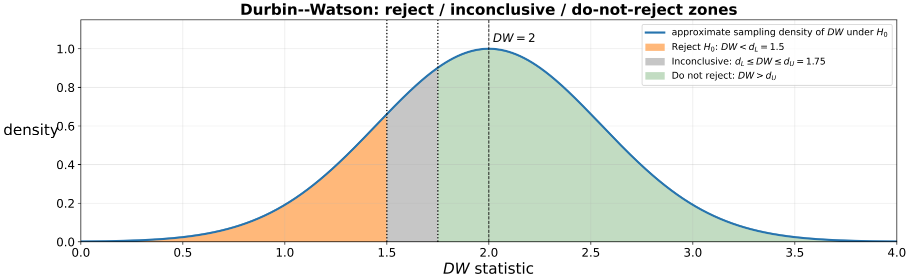
- $\hat{\rho} \approx 0 \implies DW \approx 2$ (no autocorrelation).
- $\hat{\rho} \approx 1 \implies DW \approx 0$ (strong positive autocorrelation).
- $\hat{\rho} \approx -1 \implies DW \approx 4$ (strong negative autocorrelation).

Hypotheses and the Inconclusive Zone

Null and alternative (one-sided, against positive autocorrelation):

$$H_0 : \rho = 0 \quad \text{(no autocorrelation, } DW \approx 2\text{)}$$
$$H_1 : \rho > 0 \quad \text{(positive autocorrelation, } DW < 2\text{)}$$

Unlike a t or χ^2 test, the exact distribution of DW depends on the particular values of the regressors in your sample. Durbin and Watson tabulated **two** critical values, d_L (lower) and d_U (upper):



The Decision Rule

For the lower-tail test (against $\rho > 0$):

Region	**Decision**
$DW < d_L$	Reject H_0: positive autocorrelation
$d_L \leq DW \leq d_U$	**Inconclusive**
$DW > d_U$	Do not reject H_0

The mirror image holds for the upper tail (against $\rho < 0$): use $4 - d_L$ and $4 - d_U$.

Why two critical values? The sampling distribution of DW under H_0 shifts left or right depending on the regressors. d_L and d_U bracket the worst-case and best-case designs $\Longrightarrow$ any DW between them is consistent with both rejection and non-rejection for *some* design.

$\Longrightarrow$ The inconclusive zone is the price of avoiding a critical value that depends on the specific regressors used.

Limitations of DW

The DW test is classic and easy to compute, but it imposes strong restrictions:

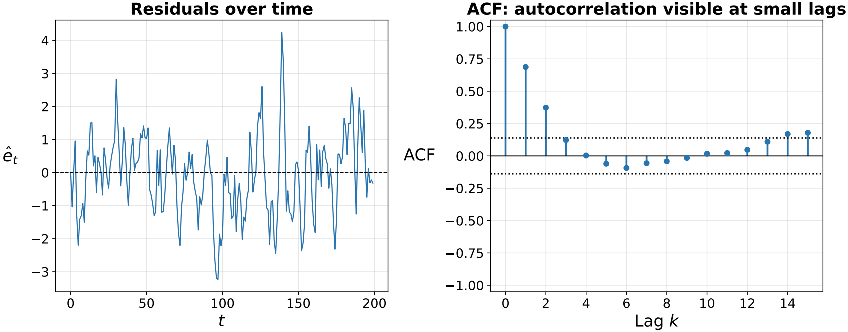
1. **Only AR(1):** it has power against $e_t = \rho e_{t-1} + u_t$ and not much else. MA(1) can sometimes be detected; AR(2) and seasonal autocorrelation often slip past.
2. **Strictly exogenous regressors:** $\mathrm{E}[e_t \mid X] = 0$ for all t. If a lagged dependent variable is on the right-hand side, DW is biased toward 2 (Durbin's h statistic is the modified version for that case).
3. **Non-stochastic / fixed regressors:** the tabulated critical values assume non-random X. Most modern textbooks treat the tables as approximate when X is random.
4. **The inconclusive zone:** a real fraction of samples land in the gap, leaving the user without a decision.

