Serial correlation (or autocorrelation) is the violation of **Assumption 4** (observations of the error term are uncorrelated with each other).

Pure Serial Correlation

This type of correlation tends to be seen in time series data. To denote a time series data set we will use a t subscript.

This type of serial correlation occurs when the error in one period is correlated with the errors in other periods. The model is assumed to be correctly specified.

The most common form of serial correlation is called **first-order serial correlation** in which the error in time t is related to the previous $(t-1)$ period's error:

$$\varepsilon_t = \rho\varepsilon_{t-1} + u_t, \ -1 < \rho < 1$$

The new parameter $\boldsymbol{\rho}$ is called the **first-order autocorrelation coefficient.** The process for the error term is referred to as a **first-order autoregressive process** or **AR(1)**.

The magnitude of ρ tells us about the strength of the serial correlation and the sign indicates the nature of the serial correlation.

- $\rho = 0$ indicates **no** serial correlation

- $\rho > 0$ indicates **positive** serial correlation – the error terms will tend to have the same sign from one period to the next

- $\rho < 0$ indicates **negative** serial correlation – the error terms will tend to have a different sign from one period to the next

Impure Serial Correlation

This type of serial correlation is caused by a specification error such as an omitted variable or ignoring nonlinearities.

Suppose the true regression equation is given by

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \varepsilon_t$$

but in our model we do not include X_{2t}. The error term ε_t will capture the effect of X_{2t}. Since many economic variables exhibit trends over time, X_{2t} is likely to depend on $X_{2,t-1}, X_{2,t-2}, \ldots$. This will translate into a seeming correlation between ε_t and $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$ and this serial correlation would violate Assumption 4.

A specification error of the functional form can also cause this type of serial correlation. Suppose the true regression equation between Y and X is quadratic but we assume it's linear. The error term will depend on X^2.

The Consequences of Serial Correlation

1. Pure serial correlation does not cause bias in the regression coefficient estimates.

2. Serial correlation causes OLS to no longer be a minimum variance estimator.

3. Serial correlation causes the estimated variances of the regression coefficients to be biased, leading to unreliable hypothesis testing. The *t*-statistics will actually appear to be more significant than they really are.

Testing for First-Order Serial Correlation

Plotting the residuals is always a good first step.

The most common formal test is the Durbin-Watson d test.

The Durbin-Watson Test

Consider the regression equation

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \varepsilon_t.$$

We are interested in determining if there is first-order autocorrelation in the error term

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

where u_t is not autocorrelated.

The test of the null hypothesis of no autocorrelation ($\rho = 0$) is based on the Durbin-Watson statistic

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

where the e_t's are the residuals from the regression equation estimated by least squares.

The value of this statistic is automatically reported in EViews regression output.

Let's consider a few cases:

- Extreme positive correlation
- Extreme negative correlation
- No serial correlation

For an alternative of **positive** autocorrelation, $H_A: \rho > 0$, look up the critical values in tables B-4, B-5 or B-6. The decision rule is as follows:

Reject H_0 if $d < d_L$.
Do not reject H_0 if $d > d_U$.
Test inconclusive if $d_L < d < d_U$.

Remedies for Serial Correlation

1. Cochrane-Orcutt Procedure

Transform the equation with AR(1) error structure

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \varepsilon_t, \quad \varepsilon_t = \rho \varepsilon_{t-1} + u_t$$

into one that is not autocorrelated.

Consider the same model in period $t-1$:

Multiply both sides of this equation by ρ:

Subtract this equation from the original equation:

If ρ were known, we could use OLS to obtain estimates that are BLUE. Unfortunately ρ is not known and therefore must be estimated.

Step 1: Compute the residuals from the OLS estimation of

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \ldots + \beta_k X_{kt} + \varepsilon_t$$

Step 2: Estimate the first-order serial correlation coefficient using $e_t = \rho e_{t-1} + u_t$. Call this estimate $\hat{\rho}$.

Step 3: Transform the variables: $Y_t^* = Y_t - \hat{\rho} Y_{t-1}$, $X_{1t}^* = X_{1t} - \hat{\rho} X_{1,t-1}$, etc.

Step 4: Regress Y_t^* on a constant, $X_{1t}^*, X_{2t}^*, \ldots, X_{kt}^*$ and obtain the OLS estimates of the transformed equation.

Step 5: Use these estimates to obtain a new set of estimates for the residual and then go back to Step 2.

Step 6: This iterative procedure is stopped when the estimates of ρ differ by no more than some pre-specified amount (eg. 0.0001) in two successive iterations.

2. Newey-West Standard Errors

Adjust the standard errors of the estimated regression coefficients but not the estimates themselves since they are still unbiased.

