

Lecture 31: Autocorrelation: concept, measures and causes; the AR(1) scheme

Econometrics ■ Unit III: Autocorrelation and instrumental variables

Kapil Kumar

Department of Statistics
Chaudhary Charan Singh University, Meerut

B.Sc. (Hons) Statistics, Semester V
Academic Session 2026–2027

Unit III: Autocorrelation and instrumental variables

Lecture 31 begins Unit III: Autocorrelation and instrumental variables.

Today

- Concept
- The first-order autoregressive scheme

Consider a regression on time-series data,

$$y_t = x_t' \beta + u_t, \quad t = 1, 2, \dots, n.$$

The classical model assumes $\text{Cov}(u_t, u_s) = 0$ for $t \neq s$. If this fails, so that disturbances of different periods are correlated, we have **autocorrelation** (or **serial correlation**).

- **Positive autocorrelation:** a positive disturbance tends to be followed by a positive one, and a negative by a negative. Residuals show long runs of the same sign.
- **Negative autocorrelation:** the signs tend to alternate.

Autocorrelation is common in time series and rare in cross-section data (where the order of observations has no meaning).

Measures. If the disturbances have mean zero and constant variance, define the autocovariance and the autocorrelation at lag s :

$$\gamma_s = \text{Cov}(u_t, u_{t-s}), \quad \rho_s = \frac{\gamma_s}{\gamma_0} = \frac{\text{Cov}(u_t, u_{t-s})}{\text{Var}(u_t)}.$$

Causes.

1. **Inertia.** Economic series such as GDP, prices and employment move slowly; the effect of a shock lasts for several periods.
2. **Omitted variables.** If a relevant variable that is itself autocorrelated is left out, its effect enters the disturbance, which then becomes autocorrelated.
3. **Wrong functional form.** Fitting a straight line to a curved relation gives residuals with long runs of the same sign.
4. **Lags.** Consumption this period depends partly on consumption last period. If the lagged term is omitted, the disturbance becomes autocorrelated.
5. **Data manipulation.** Averaging monthly data into quarters, or interpolating missing values, smooths the data and introduces correlation.

The most commonly used model for autocorrelated disturbances is the **first-order autoregressive scheme**, AR(1):

$$u_t = \rho u_{t-1} + v_t, \quad |\rho| < 1, \quad (5.1)$$

where

- the v_t are independent with $E(v_t) = 0$, $\text{Var}(v_t) = \sigma_v^2$, and
- v_t is independent of $u_{t-1}, u_{t-2}, \dots$

The constant ρ is the *coefficient of autocorrelation*.

Practice

Re-work the derivations and examples of this lecture without looking at the slides.

Next lecture

Lecture 32: Moments of the $AR(1)$ disturbances.