$\Longrightarrow$ DW is fine for a quick screen. For anything beyond AR(1), use Breusch–Godfrey (next section).

Outline

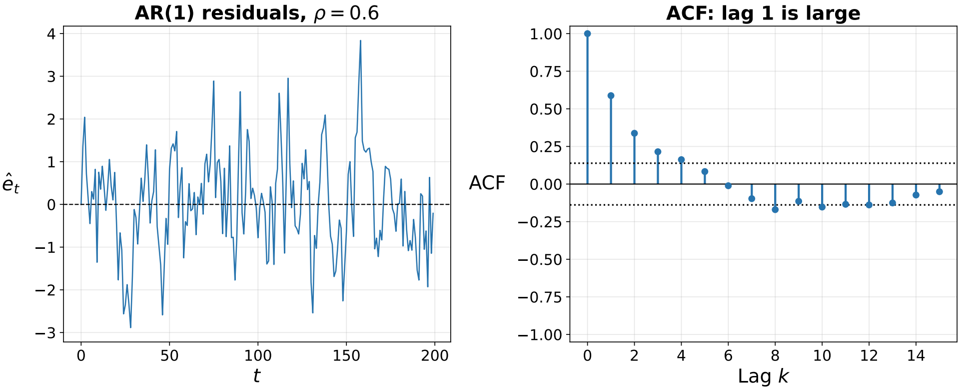
BG Test: The Idea

Premise: If autocorrelation is present, OLS residuals at time t are related to residuals at earlier times. The BG test regresses $\hat{e}_t$ on its own lags (and the original regressors) and asks whether the lag coefficients are jointly zero.



If past residuals predict current residuals $\Longrightarrow$ reject no-autocorrelation.

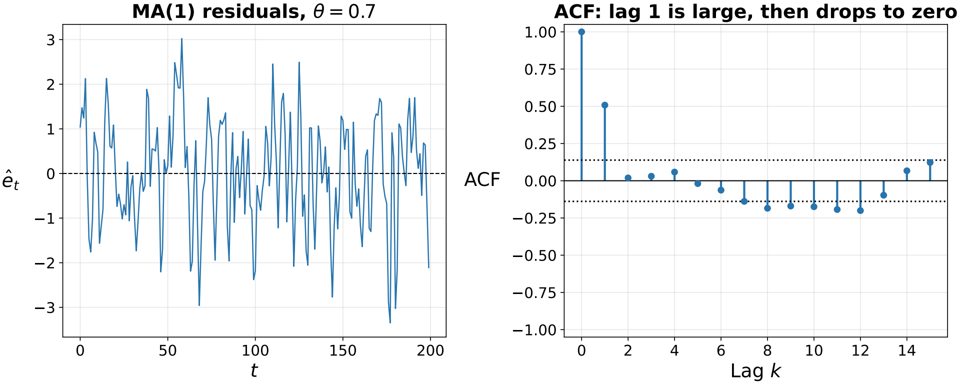
When the BG Test Works: AR(1) Errors



Residuals follow $e_t = 0.6\, e_{t-1} + u_t$. The ACF spike at lag 1 is exactly what BG with $p = 1$ is built to catch.

BG Test successfully detects autocorrelation!

