CHAPTER - V

SOME NEW INFERENTIAL METHODS FOR MODEL SPECIFICATION
5.0 INTRODUCTION

In empirical research one often faces the problem of estimating a mispecified model. Misspecification can arise either because of omission of a variable specified by the truth, the case of the left out variable, or because of inclusion of a variable not specified by the truth, the case of the irrelevant variable. Misspecification is generally interpreted as a case of left out variables, and many econometricians are concerned only with the bias resulting from it, the specification bias. Econometricians seldom pay attention to the other aspects of misspecification.

A problem which arises frequently in regression analysis is that of testing Whether the functional form employed for the \( k \)-regressors \( X_1, X_2, \ldots, X_k \), is appropriate. There are two basic approaches to testing for functional misspecification. The first is to run a new regression in which an extra term in \( X_k^2 \) is included. A test of significance on the coefficient of this variable then provides a test on the specification of \( X_k \). A number of procedures based on OLS, BLUS and Recursive residuals have been discussed in literature for testing against functional misspecification.
In the present study, some new tests for model specification of linear models have been proposed besides tests for choosing functional form of regression.

5.1 DISTRIBUTION OF OLS RESIDUAL VECTOR UNDER SPECIFICATION ERRORS

Consider the linear regression model in the case of simple heteroscedasticity as

\[ Y = X\beta + \epsilon \]  \hspace{1cm} \ldots (5.1.1)

Where \[ \epsilon \sim N(0, \Sigma) \]

Here, \( Y \) is the \((n \times 1)\) regressand vector; \( X \) is the \((n \times k)\) regressor matrix of rank ‘k’; \( \beta \) is the \((k \times 1)\) vector of regression co-efficients;

\( \epsilon \) is the \((n \times 1)\) vector of disturbance terms and \( \Sigma \) is the \(n \times n\) diagonal covariance matrix of \( \epsilon \) with \( \sigma_i^2, i = 1, 2, \ldots, n \), diagonal elements.

Define the OLS residual vector as

\[ e = M Y = M \epsilon, \]
Where, \( e_i = m_i^j \epsilon, \quad i = 1, 2, \ldots, n. \)

\[
M = \begin{bmatrix} I - X (X^\top X)^{-1} X^\top \end{bmatrix}
\] ... (5.1.2)

and \( m_i^j \) is the \( i^{th} \) row of \( M \).

Further, it can be seen that the mean vector and covariance of matrix of \( e \), which is distributed as singular normal,

i.e.,

\[
E [e / X] = 0,
\]

\[
E [e e^\top / X] = M \Sigma M
\] ... (5.1.3)

Let the non-zero characteristic roots of \( M \Sigma M \) be denoted by \( \lambda_j, j = 1, 2, \ldots, (n - k). \) Now, the conditional distribution of \( e_i e \) is a weighted sum of \( (n - k) \) central \( \chi^2 \) variables with one degree of freedom each, where the weights are proportional to the characteristic roots \( \lambda_j, j = 1, 2, \ldots, (n - k). \)

The distribution of \( e_i, i = 1, 2, \ldots, n \) conditional to \( X \) is normal with zero mean and variance \( \sigma_i^2 = \sum_{j=1}^{n} m_i^j \sigma_j^2 \)
Where, \( m_{ij} \) is the \((i, j)\)th element of \( M \).

Thus the distribution of each \( \frac{e_i^2}{\sigma_i^2} \) for \( i = 1, 2, \ldots, n \), is central \( \chi^2 \) with one degree of freedom.

It is known that under full ideal conditions, the OLS residual vector \( e \) follows normal with zero mean vector and the covariance matrix \( \sigma^2 M \).

Where, \[ M = [I - X(X'X)^{-1} X'] \]

Thus, the OLS residuals are heteroscedastic and are not distributed independently.

Under a variety of specification errors, the distribution of the OLS residual vector is singular normal with non-null mean vector of the form \( M\phi \), Where \( \phi \) is a vector which summarizes the effect of the specification error on the mean of the disturbance vector in the misspecified model. The precise formulation of the vector \( \phi \) depends upon the exact specification of the error.
A brief review about specification analysis reveals that most of the model specification tests have been described by using any of the three types of residuals namely, OLS residuals; BLUS (Best Linear Unbiased Scalar) residuals and Recursive residuals.

In the present study, some inferential procedures have been proposed with regard to model specification tests, based on studentized and predicted residuals.

5.2 STUDENTIZED RESIDUALS

In the inferential procedures for model specification, the various transformations of the OLS residuals have been suggested to overcome partially some of their shortcomings.

The OLS residuals have a distribution that is scale dependent. For many model specification tests, it is useful to define a studentized version of the residuals that does not depend on either of these quantities. Margolin (1977) suggested the studentized residuals to describe the division of a scale dependent statistic say $U$, by a scale dependent estimate $T$ so that the ratio $S = \frac{U}{T}$ has a distribution that is free of the nuisance scale parameters. David (1981) made distinction between Internal and External studentized residuals.
Consider the OLS residual vector as
\[
e = Y - \hat{Y} = Y - X\beta
\]
\[
= Y - X(X'X)^{-1}X'Y
\]
\[
\Rightarrow e = [I - X(X'X)^{-1}X']Y \quad \ldots (5.2.1)
\]
\[
\Rightarrow e = [I - V]Y \quad \ldots (5.2.2)
\]

Where,
\[
V = (V_0) = X(X'X)^{-1}X \quad \ldots (5.2.3)
\]

One can write,
\[
e = (I - V)(X\beta + \epsilon)
\]
\[
\Rightarrow e = (I - V)\epsilon \quad \ldots (5.2.4)
\]

In scalar form, for i = 1, 2, ..., n, one can have
\[
e_i = \epsilon_i - \sum_{j=1}^{n} V_{ij}\epsilon_j \quad \ldots (5.2.5)
\]

This reveals that the relationship between e and \(\epsilon\) depends only

on V. The matrix V is a symmetric idempotent matrix. One can have

\[
\text{trace} (V) = \text{rank} (V) = k, \sum_{j} V_{ij}^2 = V_{ii}
\]

.208
V is invariant under nonsingular linear transformations. In the
case two – variable linear regression model, \( V_{ii} \)'s are defined as

\[
V_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_j - \bar{x})^2}, \quad i = 1, 2, \ldots, n \quad \text{... (5.2.6)}
\]

For the linear regression models with more than one independent
variable, contours of constant \( V_{ii} \) in \( k^1 \) – dimentional space are ellipsoids,
centered at the vector of sample averages. Here \( k^1 \) is the number of
centered explanatory variables in the model.

