Multicollinearity

Exact Multicollinearity (1)

What if $[\mathbf{X}^T\mathbf{X}]^{-1}$ does not exist or $\det[\mathbf{X}^T\mathbf{X}]$ is zero? β cannot be estimated.

It is a data problem not the technique. Remember that **X** has been pre-selected.

If it really happens, get a new data set. Nothing else can be done.

Exact multicollinearity is a rare event. Easy to "fix".

Exact Multicollinearity (2)

How could Exact multicollinearity happen?

A column in **X** is exactly a multiple of one column, i.e., $X_{2i} = 2X_{3i}, \forall i = 1,..n$

or a column in **X** is equal to a linear combination of one or more other columns in **X**, i.e.,

$$X_{2i} = 2X_{3i} - X_{4i} + 2.7X_{5i} - 1.2X_{6i}, \forall i = 1,..n$$

==> Rank(**X**) <*K*.

Near Multicollinearity (1)

Real problem is the Near Multicollinearity.

How could it happen?

A column in **X** is almost a multiple of one column or almost equal to a linear combination of one or more other columns in **X**

=>Rank(X) is still *K* but det[$\mathbf{X}^T\mathbf{X}$] is almost zero.

Near Multicollinearity (2)

<u>Results</u>

Estimator of β's involved in the problem will have a very large standard error (<u>bad</u> accuracy) even though the hypothesis that these β's are zero has been rejected using an F-test with very high confidence (at very low significant level).

==> accept that the coeff. is zero but in fact it is not.

Near Multicollinearity (3)

For example, given that

$$X_{2i} \approx 2X_{3i} - X_{4i} + 2.7X_{5i} - 1.2X_{6i}, \forall i = 1,..n$$

it is likely that s.e.'s of $\hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5, \hat{\beta}_6$

are large. t-tests will accept that each parameter is zero but the F-test will reject

$$H_0 : \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = 0$$

$$H_1 : \beta_2 \neq \beta_3 \neq \beta_4 \neq \beta_5 \neq \beta_6 \neq 0$$

Near Multicollinearity (4)

<u>Symptoms</u>

1) low t-stat for at least 2 β's but high R^2.

2) high correlation among X's

3) non-robust of t-stat when removing variables. t-stat changes in sign or magnitude or both.

Symptoms do not guarantee the existence or pattern of multicollinearity.

Near Multicollinearity (5)

<u>Detection</u> (identification)

1) Ratio of max. and min. eigenvalues of $\mathbf{X}^T\mathbf{X}$ is greater than 100

2) Auxiliary regression among X's has R^2 greater than 0.9

Near Multicollinearity (6)

Auxiliary Regressions

$$X_{1i} = \gamma_{12} X_{2i} + \gamma_{13} X_{3i} + \ldots + \gamma_{1K} X_{Ki} + \varepsilon_{1i}$$

$$X_{2i} = \gamma_{21} X_{1i} + \gamma_{23} X_{3i} + \ldots + \gamma_{2K} X_{Ki} + \varepsilon_{2i}$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$X_{Ki} = \gamma_{K1} X_{1i} + \gamma_{K3} X_{3i} + \ldots + \gamma_{K,K-1} X_{K-1,i} + \varepsilon_{K-1,i}$$

High R^2 implies that the X on the left hand side is almost a linear combination of the remaining X's on the right-hand side.

Near Multicollinearity (7)

<u>Remedies</u>

1) neglect if no specific coefficient is of interest. Even with multicollinearity, no problem in prediction of Y. Why?

   Note that OLS is still BLUE.

2) get more data or get a new data set

3) change model, i.e., use log-log model

Near Multicollinearity (8)

Remedies (cont’d)

4) drop variables that cause the problem. Drop one-by-one the dependent X’s that yield the highest auxiliary R^2 until no high auxiliary R^2. However, dropping a variable will create another problem. For example, the new estimator will be biased and inconsistent. See Model specification.

Near Multicollinearity (9)

Remedies (cont’d)

5) If all the variables have to be maintained in the model, no chance to improve on OLS. Try other methods, such as, Factor Analysis, Determinant Analysis, LISREL

Polynomial model (1)

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3$$
$$+ \ldots + \beta_P X_i^P + \varepsilon_{1i}$$

It is very easy to have a multicollinearity in this model due to high correlation between the independent variables because they are all related to the same variables

Polynomial model (2)

To <u>alleviate</u> the problem, the polynomial model should be re-written

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3$$
$$+ \ldots + \beta_P x_i^P + \varepsilon_{1i}$$

where $x_i = X_i - \overline{X}$

Heteroscedasticity

Covered Topics

- What is Heteroscedasticity?
- Weighted Least Square (WLS)
- Tests for Variance equation
- Remedies
- Generalized Heteroscedasticity

What is Heteroscedasticity?

