

Stata 12 Tutorial 7

TOPIC: Linear Regression Models with Heteroskedastic Errors: Inference

DATA: **auto1.raw** (a text-format (ASCII) data file)
auto1.dta (a *Stata*-format dataset first created in *Stata 12 Tutorial 1*)

TASKS: *Stata 12 Tutorial 7* demonstrates how to use *Stata* to compute *heteroskedasticity-consistent estimates of the OLS coefficient covariance matrix* and how to perform *heteroskedasticity-robust hypothesis tests* based on OLS coefficient estimates. It uses for these purposes the regression models and hypothesis tests introduced in *Stata 12 Tutorial 2*. In addition, *Stata 12 Tutorial 7* demonstrates how to perform the **Breusch-Pagan-Koenker (BPK) LM tests for mixed heteroskedastic errors** in linear regression models.

- The **Stata commands** that constitute the primary subject of this tutorial are:

regress	Used with robust option to compute <i>heteroskedasticity-consistent</i> estimates of the OLS coefficient covariance matrix.
test	Used to compute <i>heteroskedasticity-robust F-tests</i> of linear coefficient equality restrictions after OLS estimation.
lincom	Used after estimation to compute <i>heteroskedasticity-robust two-tail t-tests</i> of exclusion (zero) restrictions on individual regression coefficients.

- The **Stata statistical functions** used in this tutorial are:

ttail (df, t_0)	Computes <i>right-tail p-values of calculated t-statistics</i> .
invttail (df, p)	Computes <i>critical values of t-distributions</i> .
Ftail (df_1, df_2, F_0)	Computes <i>right-tail p-values of calculated F-statistics</i> .
invFtail (df_1, df_2, α)	Computes <i>critical values of F-distributions</i> .
chi2tail (df, c_0)	Computes <i>right-tail p-values of calculated chi-square statistics</i> .
invchi2tail (df, α)	Computes <i>critical values of chi-square distributions</i> .

NOTE: *Stata* commands are *case sensitive*. All *Stata* command names must be typed in the Command window in **lower case letters**.

❑ Preparing for Your *Stata* Session

Before beginning your *Stata* session, use Windows Explorer to copy the *Stata*-format data set **auto1.dta** to the *Stata working directory* on the C:-drive or D:-drive of the computer at which you are working.

- **On the computers in Dunning 350**, the default *Stata* working directory is usually C:\data.
- **On the computers in MC B111**, the default *Stata* working directory is usually D:\courses.

❑ Start Your *Stata* Session

To start your *Stata* session, double-click on the ***Stata* icon** on the Windows desktop.

After you double-click the ***Stata* icon**, you will see the familiar screen of four *Stata* windows.

❑ Record Your *Stata* Session and *Stata* Commands -- log using, cmdlog using

To record your *Stata* session, including all the *Stata* commands you enter and the results (output) produced by these commands, make a text-format **.log** file named **452tutorial7.log**. To open (begin) the log file **452tutorial7.log**, enter in the Command window:

```
log using 452tutorial7.log
```

This command opens a text-format (ASCII) file called **452tutorial7.log** in the current *Stata* working directory.

Note: It is important to include the **.log** file extension when opening a log file; if you do not, your log file will be in smcl format, a format that only *Stata* can read. Once you have opened the **452tutorial7.log** file, a copy of all the commands you enter during your *Stata* session and of all the results they produce is recorded in that **452tutorial7.log** file.

To record only the *Stata* commands you type during your *Stata* session, use the *Stata* **cmdlog using** command. To start (open) the command log file **452tutorial7.txt**, enter in the Command window:

```
cmdlog using 452tutorial7
```

This command opens a plain text-format (ASCII) file called **452tutorial7.txt** in the current *Stata* working directory. All commands you enter during your *Stata* session are recorded in this file.

❑ Loading a *Stata*-Format Dataset into *Stata* -- use

In *Stata 12 Tutorial 1*, you created the *Stata*-format dataset **auto1.dta**. If you saved the dataset **auto1.dta** on your own diskette and copied this dataset to the *Stata* working directory before beginning your current *Stata* session, you can simply use the **use** command to read or load **auto1.dta** into memory. If, however, you did not save the *Stata*-format dataset **auto1.dta** you created during *Stata 12 Tutorial 1* and bring it with you on a flash memory stick (or other electronic portable storage device), you will have to repeat most of the section of *Stata 12 Tutorial 1* that explains how to create a *Stata*-format dataset from a text-format data file.

To load, or read, into memory the *Stata*-format dataset **auto1.dta**, type in the Command window:

```
use auto1
```

This command loads into memory the *Stata*-format dataset **auto1.dta**.

To summarize the contents of the current dataset, use the **describe** command. Recall from *Stata 12 Tutorial 1* that the **describe** command displays a summary of the contents of the current dataset in memory, which in this case is the *Stata*-format data file **auto1.dta**. Type in the Command window:

```
describe
```

To compute summary statistics for the variables in the current dataset, use the **summarize** command. Type in the Command window:

```
summarize
```

❑ **OLS Linear Regression with Heteroskedasticity-Robust Coefficient Covariance Matrix and Hypothesis Tests -- *regress* with *robust* option**

Model 1:

$$\text{price}_i = \beta_0 + \beta_1 \text{wgt}_i + \beta_2 \text{mpg}_i + u_i \quad (1)$$

where

price_i = the price of the i -th car (in US dollars);

wgt_i = the weight of the i -th car (in pounds);

mpg_i = the miles per gallon (fuel efficiency) for the i -th car (in miles per gallon).

- To estimate by OLS the linear regression model given by PRE (1) and display the OLS estimate of the coefficient covariance matrix, enter the following commands:

```
regress price wgt mpg
vce
```

- To save and list the OLS estimate of the coefficient covariance matrix, enter the following **matrix** commands:

```
matrix V1OLS = e(V)
matrix list V1OLS
```

- To estimate by OLS the linear regression model given by PRE (1) and compute a ***heteroskedasticity-consistent estimate of the OLS coefficient covariance matrix***, enter the following **regress** command with the **robust** option:

```
regress price wgt mpg, robust
```

- To display the ***heteroskedasticity-consistent estimate of the OLS coefficient covariance matrix*** for Model 1, enter the following command:

```
vce
```

- To save and list the ***heteroskedasticity-consistent estimate of the OLS coefficient covariance matrix*** for Model 1, enter the following commands:

```
matrix V1HC1 = e(V)
matrix list V1HC1
```

- Enter the following **display** commands to display in the Results window the values of the *heteroskedasticity-robust coefficient standard errors* $\hat{s}e(\hat{\beta}_0)$, $\hat{s}e(\hat{\beta}_1)$ and $\hat{s}e(\hat{\beta}_2)$ from the OLS sample regression equation for Model 1:

```
display _se[_cons]
display _se[wgt]
display _se[mpg]
```