From (5.2.2) if \( \varepsilon \) follows \( N(0, \sigma^2 I) \) then the OLS residual vector
\( \varepsilon \) follows a singular normal distribution with \( E(\varepsilon) = 0 \) and
\( \text{Var}(\varepsilon) = \sigma^2 (I-V) \), and the variation in \( \varepsilon \) is controlled by \( V \).

Many statisticians and econometricians have hinted that the
diagonal elements \( V_{ii} \) of the that matrix \( V = ((V_{ii})) \) may play an
important role in understanding an analysis based on the general linear
model.

209
5.2.1 Internally Studentized Residuals

Using the OLS residuals $e_i$'s, the internally studentized residuals are defined by

$$ e_i^* = \frac{e_i}{\hat{\sigma} \sqrt{1-V_{ii}}} , \; i = 1, 2, \ldots, n \quad \ldots (5.2.7) $$

Where, $$ \hat{\sigma}^2 = \frac{\sum e_i^2}{n - k} $$ is the residual mean square.

This transformation of residuals has been studied by Srikantan (1961), Anscombe and Tukey (1963) Beckman and Trussell (1974) and many others.

The sampling distribution of $e_i^*$'s is given by:

$$ e_i^* \sim \text{Beta distribution with parameters } \frac{1}{2} \text{ and } \left( \frac{n-k-1}{2} \right) $$

It follows that, $\mathbb{E} [e_i^*] = 0$, $\text{Var} (e_i^*) = 1$ and

$$ \text{Cov} (e_i^*, e_j^*) = \frac{-V_{ij}}{\left[ (1-V_{ii})(1-V_{jj}) \right]^2} , \; i \neq j \quad \ldots (5.2.8) $$

210
5.2.2 Externally Studentized Residuals

Using the OLS residuals, the externally studentized residuals are defined by

$$\hat{\sigma}_{(i)} \sqrt{1 - V_{ii}} \left( 1 - V_{ii} \right)^{1/2} \quad \ldots \ (5.2.9)$$

Where, $\hat{\sigma}_{(i)}^2$ is the residual mean square to be computed without the $i^{th}$ observation.

It is defined by

$$\left( n - k \right) \hat{\sigma}^2 - \frac{c_i^2}{1 - V_{ii}} \frac{\hat{\sigma}_{(i)}^2}{n - k - 1}$$

or

$$\hat{\sigma}_{(i)}^2 = \frac{n - k - c_i^2}{n - k - 1} \hat{\sigma}^2 \quad \ldots \ (5.2.10)$$

It should be noted that, under normality of the errors, $\hat{\sigma}_{(i)}^2$ and $c_i$ are independently distributed.

The sampling distribution of $e_i^{**}$ is the student's $t$ - distribution with $(n - k - 1)$ degrees of freedom.
A relationship between both the studentized residuals is given by

\[ e_i = e_i \frac{n - k - 1}{(n - k - e_i^*)^2} \] ...

(5.2.11)

This shows that \( e_i^* \) is a monotonic transformation of \( e_i^2 \).

5.3. PREDICTED RESIDUALS

Suppose \( \hat{\beta} \) and \( \hat{\beta}_{(i)} \) be the OLS estimates of \( \beta \) based on the full data and the data with the \( i^{th} \) observation excluded. The predicted residuals are defined by

\[ e_{(i)} = Y_i - X_i^t \hat{\beta}_{(i)}, \quad i = 1, 2, \ldots, n \] ...

(5.3.1)

Where, \( X_i \) is the \( i^{th} \) row of \( X \).

The predicted residual sum of squares (PRESS) is given by

\[ \text{PRESS} = \sum e_{(i)}^2 \] ...

(5.3.2)
A relationship between predicted and OLS residuals is given by

\[ e_{(i)} = \frac{-1}{1 - V_{ii}} , \ i = 1, 2, \ldots, n \quad \ldots (5.3.3) \]

The relationships between predicted and studentized residuals are given by

\[ e_i^* = \frac{e_{(i)}}{\hat{\sigma}/(1 - V_{ii})^{1/2}} \quad \ldots (5.3.4) \]

\[ e_i^{**} = \frac{e_{(i)}}{\hat{\sigma}_{(i)}/(1 - V_{ii})^{1/2}} \quad \ldots (5.3.5) \]

It should be noted that the \( e_{o_i} \)'s are normally distributed (if the \( \epsilon_i \)'s are normally distributed) with mean zero and variance \( \frac{\epsilon_i^2}{1 - V_{ii}} \) and have the same correlation structure as the OLS residuals \( e_i \)'s.
5.4 A TEST FOR MISSPECIFICATION USING COEFFICIENT OF DETERMINATION

Consider the general linear regression model

\[ Y = X \beta + \varepsilon \quad \cdots (5.4.1) \]

Where,
- \( Y \) is an \( n \times 1 \) vector of observable random variables;
- \( X \) is an \( n \times k \) observed regressor matrix of full rank \( k \);
- \( \beta \) is a \( k \times 1 \) vector of unknown parameters;
- and \( \varepsilon \) is an \( n \times 1 \) vector of independent normal random disturbances with \( \varepsilon \sim N(0, \sigma^2 I_n) \).

It is assumed that the elements of \( X \) are fixed constants. For each of the specification tests to be given below, the specification of the null hypothesis is that given in (5.4.1). Under the alternative hypotheses of misspecification that the true model has some specifications other than that given in (5.4.1).

One may write the alternative hypothesis under the specification tests as

(i) \( H_1 : Y = X \beta + Z\Gamma + \varepsilon, \) \( Z \) is an \( n \times 1 \) non-stochastic regressor vector; \( \Gamma \) is the corresponding coefficient vector; and \( \varepsilon \) is distributed as \( N(\phi, \sigma^2 I) \). If one uses (5.4.1) to estimate the coefficient vector \( \beta \); one has a misspecified model where the misspecification is that of omitting the variable \( Z \).
(ii) \( H_2 : Y = Z \Gamma + \epsilon, \) where \( Z \) is non-stochastic. If one considers a matrix \( X \) whose elements are obtained from the elements of \( Z \) by a non-stochastic transformation, then one would have a misspecified model in which the misspecification is that of incorrect functional form of the regressors.