Violation of constant variance of ε_i's but they are still independent.

$$V(\varepsilon_i) = \sigma_i^2$$

$$COV(\varepsilon_i, \varepsilon_j) = 0, \forall i \neq j$$

$$\mathbf{V}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_n^2 \end{bmatrix}$$

The error term (ε) is said to be heteroscedastic.

Weighted Least Square (1)

What happened to OLS estimators?

==> OLS is still LUE but not BLUE.

==> Large sample properties ???

Weighted Least Square (WLS)

$$w_i Y_i = \beta_1 w_i X_{1i} + \beta_2 w_i X_{2i} + \ldots + \beta_K w_i X_{Ki} + w_i \varepsilon_i$$

$$\underline{Y}_i = \beta_1 \underline{X}_{1i} + \beta_2 \underline{X}_{2i} + \ldots + \beta_K \underline{X}_{Ki} + \nu_i$$

where

$$w_i = \frac{1}{\sigma_i} \quad \underline{Y}_i = w_i Y_i \quad \underline{X}_{ki} = w_i X_{ki}, k = 1, \ldots, K \quad \nu_i = w_i \varepsilon_i$$

Weighted Least Square (2)

Matrix form

$$\underline{\mathbf{Y}} = \mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{Y},\ \underline{\mathbf{X}} = \mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{X}, \boldsymbol{\nu} = \mathbf{\Sigma}^{-\frac{1}{2}}\boldsymbol{\varepsilon}$$

$$\mathbf{\Sigma}^{-\frac{1}{2}} = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{\sigma_n} \end{bmatrix}$$

Weighted Least Square (3)

Note that $\underline{\mathbf{Y}} = \underline{\mathbf{X}}\boldsymbol{\beta} + \underline{\boldsymbol{\varepsilon}}$ is CLNRM and $\boldsymbol{\nu}$ is homoscedastic.

$$\mathrm{V}(\nu_i) = 1, \forall i = 1,\ldots,n$$

$$\mathbf{V}(\boldsymbol{\nu}) = \mathbf{I}_n$$

WLS estimator $\hat{\boldsymbol{\beta}} = \left[\underline{\mathbf{X}}^{\mathrm{T}}\underline{\mathbf{X}}\right]^{-1}\underline{\mathbf{X}}^{\mathrm{T}}\underline{\mathbf{Y}}$

$$\mathrm{V}(\hat{\boldsymbol{\beta}}) = \left[\underline{\mathbf{X}}^{\mathrm{T}}\underline{\mathbf{X}}\right]^{-1}$$

Note that, given Σ, the estimator $\hat{\boldsymbol{\beta}}$ is BLUE.

Variance Equation (1)

Explanatory variables

- observation index *i*
- some or all the **X**'s
- conditional mean of Y given X
- variable *Z*'s not in the model
- lagged variables (to be discussed in Time Series Analysis)

Variance Equation (2)

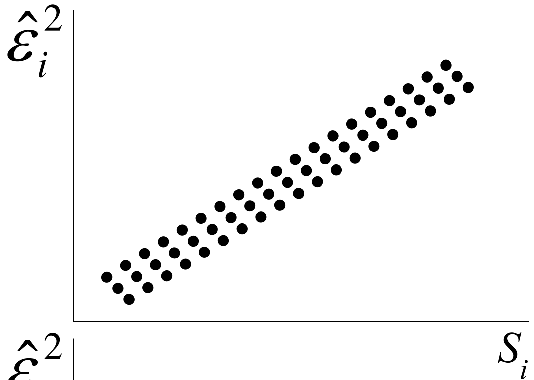
Detection

- squared residual plots
- White's Test
- Other LM tests

Squared Residual Plots (1)

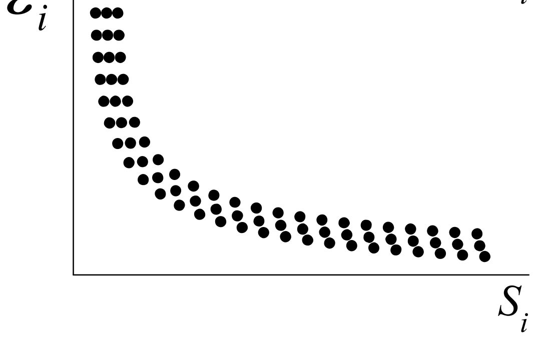
- squared residuals against the suspected explanatory variable (S)
- test against only one variable at a time
- functional forms, e.g., linear, log-lin, log-log, quadratic, polynomial, reciprocal, etc.

