

MULTICOLLINEARITY

Perfect multicollinearity:

$$\lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_k x_k = 0.$$

λ_i 's are not all simultaneously zero.

Partial collinearity:

$$\lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_k x_k + v_i = 0.$$

$v_i \rightarrow$ Stochastic error term.

Multicollinearity only refers linear relationships among the x variables. It does not rule out nonlinear relationships among them.

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + u_i$$

This type of model does not violate the assumption of no multicollinearity.

If multicollinearity is perfect, in the regression coefficients of the x variables are indeterminate & their standard errors are infinite. If there is partial multicollinearity, the regression coefficients, although determinate, possess large standard errors (in relation to the coefficients themselves) which means the coefficient cannot be estimated with great precision or accuracy.

Sources of multicollinearity

- 1) Sampling over a limited range of values
- 2) Constraints ~~on~~ on the model or in the population being sampled.
- 3) Adding polynomial terms to a regression model, especially when the range of the x variable is small.
- 4) More explanatory variables than the number of observations.

5) Common trend among explanatory variables.

Consequences of Multicollinearity

- 1) Even in the case of near multicollinearity, the OLS estimators are unbiased and BLUE, but the OLS estimators have large variances, making ~~less~~ precise estimation difficult.
- 2) Confidence intervals tend to be much wider, leading to the acceptance of the "zero null hypothesis" (i.e. the true population coefficient is zero) more readily.
- 3) t ratio of one or more coefficients tends to be statistically insignificant.
- 4) But, the R^2 of the overall measure of goodness of fit, can be very high.
- 5) The OLS estimators and their standard errors can be sensitive to small changes in the data.

$$\text{var}(\hat{\alpha}_2) = \frac{\sigma^2}{\sum x_{2i}^2 (1 - \rho_{12}^2)}$$

$$\text{var}(\hat{\alpha}_1) = \frac{\sigma^2}{\sum x_{1i}^2 (1 - \rho_{12}^2)} \quad \text{as } \rho_{12} \rightarrow 1 \quad \text{var}(\hat{\alpha}_i) \rightarrow \infty$$

$$\text{cov}(\hat{\alpha}_1, \hat{\alpha}_2) = \frac{-\rho_{12} \sigma^2}{(1 - \rho_{12}^2) \sqrt{\sum x_{1i}^2 \sum x_{2i}^2}}$$

$$\boxed{\text{VIF} = \frac{1}{(1 - \rho_{12}^2)}}$$

Variance-Inflating Factor
(VIF)

VIF shows how the variance of an estimator is inflated by the presence of multicollinearity.

As R_{ij}^2 approaches 1, the VIF approaches infinity. That is, as the extent of multicollinearity increases, the variance of an estimator increases, and in the limit it can become infinite.

$$\text{Var}(\hat{\alpha}_j) = \frac{\sigma^2}{\sum x_{ji}^2} \text{VIF}$$

$$\text{Var}(\hat{\alpha}_k) = \frac{\sigma^2}{\sum x_{ki}^2} \text{VIF}$$

$$\boxed{\text{Var}(\hat{\alpha}_j) = \frac{\sigma^2}{\sum x_{ji}^2} \left(\frac{1}{1 - R_j^2} \right)}$$

$\hat{\alpha}_j$ = (estimated) partial regression coefficient of regressor x_j

$R_j^2 = R^2$ in the regression of x_j on the remaining $(k-2)$ regressions

∴ The variance of the slope coefficients depend on 3 factors:

- 1) σ^2
- 2) VIF
- 3) $\sum x_{ji}^2$

The larger the variability in a regressor, the ~~smaller~~ smaller the variance of the coefficient of that regressor, assuming the other 2 components are constant, ∴ the greater the precision with which that coefficient can be estimated.

$$\boxed{TOL_j = \frac{1}{VIF_j} = (1 - R_j^2)} \quad \text{Tolerance.}$$

Threshold : 5

→ Threshold for σ^2 is 0.8 which is generally very high.

the fact that the F-test is significant but the t-values of the individual coefficients are not means that the two variables are so highly correlated that it is impossible to isolate the individual impact of either variables.

* Detection of Multicollinearity

- 1) Multicollinearity is a question of degree and not of kind.
- 2) Since multicollinearity refers to the condition of the explanatory variables that are assumed to be nonstochastic, it is a feature of the sample and not of the population.