- The following **scalar** commands name and save the values of the *OLS coefficient estimates* $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ for Model 1. Enter the commands:

```
scalar b0 = _b[_cons]
scalar b1 = _b[wgt]
scalar b2 = _b[mpg]
```

- The following **scalar** commands name and save the values of the *estimated heteroskedasticity-robust standard errors* for the OLS coefficient estimates $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ in Model 1. Enter the commands:

```
scalar seb0 = _se[_cons]
scalar seb1 = _se[wgt]
scalar seb2 = _se[mpg]
```

- You may also wish to generate the *heteroskedasticity-robust variance estimates* of the OLS coefficient estimates $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ for Model 1; these are equal to the squares of the corresponding estimated standard errors. Enter the following **scalar** commands to do this:

```
scalar varb0 = seb0^2
scalar varb1 = seb1^2
scalar varb2 = seb2^2
```

- To list the values of *all* currently-defined scalars, enter the following command:

```
scalar list
```

- To list only the values of the heteroskedasticity-consistent standard errors and variance estimates of the slope coefficient estimates in Model 1, enter the following command:

```
scalar list seb1 seb2 varb1 varb2
```

❑ Computing Heteroskedasticity-Robust Wald F-Tests of Restrictions on Individual Regression Coefficients -- test

Consider again the following two-tail hypothesis test on the regression coefficient β_j :

Null hypothesis is $H_0: \beta_j = b_j$

Alternative hypothesis is $H_1: \beta_j \neq b_j$

where b_j is a specified constant.

t-statistics for individual coefficient estimates

The **t-statistic** for regression coefficient estimate $\hat{\beta}_j$ takes the general form:

$$t(\hat{\beta}_j) = \frac{\hat{\beta}_j - \beta_j}{\widehat{se}(\hat{\beta}_j)} \sim t[N - K]. \quad (2)$$

The **calculated t-statistic** for testing $H_0: \beta_j = b_j$ against $H_1: \beta_j \neq b_j$ is obtained by setting $\beta_j = b_j$ in formula (2) for $t(\hat{\beta}_j)$:

$$t_0(\hat{\beta}_j) = \frac{\hat{\beta}_j - b_j}{\widehat{se}(\hat{\beta}_j)} \sim t[N - K] \text{ under the null hypothesis } H_0: \beta_j = b_j \quad (3)$$

F-statistics for individual coefficient estimates

The **F-statistic** for regression coefficient estimate $\hat{\beta}_j$ takes the general form:

$$F(\hat{\beta}_j) = \frac{(\hat{\beta}_j - \beta_j)^2}{\widehat{Var}(\hat{\beta}_j)} \sim F[1, N - K]. \quad (4)$$

The **calculated F-statistic** for testing $H_0: \beta_j = b_j$ against $H_1: \beta_j \neq b_j$ is obtained by setting $\beta_j = b_j$ in formula (4) for $F(\hat{\beta}_j)$:

$$F_0(\hat{\beta}_j) = \frac{(\hat{\beta}_j - b_j)^2}{\text{Var}(\hat{\beta}_j)} \sim F[1, N - K] \text{ under the null hypothesis } H_0: \beta_j = b_j \quad (5)$$

Equivalence of Two-Tail t-tests and F-tests of Individual Coefficients

Two-tail t-tests and F-tests of $H_0: \beta_j = b_j$ against $H_1: \beta_j \neq b_j$ are equivalent.

Reasons:

1. The t-statistic and F-statistic for the coefficient estimate $\hat{\beta}_j$ are related as follows:

$$F(\hat{\beta}_j) = \frac{(\hat{\beta}_j - \beta_j)^2}{\text{Var}(\hat{\beta}_j)} = \frac{(\hat{\beta}_j - \beta_j)^2}{(\text{sê}(\hat{\beta}_j))^2} = \left(\frac{\hat{\beta}_j - \beta_j}{\text{sê}(\hat{\beta}_j)} \right)^2 = (t(\hat{\beta}_j))^2. \quad (6)$$

2. The **calculated sample values** of the t- and F-statistics under the null hypothesis $H_0: \beta_j = b_j$ are therefore related as follows:

$$F_0(\hat{\beta}_j) = \frac{(\hat{\beta}_j - b_j)^2}{\text{Var}(\hat{\beta}_j)} = \frac{(\hat{\beta}_j - b_j)^2}{(\text{sê}(\hat{\beta}_j))^2} = \left(\frac{\hat{\beta}_j - b_j}{\text{sê}(\hat{\beta}_j)} \right)^2 = (t_0(\hat{\beta}_j))^2.$$

3. The **two-tail critical values** of the t- and F-statistics at **significance level α** are related as follows:

$$F_{\alpha}[1, N - K] = t_{\alpha/2}[N - K]. \quad (7)$$

To demonstrate the equivalence of **two-tail t-tests** and **F-tests**, use the *Stata* **test** and **lincom** commands to perform F-tests and t-tests of some two-tail hypothesis tests on the coefficients of Model 1.

- ♦ **Test 1:** $H_0: \beta_1 = 0$ versus $H_1: \beta_1 \neq 0$ in Model 1
- To calculate a ***heteroskedasticity-robust F-test*** of this hypothesis and the corresponding p-value for the calculated F-statistic, enter the following **test** and **return list** commands:

```
test wgt      or      test wgt = 0
return list
```

- To calculate a ***heteroskedasticity-robust two-tail t-test*** of H_0 against H_1 , enter the following **lincom** and **return list** commands:

```
lincom _b[wgt]
return list
```

Compare the results of this two-tail t-test with those of the previous F-test. You should be able to explain why the two tests are equivalent.

- To see explicitly how the above **lincom** command computes the ***heteroskedasticity-robust two-tail t-test*** of H_0 against H_1 , use a **scalar** command to compute the appropriate t-statistic and a **display** command with the **ttail(df, t₀)** function to compute the two-tail p-value of the calculated t-statistic. Enter the commands:

```
scalar tforb1eq0 = b1/seb1
scalar list b1 seb1 tforb1eq0
display 2*ttail(e(df_r), abs(tforb1eq0))
display e(df_r)
```

Note that **e(df_r)** is the temporarily-saved value of the degrees of freedom for the residual sum-of-squares (RSS) computed by the previous **regress** command; this is the value of $N - K = 74 - 3 = 71$, the degrees of freedom for the null distribution of the t-statistic in Model 1. Compare the results of these commands with those produced by the preceding **lincom** command. You will see that they are identical.

- ♦ **Test 2:** $H_0: \beta_2 = 0$ versus $H_1: \beta_2 \neq 0$ in Model 1
- To calculate a ***heteroskedasticity-robust F-test*** of this hypothesis and the corresponding p-value for the calculated F-statistic, enter the following **test** and **return list** commands:

```
test mpg      or    test mpg = 0
return list
```

- To calculate a ***heteroskedasticity-robust two-tail t-test*** of H_0 against H_1 , enter the following **lincom** and **return list** commands:

```
lincom _b[mpg]
return list
```

Compare the results of this two-tail t-test with those of the previous F-test. Again, you should be able to explain why the two tests are equivalent.

