# HOW TO TEST ENDOGENEITY OR EXOGENEITY: AN E-LEARNING HANDS ON SAS

*N. Uttam Singh, **Kishore K Das and *Aniruddha Roy
*ICAR Research Complex for NEH Region, Umiam, Meghalaya
**Department of Statistics, Gauhati University, Guwahati, Assam
uttamba@gmail.com

## 0.1 INTRODUCTION

Simple linear regression model (OLS) is based on the assumption that the independent variables are exogenous ($E(\varepsilon|X)=0$). That is, the error terms in the linear regression model are uncorrelated or independent of the explanatory variables. Under the exogeneity assumption, the least squares estimator $b$ is an unbiased and consistent estimator of $\beta$. Let us consider a typical regression model specification: $Y_t=b_0+b_1X_{1t}+b_2X_{2t}+b_3X_{3t}+\varepsilon_t$. Exogeneity means that each $X$ variable does not depend on the dependent variable $Y$; rather $Y$ depends on the $Xs$ and on $\varepsilon$, the model error. Since $Y$ depends on $\varepsilon$, this means that the $Xs$ are assumed to be independent of $Y$ hence $\varepsilon$. It is a standard assumption we make in regression analysis because if the 'independent variables' are not independent of $\varepsilon$ and $Y$, then the estimated regression coefficients are not consistent if we use the OLS estimating equations. Explanatory variables that are not exogenous are called endogenous variables. Therefore, if, for whatever reason, an explanatory variable $x_j$ is correlated with the model error term $u_i$, $Cov(x_j,u_i)\neq 0$, then $x_j$ is said to be an endogenous explanatory variable and we say the model suffers from endogeneity. All OLS estimators will be biased and inconsistent in the presence of endogenous regressors. Endogeneity can arise as a result of measurement error, reverse casualty/simultaneity, omitted variable or unobserved variables, omitted selection, lagged dependent variables. These are the main reasons why $X$ and e might be correlated or main causes of failure of exogeneity. The problem of endogeneity is very serious and often ignored by researchers conducting non-experimental research *(*Hamilton et al 2016).

## 0.2 MAIN CAUSES OF FAILURE OF EXOGENEITY

**Measurement Error** (Errors-in-variables): This is when we cannot measure the true $X$ variable, so that there is uncertainty attached to the measured value.

**Reverse Causality/ Simultaneity**: This occurs where the explanatory variable is jointly determined with the dependent variable. In other words, $X$ causes $Y$ but $Y$ also causes $X$. Since $Y$ contains $\varepsilon$, then variation in $\varepsilon$ would show up in $X$. Suppose we are interested in a causal relationship between $x$ and $y$, but it is possible that $y$ also affects $x$. Examples: Agricultural growth → Farmers' income → Agricultural growth, Cancer → Smoking → Cancer.

**Omitted Variable** (Correlated Missing Regressors): An important source of endogeneity is omitted variables. Consider a multivariate regression model, $y_i=\alpha_0+\alpha_1s_i+\alpha_2a_i+u_i$ where $\alpha_1$ is our coefficient of interest and $a_i$ denotes an unobservable variable with $Cov(s_i,a_i)\neq 0$. Since we cannot observe $a_i$, the model we estimate is $y_i=\beta_0+\beta_1s_i+\varepsilon_i$. Because $Cov(s_i,a_i)\neq 0$, $a_i$ will appear in the error term $\varepsilon_i$ in equation. Thus, $Cov(s_i,\varepsilon_i)\neq 0$, which violates the zero conditional mean assumption.

**Omitted Sample Selection**: What if we select the sample on the basis of something correlated with $\varepsilon$? For example, if we only look at workers, and exclude non-workers, the latter might have different $\varepsilon$ from the former.

**Lagged Dependent Variables**: Regression model (time series) includes a lagged dependent variable and the error term is serially correlated. Recall that estimate biased but consistent with a lagged dependent variable, but this assumes that the errors are independent of each other over time. The model includes a lagged dependent variable and has a serially correlated disturbance. Then dependent variable $Y_{t-1}$ is correlated with $\varepsilon_t$. This correlation will remain as time increases.

## 0.3 CORRECTING ENDOGENEITY

Solving the problem of endogeneity is the real art of econometrics. All sources of endogeneity lead to inconsistent OLS estimation. Ideally we should eliminate measurement error, introduce omitted or unobserved variables for mending the correlated missing regressors, narrow the generality of interpretation for mending sample selection bias as it does no longer apply to the population, but rather to the chunk of the population satisfying sample restrictions, estimate a system of simultaneous equations etc. Often these solutions are not achievable in practice. Reverse Causality is also tougher to handle. Thus the solution is to use an alternative estimation method known as instrumental variables (IV) or equivalently two-stage least squares (2SLS). This involves introduction of a variable $Z$ that induces changes in the explanatory variable but has no independent effect on the dependent variable, allowing a researcher to uncover the causal effect of the explanatory variable on the dependent variable.

### 0.3.1 Instrumental Variables (IV)

This newly introduced variable *Z* is called instrumental variable or instrument. Instruments are correlated with *X*, but uncorrelated with the model error term by assumption or by construction. An instrument variable is used to create a new variable by replacing the problematic variable. More formally, an instrument *Z* for the variable of concern *X* satisfies
1) $Corr(Z_i,X) \neq 0$
correlated with the problem variable i.e instrument should have relevance to the *X*
2) $Corr(Z_i,e_i) = 0$
but uncorrelated with the residual (so does not suffer from measurement error and also is not correlated with any unobservable factors influencing the dependent variable) i.e. instruments should be exogenous.
(Wooldridge, J. M. 2015).

### 0.3.2 Choice of Instrument

There is no standard method for choosing instruments for nonlinear regression. Few econometric textbooks discuss the selection of instruments for nonlinear models. Refer to Bowden, R.J. and Turkington, D.A. (1984, p. 180-182) for more information.

The purpose of the instrumental projection is to purge the regressors of their correlation with the residual. For nonlinear systems, the regressors are the partials of the residuals with respect to the parameters.

Possible instrumental variables include:

- any variable in the model that is independent of the errors
- lags of variables in the system
- derivatives with respect to the parameters, if the derivatives are independent of the errors
- low degree polynomials in the exogenous variables
- variables from the data set or functions of variables from the data set.

Selected instruments must not

- depend on any variable endogenous with respect to the equations estimated
- depend on any of the parameters estimated
- be lags of endogenous variables if there is serial correlation of the errors.

If the preceding rules are satisfied and there are enough observations to support the number of instruments used, the results should be consistent and the efficiency loss held to a minimum.

We need at least as many instruments as the maximum number of parameters in any equation, or some of the parameters cannot be estimated. Note that number of instruments means linearly independent instruments. If we add an instrument that is a linear combination of other instruments, it has no effect and does not increase the effective number of instruments.

We can, however, use too many instruments. In order to get the benefit of instrumental variables, we must have more observations than instruments. Thus, there is a trade-off; the instrumental variables technique completely eliminates the simultaneous equation bias only in large samples. In finite samples, the larger the excess of observations over instruments, the more the bias is reduced. Adding more instruments may improve the efficiency, but after some point efficiency declines as the excess of observations over instruments becomes smaller and the bias grows.

Consider the structural model, with one endogenous ($y_2$), and one exogenous ($z_1$) RHS variable. Suppose that we have two valid instruments, $z_2$ and $z_3$. Since $z_1$, $z_2$ and $z_3$ are uncorrelated with error $u_1$, so is any linear combination of these Thus, any linear combination is also a valid instrument. The best instrument is the one that is most highly correlated with $y_2$. This turns out to be a linear combination of the exogenous variables. The reduce form equation is: $y_2 = p_0 + p_1 z_1 + p_2 z_2 + p_3 z_3 + v_2$ *or* $y_2 = y_2^* + v_2$. We can think of $y_2^*$ as the part of $y_2$ that is uncorrelated with $u_1$ and $v_2$ as the part that might be correlated with u1. Thus the best IV for $y_2$ is $y_2^*$. (http://web.uvic.ca/~hschuetz/econ499/iv.pdf)

### 0.3.3 Two-Stage Least Squares (2SLS)

This procedure calculates the two-stage least squares (2SLS) estimate. This method is used fit models that include instrumental variables. 2SLS includes four types of variable(s): dependent, exogenous, endogenous, and instrument. These are defined as follows:

**Dependent Variable:** This is the response variable that is to be regressed on the exogenous and endogenous (but not the instrument) variables.

**Exogenous Variables:** These independent variables are included in both the first and second stage regression models. They are not correlated with the random error values in the second stage regression

**Endogenous Variables:** Each endogenous variable becomes the dependent variable in the first stage regression equation. Each is regressed on all exogenous and instrument variables. The predicted values from these regressions replace the original values of the endogenous variables in the second stage regression model.

**Instrument Variables:** Each endogenous variable becomes the dependent variable in the first stage regression equation. Each is regressed on all exogenous and instrument variables. The predicted values from these regressions replace the original values of the endogenous variables in the second stage regression model.

2SLS is used in econometrics, statistics, and epidemiology to provide consistent estimates of a regression equation when controlled experiments are not possible. They are discussed in every modern econometrics text. We have used Wooldridge, J. M. (2002, 2015) and Vivek B. Ajmani (2011).

## SECTION 1: DEFINING DATASET AND DESCRIPTIVE STATISTICS

### 1.1 Defining the Dataset

Throughout this learning module, we will use an example data set named "*endogeneity_test*" available at the Appendix. There are 8 variables in the dataset. We want to see the effect of variables *Min_Tem* (minimum temperature) and *Rain* (average rainfall) on the dependent variable *Foodgrain_Yld* (yield of foodgrain). The dependent variable *Foodgrain_Yld* is yield of foodgrains for different locations over a period of years. Other variables in the dataset are *Latitude* (Latitude of the particular location), *Longitude* (Longitude of the particular location), *Foodgrain_Yld_FD* (First difference of *Foodgrain_Yld*), *Min_Tem_FD* (First difference of *Mix_Tem*), *Rain_FD* (First difference of *Rain*). These variables are chosen with the assumption that they can be used as instrumental variables for testing endogeneity of the two regressors *Min_Tem* and *Rain*. Selection of instruments is not easy as they should be exogenous and should have a relevance to the endogenous regressors. Descriptive statistics of the variables can be got using the following SAS code.

### 1.2 Analysis Using SAS

```
proc means data=endogeneity_test;
var Foodgrain_Yld Min_Tem Rain Latitude Longitude Foodgrain_Yld_FD Min_Tem_FD  Rain_FD;
run;
```

The MEANS Procedure

| Variable | N | Mean | Std Dev | Minimum | Maximum |
|---|---|---|---|---|---|
| **Foodgrain_Yld** | 353 | 1347.05 | 467.9570664 | 251.8518519 | 2990.82 |
| **Min_Tem** | 376 | 17.9062659 | 0.8648703 | 15.0433157 | 22.9132395 |

| Variable | N | Mean | Std Dev | Minimum | Maximum |
|---|---|---|---|---|---|
| **Rain** | 376 | 6.4429678 | 2.2700222 | 2.3913140 | 19.8587888 |
| **Latitude** | 376 | 25.5958718 | 1.3202667 | 23.7103990 | 27.3333303 |
| **Longitude** | 376 | 92.5495664 | 1.8014092 | 88.6166475 | 94.2166674 |
| **Foodgrain_Yld_FD** | 345 | 23.5414239 | 201.9374671 | -932.1562438 | 1000.08 |
| **Min_Tem_FD** | 345 | 0.0154531 | 0.4993911 | -3.1689296 | 2.3964129 |
| **Rain_FD** | 345 | -0.0082422 | 1.8269921 | -8.7987599 | 7.4485088 |

OUTPUT 1.1. Descriptive statistics of variables in the dataset

# SECTION 2: HOUSMAN'S TEST FOR ENDOGENEITY

## 2.1 Test for Endogeneity

Housman's test (also known as Housman Specification test or Durbin, Hausman and Wu Test) can be used to determine if it is necessary to use an instrumental variables method (2SLS) rather than a more efficient OLS estimation. Hausman's statistic can also be used to compare 2SLS with 3SLS for a class of estimators for which 3SLS is asymptotically efficient (similarly for OLS and SUR). Hausman's statistic can also be used, in principle, to test the null hypothesis of normality when comparing 3SLS to FIML. The Hausman's test is also described as a test for model misspecification. In panel data analysis (the analysis of data over time), the Hausman's test can help us to choose between fixed effects model and a random effects model.