Testing

- 1) High R^2 but few significant t-ratios.
- 2) High pair-wise correlations among regressors.
High zero-order correlations are a sufficient but not a necessary condition for the existence of multicollinearity because it can exist even though the zero-order or simple correlations are comparatively low.
- 3) Examination of partial correlations

Fairer-Glauber Test:

If $R^2_{1,2,3,4}$ is very high $r^2_{12,34}$, $r^2_{13,24}$ and $r^2_{14,23}$ are comparatively low may suggest that the variables x_2, x_3, x_4 are highly intercorrelated.

4) Auxiliary Regressions

$$F_i = \frac{R^2_{x_1, x_2, x_3, \dots, x_K} / (k-2)}{(1 - R^2_{x_1, x_2, \dots, x_K}) / (n-k+1)}$$

$R^2_{x_1, x_2, \dots, x_k}$ → coefficient of determination in the regression of variable x_i on the remaining X variables.

If the computed F exceeds the critical F_i at the chosen level of significance, it is taken to mean that the particular x_i is collinear with other X 's.

Klein's rule of thumb: Multicollinearity may be a troublesome problem only if the R^2 obtained from an auxiliary regression is greater than the overall R^2 , i.e., that obtained from the regression of Y on all the regressors.

Multicollinearity

$\text{var}(\hat{\alpha}_j)$ depends on 3 factors: σ^2 , $\sum x_j^2$ and VIF_j . A high ~~VIF~~ can be counterbalanced by a low σ^2 or a high $\sum x_j^2$. A high VIF is neither necessary nor sufficient to get high variances and high standard errors. High multicollinearity, as measured by a high VIF, may not necessarily cause high standard errors.

Dropping

Dropping a variable and specification bias

Specification bias arises from incorrect specification of the model used in the analysis.

Let the model be:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

And we mistakenly take the model

$$Y_i = b_1 + b_2 X_{2i} + \hat{u}_i$$

$$E(b_{12}) = \beta_2 + \beta_3 b_{32}$$

$b_{32} \rightarrow$ Slope coefficient in the regression of X_3 on X_2 .

b_{12} will be a biased estimate of β_2 as long as b_{32} is different from zero.

Comment on the following statements:

i) High pair-wise correlations ~~across~~ across the independent variables suggest that there is severe multicollinearity problem.

correlations

High zero-order ~~correlations~~ may suggest collinearity, it is not necessary that they be high to have collinearity in any specific case.

High zero-order correlations are a sufficient but not a necessary condition for the existence of multicollinearity because it can exist even though the zero-order or simple correlations are comparatively low.

Let the model be:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$$

And suppose:

$$X_{4i} = \lambda_2 X_{2i} + \lambda_3 X_{3i}$$

where λ_2 and λ_3 are constants, not both zero.

X_4 is an exact linear combination of X_2 and X_3 , giving

$R^2_{4.23} = 1 \rightarrow$ the coefficient of determination

in the regression of X_4 on X_2 and X_3 .

$$R^2_{4.23} = \frac{\sigma_{42}^2 + \sigma_{43}^2 - 2\sigma_{42}\sigma_{43}\sigma_{23}}{1 - \sigma_{23}^2}$$

$$R^2_{4.23} = 1 \Rightarrow \sigma_{42} = 0.5, \sigma_{43} = 0.5, \sigma_{23} = -0.5$$

which are not high values.

Therefore, in models involving more than two explanatory variables, the simple or zero-order correlation will not provide an infallible guide to the presence of multicollinearity. If there are only 2 explanatory variables, the zero-order correlations will suffice.

2) Running first difference regression model is a better approach as compared to ratio transformation to reduce the severity of multicollinearity.

AUTOCORRELATION

In cross-sectional studies, data are often collected on the basis of a random sample of cross-sectional units, such as households (in a consumption function analysis) or firms (in an investment study analysis) so that there is no prior reason to believe that the error term pertaining to one household or firm is correlated with the error term of another household or firm. If by chance such a correlation is observed in cross-sectional units, it is called spatial ~~autocorrelation~~ autocorrelation, i.e., correlation in space rather than over time.