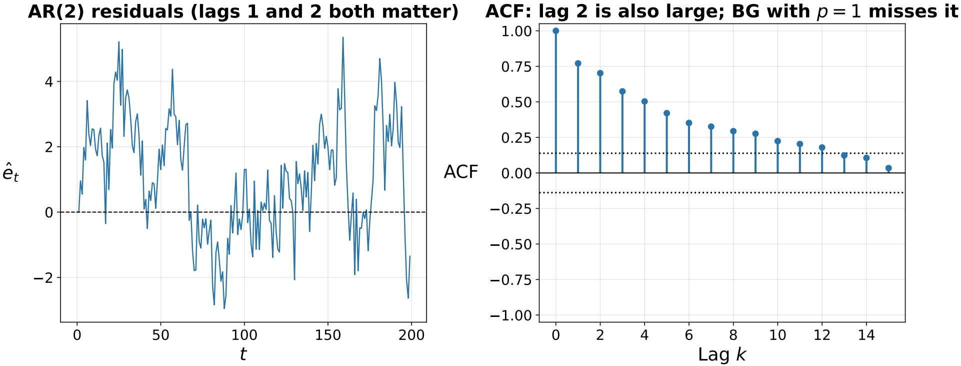
When the BG Test Works: MA(1) Errors



Residuals follow $e_t = u_t + 0.7\,u_{t-1}$. ACF lag 1 is large; lags 2+ drop to zero. BG with $p = 1$ catches this.

BG Test successfully detects autocorrelation!

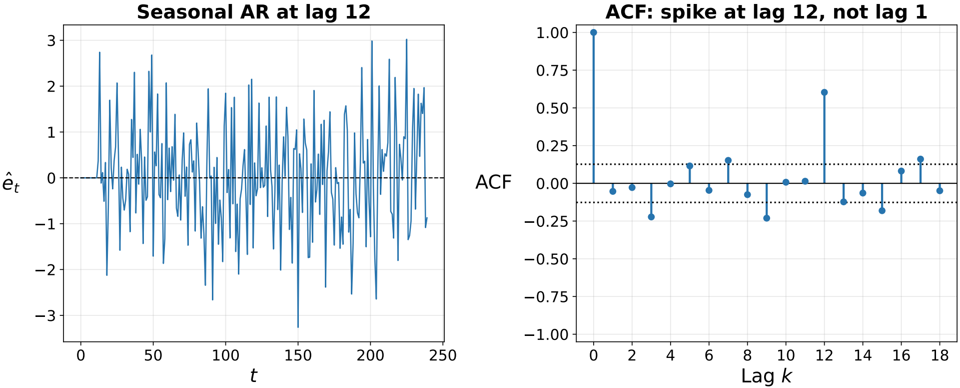
When the BG Test Fails: Wrong Lag Order



Residuals follow $e_t = 0.5\, e_{t-1} + 0.4\, e_{t-2} + u_t$ (AR(2)). Lag 2 carries real signal but BG with $p = 1$ ignores it. Either pick a larger p, or the test understates the autocorrelation.

BG Test fails to detect this kind of autocorrelation

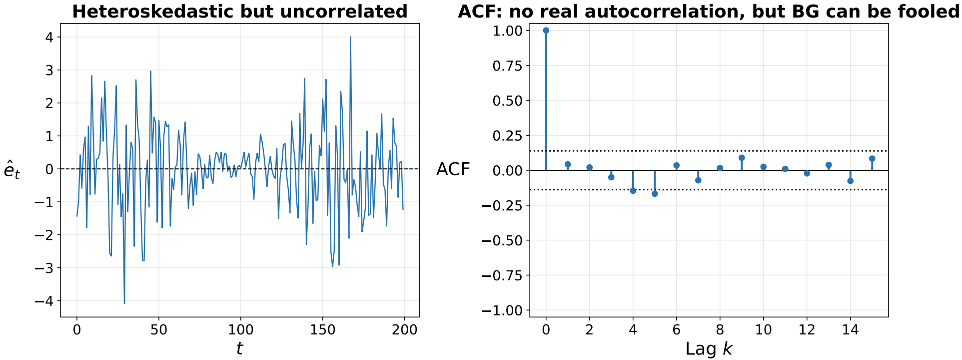
When the BG Test Fails: Seasonal Autocorrelation



Residuals follow $e_t = 0.7\, e_{t-12} + u_t$ (seasonal AR at lag 12). The ACF spike is at lag 12; lags 1–11 are roughly zero. BG with $p \leq 11$ misses it entirely.