What is interested at present is Hausman's test for endogeneity. This section introduces methods to determine if endogeneity is indeed a problem. In other words we test the exogeneity of the regressors. Wooldridge, 2002, has given the total concept and we can simplify the test mechanism in the following way. Suppose our linear model is y = b0 + b1X + u. Then the steps of the test are:

1. Seek out an appropriate instrument *Z*. Regress the endogenous variable *X* on the instrument(s) *Z* i.e. $X=d_0+d_1Z+v$. Save the residuals v.
2. Include this residual as an extra term in the original model and estimate $y= b_0+b_1X+b_2v+e$.
3. Step 3: Test whether $b_2 = 0$ (using a t test).
   If $b_2 = 0$ conclude there is no correlation between *X* and *u*.
   If $b_2 \neq 0$ conclude there is correlation between *X* and *u*.

   Why ? because $X=d_0+d_1Z+v$ and so only way *X* could be correlated with u is through v. This means the residual in original model ($y=b_0+b_1X+u$) depends on *v+some residual* i.e. $u=b_2v+e$. So estimate $y=b_0+b_1X+b_2v+e$ instead of $y=b_0+b_1X+u$ and test whether coefficient on *v* is significant. If it is, conclude that *X* and error term are indeed correlated; there is endogeneity.

```
Note: This test is only as good as the instruments used and is only valid asymptotically.
This may be a problem in small samples and so we should generally use this test only with
sample sizes well above 100.
```

## 2.2 Analysis Using SAS

Our regression equation is *Foodgrain_Yld=Min_Tem Rain*. The objective here is to determine if the variable *Min_Tem* and *Rain* are endogenous. The first step is to regress independent variables *Min_Tem* (suspected to be endogenous) on the instruments *Latitude*, *Longitude*, *Foodgrain_Yld_FD*, *Min_Tem_FD*, *Rain_FD*. The residuals (*v*) from this regression is saved and used as an explanatory variable in the regression of *Foodgrain_Yld* against *Min_Tem*, *Rain* and *v*. Recall $H_0$:*x* is exogenous while $H_1$:*x* is endogenous. If the t-statistic corresponding to *v* is significant, then the null hypothesis is rejected and we conclude that the variable *Min_Tem* is endogenous. In the same way the other explanatory variable *Rain* can also be tested for its endogeneity. The following SAS statements can be used to do the analysis. Notice that the first *Proc Reg* statements save the residuals in a temporary SAS data set called *endogeneity_test1*.

```
proc reg data=endogeneity_test noprint;
model Min_Tem=Latitude Longitude Foodgrain_Yld_FD  Min_Tem_FD Rain_FD;
output out=endogeneity_test1 r= v ;
run;

proc reg data=endogeneity_test1;
model Foodgrain_Yld=Min_Tem Rain v ;
run;
```

```
Some parts of the output are omitted. Only the portion which has majorvrelevance to the current interpretation is presented here. Full output can be got by running the program code with the dataset given in the appendix.
```

**Parameter Estimates**

| Variable | Label | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|
| **Intercept** | Intercept | **1** | 3729.45363 | 854.83696 | 4.36 | <.0001 |
| **Min_Tem** | | **1** | -115.94673 | 46.55075 | -2.49 | 0.0132 |
| **Rain** | | **1** | -44.31022 | 10.41817 | -4.25 | <.0001 |
| **v** | Residual | **1** | 345.56202 | 60.91608 | 5.67 | <.0001 |

OUTPUT 2.1. Using Proc Reg to check for endogeneity of regressor Min_Tem

## 2.3 Interpretation

The analysis results are given in Output 2.1. The results indicate that we have evidence of endogeneity of *Min_Tem* at the 1% significance level (p-value <0.0001) because we reject the null hypothesis of exogeneity as the parameter estimate of the residual is highly significant as highlighted in the output. In other words the explanatory variable *Min_Tem* is correlated with the error of the original model (*Foodgrain_Yld= Min_Tem Rain*).

The same is done for the independent variable *Rain* also to test endogeneity. The following SAS code performs this. The analysis results are given in Output 2.2. In this case also the results indicate that we have evidence of endogeneity of *Rain* at the 1% significance level (*p-value* <0.0001). Hence endogeneity of both the regressors *Min_Tem* and *Rain* in explaining the dependent variable *Foodgrain_Yld* is concluded.

```
proc reg data=endogeneity_test noprint;
model Rain=Latitude Longitude Foodgrain_Yld_FD  Min_Tem_FD Rain_FD;
output out=endogeneity_test2 r= v1 ;
run;

proc reg data=endogeneity_test2;
model Foodgrain_Yld=Min_Tem Rain v1 ;
run;
```

```
Some parts of the output are omitted. Only the portion which has major relevance to the current interpretation is presented here. Full output can be got by running the program code with the dataset given in the appendix.
```

**Parameter Estimates**

| Variable | Label | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|

| Parameter Estimates | | | | | | |
|---|---|---|---|---|---|---|
| **Variable** | **Label** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > \|t\|** |
| **Intercept** | Intercept | **1** | -1253.33990 | 609.44543 | -2.06 | 0.0405 |
| **Min_Tem** | | **1** | 121.46950 | 31.05620 | 3.91 | 0.0001 |
| **Rain** | | **1** | 65.73221 | 21.03794 | 3.12 | 0.0019 |
| **v1** | Residual | **1** | -134.42554 | 24.03614 | -5.59 | <.0001 |

OUTPUT 2.2. Using Proc Reg to check for endogeneity of regressor Rain

```
Note: The model we have been discussing is Foodgrain_Yld=Min_Tem Rain. If our model were
having an already known exogenous variable i.e if the model were of the form
Foodgrain_Yld= Min_Tem Rain Farmer_age where Farmer_age is the assumed exogenous
variable. Then the code would be as follows

proc reg data=endogeneity_test noprint;
model Min_Tem=Farmer_age Latitude Longitude Foodgrain_Yld_FD  Min_Tem_FD Rain_FD;
output out=endogeneity_test1 r= v ;
run;

proc reg data=endogeneity_test1;
model Foodgrain_Yld=Min_Tem Rain Farmer_age v ;
run;
```

```
Note: Predicted values can also be used in place of residual in the second stage
regression. The result of importance will be the same even if other accompanied outputs
are different. The corresponding SAS code is as follows.

proc reg data=endogeneity_test noprint;
model Rain=Latitude Longitude Foodgrain_Yld_FD  Min_Tem_FD Rain_FD ;
output out=endogeneity_test2 p=f p=f1;
run;

proc reg data=endogeneity_test2;
model Foodgrain_Yld=Min_Tem Rain f f1;
run;
```

## SECTION 3: TEST FOR OVERIDENTIFICATION

### 3.1 Testing Overidentifying Restrictions

We now turn our attention to addressing the problem of determining if the regression model has more instruments than is necessary. The question we address here is, “Are the extra instrument variables truly exogenous?” That is, are the extra instruments uncorrelated with the error term? Wooldridge (2002, p. 123) gives details on a simple regression-based Sargan’s hypothesis test (1958) to determine whether the regression model has more instruments than is required. The steps are as follows.

1. Consider, the linear model given by $y=X\beta+\Gamma\delta+\varepsilon$ where $y$ is $n\times1$, $X$ is $n\times L_1$, $\beta$ is $L_1\times1$, $\Gamma$ is $n\times G$, $\delta$ is $G\times1$, and $\varepsilon$ is $n\times1$. Here, Γ contains variables that are suspected of being endogenous. As before, let $W=(X, W^*)$ be the set of all instrumental variables. Here, $W^*$ is $n\times L_2$ so that $W$ is $n\times L$ with $L= L_1+L_2$ and $L_2 > G$.
2. Conduct a 2SLS and obtain $\hat{\varepsilon}$.
3. Conduct an OLS of $\hat{\varepsilon}$ on W and obtain $R^2$.
4. Sargan’s test statistic is $nR_u^2$. Under the null hypothesis of exogenous extra instruments, the test statistic is distributed as a chi-squared random variable with $L_2 - G$ degrees of freedom.

If the null hypothesis is rejected, then we need to reexamine the instruments that were selected for the analysis. The general idea is that if the instruments are truly exogenous, then they should not be correlated with the disturbance term.

### 3.2 Analysis Using SAS

We will now illustrate the computations by using the original of *Foodgrain_Yld* equation with *Latitude*, *Longitude*, *Foodgrain_Yld_FD*, *Min_Tem_FD* and *Rain_FD* as instruments. The first step is to estimate the true model by using 2SLS and to store the residuals. The following SAS statements can be used. Note that the output has been suppressed because we are interested only in storing the residuals from this analysis.

```
proc syslin 2SLS noprint data=endogeneity_test out=step1_resid;
endogenous Min_Tem Rain;
instruments Latitude Longitude Foodgrain_Yld_FD  Min_Tem_FD Rain_FD;
model Foodgrain_Yld=Min_Tem Rain;
output residual=out1_resid;
run;
```

The next step is to regress the residuals from the 2SLS analysis on all exogenous variables in the model. The following SAS statements can be used. The results of the analysis are given in Output 3.1.

```
proc reg data=step1_resid;
model out1_resid=Latitude Longitude Foodgrain_Yld_FD  Min_Tem_FD Rain_FD ;
run;
```

Note: The model we have been discussing is Foodgrain_Yld= Min_Tem Rain. If our model were having an already known exogenous variable i.e if the model were of the form Foodgrain_Yld= Min_Tem Rain Farmer_age where Farmer_age is the assumed exogenous variable. Then the code would be as follows

```
proc syslin 2SLS noprint data=endogeneity_test out=step1_resid;
endogenous Min_Tem Rain;
instruments Farmer_age Latitude Longitude Foodgrain_Yld_FD  Min_Tem_FD Rain_FD;
model Foodgrain_Yld=Min_Tem Rain Farmer_age;
output residual=out1_resid;
run;
```

The next step is to regress the residuals from the 2SLS analysis on all exogenous variables in the model. The following SAS statements can be used.

```
proc reg data=step1_resid;
model out1_resid= Farmer_age Latitude Longitude Foodgrain_Yld_FD  Min_Tem_FD Rain_FD ;
run;
```

Model: MODEL1
Dependent Variable: out1_resid Residual Values

| | |
|---|---|
| **Number of Observations Read** | 376 |
| **Number of Observations Used** | 345 |
| **Number of Observations with Missing Values** | 31 |

**Analysis of Variance**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
|---|---|---|---|---|---|
| **Model** | 5 | 9504165 | 1900833 | 8.53 | <.0001 |
| **Error** | 339 | 75512282 | 222750 | | |

| Analysis of Variance | | | | | |
|---|---|---|---|---|---|
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Corrected Total** | 344 | 85016447 | | | |

| | | | |
|---|---|---|---|
| **Root MSE** | 471.96408 | **R-Square** | 0.1118 |
| **Dependent Mean** | -5.602E-13 | **Adj R-Sq** | 0.0987 |
| **Coeff Var** | -8.42498E16 | | |

| Parameter Estimates | | | | | | |
|---|---|---|---|---|---|---|
| **Variable** | **Label** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > \|t\|** |
| **Intercept** | Intercept | **1** | 1848.32895 | 1510.08467 | 1.22 | 0.2218 |
| **Latitude** | | **1** | -88.21001 | 19.61092 | -4.50 | <.0001 |
| **Longitude** | | **1** | 4.22921 | 15.34315 | 0.28 | 0.7830 |
| **Foodgrain_Yld_FD** | | **1** | 0.53127 | 0.12653 | 4.20 | <.0001 |
| **Min_Tem_FD** | | **1** | -5.60322 | 51.45307 | -0.11 | 0.9133 |
| **Rain_FD** | | **1** | -23.55952 | 14.10741 | -1.67 | 0.0958 |

OUTPUT 3.1. Testing overidentifying restrictions in the data.

### 3.3 Interpretation

There are 345 used observations in the data set and $R^2$ =0.1118. Therefore, the test statistic value is $NR^2$=38.57. The critical value is $\chi^2_{3,0.05}$=7.81. The degrees of freedom were calculated using the formula *L2 – G*, where *L2=5* because we used *Latitude*, *Longitude*, *Foodgrain_Yld_FD*, *Min_Tem_FD* and *Rain_FD* as instruments. We suspect only two variables (*Min_Tem* and *Rain*) as being endogenous, *G*=2. Thus, the degree of freedom is 3. The null hypothesis is rejected because the test statistic value is larger than the critical value. Hence the regression model has more instruments than is necessary. F value is also highly significant indicating the significant parameter estimates. In such case we need to revisit the assignment of instruments as discussed in Section 4.