However, in cross-sectional analysis, the ordering of the data must have some logic, or economic interest, to make sense of any determination of whether (spatial) autocorrelation is present or not.

In time series data, the observations in such data follow a natural ordering over time so that successive observations are likely to exhibit intercorrelations, especially ~~if~~ if the time interval between successive observations is short.

Under both heteroscedasticity and autocorrelation the usual OLS estimators, although linear, unbiased, and asymptotically (i.e. in large samples) normally distributed, are no longer minimum variance among all linear unbiased estimators. They are not efficient relative to other linear and unbiased estimators. They may not be best linear unbiased estimators (BLUE). As a result, the usual t , F and χ^2 may not be valid.

Definition

Correlation between members of series of observations ordered in time [as in time series data] or space [as in cross-sectional data].

$$E(u_i u_j) \neq 0 \quad i \neq j$$

* Reasons for autocorrelation

1) Inertia:

Time series such as GNP, price indexes, production, employment and unemployment exhibit business cycles. Starting at the bottom of the recession, when economic recovery starts, most of these series starts moving upward. In this upswing, the value of a series at one point in time is greater than its previous value. Thus there is a "momentum" built into them, and it continues until ~~somet~~ something happens. (e.g. increase in interest rate or taxes or both) to slow them down.

2) Specification Bias: Excluded variable case.

The residuals may suggest that some variables that were originally candidates but were not included in the model for a variety of reasons should be included. This is the case of excluded variable specification bias.

Example:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t \quad \text{--- (1)}$$

Y = quantity of beef demanded.

X_2 = price of beef.

X_3 = consumer income

X_4 = price of pork.

t = time.

If we run the regression

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + v_t \quad \text{--- (2)}$$

$$v_t = \beta_4 X_{4t} + u_t$$

and to the extent the price of pork affects the consumption of beef, the error or disturbance term v will reflect a systematic pattern, thus creating (false) autocorrelation.

3) Specification Bias: Incorrect Functional Form

True Model:

$$\text{Marginal cost}_i = \beta_1 + \beta_2 \text{output}_i + \beta_3 \text{output}_i^2 + u_i$$

If we run:

$$\text{Marginal cost}_i = \alpha_1 + \alpha_2 \text{output}_i + v_i$$

$$v_i = \beta_3 \text{output}_i^2 + u_i \quad (\text{Autocorrelation})$$

4) Cobweb Phenomena

The supply of many agricultural ~~commodities~~ commodities reflects the cobweb-phenomenon, where supply reacts to price with a lag of 1 time period, because supply decisions take time to implement (the gestation period).

$$\text{Supply}_t = \beta_1 + \beta_2 P_{t-1} + u_t$$

In this situation, the disturbances u_t are not expected to be random because if the farmers over produce in year t , they are likely to ~~produce~~ & reduce their production in $t+1$, and so on, leading to a cobweb pattern.

5) Lags.

$$\text{Consumption}_t = \beta_1 + \beta_2 \text{Income} + \beta_3 \text{consumption}_{t-1} + u_t$$

→ Autoregression

Consumers do not change their consumption habits readily

for psychological, technological or institutional reasons.
 If we neglect the lagged term, the resulting error term will reflect a systematic pattern due to the influence of lagged consumption on current consumption.

6) ~~Data~~ Manipulation of Data

Averaging of time series introduces smoothness into the data by ~~damping~~ dampening the fluctuations in the monthly data. This smoothness may itself lead to a systematic pattern in the disturbances, thereby introducing autocorrelation.

7) Data Transformation

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + u_{t-1}$$

$$\Delta Y_t = \beta_2 \Delta X_t + \Delta u_t$$

$$\Delta Y_t = \beta_2 \Delta X_t + v_t \quad \{ v_t = u_t - u_{t-1} \}$$

$$\Delta Y_t = \beta_2 \Delta X_t + v_t \Rightarrow \text{First difference form.}$$

8) Nonstationarity

If Y or X or both are nonstationary, the error term u may also be nonstationary. In that case, the error term will exhibit autocorrelation.

*OLS Estimation in the Presence of autocorrelation

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

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

ρ = coefficient of autocovariance.

$$E(\varepsilon_t) = 0$$

$$\text{var}(\varepsilon_t) = \sigma_\varepsilon^2$$

$$\text{cov}(\varepsilon_t, \varepsilon_{t+s}) = 0, \quad s \neq 0$$

} white noise
error term.