Example

Let's examine the relationship between Microsoft's marketing and advertising expenditures and its revenues.

$$\text{revenues}_t = \beta_0 + \beta_1 \text{marketing}_t + \varepsilon_t$$

REVENUES: Microsoft's revenues in millions of dollars
MARKETING: Microsoft's expenditures on marketing and advertising in millions of dollars

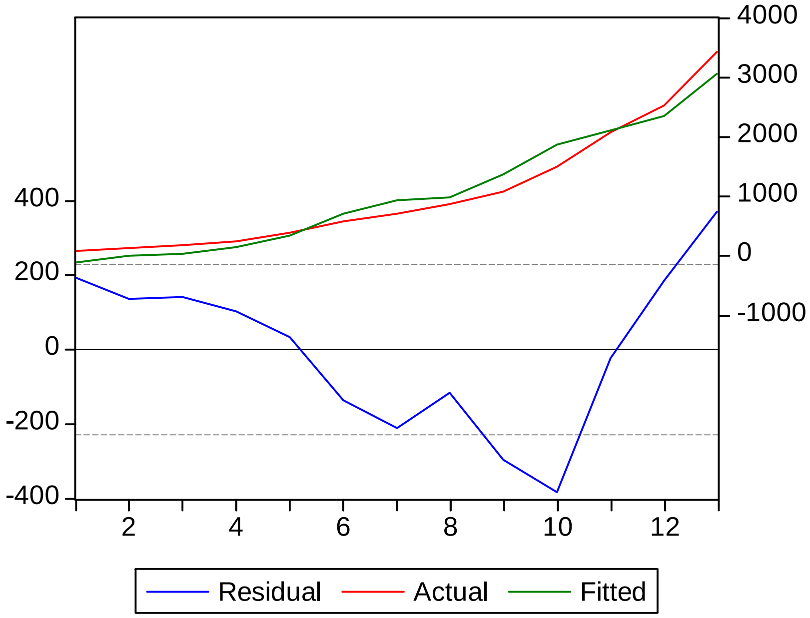
Data: real data from Microsoft's website for 1987 through to 1999

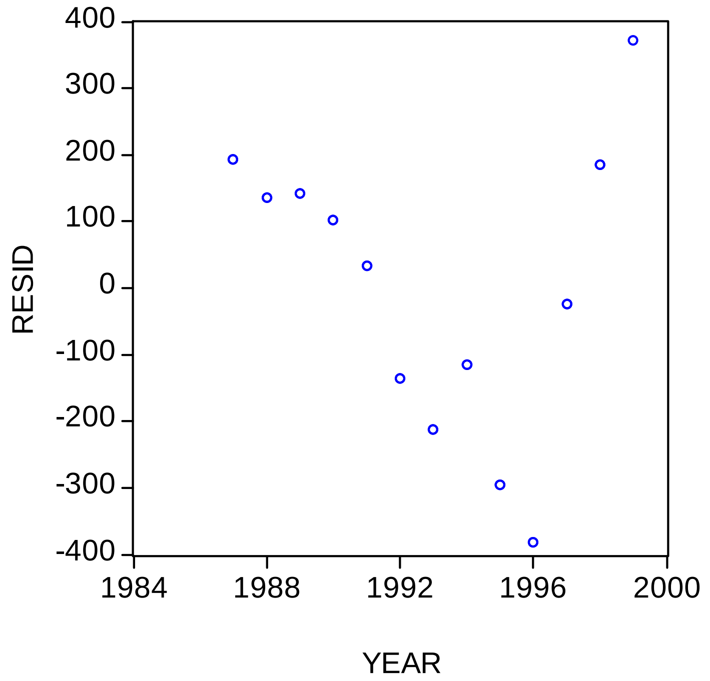
Dependent Variable: REVENUES
Method: Least Squares
Sample: 1 13
Included observations: 13

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MARKETING	6.169683	0.397549	15.51932	0.0000
C	-245.9496	105.3924	-2.333656	0.0396

R-squared	0.956323	Mean dependent var	1063.561
Adjusted R-squared	0.952352	S.D. dependent var	1043.065
S.E. of regression	227.6836	Akaike info criterion	13.83443
Sum squared resid	570238.1	Schwarz criterion	13.92134
Log likelihood	-87.92379	F-statistic	240.8493
Durbin-Watson stat	0.526471	Prob(F-statistic)	0.000000

Plot of residuals against year.





Cochrane-Orcutt Procedure

Dependent Variable: REVENUES
Method: Least Squares
Sample(adjusted): 2 13
Included observations: 12 after adjusting endpoints
Convergence achieved after 52 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MARKETING	7.019488	1.001113	7.011682	0.0001
C	-513.9911	428.6294	-1.199150	0.2611
AR(1)	0.730336	0.234503	3.114399	0.0124

R-squared	0.979561	Mean dependent var	1144.967
Adjusted R-squared	0.975020	S.D. dependent var	1045.423
S.E. of regression	165.2313	Akaike info criterion	13.26489
Sum squared resid	245712.5	Schwarz criterion	13.38611
Log likelihood	-76.58933	F-statistic	215.6720
Durbin-Watson stat	1.426942	Prob(F-statistic)	0.000000
Inverted AR Roots	.73		

Newey-West Standard Errors

Dependent Variable: REVENUES
Method: Least Squares
Sample: 1 13
Included observations: 13
White Heteroskedasticity-Consistent Standard Errors & Covariance

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MARKETING	6.169683	0.484878	12.72420	0.0000
C	-245.9496	88.11429	-2.791257	0.0175

R-squared	0.956323	Mean dependent var	1063.561
Adjusted R-squared	0.952352	S.D. dependent var	1043.065
S.E. of regression	227.6836	Akaike info criterion	13.83443
Sum squared resid	570238.1	Schwarz criterion	13.92134
Log likelihood	-87.92379	F-statistic	240.8493
Durbin-Watson stat	0.526471	Prob(F-statistic)	0.000000