# Violation of Assumptions of Classical Regression Model: Autocorrelation
# Unit 4

Prepared by

**Aditya Korekallu Srinivasa**

Scientist (Senior Scale)

ICAR-Indian Agricultural Research Institute
New Delhi

# Contents

# 1 What is Autocorrelation?

## 1.1 Definition and Concept

**Autocorrelation** (or serial correlation) refers to the correlation of a regression model's error terms across different time periods or observations. In the classical linear regression model, it is assumed that the error terms are uncorrelated:

$$\text{Cov}(u_t, u_s) = 0 \quad \text{for all} \quad t \neq s$$

Autocorrelation occurs when this assumption is violated, meaning the value of the error term in one period is correlated with its value in another.

## 1.2 Why is it a Violation?

The assumption of no autocorrelation is crucial for the validity of OLS estimators' standard errors and for hypothesis testing. When autocorrelation is present, OLS estimators remain unbiased and consistent, but they are no longer efficient (not BLUE), and standard errors are biased, leading to unreliable inference.

## 1.3 Visual Illustration

In time series data, plotting residuals against time may reveal systematic patterns (e.g., cycles, trends, or persistence), indicating autocorrelation.

# 2 Reasons for Autocorrelation

## 2.1 Nature of Data

- **Time series data:** Economic and financial time series often display persistence, cycles, or trends, making autocorrelation likely.
- **Omitted variables:** Excluding relevant variables that are themselves autocorrelated can induce autocorrelation in the error term.
- **Incorrect functional form:** Fitting a linear model to a nonlinear process can result in autocorrelated errors.

## 2.2 Economic or Institutional Causes

- **Inertia or adjustment lags:** Economic variables (e.g., consumption, investment) adjust slowly over time, leading to correlated errors.
- **Data smoothing and reporting practices:** Aggregating or smoothing data can introduce serial dependence.
- **Measurement errors:** Errors in reporting or recording time series data may persist across periods.

### 2.3 Practical Example

In a regression of GDP growth on lagged variables, if shocks to the economy persist over time, the error terms will be autocorrelated.

## 3 Practical Situations Where Autocorrelation Occurs

### 3.1 Economic Data Examples

- **Macroeconomic time series:** GDP, inflation, interest rates, and unemployment often display autocorrelation.
- **Financial returns:** Stock returns may exhibit volatility clustering (ARCH/GARCH effects).
- **Environmental and agricultural data:** Weather and yield data often show persistence over time.

### 3.2 Panel Data

Autocorrelation can occur within individual units (firms, countries) over time in panel datasets.

### 3.3 Graphical Illustration

Plotting residuals against time, or plotting the autocorrelation function (ACF), can reveal persistent positive or negative correlations.

## 4 Detection and Identification of Autocorrelation

Detecting autocorrelation is essential for time series regression diagnostics. Both informal and formal methods are used, with interpretation being critical.

### 4.1 Symptoms in Regression Output

- **Systematic patterns in residuals:** Residuals show runs of positive or negative values, or cycles, when plotted against time.
- **Slowly decaying autocorrelation function:** The sample ACF of residuals shows significant values at low lags.
- **Inefficient or misleading inference:** High $R^2$ but insignificant $t$-statistics, or inconsistent coefficient signs.

## 4.2 Formal Detection Methods

**1. Graphical Methods**

- **Residual plots:** Plot residuals or standardized residuals against time. Patterns (runs, cycles) suggest autocorrelation.
- **Correlogram/ACF plot:** Plot the sample autocorrelation function of residuals. Significant autocorrelations at lag 1 or higher suggest serial correlation.
- **Interpretation:** If residuals alternate randomly, autocorrelation is unlikely. If they show persistence, cycles, or slow decay, autocorrelation is likely.

**2. Durbin–Watson (DW) Test**

- **Procedure:** Compute the DW statistic:

$$d = \frac{\sum_{t=2}^{n}(e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}$$

  where $e_t$ are OLS residuals.

- **Decision Rule:**
  - $d \approx 2$ suggests no autocorrelation.
  - $d < 2$ suggests positive autocorrelation.
  - $d > 2$ suggests negative autocorrelation.
  - Compare $d$ to critical values ($d_L$, $d_U$) from DW tables for a formal test. If $d < d_L$, reject the null of no autocorrelation.
- **Interpretation:** A $d$ value substantially below 2 (and significant by table) indicates positive first-order autocorrelation.

**3. Breusch–Godfrey (BG) Test**

- **Procedure:**
  1. Estimate the regression and obtain residuals.
  2. Regress residuals on all regressors and $p$ lagged residuals (for testing up to $p$th order autocorrelation).
  3. Compute $nR^2$ from this auxiliary regression.
- **Decision Rule:**
  - Under the null of no autocorrelation, $nR^2$ follows a chi-square distribution with $p$ degrees of freedom.
  - **A significant p-value (<0.05) indicates autocorrelation.**
- **Interpretation:** The BG test is flexible, allowing for higher-order and conditional autocorrelation.