* \rightarrow Markov first-order autoregressive scheme.
AR(1).

\Rightarrow first-order coefficient of autocorrelation
coefficient of autocorrelation at lag 1.

* OLS Estimation Disregarding Autocorrelation

$$\text{var}(\hat{\beta}) = \frac{\sigma^2}{\sum x_t^2}$$

1) Residual variance

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_t^2}{n-2} \rightarrow \text{underestimates the true } \sigma^2$$

2) Overestimation of R^2 .

Selecting Autocorrelation

1) Graphical Method.

(i) Plotting \hat{u}_t against time(t).

(ii) Estimate the model $y_t = \alpha + \beta x_t + u_t$

(iii) Estimate \hat{u}_t from the estimated model.

(iv) Plotting \hat{u}_t against time and understanding the patterns of the line.

(v) If the line shows changes in the sign of the residual (\hat{u}_t) only for a few number of times, we suspect positive autocorrelation.

(vi) If the line shows changes in sign of the residual (\hat{u}_t) too many times, we suspect negative autocorrelation.

(vii) If the line shows changes in sign of the residual (\hat{u}_t) for moderate number of times, we suspect no autocorrelation problem.

Do we need standardization of the residuals for the plotting \hat{u}_t against time (t)?

The standardized residuals are $\frac{\hat{u}_t}{\hat{\sigma}}$.

\hat{u}_t & $\hat{\sigma}$ are measured in the units of Y .

The values of the standardized residuals will therefore be pure numbers (devoid of units of measurement) ~~so~~ and can be compared with the standardized residuals of other regressions. The standardized residuals, like \hat{u}_t , have zero mean and approximately unit variance.

(b) Plotting \hat{u}_t against \hat{u}_{t-1} .

(i) Estimate the model $y_t = \alpha + \beta X_t + u_t$.

(ii) Estimate \hat{u}_t from the estimated model.

(iii) Arrange the second to last observations as \hat{u}_t and the first to the second last observations as \hat{u}_{t-1} . (Hence, we have $n-1$ observations now)

(iv) Plotting \hat{u}_t against \hat{u}_{t-1} and understanding the pattern of the scatter plot.

(v) If the scatter plot shows increasing (decreasing) pattern we suspect positive (negative) autocorrelation.

(vi) If the scatter plot does not show any pattern, we expect no autocorrelation problem.

2) Runs Test

Run - An uninterrupted sequence of one symbol or attribute such as + or -.

Length of a run is the number of elements in it.

$$N = \text{total number of observations} \\ = N_1 + N_2.$$

N_1 = number of +ve residuals

N_2 = number of -ve residuals

R = number of runs.

* H_0 : Successive residuals are independent.

$N_1 > 10, N_2 > 10 \Rightarrow$ Number of runs is (asymptotically) normally distributed

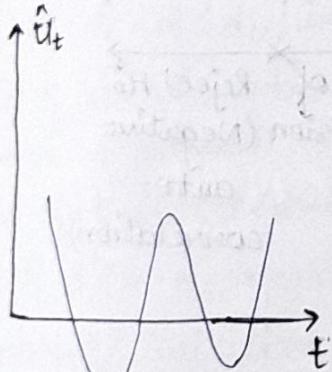
$$\text{Mean : } E(R) = \frac{2N_1 N_2}{N} + 1$$

$$\text{Variance : } \sigma_R^2 = \frac{2N_1 N_2 (2N_1 N_2 - N)}{N^2 (N-1)}$$

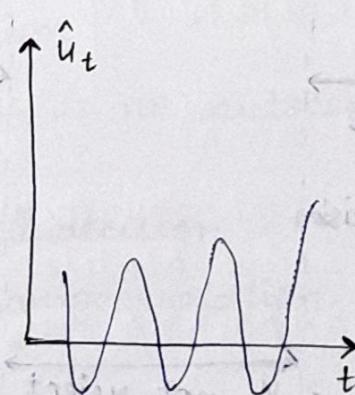
$$\text{Prob}[E(R) - 1.96\sigma_R \leq R \leq E(R) + 1.96\sigma_R] = 0.95$$

Do not reject the null hypothesis of randomness with 95% confidence if R , the number of runs, lies in the preceding confidence interval; reject the null hypothesis if the estimated R lies ~~not~~ outside these limits.