Squared Residual Plots (2)



Linear

$$V(\varepsilon_i) = \gamma_1 + \gamma_2 S_i$$



Reciprocal

$$V(\varepsilon_i) = \gamma_1 + \gamma_2 \frac{1}{S_i}$$

or log-lin

$$\ln(V(\varepsilon_i)) = \gamma_1 + \gamma_2 S_i$$

Tests for Heteroscedasticity (1)

Functional form of the variance equation must be assumed in most tests. For example,

- Breusch-Pagan's test assumes a linear function of suspected variables
- Glesjer's test sets the form of standard deviation σ_i instead of the variance
- **<u>White</u>**'s test adds squared and cross terms to the variance equation of BP test.

Tests for Heteroscedasticity (2)

- **<u>Harvey-Godfrey</u>**'s test assumes a log-lin or

  $$V(\varepsilon_i) = \exp(\gamma_1 + \gamma_2 S_{2i} + \gamma_3 S_{3i} + .. + \gamma_P S_{Pi})$$

  Note that HG assures positive fitted variances while others do not.
- Park test assumes double log form.
- Goldfeld-Quandt does not assume the form of the variance function. Instead, it checks for equality of the variances between the high group and the low group using variance ratio test(F-test). See text.

White's Test (1)

White's General Heteroscedasticity Test

<u>Concept</u>: The variables that are suspected to cause heteroscedasticity are the X's, their squared terms and their cross terms.

$$
\begin{aligned}
\mathrm{V}(\varepsilon_i) = {} & \gamma_1 + \gamma_2 X_{2i} + .. + \gamma_K X_{Ki} \\
& + \gamma_{22} X_{2i}^2 + \gamma_{23} X_{2i} X_{3i} + .. + \gamma_{2K} X_{2i} X_{Ki} \\
& \qquad + \gamma_{33} X_{3i}^2 + \gamma_{34} X_{3i} X_{4i} + .. + \gamma_{3K} X_{3i} X_{Ki} \\
& \qquad\qquad \ddots \\
& \qquad\qquad + \gamma_{KK} X_{Ki}^2 + \varepsilon_i
\end{aligned}
$$

White's Test (2)

Run OLS on the auxiliary regression of the squared residuals and all the suspected variables above. Assume $X_{1i} = 0$ for all *i*.

Perform the overall F-test or LM test for the auxiliary regression run on the following hypothesis.

$$H_0 : \gamma_2 = \gamma_3 = ... = \gamma_K = \gamma_{22} = \gamma_{23} = ... = \gamma_{2K} = ... = \gamma_{KK} = 0$$

$$H_1 : \gamma_2 \neq \gamma_3 \neq ... \neq \gamma_K \neq \gamma_{22} \neq \gamma_{23} \neq ... \neq \gamma_{2K} \neq .... \neq \gamma_{KK} \neq 0$$

EViews also provide a quick solution for White's test.

White's Test (3)

Accept H_0 ==> No heteroscedasticity

If reject H_0, no specific remedy can be done.

Without any remedy, $\widehat{V(\hat{\boldsymbol{\beta}})}$ must be re-calculated since the old one is inconsistent (wrong). White's Heteroscedasticity Consistent Covariance is

$$\widehat{V(\hat{\boldsymbol{\beta}})} = \frac{n}{n-k}\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}\left[\sum_{i=1}^{n}\hat{\varepsilon}_i^2\mathbf{X}_i^{\mathrm{T}}\mathbf{X}_i\right]\left[\mathbf{X}^{\mathrm{T}}\mathbf{X}\right]^{-1}$$

where X_i is the i^{th} observation of X.

Generalized Remedies (1)

<u>F</u>easible <u>G</u>eneralized <u>L</u>east <u>S</u>quares (FGLS)

- Estimate the variance equation using auxiliary regression of squared residuals on the explanatory variables (for the variance equation), e.g.,

$$\hat{\varepsilon}_i^2 = V(\varepsilon_i) + \xi_i = \gamma_1 + \gamma_2\frac{1}{S_i} + \xi_i$$

where ξ_i is the error term for the variance equation

Generalized Remedies (2)

- Calculate the fitted value of squared residuals or estimate V(ε_i)

$$\widehat{\mathrm{V}(\varepsilon_i)} = \widehat{\hat{\varepsilon}_i^2} = \hat{\gamma}_1 + \hat{\gamma}_2 \frac{1}{S_i}$$

- Use the weight $w_i = \dfrac{1}{\sqrt{\widehat{\mathrm{V}(\varepsilon_i)}}}$ in WLS

- FGLS is biased but consistent.

Special Cases (1)

Variance equation with single parameter, e.g.,

Example 1 $\mathrm{V}(\varepsilon_i) = \sigma^2 \dfrac{1}{S_i}$ or $\hat{\varepsilon}_i^2 = \sigma^2 \dfrac{1}{S_i} + \xi_i$

where σ^2 is a positive constant.