- An alternative way to compute the ***heteroskedasticity-robust two-tail t-test*** of H_0 against H_1 is to use a **scalar** command to compute the appropriate t-statistic and a **display** command with the **ttail(df, t₀)** function to compute the two-tail p-value of the calculated t-statistic. Enter the commands:

```
scalar tforb2eq0 = b2/seb2
scalar list b2 seb2 tforb2eq0
display 2*ttail(e(df_r), abs(tforb2eq0))
display e(df_r)
```

Compare the results of the above commands with those produced by the preceding **lincom** command. You will see that they are identical.

- ♦ **Test 3:** $H_0: \beta_1 = 1$ versus $H_1: \beta_1 \neq 1$ in Model 1
- To calculate a ***heteroskedasticity-robust F-test*** of this hypothesis and the corresponding p-value for the calculated F-statistic, enter the following **test** and **return list** commands:

```
test wgt = 1
return list
```

- One way to compute a **heteroskedasticity-robust two-tail t-test** of this hypothesis and the corresponding **two-tail p-value** for the calculated t-statistic is to use the saved results from the previous **test** command and the relationship between the t-statistic and corresponding F-statistic. To compute the t-statistic and display its value, enter the following commands:

```
scalar tb1eq1 = sqrt(r(F))
scalar list tb1eq1
```

- Use the **ttail(df, t₀)** function to compute the **two-tail p-value** for the calculated sample value of the t-statistic produced by the preceding **scalar** command. Enter the following **display** commands, which are equivalent.

```
display 2*ttail(r(df_r), abs(tb1eq1))
display 2*ttail(71, abs(tb1eq1))
```

Note that **r(df_r)** is the temporarily-saved value of the denominator degrees of freedom for the F-statistic computed by the preceding **test** command.

- An alternative way to compute the **heteroskedasticity-robust two-tail t-test** of H_0 against H_1 is to use a **scalar** command to compute the appropriate t-statistic and a **display** command with the **ttail(df, t₀)** function to compute the two-tail p-value of the calculated t-statistic. Enter the commands:

```
scalar tforb1eq1 = (b1 - 1)/seb1
scalar list b1 seb1 tforb1eq1
display 2*ttail(e(df_r), abs(tforb1eq1))
display e(df_r)
```

Compare the results of these commands with those produced by the preceding **scalar** command for **tb1eq1**. You will see that they are identical.

- The simplest way to compute this **heteroskedasticity-robust two-tail t-test** of H_0 against H_1 is to use a **lincom** command to compute the appropriate t-statistic and the two-tail p-value of the calculated t-statistic. To do this, rewrite H_0 and H_1 of Test 3 as follows:

$$\begin{aligned} H_0: \beta_1 = 1 &\Rightarrow \beta_1 - 1 = 0 \\ H_1: \beta_1 \neq 1 &\Rightarrow \beta_1 - 1 \neq 0 \end{aligned}$$

Enter the following commands:

```
lincom _b[wgt] - 1
return list
display r(estimate)/r(se)
display 2*ttail(r(df), abs(r(estimate)/r(se)))
```

Compare the results of these commands with those produced by the preceding two methods of computing a **heteroskedasticity-robust two-tail t-test** of Test 3. You will see that they are identical.

- ♦ **Test 4:** $H_0: \beta_1 = 2$ versus $H_1: \beta_1 \neq 2$ in Model 1
- To calculate a **heteroskedasticity-robust F-test** of this hypothesis and the corresponding **p-value** for the calculated F-statistic, enter the following **test** and **return list** commands:

```
test wgt = 2
return list
```

- To compare the results of the above F-test with those of a **heteroskedasticity-robust two-tail t-test** of the same hypothesis, enter the following commands:

```
scalar tb1eq2 = sqrt(r(F))
scalar list tb1eq2
display 2*ttail(r(df_r), abs(tb1eq2))
display 2*ttail(71, abs(tb1eq2))
```

- The simplest way to compute a **heteroskedasticity-robust two-tail t-test** of H_0 against H_1 is to use a **lincom** command to compute the appropriate t-statistic and the two-tail p-value of the calculated t-statistic. To do this, rewrite H_0 and H_1 of Test 4 as follows:

$$\begin{aligned} H_0: \beta_1 = 2 &\Rightarrow \beta_1 - 2 = 0 \\ H_1: \beta_1 \neq 2 &\Rightarrow \beta_1 - 2 \neq 0 \end{aligned}$$

Enter the following commands:

```
lincom _b[wgt] - 2
return list
```

```
display r(estimate)/r(se)
display 2*ttail(r(df), abs(r(estimate)/r(se)))
```

Compare the results of these commands with those produced by the preceding method of computing a ***heteroskedasticity-robust two-tail t-test*** of Test 4. You will see that they are identical.

❑ Computing Two-Tail t-tests and F-tests of *One Linear Restriction on Two Coefficients*

Nature: Tests of hypotheses that take the form of linear combinations of regression coefficients arise frequently in applied econometrics. This section presents some examples of how to perform such hypothesis tests using *Stata*.

To illustrate the nature of such hypotheses, consider again the PRE for Model 1:

$$\text{price}_i = \beta_0 + \beta_1 \text{wgt}_i + \beta_2 \text{mpg}_i + u_i. \quad (1)$$

A linear combination of the slope coefficients β_1 and β_2 in regression equation (1) takes the general form

$$c_1\beta_1 + c_2\beta_2 \text{ where } c_1 \text{ and } c_2 \text{ are specified constants.}$$

Examples:

- To re-display the OLS sample regression equation for Model 1, enter the following **regress** command:

```
regress
```

- ♦ **Test 5:** Test the proposition that the ***marginal effect of wgt_i on price_i*** equals the ***marginal effect of mpg_i on price_i*** in Model 1 (although there is no good reason to think these two marginal effects would be equal).
- The **marginal effect of wgt_i on price_i in Model 1** is obtained by partially differentiating regression equation (1) with respect to wgt_i:

$$\frac{\partial \text{price}_i}{\partial \text{wgt}_i} = \beta_1$$

- The **marginal effect of mpg_i on price_i in Model 1** is obtained by partially differentiating regression equation (1) with respect to mpg_i .

$$\frac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_2$$

- The **null and alternative hypotheses** are:

$$H_0: \beta_1 = \beta_2 \quad \Rightarrow \quad \beta_1 - \beta_2 = 0$$

$$H_1: \beta_1 \neq \beta_2 \quad \Rightarrow \quad \beta_1 - \beta_2 \neq 0$$

- The following **test** commands computes a **heteroskedasticity-robust Wald F-test** of H_0 against H_1 . Enter the commands:

```
test wgt = mpg    or    test wgt - mpg = 0
return list
```

Inspect the results generated by this command. State the inference you would draw from this F-test.