(iii) \( H_3 : Y = Z \Gamma + W \delta + \epsilon, \) where \( Z \) is an \( n \times k_1 \) non-stochastic matrix of regressors with full rank \( k_1; \) \( \Gamma \) the corresponding coefficient vector; \( W \) is an \( n \times k_2 \) stochastic matrix of regressors of rank \( k_2 \) in observed samples; and \( \delta \) is the corresponding coefficient vector. Further, the elements of each row of \( W \) and the corresponding \( \epsilon_i, i = 1, 2, \ldots, n \) are statistically dependent. In this case one has a misspecified model in which the misspecification is the 'Simultaneous Equation Problem'.

(iv) \( H_4 : Y = X \beta + \epsilon, \) \( \epsilon \sim N(\phi, \Omega). \) In this case, all the specifications made for model (5.4.1) are correct except that \( \Omega \) is assumed to be diagonal, but with unequal elements on the diagonal. Here, one has the misspecified model with the specification error of heteroscedasticity.
By augmenting the model $Y = X\beta + \epsilon$ with a set of regressors as powers of the $X$ variables, the augmented regression model can be written as

$$Y = X\beta + W\Gamma + \epsilon$$  \hspace{1cm} \ldots (5.4.2)

Where,

- $Y$ is $n \times 1$ vector of observations on the dependent variable;
- $X$ is a $n \times k$ observed regressor matrix with rank $'k'$.
- $\beta$ is a $k \times 1$ vector of unknown parameters
- $W$ is a $n \times g$ matrix of rank $g$ of test variables such as powers of $X$ variables, such that
- Rank of $(X, w) = k + g$
- $\Gamma$ is a $g \times 1$ vector of coefficients corresponds to $w$.
- $\epsilon$ is $n \times 1$ multivariate normal vector of disturbances with dispersion matrix $\sigma^2 I_n$.

Under $H_0: \Gamma = 0$, the OLS estimator and Hausman's estimators of $\beta$ are given by
\[ \hat{\beta}_0 = (X^\top X)^{-1} X^\top Y \text{ and } \hat{\beta}_1 \text{ which is composed of first } k \]

elements of

\[ [(X, w)^\top (X, w)]^{-1} (X, w)^\top Y. \]

The dispersion matrices of \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \) are given by

\[ V(\hat{\beta}_0) = \sigma^2 (X^\top X)^{-1} \quad \ldots \quad (5.4.3) \]

\[ V(\hat{\beta}_1) = \sigma^2 (X^\top M_w X)^{-1} \quad \ldots \quad (5.4.4) \]

\[ V(\hat{\beta}_1) = \sigma^2 P(w^\top M_x w)^{-1} P^\top + \sigma^2 (X^\top X)^{-1} \quad \ldots \quad (5.4.5) \]

Where, \( M_w = I - w (w^\top w)^{-1} w^\top \) and \( M_x = I - X(X^\top X)^{-1} X^\top \)

and \( P = (X^\top X)^{-1} X^\top w. \)

Clearly \( \hat{\beta}_0 \) is efficient under \( H_1 \), we have

\[ E(\hat{\beta}_0) = \beta + (X^\top X)^{-1} X^\top Z \quad \ldots \quad (5.4.6) \]

\[ E(\hat{\beta}_1) = \beta + (X^\top X)^{-1} X^\top Z - P(w^\top M_x w)^{-1} w^\top M_x Z \quad \ldots \quad (5.4.7) \]

Where, \( Z = E(\varepsilon/x). \)
Under $H_0$, $[\hat{\beta}_1 - \hat{\beta}_0] \sim N(0, M(\hat{\varrho}))$

Where, $M(\hat{\varrho}) = \sigma^2 \left[ (\chi^\top M_w \chi)^{-1} - (\chi^\top \chi)^{-1} \right]$

or $M(\hat{\varrho}) = \sigma^2 P (w^\top M_X w)^{-1} P^\top$ \hspace{1cm} ... (5.4.8)

The test statistic for testing $H_0$: $\Gamma = 0$ is given by

$$F_R = Y^\top M_X w (w^\top M_X w)^{-1} M_X \frac{Y^\top g S_1^2}{\gamma}$$ \hspace{1cm} ... (5.4.9)

Where, $S_1^2$ is the estimator of $\sigma^2$ from the OLS regression of $Y$ on $X$ and $W$.

It is equivalent to the likelihood ratio test for testing $H_0$: $\Gamma = 0$ which is based on the statistic

$$F = \frac{Y^\top \left( X, w \right) (X, w)^{-1} (X, w)^\top X (X^\top X)^{-1} X^\top Y / g}{Y^\top \left[ I - (X, w) (X, w)^\top (X, w) \right] (X, w)^\top Y / n - k - g}$$ \hspace{1cm} ... (5.4.10)
Let $R^2(X,w)$ be the squared sample multiple correlation coefficient obtained from (5.4.2) and $R^2_X$ be that obtained from (5.4.1) then one can express $F$ as

$$F = \frac{\left(R^2(X,w) - R^2_X\right)/g}{\left[1 - R^2(X,w)\right]/(n-k-g)} \quad \ldots (5.4.11)$$

Here, $F$ has a non-central $F_{g, n-k-g}$ with the noncentrality parameter

$$\mu = \Gamma^T \left[w'w - w'X(X'X)^{-1}X'w\right] \Gamma / \sigma^2 \quad \ldots (5.4.12)$$

The null hypothesis $H_0$: $\Gamma = 0$ may be tested at level $\alpha$ by rejecting if $F > F_{g, n-k-g}$, the upper $100 \alpha\%$ point of the central $F_{g, n-k-g}$ distribution. This reveals that the model $Y = X\beta + \epsilon$ is misspecified.

5.5 A TEST FOR INCORRECT FUNCTIONAL FORM

Generally an assumption of linearity can be viewed as a first order approximation to an arbitrary function. Under certain conditions Taylor's theorem states that an arbitrary function $f(X_i)$ can be approximated about the point $X_i = X_0$, by a polynomial function and a remainder. If the arbitrary function is a polynomial, the approximation is a lower order polynomial and the remainder is also a polynomial.
By the Taylor's theorem, \( f(X_i) \) can be expanded at \( X_i = X_0 \) as

\[
f(X_i) = \left[ f(X_0) + \frac{f'(X_0)(X_i - X_0)}{1!} + \frac{f''(X_0)(X_i - X_0)^2}{2!} + \frac{f'''(X_0)(X_i - X_0)^3}{3!} + \ldots + \frac{f^{(n)}(X_0)(X_i - X_0)^n}{n!} \right] + R_n
\]

\[ ... (5.5.1) \]

Where, \( f^{(i)}(X_0) \) is the \( i^{th} \) derivative of \( f(X_i) \) with respect to \( X_i \), evaluated at \( X_i = X_0; \ i = 1, 2, \ldots, n \). \( R_n \) is the remainder.