BG Test fails to detect this kind of autocorrelation

When the BG Test Fails: Heteroskedastic but Uncorrelated



Residuals are independent but their variance changes over time. The ACF should be flat, but the slow variance drift can occasionally produce spurious lag-k correlations and lead BG to falsely reject.

BG Test can be misled by non-autocorrelation problems

BG Test: Procedure

Auxiliary regression: regress the OLS residual on the original regressors *and* on p lagged residuals.

$$\hat{e}_t = \alpha_0 + \alpha_1 x_t + \rho_1 \hat{e}_{t-1} + \rho_2 \hat{e}_{t-2} + \cdots + \rho_p \hat{e}_{t-p} + v_t$$

Steps:

1. Run OLS on the original model. Save residuals $\hat{e}_t$.
2. Choose a lag order p (one for each lagged residual to include).
3. Run the auxiliary regression above. Compute R^2_{aux}.
4. Test statistic: $\text{BG} = (n-p) \cdot R^2_{\text{aux}} \sim \chi^2(p)$ under $H_0 : \rho_1 = \cdots = \rho_p = 0$.

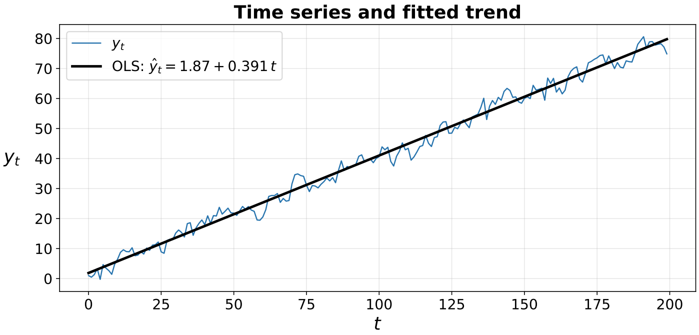
The degrees of freedom equal p, the number of *lagged-residual* terms in the auxiliary regression. **The original regressors and intercept are not counted.**

Choosing p: rule of thumb is $p = 1$ for clean monthly/yearly data, $p = 4$ for quarterly, $p = 12$ for monthly seasonal. When unsure, run several p's.

Example: Trending Series with AR(1) Errors

We have a time series y_t (e.g., a monthly sales index) over $n = 200$ periods. We model y_t as a linear trend in t:

$$y_t = \beta_0 + \beta_1 \, t + e_t$$



The fitted line captures the trend, but consecutive residuals around it tend to have the same sign: large positive deviations cluster together, large negative deviations cluster together. That is the visual signature of positive autocorrelation. We run BG with $p = 1$ to test for first-order autocorrelation.

Step 1: Run OLS

Coefficients

	Estimate	Std. Error	t-value	$\Pr(>\lvert t\rvert)$
(Intercept)	1.8656	0.3311	5.63	$< 2 \cdot 10^{-16}$ ***
t	0.3914	0.0029	135.98	$< 2 \cdot 10^{-16}$ ***

Residual standard error: 2.35 on 198 degrees of freedom
Multiple R^2*:* 0.9894 *Adjusted* R^2*:* 0.9894
F*-statistic:* 18490 on 1 and 198 DF, $p < 2 \cdot 10^{-16}$

Signif. codes: `0 *** 0.001 ** 0.01 * 0.05 . 0.1 1`

The trend is extremely significant under OLS *assumptions*. **But are the residuals independent over time?** BG will tell us.

Step 2: State the Hypothesis

The BG test asks whether $\hat{e}_t$ depends on its own past values, after controlling for t.