- The following **display** command displays the **sample value of the calculated F-statistic** computed by preceding **test** command:

```
display r(F)
```

- The following **lincom** command computes a **heteroskedasticity-robust two-tail t-test** of H_0 against H_1 . Enter the commands:

```
lincom _b[wgt] - _b[mpg]
return list
```

Inspect the results generated by this command. State the inference you would draw from this t-test.

- The following **display** command displays the *sample value of the calculated t-statistic* computed by the preceding **lincom** command:

```
display r(estimate)/r(se)
```

- The following **display** command displays the *square* of the sample value of the calculated t-statistic computed by the preceding **lincom** command:

```
display (r(estimate)/r(se))^2
```

You should understand this last **display** command confirms the equivalence of the two-tail t- and F-tests you have just performed.

- ♦ **Test 6:** Test the proposition that wgt_i and mpg_i have *equal but opposite marginal effects on price_i* in Model 1 -- i.e., that the *marginal effects on price_i* of wgt_i and mpg_i are *offsetting*.
- The *null and alternative hypotheses* are:

$$H_0: \beta_1 = -\beta_2 \quad \Rightarrow \quad \beta_1 + \beta_2 = 0$$

$$H_1: \beta_1 \neq -\beta_2 \quad \Rightarrow \quad \beta_1 + \beta_2 \neq 0$$

- The following **test** commands compute a *heteroskedasticity-robust Wald F-test* of H_0 against H_1 . Enter the commands:

```
test wgt = -mpg      or      test wgt + mpg = 0  
return list  
display r(F)
```

Inspect the results generated by these commands. State the inference you would draw from this F-test.

- The following **lincom** command computes a *heteroskedasticity-robust two-tail t-test* of H_0 against H_1 . Enter the commands:

```
lincom _b[wgt] + _b[mpg]  
return list  
display r(estimate)/r(se)  
display (r(estimate)/r(se))^2
```

Inspect the results generated by these commands. State the inference you would draw from this t-test.

Compare the results of the two-tail F-test and two-tail t-test performed above. Are these two tests equivalent? Why or why not?

□ **Interaction Variables: Squares and Cross Products of Continuous Variables**

Nature: *Interactions between two continuous variables* refer to products of explanatory variables. For example, if X_{ij} and X_{ih} are two continuous explanatory variables, the interaction term between them is the product $X_{ij}X_{ih}$. Alternatively, the interaction of the variable X_{ij} with itself is simply the product $X_{ij}X_{ij} = X_{ij}^2$. As you have learned, inclusion of these regressors in a linear regression model allows for nonconstant marginal effects of the explanatory variables on the conditional mean of the dependent variable Y .

Usage: *Interaction terms* between *continuous variables* allow the marginal effect of one explanatory variable to be a linear function of both itself and the other explanatory variable.

Consider the following generic linear regression equation:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i1}^2 + \beta_4 X_{i2}^2 + \beta_5 X_{i1}X_{i2} + u_i. \quad (8)$$

Formally, the population regression function $E(Y_i | X_{i1}, X_{i2}) = f(X_{i1}, X_{i2})$ in PRE (8) can be derived as a second-order Taylor series approximation to the function $f(X_{i1}, X_{i2})$.

Compare the marginal effects on Y of the two explanatory variables X_1 and X_2 in equation (8).

- The **marginal effect of X_1** in equation (8) is:

$$\begin{aligned} \frac{\partial Y_i}{\partial X_{i1}} &= \frac{\partial E(Y_i | X_{i1}, X_{i2})}{\partial X_{i1}} = \beta_1 + 2\beta_3 X_{i1} + \beta_5 X_{i2} \\ &= \text{a linear function of both } X_{i1} \text{ and } X_{i2} \end{aligned}$$

- The **marginal effect of X_2** in equation (8) is:

$$\frac{\partial Y_i}{\partial X_{i2}} = \frac{\partial E(Y_i | X_{i1}, X_{i2})}{\partial X_{i2}} = \beta_2 + 2\beta_4 X_{i2} + \beta_5 X_{i1}$$

= a linear function of *both* X_{i1} and X_{i2}

Example: An example of the generic regression model (8) is the following population regression equation for North American car prices:

Model 2:

$$\text{price}_i = \beta_0 + \beta_1 \text{wgt}_i + \beta_2 \text{mpg}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i^2 + \beta_5 \text{wgt}_i \text{mpg}_i + u_i. \quad (9)$$

where

price_i = the price of the i -th car (in US dollars);

wgt_i = the weight of the i -th car (in pounds);

wgt_i^2 = the square of wgt_i ;

mpg_i = the miles per gallon (fuel efficiency) for the i -th car (in miles per gallon);

mpg_i^2 = the square of mpg_i ;

$\text{wgt}_i \text{mpg}_i$ = the product of wgt_i and mpg_i for the i -th car.

- Before estimating regression equation (9), check to see that your dataset includes as variables the squared terms wgt_i^2 and mpg_i^2 , and the interaction term $\text{wgt}_i \text{mpg}_i$. If it does not, create these regressors by entering the following **generate** commands followed by a **summarize** command to check your work:

```
generate wgtsq = wgt*wgt
generate mpgsq = mpg^2
generate wgtmpg = wgt*mpg
summarize wgt mpg wgtsq mpgsq wgtmpg
```

- Now estimate by OLS regression equation (9), first without the **robust** option and then with the **robust** option. Enter the following **regress** commands:

```
regress price wgt mpg wgtsq mpgsq wgtmpg
regress price wgt mpg wgtsq mpgsq wgtmpg, robust
```

Carefully inspect the results of these two **regress** commands. How do the results of these two **regress** commands differ? In what respects are their results similar?

- ♦ **Test 7:** Test the hypothesis that the *marginal effect of wgt_i on $price_i$ is zero* in Model 2 given by regression equation (9).
- The marginal effect of wgt_i on $price_i$ is obtained by partially differentiating regression equation (9) with respect to wgt_i :

$$\frac{\partial price_i}{\partial wgt_i} = \beta_1 + 2\beta_3 wgt_i + \beta_5 mpg_i.$$

- Sufficient conditions for $\partial price_i / \partial wgt_i = 0$ for all i are $\beta_1 = 0$ and $\beta_3 = 0$ and $\beta_5 = 0$. We therefore want to test these three coefficient exclusion restrictions.
- The **null and alternative hypotheses** are:

$$H_0: \beta_1 = 0 \text{ and } \beta_3 = 0 \text{ and } \beta_5 = 0$$

$$H_1: \beta_1 \neq 0 \text{ and/or } \beta_3 \neq 0 \text{ and/or } \beta_5 \neq 0.$$

- The following **test** command computes a **heteroskedasticity-robust Wald F-test** of H_0 against H_1 . Enter the commands:

```
test wgt wgtsq wgtmpg
return list
```

Inspect the results generated by this command. State the inference you would draw from this test.