Apply WLS with $w_i = \sqrt{S_i}$

Note that V(v_i)=σ^2 for all i=1,…,n

==> WLS estimator is unbiased. No FGLS needed

Special Cases (2)

Example 2 $\quad \mathrm{V}(\varepsilon_i) = \sigma^2 (S_i)^2$

$$\hat{\varepsilon}_i^2 = \sigma^2 (S_i)^2 + \xi_i$$

where σ^2 is a positive constant.

Apply WLS with $w_i = 1/S_i$

Note that V(V_i)=σ^2 for all i=1,…,n

==> WLS estimator is unbiased. No FGLS needed.

What is Generalized Heteroscedasticity?

Violation of constant variance and/or independence of ε_i's

$$\mathrm{V}(\varepsilon_i) = \sigma_i^2$$

$$\boxed{\mathrm{COV}(\varepsilon_i, \varepsilon_j) = \sigma_{ij} \neq 0 \text{ for some } i, j}$$

$$\mathbf{V}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \sigma_{n-1,n} \\ \sigma_{1n} & \cdots & \sigma_{n-1,n} & \sigma_n^2 \end{bmatrix}$$

The error term (ε) is said to be generalized heteroscedastic.

Generalized Least Square (1)

OLS or WLS estimators are still LUE but not BLUE.

Generalized Least Square (GLS)

$$\underline{\mathbf{Y}} = \mathbf{WY}, \quad \underline{\mathbf{X}} = \mathbf{WX}, \boldsymbol{\nu} = \mathbf{W}\boldsymbol{\varepsilon}$$

where $\mathbf{W}$ is a nxn symmetric matrix such that

$$\boldsymbol{\Sigma}^{-1} = \mathbf{WW}$$

Generalized Least Square (2)

Note that $\underline{\mathbf{Y}} = \underline{\mathbf{X}}\boldsymbol{\beta} + \boldsymbol{\nu}$ is CLNRM and

$\boldsymbol{\nu}$ is homoscedastic. $\mathrm{V}(\nu_i) = 1, \forall i = 1, \ldots, n$

$$\mathbf{V}(\boldsymbol{\nu}) = \mathbf{I}_n$$

GLS estimator $\hat{\boldsymbol{\beta}} = \left[\underline{\mathbf{X}}^{\mathrm{T}}\underline{\mathbf{X}}\right]^{-1}\underline{\mathbf{X}}^{\mathrm{T}}\underline{\mathbf{Y}}$

$$\mathrm{V}(\hat{\boldsymbol{\beta}}) = \left[\underline{\mathbf{X}}^{\mathrm{T}}\underline{\mathbf{X}}\right]^{-1}$$

Note that if $\mathbf{V}(\boldsymbol{\varepsilon})$ is known, $\mathbf{W}$ can be calculated.

So is $\hat{\boldsymbol{\beta}}$. ==>the estimator $\hat{\boldsymbol{\beta}}$ is BLUE.

Generalized Least Square (3)

If Σ is known up to a proportion,

$$\boldsymbol{\Sigma} = \sigma^2\boldsymbol{\Omega} = \sigma^2\begin{bmatrix} \phi_1^2 & \phi_{12} & \cdots & \phi_{1n} \\ \phi_{12} & \phi_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \phi_{n-1,n} \\ \phi_{1n} & \cdots & \phi_{n-1,n} & \phi_n^2 \end{bmatrix}$$

where $\boldsymbol{\sigma}^2$ is unknown but $\boldsymbol{\Omega}$ matrix is known.

Generalized Least Square (4)

Choose the weighting matrix **W** such that

$$\mathbf{WW} = \boldsymbol{\Omega}^{-1}$$

Define $\underline{\mathbf{Y}} = \mathbf{WY}$, $\underline{\mathbf{X}} = \mathbf{WX}, \boldsymbol{\nu} = \mathbf{W}\boldsymbol{\varepsilon}$

Note that $\underline{\mathbf{Y}} = \underline{\mathbf{X}}\boldsymbol{\beta} + \boldsymbol{\nu}$ is CLNRM and

$\boldsymbol{\nu}$ is homoscedastic or $\mathbf{V}(\boldsymbol{\nu}) = \sigma^2\mathbf{I}_n$

==> GLS estimator is BLUE

Generalized Least Square (5)

$\mathrm{V}(\varepsilon)$ is restricted in parameters. For example,

$$\mathbf{V}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}\mathbf{P}\boldsymbol{\Sigma}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_n \end{bmatrix}, \mathbf{P} = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{12} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_{n-1,n} \\ \rho_{1n} & \cdots & \rho_{n-1,n} & 1 \end{bmatrix}$$