The expression in square brackets is the polynomial approximations, say \( P_n \) where \( n \) indicates the order of the polynomial, to the unknown function.

For \( n = 1 \), \( P_n \) results in a linear approximation, more specifically

\[
f(X_i) = [f(X_0) + f'(X_0) (X_i - X_0)] + R_1
\]

\[ ... (5.5.2) \]
Suppose the arbitrary function, is known say the quadratic function

\[ Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 \]

which is expanded about \( X = X_0 \).

We have, \( f(0) = \beta_1; f'(0) = \beta_2 + 2\beta_3 (X_i = 0) = \beta_2; f''(0) = 2\beta_3 \).

Thus, the second order expansion is given by

\[
Y_i = \frac{f(0)}{0!} + \frac{f'(0)}{1!} X_i + \frac{f''(0)}{2!} X_i^2 \quad \ldots (5.5.3)
\]

\[
\Rightarrow Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 \quad \ldots (5.5.4)
\]

Since \( n=2 \), this exactly reproduces the original function; however, if \( n = 1 \) the reminder is \( f''(0) X_i^2 / 2 \). This example shows that the higher the order of the approximation, the better the approximation to the original function. Adding terms in the powers of \( X_i \) to a linear regression can be regarded as assessing the validity of the linear approximation to an arbitrary function in \( X_i \).
To test for the functional form, it is proposed to add powers of internally studentized residuals say $e_i^2$, $e_i^3$, and $e_i^4$ to the linear regression model and test for the significance of the effects of additional variables.

Now one can write the augmented regression model as

$$ Y = X\beta + W\Gamma + \epsilon \quad \ldots (5.5.5) $$

Where, $Y$, $X$, $\beta$ and $\epsilon$ are defined as in equation (5.4.2) $W$ is (nx3) matrix of rank 3 of additional test variables such as powers of internally studentized residuals say $e_i^2$, $e_i^3$, and $e_i^4$. Since, generally there is a curvilinear relationship between $e_i$'s and $\hat{y}_i$, we introduced the powers of transformed OLS residuals say $e_i^2$, $e_i^3$, and $e_i^4$ as additional regressor $\Gamma$ is a 3 x 1 vector of coefficients corresponds to $W$.

To test $H_0 : \Gamma = 0$ first we regress $Y$ on $X$ and estimate $\hat{y}$ and hence obtain OLS residuals as $e_i = [Y_i - \hat{y}_i]$, $i = 1, 2, \ldots, n$. Also we compute the internally studentized residuals as
\[ e_i^* = \frac{e_i}{\hat{\sigma} \sqrt{1 - \psi_{ii}}} \], i = 1, 2, \ldots, n. \quad \ldots (5.5.6) \]

Where,

\[ \hat{\sigma}^2 = \frac{\Sigma e_i^2}{n - k} \] \quad \ldots (5.5.7)

and \[ V = ((\nu')) = X (X'X)^{-1} X' \]

Now we regress \( Y \) on \( X \) and \( W \) and test for the coefficients of all the powers of \( e_i^* \) are zero. For this, the F-test statistic is given by

\[
F = \frac{\left[ R^2_{(x,W)} - R^2_{X} \right]}{1 - R^2_{(x,W)}} \bigg/ \frac{1}{n - k - 3} \sim F_{3, (n-k-3)} \quad \ldots (5.5.8)
\]

If the computed F value is significant say at the 5% level, then one can accept the hypothesis that the model \( Y = X \beta + \epsilon \) is misspecified.
5.6 A MODIFIED BARTLETT'S MODEL SPECIFICATION ERROR TEST

Consider the basic linear regression model relevant to the development of test statistic as

\[ Y_{nx1} = X_{nxk} \beta_{nx1} + \epsilon_{nx1} \quad \ldots (5.6.1) \]

and \( \epsilon \sim N \left( 0, \sigma^2 I_n \right) \)

Thus, \( \epsilon \) is an \( nx1 \) vector of independent disturbance terms each distributed normally with mean zero and variance \( \sigma^2 \).

For the proposed model specification error test, the null and alternative hypotheses may be stated as

\[ H_0: Y = X \beta + \epsilon, \quad \epsilon \sim N \left( 0, \sigma^2 I_n \right) \quad \ldots (5.6.2) \]

It is the hypothesis of no specification error

\[ H_1: Y = X \beta + u, \quad u \sim N \left( 0, \Omega \right) \quad \ldots (5.6.3) \]
Where, $\Omega$ is assumed to be diagonal, but with unequal elements on the diagonal.

It is the hypothesis of the specification error of heteroscedasticity.

Define OLS residual vector $e = Y - \hat{Y}$

or $e = Y - X \hat{\beta} = Y - X (X'X)^{-1}X'Y$

or $e = MY$ where $M = [I - X(X'X)^{-1}X']$ is a symmetric idempotent matrix. By considering $e$ as an estimate of $u$, the distributional properties of the OLS residuals under the null and alternative hypotheses are given by

$$H_0 = e \sim N \left(0, \sigma^2 M\right) \quad \ldots \ (5.6.4)$$

and $H_1 = e \sim N \left(0, \psi\right) \quad \ldots \ (5.6.5)$

Where, $\psi = M \Omega M \quad \ldots \ (5.6.6)$
Here, $M$ and $\psi$ are positive semi-definite matrices, so that $e$ is distributed as singular normal.

Consider the predicted residuals as

$$e_{(i)} = \frac{c_i}{1 - v_{ii}}, \quad i = 1, 2, \ldots, n \quad \ldots (5.6.7)$$

Since, $e_{(i)}$'s follow normal distribution with mean zero and variance $\frac{\sigma^2}{1 - v_{ii}}$ and the correlation structure of $e_{(i)}$'s is the same as the OLS residuals $e_i$'s; the predicted residuals may be used in the place of the OLS residuals in the analysis. In this case, the analysis will tend to emphasize cases with large $V_{ii}$ while use of $e_i$ tends to emphasize cases with small $V_{ii}$.

Now, the null and alternative hypotheses under the specification error of heteroscedasticity can be expressed in terms of the distributed properties of the predicted residuals as

$$H_0 = e^{***} \sim N \left( 0, \sigma^2 I_n \right) \quad \ldots (5.6.8)$$
\[ H_1 = e^{\mathbf{-}} \sim N \left( 0, \Theta \right) \] \hspace{1cm} \text{(5.6.9)}

Where \( \Theta \) is a diagonal positive definite matrix.