- Use a **lincom** command to compute the estimated value of the marginal effect of wgt_i in equation (9) for a car that weighs 2,500 pounds and gets 25 miles to the gallon. Recall that the expression for $\partial price_i / \partial wgt_i$ in equation (9) is:

$$\frac{\partial price_i}{\partial wgt_i} = \beta_1 + 2\beta_3 wgt_i + \beta_5 mpg_i.$$

Evaluated at $\text{wgt}_i = 2500$ and $\text{mpg}_i = 25$, the *estimated* marginal effect of wgt_i on price_i is:

$$\text{estimate of } \frac{\partial \text{price}_i}{\partial \text{wgt}_i} = \hat{\beta}_1 + 2\hat{\beta}_3(2500) + \hat{\beta}_5(25).$$

- When you compute the estimated marginal effect of an explanatory variable for specified values of the variables on which it depends, you usually also want to test the hypothesis that the true marginal effect equals zero. In the present case, the null and alternative hypotheses are:

$$H_0: \frac{\partial \text{price}_i}{\partial \text{wgt}_i} = \beta_1 + 2\beta_3(2500) + \beta_5(25) = 0$$

$$H_1: \frac{\partial \text{price}_i}{\partial \text{wgt}_i} = \beta_1 + 2\beta_3(2500) + \beta_5(25) \neq 0$$

Either a **two-tail t-test** or an **F-test** is appropriate for performing this hypothesis test.

- The following **lincom** command computes the estimated marginal effect of wgt_i in equation (9) for a car that weighs 2,500 pounds and gets 25 miles to the gallon, and computes a **heteroskedasticity-robust two-tail t-test** of the null hypothesis that this marginal effect equals zero. Enter the **lincom** command:

```
lincom _b[wgt] + 2*_b[wgtsq]*2500 + _b[wgtmpg]*25
```

Inspect the output generated by the above **lincom** command. Would you infer that the marginal effect of wgt_i on price_i is different from zero for a car that weighs 2,500 pounds and has fuel efficiency equal to 25 miles per gallon?

- The following **test** command computes a **heteroskedasticity-robust Wald F-test** of the null hypothesis that the marginal effect of wgt_i equals zero for cars with $\text{wgt}_i = 2,500$ pounds and $\text{mpg}_i = 25$ miles per gallon. Enter the commands:

```
test wgt + 2*wgtsq*2500 + wgtmpg*25 = 0  
return list
```

This F-test is equivalent to the preceding two-tail t-test. Can you explain why?

- ♦ **Test 8:** Test the hypothesis that the *marginal effect of mpg_i on price_i is zero* in Model 2 given by regression equation (9).

- The marginal effect of mpg_i on price_i is obtained by partially differentiating regression equation (9) with respect to mpg_i.

$$\frac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_2 + 2\beta_4 \text{mpg}_i + \beta_5 \text{wgt}_i.$$

- Sufficient conditions for $\partial \text{price}_i / \partial \text{mpg}_i = 0$ for all i are $\beta_2 = 0$ and $\beta_4 = 0$ and $\beta_5 = 0$. We therefore want to test these three coefficient restrictions.
- The **null and alternative hypotheses** are:

$$H_0: \beta_2 = 0 \text{ and } \beta_4 = 0 \text{ and } \beta_5 = 0$$

$$H_1: \beta_2 \neq 0 \text{ and/or } \beta_4 \neq 0 \text{ and/or } \beta_5 \neq 0.$$

- The following **test** command computes a **heteroskedasticity-robust Wald F-test** of H_0 against H_1 . Enter the commands:

```
test mpg mpgsq wgtmpg
return list
```

Inspect the results generated by this command. State the inference you would draw from this test.

- Use a **lincom** command to compute the estimated value of the marginal effect of mpg_i in equation (9) for a car that weighs 2,500 pounds and gets 25 miles to the gallon, and to perform a **heteroskedasticity-robust two-tail t-test** of the hypothesis that this marginal effect of mpg_i equals zero.

Recall that the expression for $\partial \text{price}_i / \partial \text{mpg}_i$ in equation (9) is:

$$\frac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_2 + 2\beta_4 \text{mpg}_i + \beta_5 \text{wgt}_i.$$

Evaluated at $\text{wgt}_i = 2500$ and $\text{mpg}_i = 25$, the *estimated* marginal effect of mpg_i on price_i is:

$$\text{estimate of } \frac{\partial \text{price}_i}{\partial \text{mpg}_i} = \hat{\beta}_2 + 2\hat{\beta}_4(25) + \hat{\beta}_5(2500).$$

To test the hypothesis that the true marginal effect of mpg_i equals zero for cars for which $\text{wgt}_i = 2500$ and $\text{mpg}_i = 25$, perform a two-tail t-test or F-test of:

$$H_0: \frac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_2 + 2\beta_4(25) + \beta_5(2500) = 0$$

$$H_1: \frac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_2 + 2\beta_4(25) + \beta_5(2500) \neq 0$$

- The following **lincom** command computes the estimated marginal effect of mpg_i in equation (9) for a car that weighs 2,500 pounds and gets 25 miles to the gallon, and a **heteroskedasticity-robust two-tail t-test** of the null hypothesis that this marginal effect equals zero. Enter the **lincom** command:

```
lincom _b[mpg] + 2*_b[mpgsq]*25 + _b[wgtmpg]*2500
```

Would you infer that the marginal effect of mpg_i on price_i is different from zero for a car that weighs 2,500 pounds and has fuel efficiency equal to 25 miles per gallon?

- The following **test** command computes a **heteroskedasticity-robust Wald F-test** of the null hypothesis that the estimated marginal effect of mpg_i equals zero for a car that weighs 2,500 pounds and gets 25 miles to the gallon. Enter the **test** command:

```
test mpg + 2*mpgsq*25 + wgtmpg*2500 = 0  
return list
```

Compare the results of this F-test with those of the two-tail t-test performed above. They are equivalent tests. How do you know they are equivalent?

- ♦ **Test 9:** Test the hypothesis that the *marginal effect of wgt_i on $price_i$ is constant* in Model 2 given by regression equation (9).

- Recall that the marginal effect of wgt_i on $price_i$ is obtained by partially differentiating regression equation (9) with respect to wgt_i :

$$\frac{\partial price_i}{\partial wgt_i} = \beta_1 + 2\beta_3 wgt_i + \beta_5 mpg_i.$$

- The marginal effect of wgt_i on $price_i$ is equal to the constant β_1 if $\beta_3 = 0$ and $\beta_5 = 0$. Hence sufficient conditions for $\partial price_i / \partial wgt_i = \beta_1 = \text{a constant for all } i$ are $\beta_3 = 0$ and $\beta_5 = 0$. So we want to test these two coefficient exclusion restrictions.

- The **null and alternative hypotheses** are:

$$H_0: \beta_3 = 0 \text{ and } \beta_5 = 0$$

$$H_1: \beta_3 \neq 0 \text{ and/or } \beta_5 \neq 0.$$

- Since the above null hypothesis specifies **two** coefficient restrictions on regression equation (9), we must use an F-test. The following **test** command computes a **heteroskedasticity-robust Wald F-test** of H_0 against H_1 . Enter the commands:

```
test wgtsq wgtmpg
return list
```

Inspect the results generated by the above **test** command. State the inference you would draw from this test.