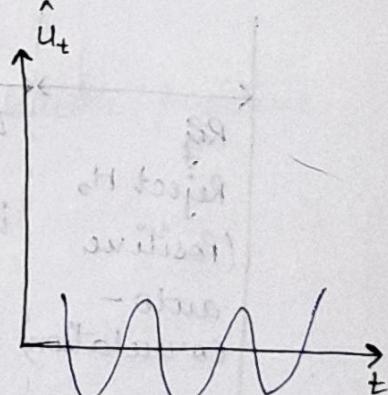
If there is positive autocorrelation, the number of runs will be few (lies at the left of confidence interval). whereas if there is negative autocorrelation, the number of runs will be many (lies at the right of confidence interval).



Positive autocorrelation



Negative autocorrelation



No autocorrelation

Standardization of \hat{u}_t would not affect the runs test.

$\frac{\hat{u}_t - \bar{u}}{\sigma_{\hat{u}}} \rightarrow^0$ (Only sign of \hat{u}_t is used which does not change even after standardization)

3) Durbin-Watson d Test

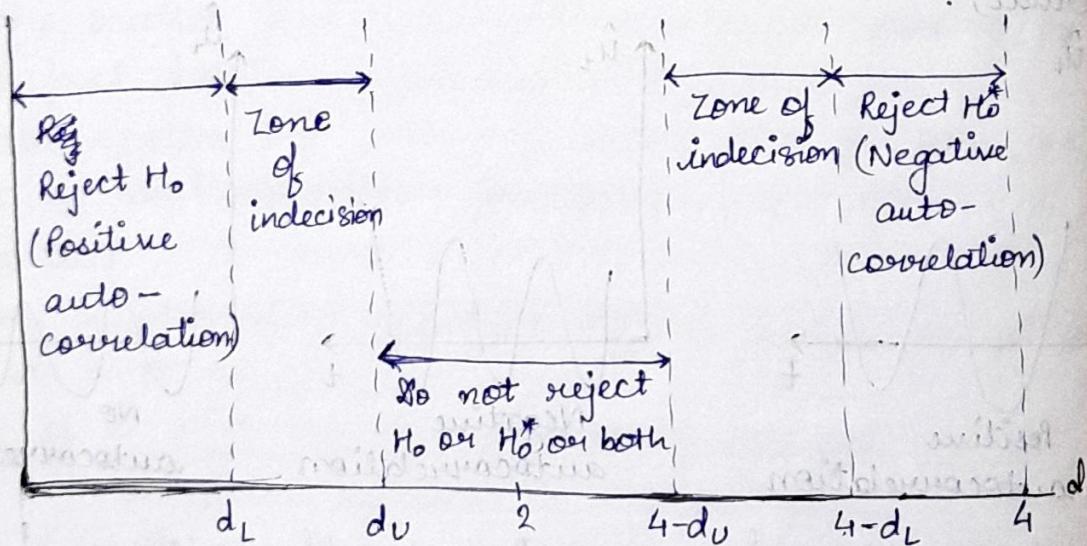
$$d = \frac{\sum_{t=2}^n (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^n \hat{u}_t^2}$$

Assumptions:

- (i) There is intercept in the model.
- (ii) The independent variables are non-stochastic.
- (iii) The random disturbance term follows first-order autoregressive scheme, i.e.,

$$u_t = \rho u_{t-1} + v_t; |\rho| < 1$$

- (iv) The random disturbance term follows normal distribution.
- (v) The model is not autoregressive, i.e., it does not include lagged value(s) of the dependent variable as independent variable(s).
- (vi) There are no missing observations in the data.