It is clear that the fundamental idea underlying the test is that the effect of specification error of heteroscedasticity is to alter the distribution of the residuals from that postulated under the null hypothesis.

Consider the Bartlett's test statistic for testing homogeneity of error variances as

\[
B = \left[ \nu \log \left( \frac{\sum_{i=1}^{m} v_i s_i^2}{\sum_{i=1}^{m} v_i} \right) - \sum_{i=1}^{m} v_i \log s_i^2 \right] / C \hspace{1cm} \text{(5.6.10)}
\]

Where

\[
C = 1 + \frac{1}{3(m-1)} \sum_{i=1}^{m} \frac{v_i}{\nu_i} \nu \hspace{1cm} \text{(5.6.11)}
\]
Here, \( S_i^2 \) is an unbiased estimate of \( \sigma_i^2 \) with \( v_i \) degrees of freedom, \( v_i = n_i - 1; \sum_{i=1}^{m} v_i = v = n - m \); and \( \sum n_i = n \) and \( m \) is the number of subgroups of squared residuals.

Under \( H_0 \), \( B \) is asymptotically distributed as central \( \chi^2 \) distribution with \( (m-1) \) degrees of freedom.

Now, it is proposed to divide the set of \( n \) predicted residuals into three non-intersecting sub-sets in order to form the statistics

\[
S_i^2, \ i = 1, 2, \ldots, 3.
\]

The value of \( m \) is chosen to be 3 for the following reason. The larger the values of \( v_i \), \( i = 1, 2, \ldots, m \), the closer is the approximation between the actual distribution of \( \beta \) and \( \chi^2_2 \) under the null hypothesis on the other hand, in order to test for heterogeneity of variances, one needs several variance estimations. \( m = 3 \) is chosen as a useful compromise. By replacing the OLS residuals with the predicted residuals, the sub-sets are chosen by setting.
\[ S_1^2 = \frac{1}{\nu_1} \sum_{i=1}^{n} e_{(i)}^2 \), \quad r_1 = \frac{n}{3} \]

\[ S_2^2 = \frac{1}{\nu_2} \sum_{i=r_1+1}^{r_2} e_{(i)}^2 \), \quad r_2 = 2 \, r_1 \] ... (5.6.12)

\[ S_3^2 = \frac{1}{\nu_3} \sum_{i=r_2+1}^{n} e_{(i)}^2 \), \quad r_3 = n - 2r_1 \]

Where \[ \nu_i = \frac{n - k}{3}, \quad \forall \, i = 1, 2, 3. \]

Here, \( r_1 \) is the integer obtained from the integer division of \( n \) by 3.

The modified Bartlett's model specification error test statistic is now given by
\[ B^* = \nu \log S^2 - \sum_{i=1}^{p} \nu_i \log S_i^2 \]  \hfill (5.6.13)

Where
\[ C = 1 + \frac{1}{3(3-1)} \left| \sum_{i=1}^{3} \left( \frac{1}{\nu_i} \right) - \frac{1}{\nu} \right| \]  \hfill (5.6.14)

\[ S^2 = \frac{1}{\nu} \sum_{i=1}^{3} \nu_i S_i^2 = \frac{1}{n-k} \sum_{j=1}^{n} e_{(i)}^2 \]  \hfill (5.6.15)

and \[ \sum_{i=1}^{p} \nu_i = \nu = n-k, \]

Here, \( B^* \) follows \( \chi^2_2 \) distribution under the null hypothesis. This test is sensitive to the alternative hypothesis of simple heteroscedasticity.
5.7 SOME NEW TESTS FOR CHOOSING FUNCTIONAL FORM OF REGRESSION

Generally, economic theory does not suggest an appropriate functional form for the relationship between economic variables. Literature about the model specification concentrated on linear and logarithmic (or log-linear) regression models as alternative specifications. Since, misspecifying the functional form of an econometric model will, in general lead to inconsistent estimates of the parameters of the model, it has been emphasized that estimated models should be subjected to rigorous model evaluation through the use of regression and residual diagnostics.

5.7.1 Specification Of The Log-Linear And Linear Models

The choice between a log-linear regression model and a linear regression model is a perennial question in econometric analysis. One can specify log-linear and linear regression models to be tested as

\[ H_0 : \log Y_j = \beta_0 + \sum_{i=1}^{k} \beta_i \log X_{ij} + \epsilon_{0j}, j = 1, 2, \ldots, n \quad \ldots (5.7.1) \]

\[ E(\epsilon_{0j}) = 0, E(\epsilon_{0j}^2) = \sigma_0^2 \]
\[ H_1 : Y_j = \beta_0 + \sum_{i=1}^{k} \beta_i X_{ij} + \epsilon_{ij}, j = 1, 2, \ldots, n \quad \ldots (5.7.2) \]

\[ E(\epsilon_{ij}) = 0, \ E(\epsilon_{0j}^2) = \sigma_1^2 \]

The OLS estimates of the parameters under \( H_0 \) and \( H_1 \) are respectively denoted by

\[ (\hat{\beta}_0, \hat{\beta}_i, \hat{\sigma}_0^2) ; \text{ and } (\hat{\beta}_0, \hat{\beta}_i, \hat{\sigma}_1^2). \]

The OLS estimated regression equations are respectively given by

\[ \hat{\log y}_j = \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i \log X_{ij} \quad \ldots (5.7.3) \]

and \[ \hat{y}_j = \hat{\beta}_0 + \sum \hat{\beta}_i X_{ij} \quad \ldots (5.7.4) \]

5.7.2 Test (1)

To test for the choice between log-linear and linear regression models one can state the hypotheses as given in (5.7.1) and (5.7.2).

The test involves the following steps:
Step (1): Estimate the models (5.7.1) and (5.7.2) and obtain the OLS residuals and hence obtain the internally studentized residuals.

OLS residuals: \( \log e_j = \log y_j - \log Y_j \) and

\[ \tilde{e}_j = [Y_j - \bar{Y}_j] \] under \( H_0 \) and \( H_1 \) respectively.

\( j = 1, 2, ..., n. \)

Internally studentized residuals:

\[ \log e_j^* = \frac{\log e_j}{\sigma_0 \sqrt{1 - \hat{v}_{jj}}} \]

and

\[ e_j = \frac{e_j}{\sigma_0 \sqrt{1 - \hat{v}_{jj}}} \] under \( H_0 \) and \( H_1 \) respectively.