## SECTION 4: FURTHER OVERIDENTIFICATION TEST FOR APPROPRIATE INSTRUMENTS

### 4.1 Appropriate number of instruments

From the result and interpretation of Section 3 it is found that the regression model has more instruments than is necessary. Therefore the extra instrument variables may not be truly exogenous or the extra instruments are correlated with the error term. After trial of different combinations of the instruments we reduce the number of instruments to *Longitude, Min_Tem_FD Rain_FD* as given in the following SAS code. However selection or identification of instruments is also an art, researchers should have a detail understanding of the variables with their cause and effect relationship involved in the model. Code is same except the reduction of number of instruments.

### 4.2 Analysis Using SAS

```
proc syslin 2SLS noprint data=endogeneity_test out=step1_resid1;
endogenous Min_Tem Rain;
instruments  Longitude Min_Tem_FD Rain_FD;
model Foodgrain_Yld=Min_Tem Rain;
output residual=out1_resid1; run;
```

```
proc reg data=step1_resid1;
model out1_resid1= Longitude Min_Tem_FD Rain_FD ;
run;
```

Model: MODEL1
Dependent Variable: out1_resid Residual Values

| | |
|---|---|
| **Number of Observations Read** | 376 |
| **Number of Observations Used** | 345 |
| **Number of Observations with Missing Values** | 31 |

**Analysis of Variance**

| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
|---|---|---|---|---|---|
| **Model** | 3 | 16951 | 5650.29642 | 0.02 | 0.9947 |
| **Error** | 341 | 77818118 | 228206 | | |
| **Corrected Total** | 344 | 77835068 | | | |

| | | | |
|---|---|---|---|
| **Root MSE** | 477.70872 | **R-Square** | 0.0002 |
| **Dependent Mean** | -1.7729E-12 | **Adj R-Sq** | -0.0086 |
| **Coeff Var** | -2.69457E16 | | |

**Parameter Estimates**

| **Variable** | **Label** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > \|t\|** |
|---|---|---|---|---|---|---|
| **Intercept** | Intercept | **1** | 170.27470 | 1440.20203 | 0.12 | 0.9060 |
| **Longitude** | | **1** | -1.83848 | 15.52928 | -0.12 | 0.9058 |
| **Min_Tem_FD** | | **1** | 12.22888 | 52.06251 | 0.23 | 0.8144 |
| **Rain_FD** | | **1** | -1.47204 | 14.23089 | -0.10 | 0.9177 |

OUTPUT 4.1. Testing overidentifying restrictions in the data.

### 4.3 Interpretation

Referring Output 4.1, there are 345 used observations in the data set and $R^2$ =0.0002. Therefore, the test statistic value is $NR^2$=0.07. The critical value is $\chi^2_{1,0.01}$=6.63. The degrees of freedom were calculated using the formula $L_2 - G$, where $L_2$=3 because we used *Longitude*, *Min_Tem_FD* and *Rain_FD* as instruments. We suspect only two variables (*Min_Tem* and *Rain*) as being endogenous, *G*=2. Thus, the degree of freedom is 1. The null hypothesis is not rejected because the test statistic value is smaller than the critical value. Hence we fail to reject that the regression model has more instruments than is necessary. That is how we can use the instruments *Longitude*, *Min_Tem_FD* and *Rain_FD* to identify the model for *Foodgrain_Yld* and we conclude these instrumental variables are truly exogenous and they are uncorrelated with the error term of the model. This can be intensified when we look at the large *p-value* of estimated parameters i.e. none of the parameter estimates are significant as the *F-value* is insignificant.

## SECTION 5: TEST FOR WEAK INSTRUMENTS

### 5.1 Weak Instruments Testing

We now turn our attention to the problem of weak instruments, that is, the case when the selected instrumental variables used in estimation have a poor correlation with the endogenous variable. We will discuss a general method for determining if weak instruments have been used in the model. Consider the model $y=X\beta+\alpha x+\varepsilon$ where $x$ is suspected of being endogenous. Assume that we have a set of instrumental variables $W$, which includes the explanatory variables in $X$. The reduced form equation relating $x$ to $X$ and $W$ is written as $x=W\delta+\gamma$ (Wooldridge, 2002). If $\delta=0$, the instruments in $W$ have no predictive power in explaining $x$. A value of $\delta$ close to zero implies that the instruments are weak. A rule of thumb proposed in the literature is that the weak instruments problem is a non-issue if the F statistic of the regression in the reduced form equation exceeds 10 (Glewwe, 2006).

### 5.2 Analysis Using SAS

We will illustrate the computations with the same dataset used in the previous sections. The variables *Min_Tem* and *Rain* were suspected of being endogenous. The variables *Longitude*, *Min_Tem_FD* and *Rain_FD* were considered as instruments and therefore the reduced regression equation for the *Foodgrain_Yld* equation is $Min\_Tem\ \ Rain = \alpha_0 + \alpha_1 Longitude + \alpha_2 Min\_Tem\_FD + \alpha_3 Rain\_FD + \gamma$

```
Note: The model we have been discussing is Foodgrain_Yld= Min_Tem Rain. If our model were having an already known exogenous variable i.e if the model were of the form Foodgrain_Yld= Min_Tem Rain Farmer_age where Farmer_age is the assumed exogenous variable. Then the Foodgrain_Yld equation would be as follows
```

$$Min_{Tem}\ \ Rain = \alpha_0 + \alpha_1 Farmer\_age + \alpha_2 Longitude + \alpha_3 Min\_Tem\_FD + \alpha_4 Rain\_FD + \gamma$$

The reduced form parameters are estimated by OLS regression. The following SAS statements can be used.

```
proc reg data=endogeneity_test;
model Min_Tem rain= Longitude Min_Tem_FD Rain_FD;
run;
```

```
Some parts of the output are omitted. Only the portion which has major relevance to the current interpretation is presented here. Full output can be got by running the program code with the dataset given in the appendix.
```

**Analysis of Variance**

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| **Model** | 3 | 90.25318 | 30.08439 | 85.75 | <.0001 |
| **Error** | 341 | 119.64199 | 0.35086 | | |
| **Corrected Total** | 344 | 209.89518 | | | |

**Parameter Estimates**

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| **Intercept** | **1** | -7.28679 | 1.78577 | -4.08 | <.0001 |
| **Longitude** | **1** | 0.27277 | 0.01926 | 14.17 | <.0001 |
| **Min_Tem_FD** | **1** | 0.48042 | 0.06455 | 7.44 | <.0001 |
| **Rain_FD** | **1** | 0.00074581 | 0.01765 | 0.04 | 0.9663 |

OUTPUT 5.1. Weak instruments analysis of the instrument Min**_Tem_FD**.

| Analysis of Variance | | | | | |
|---|---|---|---|---|---|
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 3 | 396.53789 | 132.17930 | 31.98 | <.0001 |
| **Error** | 341 | 1409.55738 | 4.13360 | | |
| **Corrected Total** | 344 | 1806.09527 | | | |

| Parameter Estimates | | | | | |
|---|---|---|---|---|---|
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > \|t\|** |
| **Intercept** | **1** | 38.45145 | 6.12949 | 6.27 | <.0001 |
| **Longitude** | **1** | -0.34542 | 0.06609 | -5.23 | <.0001 |
| **Min_Tem_FD** | **1** | -0.02320 | 0.22158 | -0.10 | 0.9167 |
| **Rain_FD** | **1** | 0.49668 | 0.06057 | 8.20 | <.0001 |

OUTPUT 5.2. Weak instruments analysis of the instrument Rain_FD.

### 5.3 Interpretation

The analysis results are given in Output 5.1 and 5.2. Note that the F statistic value in both the cases are very large (85.75 and 31.98) which are larger than 10 and therefore we cannot reject the hypothesis that we have weak instruments. Meaning is that we cannot conclude the variables *Longitude, Min_Tem_FD* and *Rain_FD* are weak.

## SECTION 6: TEST FOR ENDOGENEITY WITH IDENTIFIED INSTRUMENTS

### 6.1 Retesting Endogeneity

Thus we have identified the appropriate instruments viz *Longitude*, *Min_Tem_FD* and *Rain_FD* out of the pool of instruments in Section 4 and they are not found weak in Section 5. Now we can retest endogeneity of the regressors *Min_Tem* and *Rain* using the identified instrumental variables (*Longitude*, *Min_Tem_FD* and *Rain_FD*). The SAS code is given below followed by corresponding outputs.

### 6.2 Analysis Using SAS

```
proc reg data=endogeneity_test noprint;
model Min_Tem=Longitude   Min_Tem_FD Rain_FD ;
output out=endogeneity_test1 r= v ;
run;

proc reg data=endogeneity_test1;
model Foodgrain_Yld=Min_Tem Rain v ;
run;
```

| Parameter Estimates | | | | | | |
|---|---|---|---|---|---|---|
| **Variable** | **Label** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > \|t\|** |
| **Intercept** | Intercept | **1** | 3698.87196 | 855.00735 | 4.33 | <.0001 |
| **Min_Tem** | | **1** | -114.36108 | 46.57182 | -2.46 | 0.0146 |
| **Rain** | | **1** | -43.99562 | 10.41971 | -4.22 | <.0001 |

**Parameter Estimates**

| Variable | Label | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|
| **v** | Residual | **1** | 342.82216 | 60.93134 | 5.63 | <.0001 |

OUTPUT 6.1. Using Proc Reg to check for endogeneity of regressor Min_Tem.

```
proc reg data=endogeneity_test noprint;
model Rain=Longitude   Min_Tem_FD Rain_FD  ;
output out=endogeneity_test2 r= v1 ;
run;

proc reg data=endogeneity_test2;
model Foodgrain_Yld= Min_Tem Rain v1 ;
run;
```

**Parameter Estimates**

| Variable | Label | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|
| **Intercept** | Intercept | **1** | -674.54119 | 639.25252 | -1.06 | 0.2921 |
| **Min_Tem** | | **1** | 105.20425 | 32.22763 | 3.26 | 0.0012 |
| **Rain** | | **1** | 21.19310 | 23.43236 | 0.90 | 0.3664 |
| **v1** | Residual | **1** | -72.74688 | 26.19406 | -2.78 | 0.0058 |

OUTPUT 6.2. Using Proc Reg to check for endogeneity of regressor Rain.

### 6.3 Interpretation

It is evident from Output 6.1 and 6.2 that with the identified and appropriate instrumental variables *Longitude*, *Min_Tem_FD* and *Rain_FD* too, both the explanatory variables *Min_Tem* and *Rain* are found endogenous. (Remember also the fact in Section 2 that the regressors *Min_Tem* and *Rain* were found endogenous when tested with all the available instruments viz *Latitude*, *Longitude*, *Foodgrain_Yld_FD*, *Min_Tem_FD* and *Rain_FD*).

Now we can estimate parameters of our model of interest (*Foodgrain_Yld=Min_Tem Rai*) using 2SLS instead of OLS because OLS estimates will be bias as the regressors are endogenous as confirmed in the previous sections specially in Section 6 which confirmed the endogeneity with the appropriate instruments. The 2SLS estimate is as follows.

## SECTION 7: TWO STAGE LEAST SQUARES ESTIMATE

### 7.1 Two Stage Least Squares (TSLS)

In the discussion so far, we assumed that the regression models suffer from the presence of endogenous explanatory variable(s). There are techniques to estimate the model parameters under the presence of endogenous variables. 2SLS is one. Wooldridge, J. M. (2002, 2015) and Vivek B. Ajmani (2011) has given the total concept and we can simplify the steps of test mechanism in the following way.

**Stage 1:**
Isolate the part of *X* that is uncorrelated with error. We do this by regressing *X* on *Z* using OLS i.e. $X_i = \pi_0 + \pi_1 Z_i + v_i$ because $Z_i$ is uncorrelated with $e_i$ and $\pi_0 + \pi_1 Z_i$ is uncorrelated with $e_i$. We don't know $\pi_0$ or $\pi_1$ but we have estimated them so as to obtain the predicted values of *X* given *Z* i.e. $\hat{X}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i$. This is "clean". Since *Z* is uncorrelated with the model error term, so is any linear function of *Z*.