Null and alternative (here $p = 1$):

$$H_0 : \rho_1 = 0 \quad \text{(no first-order autocorrelation)}$$
$$H_1 : \rho_1 \neq 0 \quad \text{(autocorrelation at lag 1)}$$

Test statistic:

$$\text{BG} = (n - p) \cdot R^2_{\text{aux}} \ \sim \ \chi^2(p) \text{ under } H_0$$

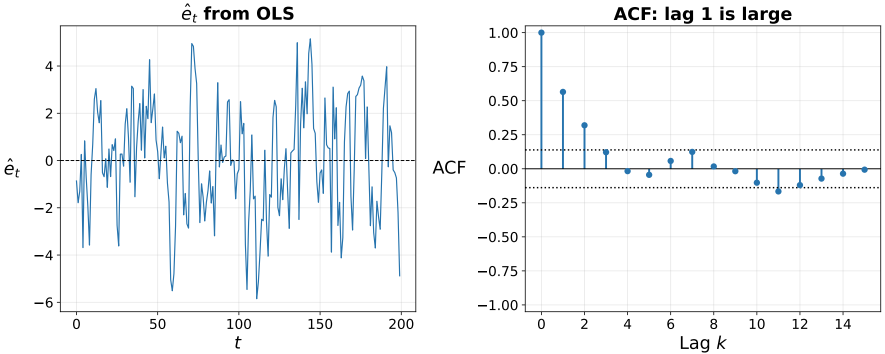
where p is the number of *lagged-residual* terms (the regressors and intercept are not counted). Here $p = 1$.

Step 3: Auxiliary Regression

Compute residuals $\hat{e}_t$ and regress on t plus one lag of itself:

$$\hat{e}_t = \alpha_0 + \alpha_1\, t + \rho_1\, \hat{e}_{t-1} + v_t$$

Estimated: $\hat{\rho}_1 = 0.577, \quad R^2_{\text{aux}} = 0.326.$



The residual time series wanders rather than scattering randomly; the ACF lag-1 spike is exactly what BG is built to catch.

Step 4: Compute the Test Statistic

Plug in:

$$\begin{aligned} \text{BG} &= (n - p) \cdot R^2_{\text{aux}} \\ &= (200 - 1) \cdot 0.326 \\ &= 199 \cdot 0.326 \\ &= \boxed{64.92} \end{aligned}$$

Compare to the critical value of $\chi^2(1)$ at $\alpha = 0.05$.

Step 5: Compare to χ^2 Critical Value

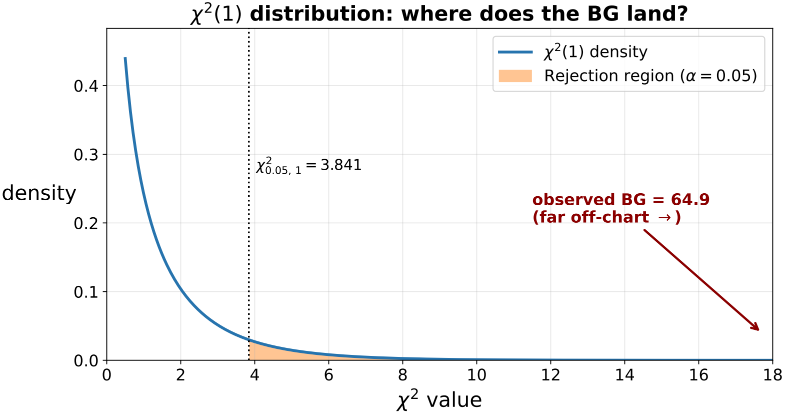
df	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.025$	$\alpha = 0.01$
1	2.706	**3.841**	5.024	6.635
2	4.605	5.991	7.378	9.210
3	6.251	7.815	9.348	11.345
4	7.779	9.488	11.143	13.277
5	9.236	11.070	12.833	15.086

We chose $p = 1$ lagged residual in the auxiliary $\implies$ df $= 1$. The intercept and original regressors are not counted. At $\alpha = 0.05$:

$$\chi^2_{0.05,\,1} = 3.841$$

Reject H_0 if BG > 3.841.