- ♦ **Test 10:** Test the hypothesis that the *marginal effect of mpg_i on $price_i$ is constant* in Model 2 given by regression equation (9).

- Recall that the marginal effect of mpg_i on $price_i$ is obtained by partially differentiating regression equation (9) with respect to mpg_i :

$$\frac{\partial price_i}{\partial mpg_i} = \beta_2 + 2\beta_4 mpg_i + \beta_5 wgt_i.$$

- The marginal effect of mpg_i on price_i is equal to the constant β_2 if $\beta_4 = 0$ and $\beta_5 = 0$. Hence sufficient conditions for $\partial \text{price}_i / \partial \text{mpg}_i = \beta_2 = \text{a constant for all } i$ are $\beta_4 = 0$ and $\beta_5 = 0$. We therefore want to test these three coefficient restrictions.

- The **null and alternative hypotheses** are:

$$H_0: \beta_4 = 0 \text{ and } \beta_5 = 0$$

$$H_1: \beta_4 \neq 0 \text{ and/or } \beta_5 \neq 0.$$

- Since the above null hypothesis specifies two coefficient restrictions on regression equation (9), we must use an F-test. The following **test** command computes a **heteroskedasticity-robust Wald F-test** of H_0 against H_1 . Enter the commands:

```
test mpgsq wgtmpg
return list
```

Inspect the results generated by this **test** command. State the inference you would draw from this test.

- ♦ **Test 11:** Test the hypothesis that the **marginal effect of wgt_i on price_i is unrelated to wgt_i** in Model 2 given by regression equation (9).

- Recall that the marginal effect of wgt_i on price_i is obtained by partially differentiating regression equation (9) with respect to wgt_i :

$$\frac{\partial \text{price}_i}{\partial \text{wgt}_i} = \beta_1 + 2\beta_3 \text{wgt}_i + \beta_5 \text{mpg}_i.$$

- A sufficient condition for $\partial \text{price}_i / \partial \text{wgt}_i$ to be unrelated to wgt_i for all i is $\beta_3 = 0$. We therefore want to test this coefficient exclusion restriction on PRE (9).

- The **null and alternative hypotheses** are:

$$H_0: \beta_3 = 0$$

$$H_1: \beta_3 \neq 0.$$

- The following **test** command computes a *heteroskedasticity-robust Wald F-test* of H_0 against H_1 . Enter the commands:

```
test wgtsq      or      test wgtsq = 0
return list
```

Inspect the results generated by this **test** command. State the inference you would draw from this F-test.

- The following **lincom** command computes a *heteroskedasticity-robust two-tail t-test* of H_0 against H_1 . Enter the commands:

```
lincom _b[wgtsq]
return list
```

Inspect the results generated by these commands. State the inference you would draw from this t-test.

Remarks: The above **lincom** command simply replicates the t-test of this same hypothesis that is reported by the **regress** command used to estimate regression equation (9) by OLS. You should be able to explain and verify from the above results why the two-tail t-test of $H_0: \beta_3 = 0$ versus $H_1: \beta_3 \neq 0$ is equivalent to the preceding F-test of the same hypothesis.

- ♦ **Test 12:** Consider the following two propositions:
 1. The *marginal effect of wgt_i on $price_i$ is unrelated to mpg_i* in Model 2 given by regression equation (9).
 2. The *marginal effect of mpg_i on $price_i$ is unrelated to wgt_i* in Model 2 given by regression equation (9).
- Recall that the marginal effects on $price_i$ of wgt_i and mpg_i in regression equation (9) are respectively:

$$\frac{\partial price_i}{\partial wgt_i} = \beta_1 + 2\beta_3 wgt_i + \beta_5 mpg_i = \text{marginal effect of } wgt_i \text{ in Model 2}$$

$$\frac{\partial \text{price}_i}{\partial \text{mpg}_i} = \beta_2 + 2\beta_4 \text{mpg}_i + \beta_5 \text{wgt}_i = \text{marginal effect of mpg}_i \text{ in Model 2}$$

Hypotheses 1 and 2 imply the same coefficient exclusion restriction on regression equation (9), namely the restriction $\beta_5 = 0$.

1. The **marginal effect of wgt_i on price_i is unrelated to mpg_i** if $\beta_5 = 0$ in Model 2 given by regression equation (9).
 2. The **marginal effect of mpg_i on price_i is unrelated to wgt_i** if $\beta_5 = 0$ in Model 2 given by regression equation (9).
- The **null and alternative hypotheses** for both these propositions are:

$$H_0: \beta_5 = 0$$

$$H_1: \beta_5 \neq 0.$$

- The following **test** command computes a **heteroskedasticity-robust Wald F-test** of H_0 against H_1 . Enter the commands:

```
test wgtmpg      or      test wgtmpg = 0
return list
```

Inspect the results generated by this **test** command. State the inference you would draw from this F-test.

- The following **lincom** command computes a **heteroskedasticity-robust two-tail t-test** of H_0 against H_1 . Enter the commands:

```
lincom _b[wgtmpg]
return list
```

Inspect the results generated by these commands. State the inference you would draw from this t-test.

Remarks: The above **lincom** command simply replicates the two-tail t-test of this same hypothesis that is reported by the **regress** command used to estimate regression equation (9) by OLS. You should be able to explain and verify from the

above results why the two-tail t-test of $H_0: \beta_5 = 0$ versus $H_1: \beta_5 \neq 0$ is equivalent to the F-test of the same hypothesis.

□ LM Tests for Heteroskedastic Errors

All the preceding tests have been performed on the assumption that the error terms in Model 2 exhibit heteroskedasticity of some unknown form. In this section, we demonstrate how this assumption can be tested using LM (Lagrange Multiplier) tests for mixed heteroskedastic errors.

Model 2

$$\text{price}_i = \beta_0 + \beta_1 \text{wgt}_i + \beta_2 \text{mpg}_i + \beta_3 \text{wgt}_i^2 + \beta_4 \text{mpg}_i^2 + \beta_5 \text{wgt}_i \text{mpg}_i + u_i. \quad (9)$$

We illustrate LM tests for mixed heteroskedasticity of the error terms u_i in Model 2 by testing two hypotheses about the error variance in equation (9): first, that $\text{Var}(u_i)$ varies with wgt_i and mpg_i ; second, that $\text{Var}(u_i)$ varies with all five regressors in equation (9).