\( j = 1, 2, ..., n. \)
Where, $\hat{v}_{jj}'s$ and $\tilde{v}_{jj}$ are the elements of the corresponding Hat matrices.

Step (2) : Consider the linear regression models as

(a) 

(b) \[ \log Y_j = \beta_0 + \sum_{i=1}^{\infty} \beta_i \log X_{ij} + \gamma_0 \log e_j^* - \exp \log e_j^* + \varepsilon_j \ldots (5.7.5) \]

(b) \[ Y_j = \beta_0 + \sum_{i=1}^{\infty} \beta_i X_{ij} + \gamma_1 \log e_j^* - \log e_j^* + \varepsilon_j \ldots (5.7.6) \]

Step (3) : Given that (5.7.5) and (5.7.6) simply involving augmenting the original log-linear and linear models with an additional variable, the test statistics are straightforward to compute. The tests of $H_0$ and $H_1$ are equivalent to testing $\gamma_0 = 0$ and $\gamma_1 = 0$ respectively in (5.7.5) and (5.7.6). The test statistics will be distributed asymptotically as $N(0,1)$ variates under $H_0$ and $H_1$, respectively, as long as the disturbances under both hypotheses satisfy certain technical conditions for the application of a central limit theorem so that normality need not be assumed.
If the disturbance under $H_0$ is normally distributed, the test of $H_0: \gamma_0 = 0$ is distributed as $t$ distribution with $(n-k-2)$ degrees of freedom under $H_0$. Similarly the test of $H_1: \gamma_1 = 0$ is distributed as $t$ distribution with $(n-k-2)$ degrees of freedom under $H_1$, in finite samples.

5.7.3 Test (2): Consider the test procedure simply involving augmenting the original regressors with powers of the internally studentized residuals for the null model and then testing the joint significance of the estimated coefficients of these additional variables.

This test is based on the auxiliary regressions

$$\log Y_j = \beta_0 + \sum_{i=1}^k \beta_i \log X_{ij} + \sum_{p=2}^m \gamma_{0p} \log e_j^p + \epsilon, \ldots (5.7.7)$$

and

$$Y_j = \beta_0 + \sum_{i=1}^k \beta_i X_{ij} + \sum_{p=2}^m \gamma_{1p} \left( e_j^p \right)^p + \epsilon, \ldots (5.7.8)$$

$$j = 1, 2, \ldots, n,$$
The test of $H_0$ against (5.7.7) which will be a specification error test, is the test of

$$
\gamma_{02} = \gamma_{03} = \ldots = \gamma_{0m} = 0 \text{ and the test statistic is distributed as } F_{(m-1), (n-k-m)} \text{ under } H_0 \text{ is the disturbances are normal. Similarly the specification error test of } H_1 \text{ in (5.7.8) is the test of } \gamma_{12} = \gamma_{13} = \ldots \gamma_{1m} = 0 \text{ which is approximately distributed as } F_{(m-1), (n-k-m)} \text{ under } H_1 \text{ is the disturbances are approximately normal.}
$$

5.7.4 Test (3): The modified White's (1980) functional form tests of $H_0$ and $H_1$ are given as follows:

Consider the auxiliary regressions as

$$
\log Y_j = \beta_0 + \sum_{i=1}^k \beta_i \log X_{ij} + \delta_{00} \log e_j^* - 2 \\
+ \sum_{i=1}^k \delta_{0i} \left[ \log X_{ij} \left( \log e_j^* \right)^{-2} \right] + \varepsilon, \quad (5.7.9)
$$
and \( Y_j = \beta_0 + \sum_{i=1}^{k} \beta_i X_{ij} + \delta_{10} \left( \frac{1}{c_j} \right)^2 \)

\[ + \sum_{t=1}^{k} \delta_{1t} \left( X_{ij} \left( \frac{1}{c_j} \right)^2 \right)^2 + \epsilon_j \quad ...(5.7.10)\]

The tests of \( H_0 \) and \( H_1 \) are equivalent to the tests of

\( H_0 : \delta_{00} = \delta_{01} = \delta_{02} = \ldots = \delta_{0k} = 0 \) and

\( H_1 : \delta_{10} = \delta_{11} = \delta_{12} = \ldots = \delta_{1k} = 0 \)

respectively in the auxiliary regression (5.7.9) and (5.7.10).

The test statistic of \( H_0 \) being distributed as \( F_{(k+1,n-2k-2)} \) if the disturbances are normally distributed and that of \( H_1 \) being approximately distributed as \( F_{(k+1,n-2k-2)} \).

**5.7.5 Test (4):** Consider the null hypothesis of linearity as

\( H_0: Y = X \beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2 I) \).
Let \( \tilde{e}_R^1 \tilde{e}_R \) be the restricted predicted residuals sum of squares from all available ‘n’ observations and let \( \tilde{e}_{UR}^1 \tilde{e}_{UR} \) be the unrestricted predicted residuals sum of squares from the middle subset of \( t_1 \) observations. The sample data may be assumed to be split into three parts containing \( n_0, n_1 \) and \( n_2 \) observations with \((k+1) < n_1 < n\).

The test statistic for testing \( H_0 \) is given by

\[
F = \frac{\left[ \tilde{e}_R^1 \tilde{e}_R - \tilde{e}_{UR}^1 \tilde{e}_{UR} \right]}{\frac{n_0 + n_2}{n_1 - k - 1}}
\]  

... (5.7.11)

This test statistic is distributed as \( F_{(n_0 + n_2), (n_1 - k - 1)} \), under the null hypothesis of no misspecification. Generally \( n_1 \) may be chosen as

\[n_1 = n_0 + n_2 = \frac{n}{2}, \text{ so that approximately one - half } \left( \begin{array}{l} 1 \\ 2 \end{array} \right) \text{ of the data points are used for subset estimation.}\]
5.8 A TEST FOR MISSPECIFICATION OF THE LINEAR STATISTICAL MODEL

The basic problem in the specification tests is whether or not to include a subset of explanatory variables in a linear model. The common practice is to base this decision upon the standard F-statistic computed to test the null hypothesis that the coefficients associated to this subset of variables are all equal to zero. A general method to this problem of model specification was introduced by Hausman (1978). This test was based upon the difference between unconstrained and constrained estimators of the subset of regression coefficients specified under the null and under the alternative hypotheses.