H_0 : No positive autocorrelation

H_0^* : No negative autocorrelation

for large sample,

$$\sum_{t=2}^n \hat{u}_t^2 \approx \sum_{t=2}^n \hat{u}_{t-1}^2 \approx \sum_{t=1}^n \hat{u}_t^2$$

$$d \approx \frac{2 \left(\sum_{t=2}^n \hat{u}_t^2 - \sum_{t=2}^n \hat{u}_t \hat{u}_{t-1} \right)}{\sum_{t=1}^n \hat{u}_t^2} \approx 2(1 - \hat{\rho})$$

$\hat{\rho} = \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2} \rightarrow$ sample first-order coefficient
of autocorrelation,
(estimator of ρ).

$$-1 \leq \rho \leq 1 \quad 0 \leq d \leq 4.$$

- (i) When $\hat{\rho} = 1$ (perfect positive autocorrelation), $d = 0$.
- (ii) When $\hat{\rho} = -1$ (perfect negative autocorrelation), $d = 4$.
- (iii) When $\hat{\rho} = 0$ (no autocorrelation), $d = 2$.

Decision Rules:

- (i) If $d < d_L$, positive autocorrelation.
- (ii) If $d_L < d < d_U$, indecision
- (iii) If $d_U < d < 4 - d_U$, there is no autocorrelation (positive or negative)
- (iv) If $4 - d_U < d < 4 - d_L$, indecision
- (v) If $d > 4 - d_L$, negative autocorrelation.

Concerns/Limitations

- (i) The test cannot be applied if the model does not have intercept.
- (ii) The test cannot be applied if the model is ~~not~~ autoregressive.
- (iii) The test can be applied only for the first-order autoregressive scheme.

(iv) The test requires the independent variables to be non-stochastic.

(v) The test result is inconclusive if the computed value of d lies in the indecisive zone.

d_U is approximately the true significance limit.

Modified d test:

Given the level of significance α :

1. $H_0: \rho = 0$ v/s $H_1: \rho > 0$. Reject H_0 at α level if $d < d_U$.

Statistically significant positive autocorrelation.

2. $H_0: \rho = 0$ v/s $H_1: \rho < 0$. Reject H_0 at α level if the estimated $(4-d) < d_U$, i.e., statistically significant negative autocorrelation.

3. $H_0: \rho = 0$ v/s $H_1: \rho \neq 0$. Reject H_0 at 2α level if $d < d_U$ or $(4-d) < d_U$, i.e. statistically significant autocorrelation, positive or negative.

If sample size is large (technically infinite), then

$$\sqrt{n} \left(1 - \frac{1}{2} d \right) \approx N(0, 1)$$

$$\sqrt{n} \hat{\rho} \sim N(0, 1)$$

4) Durbin-Watson 'h' Test for Autocorrelation - For first-order Autoregressive Models.

$$\text{Model: } Y_t = \alpha + \beta X_t + \gamma Y_{t-1} + u_t$$

H_0 : There is no autocorrelation
(i.e. autocorrelation coefficient is zero).

Test statistic: $\Theta = \hat{\gamma} \sqrt{\frac{n}{1 - n \text{vari}(\hat{\gamma})}}$

* $\hat{\gamma}$ \Rightarrow Estimated coefficient of lagged dependent variable.

for large sample,

$$\epsilon \sim N(0, 1)$$

Rejection of null hypothesis indicates presence of autocorrelation problem.

Concerns/limitations

- (i) The denominator of the test statistic, $1 - n[\text{var}(\hat{\gamma})]$ may be negative.
- (ii) The sample size may not be sufficiently large.
- (iii) The test can be applied only for the first-order autoregressive model.

5) The Breusch-Godfrey (BG) Test

It allows for

- (i) nonstochastic regressors, such as the lagged values of the regressand.
- (ii) higher-order autoregressive schemes, such as AR(1), AR(2).
- (iii) Simple or higher-order moving averages of white noise error terms, such as ϵ_t .

$$u_t = \epsilon_t + \lambda_1 \epsilon_{t-1} + \lambda_2 \epsilon_{t-2}$$

$$\text{Model: } y_t = \beta_1 + \beta_2 x_t + u_t$$

$$\text{Assumption: AR}(p): u_t = \beta_1 u_{t-1} + \beta_2 u_{t-2} + \dots + \beta_p u_{t-p} + \nu_t$$

$$H_0: \beta_1 = \beta_2 = \beta_3 = \dots = \beta_p = 0 \quad (\text{i.e. no autocorrelation of any order}).$$

(i) Estimate the model $y_t = \alpha + \beta_1 x_t + \beta_2 x_t + u_t$.

(ii) Estimate \hat{u}_t .