**Stage 2:**
Replace $X_i$ by the predicted values of $X_i$ in the regression of interest. Next regress $Y$ on $X$ (the predicted $X$ from the first stage regression), $Y = b_0 + b_1\hat{X} + e$ because $\hat{X}$ is uncorrelated with $e$ in large samples then $b_1$ can be estimated consistently by OLS using this second stage regression. This regression does not suffer from endogeneity. But it does suffer from having less variance in its regressor.

**Two-Stage Least Squares Summary**

**Preliminaries:** Seek out an appropriate instrument $Z$ Generally this is not easy because 1)It has to be exogenous, that is uncorrelated with the error term 2) It has to be relevant to the endogenous variable

**Stage 1:** Regress $X_i$ on $Z_i$ using OLS to obtain predicted values $\hat{X}_i$

**Stage 2:** Using OLS, regress $Y_i$ on $\hat{X}_i$; the estimated coefficient on $\hat{X}_i$ is the 2sls estimator of 1

**Postscript:** Generally we want more than one instrument, so as to improve the prediction $\hat{X}_i$. Also, there may be more than one endogenous variable, e.g. $X_{1i}$, $X_{2i}$ ...

## 7.2 Analysis Using SAS

2SLS estimate can be achieved using the following SAS code. In the first *proc reg* the predicted value of endogenous regressors *Min_Tem* and *Rain* designated by *f* and *f1* respectively are saved in the temporary dataset *endogeneity_test1*. In the second *proc reg* endogenous regressors *Min_Tem* and *Rain* are replaced by *f* and *f1* as regressors in the original model.

```
proc reg data=endogeneity_test noprint;
model Min_Tem Rain=Longitude   Min_Tem_FD Rain_FD  ;
output out=endogeneity_test1 p= f p= f1 ;
run;

proc reg data=endogeneity_test1;
model Foodgrain_Yld=f f1 ;
run;
```

The REG Procedure
Model: MODEL1
Dependent Variable: Foodgrain_Yld

| | |
|---|---|
| **Number of Observations Read** | 376 |
| **Number of Observations Used** | 345 |
| **Number of Observations with Missing Values** | 31 |

**Analysis of Variance**

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| **Model** | 2 | 624702 | 312351 | 1.44 | 0.2375 |
| **Error** | 342 | 74004155 | 216386 | | |
| **Corrected Total** | 344 | 74628857 | | | |

| | | | |
|---|---|---|---|
| **Root MSE** | 465.17354 | **R-Square** | 0.0084 |
| **Dependent Mean** | 1356.53102 | **Adj R-Sq** | 0.0026 |
| **Coeff Var** | 34.29140 | | |

| Parameter Estimates | | | | | | |
|---|---|---|---|---|---|---|
| **Variable** | **Label** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > \|t\|** |
| **Intercept** | Intercept | **1** | 3129.37492 | 1054.28997 | 2.97 | 0.0032 |
| **f** | Predicted Value of Min_Tem | **1** | -91.80926 | 54.03377 | -1.70 | 0.0902 |
| **f1** | Predicted Value of Rain | **1** | -18.55317 | 25.77830 | -0.72 | 0.4722 |

OUTPUT 7.1. Using Proc Reg to estimate 2SLS.

2SLS estimate (Output 7.1) can be compared with the OLS estimate result given by the below SAS code and Output 7.2.

```
proc reg data=endogeneity_test;
model Foodgrain_Yld=Min_Tem Rain ;
run;
```

| Parameter Estimates | | | | | |
|---|---|---|---|---|---|
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > \|t\|** |
| **Intercept** | **1** | -105.41631 | 578.21753 | -0.18 | 0.8554 |
| **Min_Tem** | **1** | 93.45748 | 31.26842 | 2.99 | 0.0030 |
| **Rain** | **1** | -35.79684 | 10.73527 | -3.33 | 0.0009 |

OUTPUT 7.2. Using Proc Reg to estimate OLS.

### 7.3 Interpretation

From Output 7.1 and 7.2 it is evident that the 2SLS results differ from the least squares results in many ways. The OLS standard errors are smaller than the instrumental variables standard error. The parameter estimates for the intercept in instrumental variable estimation (2SLS) is significant at 1% significance level and larger in magnitude than the one obtained from OLS while estimated intercept in OLS is not significant. The OLS estimate for the parameter values of *Min_Tem* and *Rain* are highly significant and larger than that of 2SLS estimate. Hence, if we use OLS estimate when the regressors are endogenous, we would have misinterpreted the system.

## SECTION 8: HAUSMAN'S TEST FOR DETERMINING BETWEEN OLS AND 2SLS ESTIMATORS AND ACCORDINGLY CONCLUDING ENDOGENEITY OR EXOGENEITY

### 8.1 Hausman's Specification Test

In Section 6 of Hausman's test for endogeneity and Section 7 of estimation of 2SLS, various *prog reg* statements are used. But in SAS, a dedicated package of code (using *proc model*) for Hausman's test can be used to accomplish both Section 6 and Section 7 at one go. Several of the estimation methods supported by *Proc Model* are instrumental variables methods. Here, Hausman's specification test is used to determine if there are significant differences between the OLS and the IV estimators (2SLS). As discussed in Greene (2003, pp. 80–83), under the null hypothesis of no endogeneity, both the OLS and the IV estimators are consistent. Under the alternative hypothesis of endogeneity, only the IV estimator is consistent. Accordingly endogeneity can also be concluded. If significant difference is found between OLS and IV estimates, then the suggested estimate is IV and in turn it suggest evidence of at least one endogenous explanatory variable in the model. On the other hand, if OLS is suggested by the test, then all the regressors are exogenous. Hausman's test is based on the principle that if there are two estimators ($\beta^1$, $\beta^2$) which converge to β under the null hypothesis and converge to different values under the alternative hypothesis then the null hypothesis can be tested by testing whether the two estimators are different. The test statistic is given by

$$H = d^T[s_{IV}^2(\hat{X}^T\hat{X})^{-1} - s_{OLS}^2(X^TX)^{-1}]^{-1}$$

Here, $d = [b_{IV} \_ b_{OLS}]$, $s_{IV}^2(\hat{X}^T\hat{X})^{-1}$, $s_{OLS}^2(X^TX)^{-1}$ are the terms associated with the asymptotic covariance of the two estimators, respectively. Under the null hypothesis, $H_0$ is distributed as a $\chi^2$ with $k^*$ degrees of freedom. The degree of freedom, $k^*$, is the number of variables in $X$ that are suspected of being endogenous.

## 8.2 Analysis Using SAS

We will use the same dataset of *endogeneity_test* used in the previous sections to illustrate the computations involved in SAS. This Hausman's test can be performed in SAS by using the *proc model* procedure as shown in the following code. Notice that we specify the endogenous variable in the "endo" statement, the instruments in the "instruments" statement, and then write down the linear model to be estimated. *beta1*, *beta2* and *beta3* are the parameters to be estimated. This is followed by the "fit" statement using the dependent variable with the option that the Hausman's test be used to compare the OLS and the instrumental variable estimator.

```
proc model data=endogeneity_test;
endo Min_Tem Rain;
instruments  Longitude Min_Tem_FD Rain_FD ;
Foodgrain_Yld=beta1+beta2*Min_Tem+beta3*Rain;
fit Foodgrain_Yld/ols 2sls hausman;
run;
```

Note: The model we have been discussing is Foodgrain_Yld= Min_Tem Rain. If our model were having an already known exogenous variable i.e if the model were of the form Foodgrain_Yld= Min_Tem Rain Farmer_age where Farmer_age is the assumed exogenous variable. Then the SAS code would be as follows.

```
proc model data=endogeneity_test;
endo Min_Tem Rain;
instruments  Farmer_age Longitude   Min_Tem_FD Rain_FD ;
Foodgrain_Yld=beta1+beta2*Min_Tem+beta3*Rain+ beta4*Farmer_age;
fit Foodgrain_Yld/ols 2sls hausman;
run;
```

The procedure checks if the OLS estimates are more efficient than the 2SLS procedure. The degree of freedom used for the test is k, the number of columns of X. The analysis results are produced in Output 8.1.

Some parts of the output are omitted. Only the portion which has major relevance to the current interpretation is presented here. Full output can be got by running the program code with the dataset given in the appendix.

NOTE: At OLS Iteration 1 CONVERGE=0.001 Criteria Met.

| Minimization Summary | |
|---|---|
| **Parameters Estimated** | 3 |
| **Method** | Gauss |
| **Iterations** | 1 |

The MODEL Procedure

**Nonlinear OLS Summary of Residual Errors**

| Equation | DF Model | DF Error | SSE | MSE | Root MSE | R-Square | Adj R-Sq |
|---|---|---|---|---|---|---|---|
| **Foodgrain_Yld** | 3 | 342 | 70079646 | 204911 | 452.7 | 0.0610 | 0.0555 |

| Nonlinear OLS Parameter Estimates | | | | |
|---|---|---|---|---|
| **Parameter** | **Estimate** | **Approx Std Err** | **t Value** | **Approx Pr > \|t\|** |
| **beta1** | 70.22319 | 586.0 | 0.12 | 0.9047 |
| **beta2** | 84.48972 | 31.6588 | 2.67 | 0.0080 |
| **beta3** | -36.7198 | 10.7926 | -3.40 | 0.0007 |

NOTE: At 2SLS Iteration 1 CONVERGE=0.001 Criteria Met.

| Minimization Summary | |
|---|---|
| **Parameters Estimated** | 3 |
| **Method** | Gauss |
| **Iterations** | 1 |

The MODEL Procedure

| Nonlinear 2SLS Summary of Residual Errors | | | | | | | |
|---|---|---|---|---|---|---|---|
| **Equation** | **DF Model** | **DF Error** | **SSE** | **MSE** | **Root MSE** | **R-Square** | **Adj R-Sq** |
| **Foodgrain_Yld** | 3 | 342 | 77835068 | 227588 | 477.1 | -0.0430 | -0.0491 |

| Nonlinear 2SLS Parameter Estimates | | | | |
|---|---|---|---|---|
| **Parameter** | **Estimate** | **Approx Std Err** | **t Value** | **Approx Pr > \|t\|** |
| **beta1** | 3129.375 | 1081.2 | 2.89 | 0.0040 |
| **beta2** | -91.8093 | 55.4147 | -1.66 | 0.0985 |
| **beta3** | -18.5532 | 26.4371 | -0.70 | 0.4833 |

| Hausman's Specification Test Results | | | | |
|---|---|---|---|---|
| **Efficient under H0** | **Consistent under H1** | **DF** | **Statistic** | **Pr > ChiSq** |
| **OLS** | 2SLS | 3 | 25.28 | <.0001 |

OUTPUT 8.1. Hausman's test for the data using Proc model.

### 8.3 Interpretation

Output from both OLS and instrumental variable estimation are shown in Output 8.1. Note that convergence was achieved very quickly for both models. The 2SLS results differ from the least squares results in many ways. The OLS standard errors are smaller than the instrumental variables standard error. The parameter estimates for the intercept (*beta1*) and *Min_Tem* (*beta2*) in instrumental variable estimation are larger in magnitude than the ones obtained from OLS. The instrumental variable estimate for the parameter value of *Rain* (*beta3*) is smaller than that of OLS estimate. Both *beta2* and *beta3* in OLS are significant at 1% significance level while in 2SLS only *beta1* is significant at 1% significance level. The value of the test statistic for Hausman's test is 25.28 with 3 degrees of freedom and is highly significant (p-value <0.0001) indicating that the instrumental variable estimator is more efficient than the OLS estimator. This is the first point where the *proc model* statements in the current Section 8 are making a difference from *proc reg* statements in Section 6 and Section 7 that it (*proc model*) determines if there are significant differences between OLS and 2SLS estimates. This cannot be covered by *prog reg* statements as shown in Section 7 where only the parameters are estimated but not tested for significant difference between OLS and 2SLS. From output 8.1, as Housman's test suggest 2SLS estimate rather than

OLS estimate, it can also be concluded the presence of at least one endogeneous explanatory variable resulting the OLS model suffers from endogeneity. Meaning is that at least one regressor is confirmed to be endogenous. It may be either *Min_Temp* or *Rain* or both. In fact this much reason is enough for (a) why OLS estimator should not be used and (b) confirmation of endogeneity in the regressor variables. However we may want to test which regressor(s) is/are endogenous. This is the second point where the *proc model* statements are again making a difference from *proc reg* statements that it (*proc model*) does not report which regressor variables are endogenous (instead it only reports at least one is endogenous). But this can be covered by *prog reg* statements as shown in Section 6 where we have confirmed that both the regressors *Min_Temp* and *Rain* are endogenous and instruments deployed are not weak but appropriate. The discussion on comparison of the two procedures of SAS can be summarized as follows.