Test 1 for Mixed Heteroskedastic Errors

Suppose you want to test if the error variance σ_i^2 is related to wgt_i and mpg_i . Without specifying fully the form of the conditional variance function $h(\cdot)$, you can formulate the null and alternative hypotheses as follows:

$$\begin{aligned} H_0: \quad & \text{Var}(u_i | \mathbf{x}_i^T) = h(\alpha_0) = \sigma^2 > 0 \quad \forall i \\ H_1: \quad & \text{Var}(u_i | \mathbf{x}_i^T) = \sigma_i^2 = h(\alpha_0 + \alpha_1 \text{wgt}_i + \alpha_2 \text{mpg}_i) > 0 \quad \forall i \end{aligned}$$

Comparing H_0 and H_1 , we see that H_0 implies the following two coefficient restrictions: $\alpha_1 = 0$ and $\alpha_2 = 0$. So the **null and alternative hypotheses** can be written as:

$$\begin{aligned} H_0: \quad & \alpha_1 = 0 \text{ and } \alpha_2 = 0 \\ H_1: \quad & \alpha_1 \neq 0 \text{ and/or } \alpha_2 \neq 0 \end{aligned}$$

The LM-BPK test statistic for testing H_0 against H_1 is computed from OLS estimation of the following test regression:

$$\tilde{u}_i^2 = d_0 + d_1 \text{wgt}_i + d_2 \text{mpg}_i + e_i \quad (10)$$

where

$\tilde{u}_i^2$ = the *squared* OLS residual for the i -th observation from OLS estimation of Model 2

e_i = the i -th error term in the test regression.

The **LM-BPK test statistic** is:

$$\text{LM-BPK} = N R_{\hat{u}^2}^2 \quad (11)$$

where:

N = the number of sample observations;

$R_{\hat{u}^2}^2 = \text{ESS}_{\hat{u}^2} / \text{TSS}_{\hat{u}^2}$ = the R^2 from OLS estimation of LM test regression (10).

The *asymptotic null distribution* of LM-BPK under the null hypothesis that $\alpha_1 = 0$ and $\alpha_2 = 0$ is **chi-square with 2 degrees of freedom**:

$$\text{LM-BPK} \overset{a}{\sim} \chi^2[2] \quad \text{under } H_0: \alpha_1 = 0 \text{ and } \alpha_2 = 0$$

- First, we need the OLS residuals from OLS estimation of Model 2 in order to compute the BPK test statistic. We therefore need to re-estimate Model 2. Enter the **regress** command:

```
regress price wgt mpg wgtsq mpgsq wgtmpg
```

- Second, save the OLS residuals $\{\tilde{u}_i : i = 1, \dots, N\}$ from the above **regress** command. Enter the following **predict** command with the **residual** option:

```
predict uhat, residual
```

- Third, generate the squared OLS residuals $\{\tilde{u}_i^2 : i = 1, \dots, N\}$. Enter the generate command:

```
generate uhatsq = uhat^2
```

- Fourth, estimate by OLS the LM test regression (10). Enter the commands:

```
regress uhatsq wgt mpg  
ereturn list
```

- Fifth, from the temporarily-saved results of the above **regress** command, which include the value of N in the scalar **e(N)** and the value of the test regression R -squared in the scalar **e(r2)**, compute the value of the LM-BPK test statistic $LM-BPK = N R_{\hat{u}}^2$. Enter the **scalar** command:

```
scalar LM1 = e(N)*e(r2)  
scalar list LM1
```

- Finally, use the **chi2tail(df, c₀)** function to compute the p-value for the calculated LM-BPK test statistic **LM1**. The **chi2tail(df, c₀)** statistical function in *Stata* computes the **right-tail p-value** of a **chi-square statistic** with sample value c_0 that has a chi-square null distribution with **df** degrees of freedom. Recall that the null distribution of the LM-BPK test statistic **LM1** is $\chi^2[2]$, the chi-square distribution with 2 degrees of freedom; this means that **df** = 2. Enter the **display** command:

```
display chi2tail(2, LM1)
```

This completes the LM-BPK test procedure. What inference would you draw from the test results? In particular, would you reject or retain the null hypothesis of homoskedastic errors against the alternative hypothesis that the error variance in Model 2 is related to wgt_i and mpg_i ?

- To compute the **1 percent** and **5 percent critical values of the chi-square distribution with 2 degrees of freedom**, use the **invchi2tail(df, α)** statistical function in *Stata*. The **invchi2tail(df, α)** statistical function computes $\chi^2_{\alpha}[df]$, the **right-tail α-level critical value** of a chi-square distribution with **df** degrees of freedom. Set **df** = 2 and **α** = 0.01 to get the 1 percent critical value of $\chi^2[2]$; set **df** = 2 and **α** = 0.05 to get the 5 percent critical value of $\chi^2[2]$. Enter the following commands:

```
display invchi2tail(2, 0.01)
display invchi2tail(2, 0.05)
scalar list LM1
```

Would you reject or retain the null hypothesis of homoskedastic errors at the 1 percent significance level? At the 5 percent significance level?

- The **original Breusch-Pagan LM test** for mixed heteroskedastic errors can be computed using the **hettest** command in *Stata*. For this test, simply enter the command:

```
hettest wgt mpg
```

Compare the results of this BP test with those of the BPK test you previously computed. The test statistics are different, so it is entirely possible that the original BP test and the robust BPK test of the same hypothesis will yield different inferences.

Test 2 for Mixed Heteroskedastic Errors

Suppose now you want to test if the error variance σ_i^2 is related to all five of the regressors in Model 2, namely wgt_i , mpg_i , wgt_i^2 , mpg_i^2 , and $wgt_i mpg_i$. Without specifying fully the form of the conditional variance function $h(\cdot)$, you can formulate the null and alternative hypotheses as follows:

$$H_0: \sigma_i^2 = h(\alpha_0) = \sigma^2 > 0 \quad \forall i$$

$$H_1: \sigma_i^2 = h(\alpha_0 + \alpha_1 wgt_i + \alpha_2 mpg_i + \alpha_3 wgt_i^2 + \alpha_4 mpg_i^2 + \alpha_5 wgt_i mpg_i) > 0 \quad \forall i$$

Comparing H_0 and H_1 , we see that H_0 implies the following five coefficient restrictions: $\alpha_1 = 0$ and $\alpha_2 = 0$ and $\alpha_3 = 0$ and $\alpha_4 = 0$ and $\alpha_5 = 0$. So the **null and alternative hypotheses** can be written as:

$$H_0: \alpha_j = 0 \quad \text{for all } j = 1, \dots, 5$$

$$H_1: \alpha_j \neq 0 \quad j = 1, \dots, 5$$

The LM-BPK test statistic for testing H_0 against H_1 is computed from OLS estimation of the following LM test regression:

$$\tilde{u}_i^2 = d_0 + d_1 \text{wgt}_i + d_2 \text{mpg}_i + d_3 \text{wgt}_i^2 + d_4 \text{mpg}_i^2 + d_5 \text{wgt}_i \text{mpg}_i + e_i \quad (12)$$

where

$\tilde{u}_i^2$ = the *squared* OLS residual for the i-th observation from OLS estimation of Model 2

e_i = the i-th error term in the LM test regression.