(iii) Regress \hat{u}_t on the lagged \hat{u}_t (i.e. $\hat{u}_{t-1}, \hat{u}_{t-2}, \dots, \hat{u}_{t-p}$), and the original independent variables (with $n-p$ observations).

$$\hat{u}_t = \alpha_0 + \alpha_1 x_t + \lambda_1 \hat{u}_{t-1} + \lambda_2 \hat{u}_{t-2} + \dots + \lambda_p \hat{u}_{t-p} + \varepsilon_t$$

(iv) for large sample, the test statistic

$$[D = (n-p)R^2 \sim \chi_p^2]$$

Rejection of the null hypothesis indicates presence of autocorrelation of some order.

A drawback of the BG test is that the value of p , the length of the lag, cannot be specified apriori.

* Model Mis-Specification v/s Pure Autocorrelation

How to distinguish between pure autocorrelation and specification errors?

If the data is time-series, then it is possible that y & x exhibit trends.

Example:

$y \rightarrow$ wages $x \rightarrow$ productivity

$$\hat{Y}_t = \alpha + \beta_1 X_t + \beta_2 t + u_t$$

Even after allowing for the trend variable, if the d value still suggests autocorrelation, then it may be pure correlation.

We could also test for different functional form such as regressing y_t on x_t & x_t^2 .

Even after this, if the d statistic suggests autocorrelation then it may be pure autocorrelation.

• First-Difference Model:

If the autocorrelation coefficient (ρ) is not significantly different from 1 (one).

Assumption: $\rho = 1$ or $u_t = u_{t-1} + v_t$

(v_t satisfies all the usual assumptions of classical linear regression model).

$$Y_t = \alpha + \beta X_t + u_t \quad \text{--- (1)}$$

$$Y_{t-1} = \alpha + \beta X_{t-1} + u_{t-1} \quad \text{--- (2)}$$

$$(Y_t - Y_{t-1}) = \beta(X_t - X_{t-1}) + (u_t - u_{t-1})$$

$$\Delta Y_t = \beta \Delta X_t + \Delta u_t$$

$$\Delta Y_t = \beta \Delta X_t + v_t$$

estimate the model $Y_t = \alpha + \beta X_t + u_t$.

estimate \hat{u}_t and test for autocorrelation.

If there is autocorrelation, test the null hypothesis

$$H_0: \rho = 1$$

Brennblutt - Webb Test

Test statistic:

$$\Theta = \frac{\sum_{t=2}^n \hat{v}_t^2}{\sum_{t=1}^n \hat{u}_t^2}$$

$\hat{u}_t \rightarrow$ Residual of original model

$\hat{v}_t \rightarrow$ Residual of first difference model.

If $\Theta < d_L$ with the given number of observations and the number of explanatory variables and at the chosen level of significance, the null hypothesis $H_0: \rho = 1$ is not rejected and the first difference model is estimated.

Generate the variables $\Delta Y_t = Y_t - Y_{t-1}$, $\Delta X_t = X_t - X_{t-1}$

Estimate the model: $\Delta Y_t = \beta \Delta X_t + v_t$ with $n-1$ observations and test ~~for~~ for autocorrelation of v_t to confirm that the new model does not have autocorrelation problem.
Issue of intercept in the first difference model.

Model with trend component: $Y_t = \alpha + \beta X_t + \gamma t + u_t$.

$$Y_{t-1} = \alpha + \beta X_{t-1} + \gamma(t-1) + u_{t-1}$$

$$\Delta Y_t = \beta \Delta X_t + \gamma + \Delta u_t \Rightarrow \Delta Y_t = \gamma + \beta \Delta X_t + v_t$$

Thus, inclusion of trend component brings in intercept into the first-difference model. Inclusion of the trend component may also help in resolving/reducing autocorrelation problem.

σ^2 value drops when first difference model is considered as it represents the behaviour of variables around their (linear) trend values.

If $f = 1$

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

$$(u_t - u_{t-1}) = \varepsilon_t$$

∴ Taking first-order model will make the error term stationary, for it is equal to ε_t , which is a white noise error term.

• Generalized Difference Method:

If the autocorrelation coefficient is positive and significantly different from 1 (one).