Table 8.1: Summary of Comparison of Housman' Test performed by Proc reg and Proc model.

| **Action** | **Proc reg** | **Proc model** |
|---|---|---|
| Estimates parameters for both OLS and 2SLS | Yes | Yes |
| Determines if there are significant differences between OLS and 2SLS estimates | No | Yes |
| Suggests either OLS or 2SLS | No | Yes |
| Reports which regressor variables are endogenous | Yes | No (but reports at least one is endogenous) |
| Overidentification test (separately using *proc syslin*) | Yes | Yes |
| Weak instrument test (separately using *proc reg*) | Yes | Yes |

NB: This comparison is totally based as per the SAS codes used in the current study.

These two procedures *Proc Reg* and *Proc Model* are supposed to give the same results of estimates. However the results are deviated a little from each other though the underlying interpretation is the same. This deviation may be due to the sample size (<50) which is not so large, or/and the method-specific computational variations of the different procedures in estimating the parameters. The estimated results of the two procedures presented in Output 7.1, 7.2 and 8.1 are compared to view at one grasp in the following table.

Table 8.2: Estimates of OLS and 2SLS by Proc reg and Proc model.

| | Using Proc Reg | | | | Using Proc Model | | | |
|---|---|---|---|---|---|---|---|---|
| | 2SLS Parameter Estimates | | | | 2SLS Parameter Estimates | | | |
| Parameter | Estimate | Std Err | t Value | p Value | Estimate | Std Err | t Value | p Value |
| Intercept | 3129.37492 | 1054.28997 | 2.97 | 0.0032 | 3129.375 | 1081.2 | 2.89 | 0.0040 |
| Min_Tem | -91.80926 | 54.03377 | -1.70 | 0.0902 | -91.8093 | 55.4147 | -1.66 | 0.0985 |
| Rain | -18.55317 | 25.77830 | -0.72 | 0.4722 | -18.5532 | 26.4371 | -0.70 | 0.4833 |
| | OLS Parameter Estimates | | | | OLS Parameter Estimates | | | |
| Intercept | -105.41631 | 578.21753 | -0.18 | 0.8554 | 70.22319 | 586.0 | 0.12 | 0.9047 |
| Min_Tem | 93.45748 | 31.26842 | 2.99 | 0.0030 | 84.48972 | 31.6588 | 2.67 | 0.0080 |
| Rain | -35.79684 | 10.73527 | -3.33 | 0.0009 | -36.7198 | 10.7926 | -3.40 | 0.0007 |

Hausman's test can also be used to compare 2SLS with 3SLS for a class of estimators for which 3SLS is asymptotically efficient (similarly for OLS and SUR). Hausman's m-statistic can also be used, in principle, to test the null hypothesis of normality when comparing 3SLS to FIML. Because of the poor performance of this form of the test, it is not offered in the MODEL procedure. Refer to R.C. Fair (1984, pp. 246-247) for a discussion of why Hausman's test fails for common econometric models. In the following example, OLS, SUR, 2SLS, 3SLS, and FIML are used to estimate a model, and Hausman's test is requested.

```
fit Foodgrain_Yld/ols sur 2sls 3sls fiml hausman;
```

## CONCLUDING REMARK

"**How to Test Endogeneity or Exogeneity: An E-Learning Hands on SAS**" was borne during the journey of my PhD programme as I stumble at a small requirement of testing a variable for its endogeneity. Beyond my requirement of thesis I was compelled to go little further into the lesson as it was found challenging and interesting for me. To handle a model suffering from endogeneity, seeking instruments which are, in some sense, very controversial and ambiguous are real challenge. Another enjoyable challenge is when the theory gets down on the editor window of a powerful statistical package like SAS. However, whatever is presented here is less than a drop of the ocean of the concerning topic. The intension is to enable researchers use a ready to serve template for testing of endogeneity in SAS because I was stuck during my time. Vivek B. Ajmani's *Applied Econometrics Using The SAS System* (2011, chapter 4 page 52-69) was really an eye opener for me regarding endogeneity test. Many techniques are followed as per this incredible book. I am indebted to my co-authors Professor Kishore K. Das and Dr. Aniruddha Roy for their constant contribution of statistical and econometric concepts in materializing this "Hands on SAS". For any communication please mail me at uttamba@gmail.com.

## REFERENCES


Bowden, R.J. and Turkington, D.A. (1984). *Instrumental Variables*, Cambridge University Press.

Fair, R.C. (1984). *Specification, Estimation, and Analysis of Macroeconometric Models*, Cambridge: Harvard University Press.

Glewwe, P. (2006). Department of Applied Economics, St. Paul, MN, personal communication, January 31, 2006.

Greene, W. H. (2003). *Econometric Analysis*, Prentice Hall, New Jersey.

Hamilton, Barton H.; Nickerson, Jackson A. (2016-07-31). *Correcting for Endogeneity in Strategic Management Research".* Strategic Organization. 1 (1): 51–78.

Sargan, J. D. (1958). *The Estimation of Economic Relationships Using Instrumental Variables*. *Econometrica*, 26: 393–415.

SAS 9.3 Help and Documentation (2002-2011). SAS Institute Inc. Cary, NC, USA.

Vivek B. Ajmani (2011). *Applied Econometrics Using The SAS System*. John Wiley & Sons, Inc., Hoboken, New Jersey. ISBN: 1118210328, 9781118210321

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*, Massachusetts Institute of Technology, Cambridge, MA.

Wooldridge, J. M. (2015). *Introductory Econometrics: A Modern Approach*, 5th ed., South Western College Publishing.


## APPENDIX

The following is the complete program code and the accompanied dataset which have been used to illustrate various analyses discussed in this learning module. One can directly copy and paste in the editor window of SAS system and run the procs one by one. On the top of every proc, comments are given for understanding of the code for which it is meant. Interpretation can be understood from the discussions made in every section.

```
/******************************************************************************************************/
/*© 2017 N. Uttam Singh */
/*N. Uttam Singh , Scientist, ICAR Research Complex for NEH Region, Umiam, Meghalaya, India 793103*/
/******************************************************************************************************/
Data endogeneity_test;
input Foodgrain_Yld Min_Tem Rain Latitude Longitude  Foodgrain_Yld_FD Min_Tem_FD Rain_FD;
cards;
821.84  17.79   6.56    27.10   93.62   .       .       .
793.56  19.51   6.64    27.10   93.62   -28.28  1.72    0.08
1428.57 18.27   6.35    27.10   93.62   635.01  -1.24   -0.29
827.84  18.18   9.54    27.10   93.62   -600.73 -0.10   3.19
847.40  17.81   7.26    27.10   93.62   19.56   -0.36   -2.28
864.20  17.89   6.85    27.10   93.62   16.80   0.07    -0.41
876.90  18.50   7.50    27.10   93.62   12.70   0.61    0.64
882.83  18.16   9.78    27.10   93.62   5.93    -0.34   2.29
926.14  18.07   8.25    27.10   93.62   43.32   -0.10   -1.53
896.28  18.36   7.88    27.10   93.62   -29.86  0.29    -0.37
900.28  18.23   9.70    27.10   93.62   4.00    -0.13   1.82
1013.51 18.18   6.76    27.10   93.62   113.23  -0.05   -2.94
1052.72 18.26   7.58    27.10   93.62   39.21   0.08    0.82
1263.86 18.43   9.38    27.10   93.62   211.14  0.17    1.81
1050.52 18.14   7.13    27.10   93.62   -213.33 -0.29   -2.26
1043.25 18.13   7.32    27.10   93.62   -7.27   0.00    0.19
1045.49 17.85   8.38    27.10   93.62   2.24    -0.29   1.07
1044.46 17.89   8.70    27.10   93.62   -1.03   0.04    0.32
```