The **LM-BPK test statistic** is:

$$\text{LM-BPK} = N R_{\hat{u}^2}^2 \quad (13)$$

where:

N = the number of sample observations;

$R_{\hat{u}^2}^2 = \text{ESS}_{\hat{u}^2} / \text{TSS}_{\hat{u}^2}$ = the R^2 from OLS estimation of LM test regression (12)

The **asymptotic null distribution** of LM-BPK under the null hypothesis that $\alpha_j = 0$ for all $j = 1, \dots, 5$ is **chi-square with 5 degrees of freedom**:

$$\text{LM-BPK} \overset{a}{\sim} \chi^2[5] \quad \text{under } H_0: \alpha_j = 0 \quad \forall j = 1, \dots, 5$$

- First, we need the OLS residuals from OLS estimation of Model 2 in order to compute the BPK test statistic (13). However, you previously saved these OLS residuals for Model 2 as **uhat** and generated their squared values as **uhatsq**, so they do not have to be created again to perform this test. But if you want to re-estimate Model 2, enter the command:

```
regress price wgt mpg wgtsq mpgsq wgtmpg
```

- Estimate by OLS the LM test regression (12). Enter the commands:

```
regress uhatsq wgt mpg wgtsq mpgsq wgtmpg  
ereturn list
```

- Next, from the temporarily-saved results of the above **regress** command, which include the value of N in the scalar **e(N)** and the value of the test regression R -squared in the scalar **e(r2)**, compute the value of the LM-BPK test statistic $LM-BPK = N R_{\hat{u}}^2$. Enter the **scalar** commands:

```
scalar LM2 = e(N)*e(r2)
scalar list LM2
```

- Finally, use the **chi2tail(df, c₀)** function to compute the p-value for the calculated LM-BPK test statistic **LM2**. Recall that the null distribution of the LM-BPK test statistic **LM2** is $\chi^2[5]$, the chi-square distribution with 5 degrees of freedom; this means that **df** = 5. Enter the **display** command:

```
display chi2tail(5, LM2)
```

This completes the LM-BPK test procedure. What inference would you draw from the test results? In particular, would you reject or retain the null hypothesis of homoskedastic errors against the alternative hypothesis that the error variance in Model 2 is related to wgt_i , mpg_i , wgt_i^2 , mpg_i^2 , and $wgt_i mpg_i$?

- To compute the **1 percent** and **5 percent critical values of the chi-square distribution with 5 degrees of freedom**, use the **invchi2tail(df, α)** statistical function in *Stata*. Set **df** = 5 and **α** = 0.01 to get the 1 percent critical value of $\chi^2[5]$; set **df** = 5 and **α** = 0.05 to get the 5 percent critical value of $\chi^2[5]$. Enter the following commands:

```
display invchi2tail(5, 0.01)
display invchi2tail(5, 0.05)
scalar list LM2
```

Would you reject or retain the null hypothesis of homoskedastic errors at the 1 percent significance level? At the 5 percent significance level?

Test 2: The BPK F-Test for Mixed Heteroskedastic Errors

An alternative test statistic used in the BPK test for mixed heteroskedastic errors is the analysis-of-variance F-statistic from OLS estimation of the LM test regression.

This F-statistic can be written in terms of the R-squared from the LM test regression as follows:

$$F\text{-BPK} = \frac{R_{\hat{u}^2}^2 / p}{(1 - R_{\hat{u}^2}^2) / (N - p - 1)} = \frac{R_{\hat{u}^2}^2}{(1 - R_{\hat{u}^2}^2)} \frac{(N - p - 1)}{p} \quad (14)$$

The null distribution of the F-BPK statistic is approximately an F distribution with *numerator* degrees of freedom = p and *denominator* degrees of freedom = $N - p - 1$:

$$F\text{-BPK} \stackrel{\text{app}}{\sim} F[p, N - p - 1] \quad \text{under } H_0. \quad (15)$$

No additional computations are required to obtain the sample value of this F-statistic and its p-value; these are routinely generated by the **regress** command used to estimate the LM test regression by OLS.

- But you can directly compute the BPK F-test for mixed heteroskedastic errors by entering the following commands:

```
scalar F2 = (e(r2)/e(df_m))/((1-e(r2))/e(df_r))
scalar list F2
display Ftail(e(df_m), e(df_r), F2)
```

Note the use in the above commands of several saved scalars from the **regress** command that estimated the LM test regression: $\mathbf{e(r2)} = R_{\hat{u}^2}^2$; $\mathbf{e(df_m)} = p$; $\mathbf{e(df_r)} = N - p - 1$. Compare the sample value of **F2** and its p-value with the value of the ANOVA F-statistic and its p-value from the LM test regression. They should be identical.

- You may prefer to compute the F-BPK test statistic by creating scalars for $R_{\hat{u}^2}^2$, the numerator degrees of freedom p , and the denominator degrees of freedom $N - p - 1$. The following commands demonstrate how to do this:

```
scalar Rsqu = e(mss)/(e(mss) + e(rss))
scalar df1 = e(df_m)
scalar df2 = e(df_r)
scalar Fbpk = (Rsqu/df1)/((1 - Rsqu)/df2)
```

```
scalar list Rsqu df1 df2 Fbpk  
display Ftail(df1, df2, Fbpk)
```

- Finally, compare the sample values of **F2** and **Fbpk** with the saved value of the ANOVA F-statistic from OLS estimation of the LM test regression. Enter the commands:

```
display e(F)  
scalar list F2 Fbpk
```

- The original Breusch-Pagan LM test for mixed heteroskedastic errors can be computed using the **hettest** command in *Stata*. For this test, simply enter the command:

```
hettest wgt mpg wgtsq mpgsq wgtmpg
```

Compare the results of this BP test with those of the BPK test you previously computed. The test statistics are different, so it is entirely possible that the original BP test and the robust BPK test of the same hypothesis will yield different inferences.

❑ Preparing to End Your *Stata* Session

Before you end your *Stata* session, you should do two things.

- First, you may want to **save the current data set**. Enter the following **save** command to save the current data set as *Stata*-format data set **auto7.dta**:

```
save auto7
```

- Second, **close the command log file** you have been recording. Enter the command:

```
cmdlog close
```

- Third, **close the log file** you have been recording. Enter the command:

```
log close
```

❑ End Your *Stata* Session -- exit

- **To end your *Stata* session**, use the **exit** command. Enter the command:

```
exit      or      exit, clear
```

❑ Cleaning Up and Clearing Out

After returning to Windows, you should copy all the files you have used and created during your *Stata* session to your own portable electronic storage device such as a flash memory stick. These files will be found in the ***Stata working directory***, which is usually **C:\data** on the computers in Dunning 350. There are three files you will want to be sure you have: the complete *Stata* log file **452tutorial7.log**; the *Stata* command log file **452tutorial7.txt**; and the saved *Stata*-format dataset **auto7.dta**. Use the Windows **copy** command to copy any files you want to keep to your own portable electronic storage device (e.g., a flash memory stick).

Finally, **as a courtesy to other users** of the computing classroom, please delete all the files you have used or created from the *Stata* working directory.