Consider the classical linear model

\[ Y = X \beta + \epsilon \]  \hspace{1cm} (5.8.1)

or \[ Y = X_1 \beta_1 + X_2 \beta_2 + \epsilon \]  \hspace{1cm} (5.8.2)

Where \( Y \) is a \((n \times 1)\) observation vector; \( X, X_1 \) and \( X_2 \) are non-stochastic design matrices of the respective sizes \((n \times k), (n \times k_1)\) and \((n \times k_2)\) where \( k = k_1 + k_2 \); \( \beta, \beta_1 \) and \( \beta_2 \) are \((k \times 1), (k_1 \times 1)\) and \((k_2 \times 1)\)
vectors of unknown parameters; is (n x 1) vector of disturbances, which is normally distributed with zero mean and covariance matrix \( \sigma^2 I_n \). It is assumed that rank of \( \left( X_1^\top X_2 \right) \) is \( k_1 \).

Now, the OLS estimator of \( \beta \) is given by

\[
\hat{\beta} = \frac{\hat{\beta}_1}{\hat{\beta}_2} = \left( X^\top X \right)^{-1} X^\top Y
\]

\[\cdots (5.8.3)\]

\[
\begin{pmatrix}
\hat{\beta}_1 \\
\hat{\beta}_2
\end{pmatrix}
= 
\begin{bmatrix}
X_1^\top X_1 & X_1^\top X_2 & X_1^\top Y \\
X_2^\top X_1 & X_2^\top X_2 & X_2^\top Y
\end{bmatrix}
\]

\[\cdots (5.8.4)\]

Here, \( \hat{\beta} \sim N(\beta, \sigma^2 (X^\top X)^{-1}) \)

\[\cdots (5.8.5)\]

Consider the \( X_1 \) subset of \( X \) and assume that the \( X_2 \) set of variables appears in the model with a zero coefficient vector, that is, these variables are extraneous and are omitted from the linear model.

Now, one may write this in the set of restrictions as

\[ R \beta = r \]

Where

\[ R = \begin{bmatrix} 0 & I_{K_2} \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}, r = \Omega_{K_2} \]
Where \( R \) is a \((k_2 \times k)\) known matrix of rank \( k_2 \);
\( O_{K_2} \) is a \((k_2 \times 1)\) vector of zeros.

The general restricted least squares estimator as

\[
\beta^* = \hat{\beta} - \left( X'X \right)^{-1} R^1 \left[ R \left( X'X \right)^{-1} R^1 \right]^{-1} \left( R \hat{\beta} - r \right) \quad (5.8.6)
\]

If the elements of \( \beta_2 \) are restricted to zero, the restricted least squares estimator for \( \beta \) is given by

\[
\beta^* = \begin{bmatrix} \beta^*_1 \\ O_{K_2} \end{bmatrix} = \begin{bmatrix} \left( X_1'X_1 \right)^{-1} X_1'Y \\ O_{K_2} \end{bmatrix} \quad (5.8.7)
\]

If the restrictions are correct, \( \beta_2 = 0 \) then \( \delta \) is a null vector.

Here, \( R \beta - r = \delta \), where \( \delta \) is a vector of parameter specification errors of dimension \( k_2 \).
The mean and covariance matrices of $\beta^*$ are given by

$$E[\beta^*] = E \left( \hat{\beta} - \left( X'X \right)^{-1} R \left[ R \left( X'X \right)^{-1} \right] R \right) \left( K \hat{\beta} - 0 \right)$$

$$= \beta - \left( X'X \right)^{-1} R \left[ R \left( X'X \right)^{-1} \right] R \left( \hat{\beta} - 0 \right)$$

$$E[\beta^*] = \beta_1 + \left( X'X \right)^{-1} \left( X'X_2 \right) \beta_2$$

$$\beta_2 - \delta \quad \ldots \quad (5.8.8)$$

If $\beta_2 = 0$ and thus $\delta = 0$, then $\beta_1^*$ is an unbiased estimate of $\beta_1$.

The covariance matrix of $\beta^*$ is given by

$$E[\beta^* - E(\beta^*)] \left( \beta^* - E(\beta^*) \right)^t$$

$$= \sigma^2 \left( X'X \right)^{-1} - \left( X'X \right)^{-1} R \left[ R \left( X'X \right)^{-1} \left( X'X \right)^{-1} \right] R \left( X'X \right)$$

242
\[ \begin{align*}
\mathbb{E} \begin{bmatrix} \hat{\beta}_1^* - \mathbb{E}\hat{\beta}_1^* \\ \hat{\beta}_2^* - \mathbb{E}\hat{\beta}_2^* \\ \hat{\beta}_2 - \mathbb{E}\hat{\beta}_2, 
\end{bmatrix} &= \begin{bmatrix} \sigma^2(\mathbf{x}'_1 \mathbf{x}_1) \\ 0 
\end{bmatrix} O_{K'} \\
\text{Var} \left( \hat{\beta}_1^* \right) &= \sigma^2 \left( \mathbf{x}'_1 \mathbf{x}_1 \right) \\
\text{i.e.,} \quad \mathbb{E} \left[ \hat{\beta}_1^* \right] &= \beta_1 + (\mathbf{x}'_1 \mathbf{x}_1)^{-1} \mathbf{x}'_1 \mathbf{x}_2 \beta_2 = \beta_1 + \eta \\
\text{Where,} \quad \eta &= (\mathbf{x}'_1 \mathbf{x}_1)^{-1} \mathbf{x}'_1 \mathbf{x}_2 \beta_2 = \text{Bias}; \\
\text{and} \quad \text{Var} \left( \hat{\beta}_1^* \right) &= \sigma^2 \left( \mathbf{x}'_1 \mathbf{x}_1 \right) \\
\text{Whether} \ \beta_2 = 0 \ \text{or not}, \ \hat{\beta}_1^* \ \text{is normally distributed} \\
\hat{\beta}_1^* &\sim N \left( \beta_1 + \eta, \sigma^2 \left( \mathbf{x}'_1 \mathbf{x}_1 \right) \right) \\
\beta_1^* \text{ is an unbiased and efficient estimator of } \beta, \ \text{when } \eta = 0, \ \text{but otherwise is biased. However, } \hat{\beta}_1 \ \text{is unbiased regardless of Bias.}
\end{align*} \]
Consider the covariance matrix of OLS estimator $\hat{\beta}$ as

$$E[\hat{\beta} - \beta][\hat{\beta} - \beta]^\top = \sigma^2 (X^\top X)^{-1}$$

$$= \sigma^2 \begin{bmatrix} x_1^\top x_1 & x_1^\top x_2 \\ x_2^\top x_1 & x_2^\top x_2 \end{bmatrix} \quad \cdots (5.8.12)$$