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1070.54 | 18.02 | 8.44 | 27.10 | 93.62 | 26.09 | 0.13 | -0.26 |
| 1579.43 | 17.79 | 6.78 | 27.10 | 93.62 | 508.89 | -0.24 | -1.66 |
| 1067.96 | 18.22 | 9.57 | 27.10 | 93.62 | -511.47 | 0.43 | 2.79 |
| 1083.19 | 18.63 | 8.66 | 27.10 | 93.62 | 15.23 | 0.41 | -0.91 |
| 1119.05 | 18.10 | 7.53 | 27.10 | 93.62 | 35.85 | -0.53 | -1.13 |
| 1151.82 | 18.18 | 8.30 | 27.10 | 93.62 | 32.77 | 0.08 | 0.77 |
| 1161.00 | 19.36 | 8.20 | 27.10 | 93.62 | 9.18 | 1.18 | -0.10 |
| 1164.81 | 17.93 | 6.48 | 27.10 | 93.62 | 3.81 | -1.43 | -1.72 |
| 1060.56 | 19.22 | 7.52 | 27.10 | 93.62 | -104.26 | 1.30 | 1.04 |
| 1221.70 | 19.49 | 6.32 | 27.10 | 93.62 | 161.14 | 0.26 | -1.20 |
| 1086.61 | 19.69 | 7.80 | 27.10 | 93.62 | -135.08 | 0.21 | 1.48 |
| 1111.41 | 19.18 | 6.72 | 27.10 | 93.62 | 24.80 | -0.51 | -1.09 |
| 1171.30 | 18.35 | 5.83 | 27.10 | 93.62 | 59.88 | -0.83 | -0.88 |
| 1145.66 | 19.42 | 8.18 | 27.10 | 93.62 | -25.64 | 1.07 | 2.34 |
| 1040.42 | 19.92 | 5.93 | 27.10 | 93.62 | -105.24 | 0.50 | -2.25 |
| 1124.67 | 19.15 | 6.14 | 27.10 | 93.62 | 84.24 | -0.77 | 0.21 |
| 1147.65 | 19.49 | 5.16 | 27.10 | 93.62 | 22.99 | 0.33 | -0.98 |
| 1153.93 | 19.63 | 6.86 | 27.10 | 93.62 | 6.27 | 0.15 | 1.70 |
| 1226.21 | 20.52 | 7.50 | 27.10 | 93.62 | 72.29 | 0.88 | 0.64 |
| 1277.34 | 22.91 | 7.69 | 27.10 | 93.62 | 51.13 | 2.40 | 0.19 |
| 1178.09 | 21.54 | 7.74 | 27.10 | 93.62 | -99.25 | -1.38 | 0.05 |
| 1212.38 | 18.37 | 4.40 | 27.10 | 93.62 | 34.29 | -3.17 | -3.34 |
| 1216.34 | 18.80 | 5.50 | 27.10 | 93.62 | 3.96 | 0.44 | 1.11 |
| 1240.64 | 18.80 | 4.87 | 27.10 | 93.62 | 24.30 | 0.00 | -0.63 |
| 1255.15 | 20.96 | 5.68 | 27.10 | 93.62 | 14.51 | 2.16 | 0.81 |
| 1555.49 | 19.38 | 7.92 | 27.10 | 93.62 | 300.34 | -1.58 | 2.23 |
| 1663.28 | 18.61 | 4.38 | 27.10 | 93.62 | 107.79 | -0.77 | -3.54 |
| 1766.57 | 18.10 | 3.94 | 27.10 | 93.62 | 103.29 | -0.51 | -0.44 |
| 1771.51 | 18.50 | 5.98 | 27.10 | 93.62 | 4.94 | 0.40 | 2.04 |
| 865.64 | 17.61 | 5.14 | 26.75 | 94.22 | . | . | . |
| 923.02 | 17.62 | 5.77 | 26.75 | 94.22 | 57.38 | 0.02 | 0.63 |
| 989.31 | 18.16 | 5.40 | 26.75 | 94.22 | 66.29 | 0.54 | -0.37 |
| 889.57 | 18.11 | 6.88 | 26.75 | 94.22 | -99.74 | -0.05 | 1.48 |
| 974.89 | 17.68 | 6.09 | 26.75 | 94.22 | 85.31 | -0.43 | -0.79 |
| 945.44 | 17.75 | 5.73 | 26.75 | 94.22 | -29.45 | 0.07 | -0.36 |
| 1042.10 | 18.34 | 6.23 | 26.75 | 94.22 | 96.65 | 0.59 | 0.50 |
| 970.63 | 17.93 | 7.48 | 26.75 | 94.22 | -71.47 | -0.41 | 1.24 |
| 943.31 | 17.84 | 5.94 | 26.75 | 94.22 | -27.32 | -0.09 | -1.53 |
| 996.28 | 18.17 | 5.40 | 26.75 | 94.22 | 52.97 | 0.33 | -0.55 |
| 913.19 | 18.06 | 6.57 | 26.75 | 94.22 | -83.09 | -0.10 | 1.17 |
| 998.29 | 17.98 | 4.94 | 26.75 | 94.22 | 85.10 | -0.08 | -1.62 |
| 941.28 | 18.09 | 5.12 | 26.75 | 94.22 | -57.01 | 0.12 | 0.18 |
| 865.66 | 18.28 | 6.06 | 26.75 | 94.22 | -75.62 | 0.19 | 0.94 |
| 1073.09 | 17.90 | 5.58 | 26.75 | 94.22 | 207.44 | -0.38 | -0.48 |
| 965.59 | 17.89 | 5.86 | 26.75 | 94.22 | -107.50 | -0.01 | 0.27 |
| 1082.56 | 17.70 | 6.36 | 26.75 | 94.22 | 116.97 | -0.20 | 0.51 |
| 1054.05 | 17.66 | 6.22 | 26.75 | 94.22 | -28.51 | -0.03 | -0.14 |
| 1009.98 | 17.83 | 5.99 | 26.75 | 94.22 | -44.06 | 0.17 | -0.23 |
| 1111.34 | 17.67 | 5.21 | 26.75 | 94.22 | 101.35 | -0.16 | -0.78 |
| 1003.18 | 18.10 | 6.24 | 26.75 | 94.22 | -108.16 | 0.44 | 1.03 |
| 1117.32 | 18.50 | 6.53 | 26.75 | 94.22 | 114.14 | 0.40 | 0.29 |
| 1028.00 | 17.84 | 5.91 | 26.75 | 94.22 | -89.31 | -0.65 | -0.62 |
| 1104.83 | 18.02 | 6.61 | 26.75 | 94.22 | 76.82 | 0.18 | 0.70 |
| 1266.07 | 17.87 | 6.46 | 26.75 | 94.22 | 161.24 | -0.15 | -0.15 |
| 1202.05 | 17.73 | 5.55 | 26.75 | 94.22 | -64.01 | -0.14 | -0.90 |
| 1235.01 | 17.86 | 6.53 | 26.75 | 94.22 | 32.96 | 0.13 | 0.97 |
| 1266.06 | 18.37 | 5.12 | 26.75 | 94.22 | 31.05 | 0.51 | -1.40 |
| 1283.25 | 18.47 | 6.14 | 26.75 | 94.22 | 17.19 | 0.09 | 1.02 |
| 1281.21 | 18.40 | 5.27 | 26.75 | 94.22 | -2.04 | -0.07 | -0.88 |
| 1294.43 | 17.85 | 5.17 | 26.75 | 94.22 | 13.22 | -0.55 | -0.10 |
| 1314.28 | 19.00 | 5.83 | 26.75 | 94.22 | 19.85 | 1.16 | 0.66 |
| 1287.64 | 19.13 | 6.29 | 26.75 | 94.22 | -26.64 | 0.13 | 0.45 |
| 1408.85 | 18.28 | 5.90 | 26.75 | 94.22 | 121.22 | -0.85 | -0.38 |
| 1457.38 | 18.81 | 5.17 | 26.75 | 94.22 | 48.53 | 0.53 | -0.73 |
| 1460.25 | 18.79 | 5.70 | 26.75 | 94.22 | 2.88 | -0.02 | 0.53 |
| 1416.52 | 18.58 | 6.15 | 26.75 | 94.22 | -43.74 | -0.21 | 0.45 |
| 1471.55 | 18.79 | 6.80 | 26.75 | 94.22 | 55.04 | 0.21 | 0.65 |
| 1404.85 | 18.98 | 5.84 | 26.75 | 94.22 | -66.70 | 0.19 | -0.95 |
| 1415.90 | 18.23 | 4.60 | 26.75 | 94.22 | 11.05 | -0.75 | -1.25 |
| 1285.71 | 18.58 | 5.95 | 26.75 | 94.22 | -130.19 | 0.35 | 1.35 |
| 1378.08 | 18.51 | 5.56 | 26.75 | 94.22 | 92.36 | -0.07 | -0.39 |
| 1551.22 | 19.04 | 4.63 | 26.75 | 94.22 | 173.14 | 0.54 | -0.93 |
| 1662.44 | 18.87 | 6.57 | 26.75 | 94.22 | 111.22 | -0.17 | 1.94 |
| 1762.69 | 18.27 | 4.51 | 26.75 | 94.22 | 100.25 | -0.60 | -2.05 |
| 1704.33 | 17.70 | 4.97 | 26.75 | 94.22 | -58.36 | -0.57 | 0.46 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1961.95 | 18.19 | 4.88 | 26.75 | 94.22 | 257.62 | 0.48 | -0.10 |
| 1743.62 | 17.56 | 3.64 | 24.80 | 93.95 | . | . | . |
| 1402.23 | 17.66 | 3.44 | 24.80 | 93.95 | -341.39 | 0.10 | -0.20 |
| 1826.36 | 17.90 | 3.58 | 24.80 | 93.95 | 424.13 | 0.23 | 0.14 |
| 1628.74 | 18.06 | 3.72 | 24.80 | 93.95 | -197.62 | 0.16 | 0.14 |
| 1153.74 | 17.58 | 5.65 | 24.80 | 93.95 | -475.00 | -0.47 | 1.93 |
| 1239.83 | 17.82 | 2.39 | 24.80 | 93.95 | 86.09 | 0.24 | -3.26 |
| 1083.90 | 18.44 | 4.44 | 24.80 | 93.95 | -155.93 | 0.61 | 2.05 |
| 1438.14 | 18.09 | 5.72 | 24.80 | 93.95 | 354.23 | -0.35 | 1.28 |
| 1532.38 | 17.84 | 5.31 | 24.80 | 93.95 | 94.25 | -0.25 | -0.41 |
| 1559.96 | 18.11 | 4.96 | 24.80 | 93.95 | 27.58 | 0.27 | -0.35 |
| 1518.97 | 18.06 | 5.40 | 24.80 | 93.95 | -40.99 | -0.05 | 0.44 |
| 1662.67 | 17.93 | 3.23 | 24.80 | 93.95 | 143.70 | -0.13 | -2.17 |
| 1611.41 | 18.05 | 3.02 | 24.80 | 93.95 | -51.26 | 0.12 | -0.21 |
| 1468.85 | 18.11 | 4.38 | 24.80 | 93.95 | -142.56 | 0.07 | 1.36 |
| 1464.99 | 17.99 | 3.41 | 24.80 | 93.95 | -3.86 | -0.12 | -0.96 |
| 1508.34 | 17.97 | 5.15 | 24.80 | 93.95 | 43.35 | -0.03 | 1.73 |
| 1394.99 | 17.87 | 5.39 | 24.80 | 93.95 | -113.34 | -0.10 | 0.25 |
| 1600.36 | 17.68 | 4.81 | 24.80 | 93.95 | 205.37 | -0.19 | -0.58 |
| 1997.69 | 17.96 | 4.62 | 24.80 | 93.95 | 397.32 | 0.29 | -0.19 |
| 2038.17 | 17.70 | 4.25 | 24.80 | 93.95 | 40.48 | -0.27 | -0.37 |
| 1500.00 | 18.08 | 4.34 | 24.80 | 93.95 | -538.17 | 0.38 | 0.09 |
| 1678.59 | 18.37 | 4.06 | 24.80 | 93.95 | 178.59 | 0.29 | -0.29 |
| 1662.62 | 17.80 | 4.76 | 24.80 | 93.95 | -15.97 | -0.57 | 0.70 |
| 1555.02 | 17.93 | 5.06 | 24.80 | 93.95 | -107.60 | 0.13 | 0.30 |
| 1762.96 | 17.93 | 6.34 | 24.80 | 93.95 | 207.94 | 0.00 | 1.28 |
| 2130.86 | 17.76 | 3.92 | 24.80 | 93.95 | 367.90 | -0.17 | -2.42 |
| 1819.02 | 17.82 | 5.33 | 24.80 | 93.95 | -311.84 | 0.06 | 1.41 |
| 2166.46 | 18.37 | 3.57 | 24.80 | 93.95 | 347.45 | 0.55 | -1.76 |
| 2990.82 | 18.45 | 4.25 | 24.80 | 93.95 | 824.36 | 0.08 | 0.68 |
| 2158.60 | 18.42 | 3.95 | 24.80 | 93.95 | -832.22 | -0.03 | -0.31 |
| 2264.93 | 17.75 | 3.94 | 24.80 | 93.95 | 106.33 | -0.67 | 0.00 |
| 2258.82 | 18.77 | 3.96 | 24.80 | 93.95 | -6.10 | 1.02 | 0.01 |
| 2309.01 | 19.19 | 3.87 | 24.80 | 93.95 | 50.18 | 0.41 | -0.08 |
| 2327.76 | 18.27 | 3.98 | 24.80 | 93.95 | 18.75 | -0.91 | 0.11 |
| 2358.76 | 18.56 | 3.87 | 24.80 | 93.95 | 31.00 | 0.29 | -0.11 |
| 2305.70 | 18.99 | 3.83 | 24.80 | 93.95 | -53.06 | 0.42 | -0.04 |
| 2217.42 | 18.56 | 3.56 | 24.80 | 93.95 | -88.28 | -0.42 | -0.27 |
| 2355.17 | 18.82 | 5.92 | 24.80 | 93.95 | 137.75 | 0.26 | 2.35 |
| 2416.62 | 19.00 | 4.09 | 24.80 | 93.95 | 61.45 | 0.18 | -1.83 |
| 2241.28 | 18.25 | 3.29 | 24.80 | 93.95 | -175.34 | -0.75 | -0.80 |
| 2241.28 | 18.53 | 5.93 | 24.80 | 93.95 | 0.00 | 0.28 | 2.64 |
| 2297.39 | 18.47 | 5.93 | 24.80 | 93.95 | 56.10 | -0.06 | 0.00 |
| 2235.99 | 19.00 | 3.11 | 24.80 | 93.95 | -61.39 | 0.53 | -2.82 |
| 1795.99 | 18.85 | 4.79 | 24.80 | 93.95 | -440.00 | -0.16 | 1.68 |
| 2244.06 | 18.23 | 4.43 | 24.80 | 93.95 | 448.07 | -0.62 | -0.35 |