Assumption: $u_t = fu_{t-1} + v_t$; $0 < f < 1$.

v_t satisfies all the usual assumptions of CLRM.

$$Y_t = \alpha + \beta X_t + u_t$$

~~$$fY_t = \alpha f + f\beta X_t + fu_t$$~~

$$fY_{t-1} = \alpha f + \beta f X_{t-1} + fu_{t-1}$$

$$Y_t - fY_{t-1} = \alpha(1-f) + \beta(X_t - fX_{t-1}) + u_t - fu_{t-1}$$

$$Y_t^* = Y_t - fY_{t-1}, \text{ or } \boxed{Y_t^* = \alpha^* + \beta^* X_t^* + v_t}$$

$$Y_t^* = Y_t - fY_{t-1}$$

$$X_t^* = X_t - fX_{t-1}$$

Estimate the model $y_t = \alpha + \beta x_t + u_t$

Estimate \hat{u}_t and test for autocorrelation.

If there is autocorrelation, get the estimated value of ρ (autocorrelation coefficient) as

$$\hat{\rho} \approx \left(1 - \frac{d}{2}\right) \text{ for large sample.}$$

Generate the variables $y_t^* = (y_t - \hat{\rho} y_{t-1})$

$$x_t^* = (x_t - \hat{\rho} x_{t-1})$$

Estimate the model $y_t^* = \alpha^* + \beta x_t^* + v_t$ with $n-1$ observations and test for autocorrelation of v_t to confirm that the new model does not have autocorrelation problem.

Prais-Winsten Transformation

The following transformation can be used for the first observation:

$$y_1^* = y_1 \sqrt{1 - \hat{\rho}^2} \quad x_1^* = x_1 \sqrt{1 - \hat{\rho}^2}$$

Estimated value of ρ for small sample:

Thiel-Nagar Estimation:

Estimated value of ρ (autocorrelation coefficient) for small sample:

$$\hat{\rho} = \frac{n^2 \left(1 - \frac{d}{2}\right) + k^2}{n^2 - k^2}$$

n = number of observations

d = value of the Durbin-Watson statistic

k = number of coefficients (including the intercept).

$$\hat{\rho} = \frac{\left(1 - \frac{d}{2}\right) + \frac{k^2}{n^2}}{1 - \frac{k^2}{n^2}}$$

$$n \rightarrow \infty, \hat{f} = \frac{n^2 \left(1 - \frac{d}{2}\right) + k^2}{n^2 - k^2} \rightarrow 1 - \frac{d}{2}$$

Alternative estimate of f

Estimation of the model: $y_t = \alpha + \beta x_t + u_t$

$$\hat{u}_t = f \hat{u}_{t-1} + v_t$$

and get the estimated value of f .

Cochrane - Orcutt Iterative Process

Estimate the model $y_t = \alpha + \beta x_t + u_t$.

Use the residuals from the regression to run the regression $\hat{u}_t = f \hat{u}_{t-1} + v_t$ and estimate f .

Estimate the generalized difference equation

$$y_t^* = \alpha^* + \beta x_t^* + v_t$$

$$y_t^* = y_t - y_{t-1}, x_t^* = x_t - \hat{x}_{t-1}$$

and re-estimate the coefficients $\hat{\alpha}$ & $\hat{\beta}$.

Estimate the new residuals using \hat{u}_t using $\hat{\alpha}$ & $\hat{\beta}$.

Re-estimate f i.e. \hat{f} using the new-residuals.

Estimate generalized difference equation and get new estimates of α & β .

Continue the process until the successive estimates for f differ only very marginally (< 0.005).

Hildreth - Lu search Procedure

Consider the first-order autoregressive scheme:

$$u_t = f u_{t-1} + v_t; |f| < 1$$

Select values of f b/w $+1$ & -1 with an interval of 0.1 (say) and estimate the generalized difference models for each value of f .

Obtain the RSS in each case.

Select that value of ρ or the model for which RSS is minimum.

The chosen value of ρ and hence the final estimates of the coefficients may differ depending on the interval taken.

~~Estimation of the Model~~

Estimation of the Model:

$$Y_t = \alpha(1-\rho) + \beta X_t - \rho \beta X_{t-1} + \rho Y_{t-1} + u_t$$

Estimate the above model.

estimate the generalized difference model using the coefficient of Y_{t-1} as an estimate of ρ .