Thus, $\text{Var}(\hat{\beta}_1) = \sigma^2 \left\{ (x_1^\top x_1)^{-1} + (x_1^\top x_2)^{-1} (x_1^\top x_2) \left[ x_2^\top x_2 - (x_2^\top x_1) \right] \right\} \quad \cdots (5.8.13)$

$$\left( x_1^\top x_1 \right)^{-1} \left( x_1^\top x_2 \right)^{-1} \left( x_2^\top x_1 \right) \left( x_1^\top x_1 \right)^{-1} \right\} \quad \cdots (5.8.13)$$

Where $\beta_2 \neq 0$, the mean squared error matrix of $\beta_1^\ast$ is given by

$$\text{MSE} (\beta_1^\ast) = \sigma^2 (x_1^\top x_1)^{-1} + (x_1^\top x_2)^{-1} x_1^\top x_2 \beta_2 \beta_2^\top x_2^\top x_1 \left( x_1^\top x_1 \right)^{-1} \quad \cdots (5.8.14)$$

or

$$\text{MSE} (\beta_1^\ast) = \text{Var}(\beta_1^\ast) + \eta \eta^\top \quad \cdots (5.8.15)$$
Also, the mean squared error matrix of the OLS estimator \( \hat{\beta}_1 \) is given by

\[
\text{MSE} (\hat{\beta}_1) = \text{Var} (\hat{\beta}_1) = \sigma^2 \left\{ (X_1'X_1)^{-1} + (X_1'X_2)^{-1} (X_1'X_2) \left[ x_{1}^2 x_2 - (x_{1}^2 x_1) \right] x_1 \right\} (X_1'X_1)^{-1} \left( X_1'X_1 \right)^{-1} \left( X_1'X_1 \right)^{-1} \left( X_1'X_1 \right)^{-1} \cdots \tag{5.8.16}
\]

Consider the difference between the MSE matrices of \( \hat{\beta}_1 \) and \( \beta_1^* \) as

\[
\Delta_1 = \text{MSE} (\hat{\beta}_1) - \text{MSE} (\beta_1^*) = \left( x_1' x_1 \quad x_1' x_2 \right)
\]

\[
\begin{bmatrix}
    x_1' x_2 - (x_1' x_1)^{-1} (x_1' x_2) & -\beta_2 \beta_2^t \\
    \quad & x_1 \end{bmatrix} (X_1'X_1)^{-1}
\]

\[
\Delta_1 = \left[ (X_1'X_1)^{-1} (X_1'X_2) \right] \left[ \text{Var} (\hat{\beta}_2) - \beta_2 \beta_2^t \right] (X_1'X_1)^{-1} \left( X_1'X_1 \right)^{-1} \left( X_1'X_1 \right)^{-1} \cdots \tag{5.8.17}
\]

The matrix \( \Delta_1 \) will be positive semidefinite if \( \left[ \text{Var} (\hat{\beta}_2) - \beta_2 \beta_2^t \right] \) is positive semidefinite.

245
Under the specification test, one may state the null and alternative hypotheses as

\[ H_0: \eta = 0 \sim H_1: \eta \neq 0 \]

or

\[ H_0: \beta_2 = 0 \sim H_1: \beta_2 \neq 0 \]

The specification test statistic is defined as

\[
Q = (\hat{\beta}_1 - \beta_1^*)\left[\text{var} (\hat{\beta}_1) \text{var} (\beta_1^*)\right]^{-1} (\hat{\beta}_1 - \beta_1^*) \ldots (5.8.18)
\]

\[
Q = (\hat{\beta}_1 - \beta_1^*)\left\{\sigma^2 \left( x_1^T X_1 \right)^{-1} \left( x_1^T X_2 \right) \left[ X_2^T X_2 - \left( x_2^T X_1 \right) \left( x_1^T X_1 \right)^{-1} \left( x_1^T X_2 \right) \right] \right. \left( x_2^T X_1 \right) \left( x_1^T X_1 \right)^{-1} \left( x_1^T X_2 \right) \right\}^{-1} (\hat{\beta}_1 - \beta_1^*) \ldots (5.8.19)
\]

When \( H_0 : \eta = 0 \) true, the test statistic \( Q \) follows a central \( \chi^2 \) distribution with \( k_1 \) degrees of freedom.
The inverse matrix \( [\text{Var}(\hat{\beta}_1) - \text{Var}(\beta_1^*)]^{-1} \) exists since rank of \( (X_1'X_2) \) is assumed to be \( k_1 \). This assumption can be relaxed and the specification test statistic can be written as

\[
Q = (\hat{\beta}_1 - \beta_1^*)' [\text{Var}(\hat{\beta}_1) - \text{Var}(\beta_1^*)]^{-1} (\hat{\beta}_1 - \beta_1^*) \quad \ldots (5.8.20)
\]

Where \( [\text{Var}(\hat{\beta}_1) - \text{Var}(\beta_1^*)]^{-1} \) is the generalized inverse matrix of \( \text{Var}(\hat{\beta}_1) - \text{Var}(\beta_1^*) \).

In the test statistic, (5.8.19), the unknown variance \( \sigma^2 \) may be replaced with any of the following estimators:

\[
(i) \quad \hat{\sigma}^2 = S^2 = \frac{(y - X\hat{\beta})' (y - X\hat{\beta})}{n - k} = \frac{I_n - X' X}{n - k} X
\]

\[
= \text{Unbiased estimate of } \sigma^2 \quad \ldots (5.8.21)
\]
(ii) \( \sigma^2 = \frac{(Y - X\hat{\beta})' (Y - X\hat{\beta})}{n - k_1} = I_{n - X_1}^{-1} X_1' \quad ... (5.8.22) \)

= Estimate of \( \sigma^2 \) based on the restricted least squares estimator \( \beta^* \). Unless \( \beta_2 = 0 \), \( \sigma^2_{\beta^*} \) yields a biased estimate of \( \sigma^2 \). \( \sigma^2_{\beta^*} \) has positive bias.

(iii) \( \hat{\sigma}^2 = \frac{(Y - X\hat{\beta})' (Y - X\hat{\beta})}{n - k + 2} = I_{n - X}^{-1} X' \quad ... (5.8.23) \)

= Minimum MSE estimate of \( \sigma^2 \).