Step 6: Visualize the Decision



The orange region is the upper 5% tail of $\chi^2(1)$. BG $= 64.92$ is far past the right edge; p-value is essentially zero.

Step 7: Conclusion

Decision: BG $= 64.92 \gg 3.841 \implies$ **reject** H_0 at $\alpha = 0.05$ (and at any conventional level).

Plain language: the residuals are strongly autocorrelated at lag 1. After fitting the trend, today's deviation tells you something about tomorrow's.

Implications:

- OLS coefficients ($\hat{\beta}_0 = 1.87$, $\hat{\beta}_1 = 0.391$) remain unbiased and consistent.
- OLS standard errors (e.g., 0.0029 on the trend) are **inconsistent**; the gap does not shrink with n.
- Use *HAC* (Newey-West) standard errors, or model the AR structure directly via GLS.

Outline

Test-and-Fix vs Correct-the-SE

A DW or BG test rejects no-autocorrelation. Two strategies:

1. **Model the structure.** Specify $e_t = \rho e_{t-1} + u_t$, then estimate by GLS (or Cochrane–Orcutt, a specific iterated GLS routine for AR(1) errors). Both inference and efficiency improve, but only if the AR(1) assumption is correct.
2. **Keep OLS, fix the SE.** Use a HAC (heteroskedasticity-and-autocorrelation consistent) standard error. Inference is valid for a wide class of error processes; the point estimate is unchanged.

$\Longrightarrow$ Newey–West HAC SEs are the workhorse: they require no parametric model for the error dynamics and are the applied-econometrics default.

The HAC Idea

Recall the diagonal-plus-off-diagonal decomposition (from the consequences section):

$$\operatorname{Var}\left(\sum_t (x_t - \bar{x})\, e_t\right) = \underbrace{\sum_t (x_t - \bar{x})^2 \operatorname{Var}(e_t)}_{\text{White-style term}} + \underbrace{2 \sum_{t<s} (x_t - \bar{x})(x_s - \bar{x}) \operatorname{Cov}(e_t, e_s)}_{\text{autocorrelation term}}.$$

HAC estimators plug in sample analogues. The diagonal piece uses $\hat{e}_t^2$. The off-diagonal piece sums $\hat{e}_t \hat{e}_{t-k}$ across lags.

Why not sum all $T - 1$ lags? Sample autocovariances at far lags are noisy. Newey–West truncates the sum at a chosen **bandwidth** L and applies triangular down-weighting to the lags that are kept (the Bartlett kernel).

The Newey–West Estimator: Schematic

Newey–West estimates the long-run variance as two pieces:

$$\widehat{S}_L = \underbrace{(\text{White-style sum})}_{\text{diagonal: } \hat{e}_t^2 \text{ terms}} + \underbrace{(\text{weighted sum of lag-}k\text{ autocovariances, } k = 1, \ldots, L)}_{\text{autocorrelation correction}}.$$

The autocorrelation piece sums sample products $\hat{e}_t \hat{e}_{t-k}$ across lags, with triangular weights that taper to zero at lag $L + 1$ (the Bartlett kernel).

- $L = 0$: no autocorrelation correction; Newey–West reduces to White (HC0).
- Larger L: more lags included; correction grows.

$\Longrightarrow$ The estimator combines a heteroskedasticity correction (the White part) with an autocorrelation correction. Software handles the explicit weights and the multi-regressor matrix form.