| 2396.64 | 17.74 | 4.05 | 24.80 | 93.95 | 152.58 | -0.49 | -0.38 |
| 1925.64 | 18.27 | 4.15 | 24.80 | 93.95 | -471.00 | 0.53 | 0.09 |
| . | 17.48 | 3.77 | 25.57 | 91.88 | . | . | . |
| . | 17.50 | 3.19 | 25.57 | 91.88 | . | . | . |
| . | 18.05 | 3.62 | 25.57 | 91.88 | . | . | . |
| . | 18.05 | 3.83 | 25.57 | 91.88 | . | . | . |
| 1022.56 | 17.58 | 4.87 | 25.57 | 91.88 | . | . | . |
| 1005.99 | 17.59 | 3.58 | 25.57 | 91.88 | -16.56 | 0.01 | -1.29 |
| 1041.59 | 18.22 | 5.08 | 25.57 | 91.88 | 35.60 | 0.63 | 1.50 |
| 1029.68 | 17.69 | 4.84 | 25.57 | 91.88 | -11.92 | -0.53 | -0.24 |
| 1009.75 | 17.68 | 4.83 | 25.57 | 91.88 | -19.93 | -0.02 | -0.01 |
| 1078.95 | 17.96 | 5.80 | 25.57 | 91.88 | 69.20 | 0.28 | 0.97 |
| 1128.27 | 17.90 | 6.00 | 25.57 | 91.88 | 49.32 | -0.06 | 0.20 |
| 1154.74 | 17.85 | 5.33 | 25.57 | 91.88 | 26.48 | -0.05 | -0.68 |
| 1166.54 | 18.00 | 4.45 | 25.57 | 91.88 | 11.79 | 0.15 | -0.88 |
| 1153.28 | 18.22 | 6.38 | 25.57 | 91.88 | -13.26 | 0.21 | 1.93 |
| 1245.18 | 17.85 | 5.14 | 25.57 | 91.88 | 91.90 | -0.36 | -1.24 |
| 1156.20 | 17.77 | 4.32 | 25.57 | 91.88 | -88.97 | -0.09 | -0.82 |
| 1140.31 | 17.73 | 6.15 | 25.57 | 91.88 | -15.89 | -0.04 | 1.83 |
| 1181.62 | 17.59 | 5.25 | 25.57 | 91.88 | 41.31 | -0.14 | -0.90 |
| 1152.85 | 17.75 | 5.57 | 25.57 | 91.88 | -28.77 | 0.16 | 0.31 |
| 1205.35 | 17.60 | 4.89 | 25.57 | 91.88 | 52.50 | -0.16 | -0.68 |
| 945.97 | 18.23 | 5.85 | 25.57 | 91.88 | -259.39 | 0.63 | 0.97 |
| 1027.44 | 18.46 | 6.25 | 25.57 | 91.88 | 81.47 | 0.23 | 0.39 |
| 1000.00 | 17.76 | 6.12 | 25.57 | 91.88 | -27.44 | -0.70 | -0.13 |
| 1128.55 | 18.02 | 7.09 | 25.57 | 91.88 | 128.55 | 0.26 | 0.97 |
| 1160.03 | 17.79 | 7.90 | 25.57 | 91.88 | 31.48 | -0.22 | 0.80 |
| 1163.30 | 17.61 | 4.96 | 25.57 | 91.88 | 3.27 | -0.18 | -2.93 |
| 1101.67 | 17.76 | 7.89 | 25.57 | 91.88 | -61.63 | 0.15 | 2.92 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1132.68 | 18.27 | 2.53 | 25.57 | 91.88 | 31.01 | 0.51 | -5.36 |
| 1096.63 | 18.25 | 5.26 | 25.57 | 91.88 | -36.05 | -0.02 | 2.73 |
| 1088.35 | 18.27 | 4.42 | 25.57 | 91.88 | -8.28 | 0.02 | -0.85 |
| 1348.71 | 17.72 | 5.68 | 25.57 | 91.88 | 260.37 | -0.55 | 1.26 |
| 1405.12 | 18.96 | 5.71 | 25.57 | 91.88 | 56.41 | 1.24 | 0.02 |
| 1398.36 | 19.06 | 5.38 | 25.57 | 91.88 | -6.76 | 0.10 | -0.32 |
| 1531.18 | 18.16 | 6.71 | 25.57 | 91.88 | 132.82 | -0.90 | 1.33 |
| 1597.63 | 18.49 | 5.99 | 25.57 | 91.88 | 66.45 | 0.34 | -0.73 |
| 1666.91 | 18.38 | 5.00 | 25.57 | 91.88 | 69.28 | -0.11 | -0.98 |
| 1685.97 | 18.45 | 5.50 | 25.57 | 91.88 | 19.05 | 0.08 | 0.50 |
| 1733.14 | 18.63 | 5.86 | 25.57 | 91.88 | 47.17 | 0.17 | 0.36 |
| 1648.61 | 18.92 | 4.52 | 25.57 | 91.88 | -84.53 | 0.29 | -1.34 |
| 1455.48 | 18.29 | 4.28 | 25.57 | 91.88 | -193.12 | -0.63 | -0.24 |
| 1800.16 | 18.54 | 6.09 | 25.57 | 91.88 | 344.67 | 0.25 | 1.81 |
| 1773.53 | 18.47 | 5.84 | 25.57 | 91.88 | -26.63 | -0.07 | -0.25 |
| 1783.40 | 18.90 | 4.38 | 25.57 | 91.88 | 9.87 | 0.43 | -1.46 |
| 1808.92 | 18.87 | 6.54 | 25.57 | 91.88 | 25.52 | -0.03 | 2.16 |
| 1802.77 | 18.17 | 4.54 | 25.57 | 91.88 | -6.15 | -0.70 | -2.00 |
| 1872.70 | 17.71 | 5.04 | 25.57 | 91.88 | 69.93 | -0.46 | 0.50 |
| 1996.58 | 18.20 | 4.83 | 25.57 | 91.88 | 123.88 | 0.49 | -0.21 |
| . | 17.14 | 4.60 | 23.71 | 92.72 | . | . | . |
| . | 17.20 | 5.96 | 23.71 | 92.72 | . | . | . |
| . | 17.56 | 5.08 | 23.71 | 92.72 | . | . | . |
| . | 17.82 | 6.99 | 23.71 | 92.72 | . | . | . |
| 968.94 | 17.17 | 10.40 | 23.71 | 92.72 | . | . | . |
| 980.49 | 17.46 | 9.74 | 23.71 | 92.72 | 11.54 | 0.29 | -0.67 |
| 963.16 | 18.06 | 12.41 | 23.71 | 92.72 | -17.33 | 0.60 | 2.67 |
| 821.94 | 17.63 | 19.86 | 23.71 | 92.72 | -141.22 | -0.43 | 7.45 |
| 742.57 | 17.48 | 12.47 | 23.71 | 92.72 | -79.36 | -0.16 | -7.39 |
| 836.13 | 17.74 | 8.58 | 23.71 | 92.72 | 93.55 | 0.26 | -3.89 |
| 649.54 | 17.67 | 11.01 | 23.71 | 92.72 | -186.59 | -0.07 | 2.43 |
| 665.40 | 17.51 | 7.23 | 23.71 | 92.72 | 15.87 | -0.16 | -3.78 |
| 435.53 | 17.72 | 10.86 | 23.71 | 92.72 | -229.87 | 0.21 | 3.63 |
| 322.39 | 17.82 | 8.69 | 23.71 | 92.72 | -113.15 | 0.10 | -2.17 |
| 1045.32 | 17.63 | 8.45 | 23.71 | 92.72 | 722.93 | -0.19 | -0.24 |
| 1042.30 | 17.60 | 11.15 | 23.71 | 92.72 | -3.02 | -0.03 | 2.70 |
| 866.67 | 17.63 | 8.40 | 23.71 | 92.72 | -175.63 | 0.03 | -2.75 |
| 887.10 | 17.35 | 12.16 | 23.71 | 92.72 | 20.43 | -0.28 | 3.75 |
| 892.31 | 17.77 | 11.97 | 23.71 | 92.72 | 5.21 | 0.42 | -0.19 |
| 1066.67 | 17.46 | 12.60 | 23.71 | 92.72 | 174.36 | -0.31 | 0.64 |
| 1001.90 | 17.84 | 14.59 | 23.71 | 92.72 | -64.77 | 0.38 | 1.99 |
| 1016.92 | 17.91 | 15.36 | 23.71 | 92.72 | 15.02 | 0.07 | 0.76 |
| 1143.61 | 17.37 | 12.37 | 23.71 | 92.72 | 126.69 | -0.55 | -2.99 |
| 1148.09 | 17.52 | 11.45 | 23.71 | 92.72 | 4.48 | 0.15 | -0.91 |
| 1316.64 | 17.64 | 11.99 | 23.71 | 92.72 | 168.55 | 0.12 | 0.53 |
| 1366.67 | 17.44 | 9.17 | 23.71 | 92.72 | 50.03 | -0.19 | -2.82 |
| 1453.54 | 17.51 | 12.37 | 23.71 | 92.72 | 86.87 | 0.07 | 3.20 |
| 1596.56 | 17.89 | 9.02 | 23.71 | 92.72 | 143.02 | 0.38 | -3.35 |
| 1556.11 | 17.96 | 12.08 | 23.71 | 92.72 | -40.45 | 0.06 | 3.06 |
| 1583.87 | 17.85 | 10.03 | 23.71 | 92.72 | 27.76 | -0.11 | -2.05 |
| 1726.45 | 17.44 | 11.05 | 23.71 | 92.72 | 142.58 | -0.41 | 1.01 |
| 1660.45 | 17.77 | 14.72 | 23.71 | 92.72 | -66.00 | 0.33 | 3.67 |
| 1638.35 | 18.56 | 5.92 | 23.71 | 92.72 | -22.10 | 0.79 | -8.80 |
| 1780.78 | 18.09 | 5.97 | 23.71 | 92.72 | 142.43 | -0.47 | 0.05 |
| 2036.12 | 18.28 | 11.37 | 23.71 | 92.72 | 255.35 | 0.20 | 5.40 |
| 1916.54 | 18.24 | 12.66 | 23.71 | 92.72 | -119.58 | -0.05 | 1.29 |
| 1865.61 | 18.18 | 11.77 | 23.71 | 92.72 | -50.93 | -0.06 | -0.89 |
| 1853.72 | 18.15 | 7.01 | 23.71 | 92.72 | -11.88 | -0.03 | -4.76 |
| 1887.88 | 18.65 | 11.34 | 23.71 | 92.72 | 34.16 | 0.49 | 4.33 |
| 1754.05 | 18.12 | 4.13 | 23.71 | 92.72 | -133.82 | -0.53 | -7.21 |
| 821.90 | 18.27 | 5.56 | 23.71 | 92.72 | -932.16 | 0.15 | 1.43 |
| 284.65 | 18.29 | 5.43 | 23.71 | 92.72 | -537.25 | 0.02 | -0.13 |
| 897.87 | 18.70 | 8.60 | 23.71 | 92.72 | 613.22 | 0.41 | 3.18 |
| 1044.78 | 18.62 | 12.15 | 23.71 | 92.72 | 146.91 | -0.08 | 3.54 |
| 1246.23 | 17.86 | 4.07 | 23.71 | 92.72 | 201.45 | -0.76 | -8.08 |
| 1381.66 | 17.59 | 5.20 | 23.71 | 92.72 | 135.43 | -0.27 | 1.13 |
| 1756.42 | 17.97 | 8.57 | 23.71 | 92.72 | 374.76 | 0.39 | 3.37 |
| 311.92 | 17.42 | 5.38 | 25.67 | 94.12 | . | . | . |
| 251.85 | 17.51 | 4.96 | 25.67 | 94.12 | -60.06 | 0.09 | -0.42 |
| 252.63 | 17.85 | 4.69 | 25.67 | 94.12 | 0.77 | 0.34 | -0.28 |
| 827.24 | 17.89 | 4.13 | 25.67 | 94.12 | 574.62 | 0.04 | -0.56 |
| 892.74 | 17.63 | 5.23 | 25.67 | 94.12 | 65.50 | -0.26 | 1.11 |
| 531.20 | 17.60 | 4.51 | 25.67 | 94.12 | -361.54 | -0.03 | -0.73 |
| 580.54 | 18.25 | 4.68 | 25.67 | 94.12 | 49.34 | 0.65 | 0.17 |
| 578.41 | 17.93 | 6.19 | 25.67 | 94.12 | -2.13 | -0.32 | 1.51 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 612.35 | 17.70 | 5.16 | 25.67 | 94.12 | 33.94 | -0.23 | -1.03 |
| 878.97 | 18.00 | 5.37 | 25.67 | 94.12 | 266.62 | 0.30 | 0.21 |
| 821.58 | 17.91 | 4.72 | 25.67 | 94.12 | -57.39 | -0.09 | -0.65 |
| 869.98 | 17.83 | 3.25 | 25.67 | 94.12 | 48.40 | -0.08 | -1.47 |
| 861.42 | 17.91 | 2.95 | 25.67 | 94.12 | -8.56 | 0.08 | -0.30 |
| 509.15 | 17.99 | 4.59 | 25.67 | 94.12 | -352.27 | 0.08 | 1.64 |
| 762.21 | 17.79 | 5.02 | 25.67 | 94.12 | 253.06 | -0.19 | 0.43 |
| 803.86 | 17.83 | 4.09 | 25.67 | 94.12 | 41.64 | 0.04 | -0.93 |
| 712.77 | 17.66 | 4.34 | 25.67 | 94.12 | -91.09 | -0.18 | 0.26 |
| 822.92 | 17.56 | 4.93 | 25.67 | 94.12 | 110.15 | -0.10 | 0.59 |
| 793.58 | 17.66 | 3.85 | 25.67 | 94.12 | -29.34 | 0.10 | -1.08 |
| 749.05 | 17.54 | 3.25 | 25.67 | 94.12 | -44.53 | -0.11 | -0.60 |
| 635.99 | 17.93 | 4.87 | 25.67 | 94.12 | -113.06 | 0.39 | 1.62 |
| 653.92 | 18.22 | 4.83 | 25.67 | 94.12 | 17.93 | 0.29 | -0.04 |
| 1013.09 | 17.64 | 5.35 | 25.67 | 94.12 | 359.17 | -0.57 | 0.52 |
| 1125.63 | 17.76 | 6.13 | 25.67 | 94.12 | 112.54 | 0.12 | 0.78 |
| 1109.39 | 17.77 | 6.09 | 25.67 | 94.12 | -16.24 | 0.01 | -0.04 |
| 1116.23 | 17.73 | 5.91 | 25.67 | 94.12 | 6.84 | -0.04 | -0.17 |
| 1219.08 | 17.69 | 5.88 | 25.67 | 94.12 | 102.86 | -0.03 | -0.04 |
| 1237.11 | 18.22 | 4.33 | 25.67 | 94.12 | 18.03 | 0.52 | -1.55 |