4. **Runs Test**

- **Procedure:** Count the number of runs (sequences of positive or negative residuals) in the residual series.
- **Decision Rule:** Compare the observed number of runs to the expected number under randomness.
- **Interpretation:** Too few or too many runs suggest non-randomness and possible autocorrelation.

### 4.3 Summary Table: Interpreting Detection Methods

| **Method** | **Procedure** | **Evidence of Autocorrelation** |
|---|---|---|
| Residual/ACF Plots | Plot residuals or ACF against time | Persistent patterns, cycles, or slow decay in ACF |
| Durbin–Watson (DW) | Compute $d$ statistic from residuals | $d \ll 2$ (and below critical value): positive autocorrelation |
| Breusch–Godfrey (BG) | Auxiliary regression of residuals on lags | Significant $nR^2$ statistic (low p-value) |
| Runs Test | Count number of runs in residuals | Too few or too many runs vs. expectation |

### 4.4 Practical Guidance

- Always begin with graphical analysis to spot patterns.
- Use DW for first-order autocorrelation; BG for higher-order or conditional autocorrelation.
- A significant test result (low p-value or $d$ far from 2) means the null of no autocorrelation is rejected.
- Non-significant results suggest no strong evidence of autocorrelation, but do not guarantee its absence.

## 5 Consequences of Autocorrelation

### 5.1 Theoretical Consequences

- OLS estimators remain unbiased and consistent (if regressors are strictly exogenous).
- OLS is no longer efficient (not BLUE); standard errors are underestimated or overestimated.
- Inference (t, F tests) becomes invalid due to incorrect standard errors.

### 5.2 Practical Consequences

- **Over- or underestimation of significance:** Variables may appear insignificant (or significant) due to incorrect standard errors.
- **Misleading policy conclusions:** Policy recommendations based on incorrect inference can be problematic.
- **Reduced predictive accuracy:** Model predictions may be less reliable, especially in dynamic settings.

### 5.3 Example

In a regression of inflation on lagged inflation and other variables, autocorrelated errors can lead to underestimated standard errors, overstating the significance of predictors.

## 6 Remedies for Autocorrelation

### 6.1 Model Specification

- **Add omitted variables:** Include relevant variables that may explain the persistence in the data.
- **Use appropriate functional form:** Nonlinear or dynamic models may better capture the data's structure.

### 6.2 Transformations and Estimation Methods

- **Generalized Least Squares (GLS):** If the autocorrelation process is known (e.g., AR(1)), GLS or Cochrane-Orcutt estimation can be used for efficient estimates.
- **Newey–West robust standard errors:** Adjust OLS standard errors to remain valid in the presence of autocorrelation (especially for large samples).
- **Prais–Winsten and Hildreth–Lu methods:** Iterative procedures to estimate models with AR(1) errors.

### 6.3 Lagged Dependent Variables

Including lagged dependent variables as regressors can help capture dynamics and reduce autocorrelation in the errors.

### 6.4 Interpretation and Cautions

If the primary goal is prediction and the model fits well, mild autocorrelation may not be a major concern. For inference, addressing autocorrelation is essential.

## Glossary

- **Autocorrelation**: Correlation of error terms across time or observations in a regression model.
- **Serial Correlation**: Another term for autocorrelation, especially in time series.
- **Durbin–Watson Statistic**: A test statistic for detecting first-order autocorrelation in regression residuals.
- **Breusch–Godfrey Test**: A general test for higher-order autocorrelation.
- **Generalized Least Squares (GLS)**: An estimation method that accounts for known forms of autocorrelation or heteroscedasticity.
- **Cochrane–Orcutt Procedure**: A method for estimating regression models with AR(1) errors.
- **Newey–West Standard Errors**: Robust standard errors valid in the presence of autocorrelation and heteroscedasticity.

## Practice Questions

1. Define autocorrelation. How does it violate the classical regression assumptions?
2. List and explain the main causes of autocorrelation in applied econometric work.
3. Give practical examples where autocorrelation is likely to occur in economic data.
4. Describe and interpret at least three methods for detecting autocorrelation in a regression model.
5. What are the consequences of autocorrelation for OLS estimation and inference?
6. Discuss at least three remedies for autocorrelation. When should each be used?
7. Is autocorrelation always a problem? Explain with reference to prediction versus inference.
8. How can the Durbin–Watson and Breusch–Godfrey tests be used to diagnose autocorrelation?
9. What is the Cochrane–Orcutt procedure and when is it appropriate to use?
10. Explain the use and interpretation of Newey–West standard errors.

## References


- Gujarati, D. N., & Porter, D. C. (2010). *Basic Econometrics* (5th ed.). McGraw-Hill. [See especially Chapter 12: Autocorrelation]
- Wooldridge, J. M. (2013). *Introductory Econometrics: A Modern Approach* (5th ed.). Cengage Learning. [See discussion in the chapter on time series and autocorrelation]