Choosing the Bandwidth L

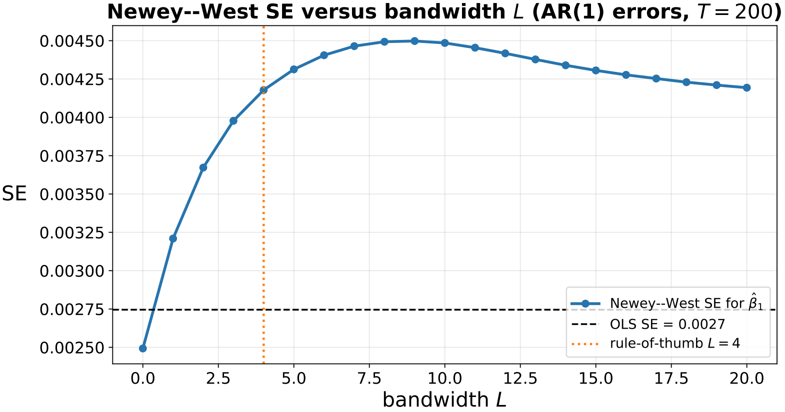
The bandwidth trades bias against variance:

- L too small $\implies$ omits real autocovariances; SEs still inconsistent.
- L too large $\implies$ includes noisy estimates of small autocovariances.

In practice:

- Software picks a sensible default (typically $L \approx 4$ for sample sizes near $T = 100$, scaling slowly with T).
- Report SEs for a small range of L values as a sanity check.
- If the SE is stable across $L \in \{2, 4, 8\}$, you're fine. If it drifts upward as L grows, the autocorrelation extends further than the default captures.

Bandwidth Sensitivity in Practice



The SE rises sharply from $L = 0$ (no correction, equals OLS) up to $L \approx 4$, then plateaus. The plateau is the signal that the AR(1) structure has been absorbed; further lags add noise but no bias.

$\Longrightarrow$ Read the value off the plateau, not off $L = 0$ and not off the largest L available.

Newey–West in R

Using the sandwich and lmtest packages:

```
library(sandwich) library(lmtest)
model <- lm(y   x, data = ts_data)
# Auto bandwidth coeftest(model, vcov. = NeweyWest(model))
# Explicit bandwidth L=4 V <- NeweyWest(model, lag = 4, prewhite = FALSE) coeftest(model, vcov. = V)
```

- Point estimate $\hat{\beta}$ is unchanged.
- Only the SE changes (and therefore CI, t, p).
- Reporting OLS and HAC SEs side-by-side is standard practice.

Outline

Summary: Detect, Test, Correct

Tool	When to Use	Limitation
Time plot of $\hat{e}_t$	Always: first look at the residuals	Subjective; no p-value
Lag-1 scatter, ACF	Diagnosing the kind of autocorrelation	Suggestive, not formal
Durbin–Watson	Quick test for AR(1) errors with strictly exogenous regressors	Only AR(1); inconclusive zone; fails with lagged y
Breusch–Godfrey	General test for AR(p); allows lagged dependent variables	Need to choose p
Newey–West HAC SEs	Default fix when autocorrelation is present	Need to choose bandwidth L

The decision flow:

1. Run OLS. Plot residuals and ACF.
2. Run BG (preferred) or DW. If significant $\implies$ OLS SEs are unreliable.
3. Report Newey–West HAC SEs alongside (or instead of) OLS SEs.

What This Topic Did Not Cover

Two adjacent topics, each big enough for its own treatment:

- **Dynamic models:** when the data-generating process actually involves lags of y (e.g., $y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 x_t + e_t$), autocorrelation in e_t makes OLS *biased*, not just inefficient. Covered in the next topic.
- **Spurious regression and cointegration:** two trending series can produce significant OLS slopes even when they share no economic relationship. The tools developed here (DW, BG, HAC) help diagnose the problem, but the fix is differencing or cointegration analysis.

The minimum you should leave with:

- Autocorrelation breaks OLS SEs, not (usually) OLS point estimates.
- Always test in time-series regressions; BG dominates DW outside of textbook AR(1) cases.
- HAC standard errors are the default correction $\iff$ you can keep OLS coefficients and still report valid t-statistics.

Thank you!

jakeanderson@g.ucla.edu