| 1175.05 | 18.29 | 4.14 | 25.67 | 94.12 | -62.06 | 0.07 | -0.19 |
| 1241.16 | 18.25 | 4.45 | 25.67 | 94.12 | 66.11 | -0.04 | 0.31 |
| 1052.55 | 17.61 | 4.64 | 25.67 | 94.12 | -188.61 | -0.64 | 0.19 |
| 1179.78 | 18.63 | 5.01 | 25.67 | 94.12 | 127.23 | 1.02 | 0.37 |
| 1331.44 | 18.94 | 5.24 | 25.67 | 94.12 | 151.66 | 0.31 | 0.23 |
| 976.80 | 18.11 | 4.89 | 25.67 | 94.12 | -354.65 | -0.83 | -0.36 |
| 1406.10 | 18.42 | 4.48 | 25.67 | 94.12 | 429.30 | 0.31 | -0.41 |
| 1380.19 | 18.61 | 4.10 | 25.67 | 94.12 | -25.91 | 0.20 | -0.38 |
| 1564.52 | 18.40 | 5.18 | 25.67 | 94.12 | 184.32 | -0.21 | 1.08 |
| 1560.55 | 18.67 | 5.66 | 25.67 | 94.12 | -3.97 | 0.27 | 0.47 |
| 1577.40 | 18.83 | 4.63 | 25.67 | 94.12 | 16.86 | 0.16 | -1.03 |
| 1614.60 | 18.05 | 3.99 | 25.67 | 94.12 | 37.19 | -0.78 | -0.64 |
| 1481.66 | 18.37 | 5.63 | 25.67 | 94.12 | -132.94 | 0.32 | 1.64 |
| 1566.89 | 18.31 | 5.09 | 25.67 | 94.12 | 85.23 | -0.06 | -0.54 |
| 1811.20 | 18.82 | 4.28 | 25.67 | 94.12 | 244.31 | 0.51 | -0.81 |
| 1255.61 | 18.66 | 6.34 | 25.67 | 94.12 | -555.59 | -0.16 | 2.06 |
| 1901.65 | 18.05 | 4.98 | 25.67 | 94.12 | 646.04 | -0.61 | -1.36 |
| 1919.75 | 17.55 | 4.35 | 25.67 | 94.12 | 18.09 | -0.50 | -0.63 |
| 1967.05 | 18.11 | 4.62 | 25.67 | 94.12 | 47.31 | 0.56 | 0.27 |
| . | 15.90 | 7.21 | 27.33 | 88.62 | . | . | . |
| . | 15.83 | 7.96 | 27.33 | 88.62 | . | . | . |
| . | 16.02 | 6.74 | 27.33 | 88.62 | . | . | . |
| . | 15.59 | 7.36 | 27.33 | 88.62 | . | . | . |
| . | 15.04 | 8.08 | 27.33 | 88.62 | . | . | . |
| . | 15.71 | 7.09 | 27.33 | 88.62 | . | . | . |
| . | 16.24 | 9.21 | 27.33 | 88.62 | . | . | . |
| . | 16.01 | 7.72 | 27.33 | 88.62 | . | . | . |
| . | 16.00 | 8.51 | 27.33 | 88.62 | . | . | . |
| . | 16.09 | 8.80 | 27.33 | 88.62 | . | . | . |
| . | 16.05 | 10.43 | 27.33 | 88.62 | . | . | . |
| . | 15.93 | 8.08 | 27.33 | 88.62 | . | . | . |
| . | 16.14 | 8.94 | 27.33 | 88.62 | . | . | . |
| . | 16.63 | 8.54 | 27.33 | 88.62 | . | . | . |
| . | 15.92 | 8.08 | 27.33 | 88.62 | . | . | . |
| 966.62 | 15.96 | 7.15 | 27.33 | 88.62 | . | . | . |
| 942.05 | 15.65 | 7.64 | 27.33 | 88.62 | -24.57 | -0.31 | 0.49 |
| 1051.25 | 15.74 | 9.19 | 27.33 | 88.62 | 109.20 | 0.09 | 1.55 |
| 1165.38 | 15.80 | 9.13 | 27.33 | 88.62 | 114.14 | 0.06 | -0.06 |
| 1212.31 | 15.37 | 7.33 | 27.33 | 88.62 | 46.93 | -0.42 | -1.81 |
| 1211.20 | 15.95 | 8.02 | 27.33 | 88.62 | -1.11 | 0.58 | 0.70 |
| 1137.39 | 16.96 | 7.40 | 27.33 | 88.62 | -73.81 | 1.01 | -0.62 |
| 1142.86 | 16.11 | 6.86 | 27.33 | 88.62 | 5.47 | -0.86 | -0.54 |
| 1271.27 | 16.20 | 9.29 | 27.33 | 88.62 | 128.41 | 0.09 | 2.43 |
| 1287.19 | 15.90 | 7.91 | 27.33 | 88.62 | 15.93 | -0.30 | -1.37 |
| 1275.55 | 15.71 | 5.35 | 27.33 | 88.62 | -11.65 | -0.19 | -2.57 |
| 1274.27 | 16.35 | 6.24 | 27.33 | 88.62 | -1.27 | 0.65 | 0.89 |
| 1302.12 | 16.62 | 5.63 | 27.33 | 88.62 | 27.85 | 0.27 | -0.61 |
| 1340.74 | 16.38 | 9.29 | 27.33 | 88.62 | 38.62 | -0.24 | 3.66 |
| 1339.46 | 16.51 | 7.24 | 27.33 | 88.62 | -1.28 | 0.13 | -2.05 |
| 1374.03 | 15.98 | 7.40 | 27.33 | 88.62 | 34.57 | -0.53 | 0.17 |
| 1335.49 | 17.26 | 9.66 | 27.33 | 88.62 | -38.53 | 1.28 | 2.25 |
| 1193.42 | 17.10 | 6.02 | 27.33 | 88.62 | -142.07 | -0.16 | -3.63 |
| 1316.73 | 16.66 | 5.29 | 27.33 | 88.62 | 123.31 | -0.43 | -0.74 |
| 1356.11 | 17.10 | 7.26 | 27.33 | 88.62 | 39.38 | 0.44 | 1.97 |
| 1288.51 | 16.77 | 7.32 | 27.33 | 88.62 | -67.60 | -0.33 | 0.06 |

```
1334.25 16.76   8.22    27.33   88.62   45.74   -0.01   0.90
1395.25 16.74   5.85    27.33   88.62   61.00   -0.01   -2.37
1405.95 16.86   6.67    27.33   88.62   10.70   0.12    0.83
1353.58 17.11   6.40    27.33   88.62   -52.38  0.25    -0.28
1353.58 17.04   6.72    27.33   88.62   0.00    -0.07   0.32
1377.78 16.90   6.97    27.33   88.62   24.20   -0.13   0.25
1350.50 17.05   6.99    27.33   88.62   -27.28  0.15    0.01
1496.17 17.34   9.20    27.33   88.62   145.67  0.28    2.21
1447.70 16.60   5.93    27.33   88.62   -48.48  -0.73   -3.26
1495.00 16.57   6.80    27.33   88.62   47.30   -0.03   0.87
1538.24 16.32   8.03    27.33   88.62   43.24   -0.26   1.23
805.69  17.17   5.74    23.84   91.28   .       .       .
820.11  17.23   6.65    23.84   91.28   14.42   0.06    0.91
779.29  17.48   6.72    23.84   91.28   -40.82  0.25    0.07
874.63  17.82   6.05    23.84   91.28   95.34   0.34    -0.67
948.42  17.22   5.51    23.84   91.28   73.79   -0.60   -0.54
972.92  17.42   4.64    23.84   91.28   24.51   0.20    -0.87
651.00  17.94   7.96    23.84   91.28   -321.93 0.52    3.33
1203.77 17.23   8.07    23.84   91.28   552.77  -0.71   0.11
1085.59 17.32   7.21    23.84   91.28   -118.18 0.09    -0.86
1215.71 17.25   8.47    23.84   91.28   130.11  -0.06   1.26
1118.93 17.65   7.69    23.84   91.28   -96.77  0.39    -0.78
1201.60 17.64   6.73    23.84   91.28   82.67   -0.01   -0.96
1233.61 17.94   5.26    23.84   91.28   32.00   0.30    -1.47
1185.41 18.01   5.86    23.84   91.28   -48.19  0.07    0.60
1351.46 17.76   5.99    23.84   91.28   166.05  -0.24   0.13
1181.49 17.64   7.97    23.84   91.28   -169.97 -0.12   1.98
1412.60 17.62   8.46    23.84   91.28   231.11  -0.03   0.48
1309.75 17.36   8.19    23.84   91.28   -102.85 -0.25   -0.27
1382.20 17.61   6.42    23.84   91.28   72.45   0.25    -1.77
1304.95 17.57   6.48    23.84   91.28   -77.25  -0.04   0.06
1459.02 17.92   6.49    23.84   91.28   154.06  0.35    0.01
1548.27 17.99   8.74    23.84   91.28   89.26   0.07    2.24
1595.46 17.45   6.89    23.84   91.28   47.18   -0.54   -1.84
1755.27 17.52   7.80    23.84   91.28   159.81  0.07    0.90
1747.22 17.64   9.71    23.84   91.28   -8.04   0.12    1.91
1758.98 17.57   5.37    23.84   91.28   11.76   -0.07   -4.34
1715.94 17.58   9.98    23.84   91.28   -43.04  0.01    4.62
1818.47 17.96   5.69    23.84   91.28   102.53  0.39    -4.29
1555.72 17.99   6.96    23.84   91.28   -262.76 0.03    1.27
1926.05 17.94   5.70    23.84   91.28   370.33  -0.05   -1.26
1016.87 17.46   6.18    23.84   91.28   -909.18 -0.48   0.48
2016.95 18.39   5.89    23.84   91.28   1000.08 0.93    -0.29
1874.62 18.59   5.92    23.84   91.28   -142.32 0.20    0.03
2119.27 18.21   6.89    23.84   91.28   244.65  -0.38   0.98
2060.26 18.31   6.70    23.84   91.28   -59.01  0.10    -0.19
2311.41 18.29   6.45    23.84   91.28   251.15  -0.02   -0.25
2288.81 18.20   6.38    23.84   91.28   -22.60  -0.09   -0.08
2120.64 17.64   7.73    23.84   91.28   -168.17 -0.56   1.35
2178.54 18.67   5.86    23.84   91.28   57.90   1.03    -1.87
2193.85 18.27   5.07    23.84   91.28   15.31   -0.40   -0.79
2399.09 18.32   7.21    23.84   91.28   205.24  0.05    2.14
2562.93 18.33   5.39    23.84   91.28   163.84  0.01    -1.82
2525.58 18.73   5.56    23.84   91.28   -37.35  0.40    0.17
2544.64 18.66   7.24    23.84   91.28   19.07   -0.07   1.68
2587.42 17.95   5.54    23.84   91.28   42.78   -0.71   -1.70
2619.89 17.68   5.84    23.84   91.28   32.46   -0.27   0.30
2711.22 18.04   4.99    23.84   91.28   91.34   0.36    -0.85
;
/*****************************************************************************************************/
/**********************************Section 1: Descriptive Statistics of the variables***************************/
/*****************************************************************************************************/
proc means data=endogeneity_test;
var Foodgrain_Yld Min_Tem     Rain Latitude Longitude Foodgrain_Yld_FD Min_Tem_FD  Rain_FD;
run;
/*****************************************************************************************************/
/**********************************Section 2: Housman Test for endogeneity**************************************/
/*****************************************************************************************************/
/*Notice that the first Proc Reg statements save the residuals in a temporary SAS data set called endogeneity_test1.*/
/*noprint is optional as we just dont want to print the result of regression because we need only residual (r) to be saved and used in the next regression*/
/*****************************************************************************************************/
proc reg data=endogeneity_test noprint;
model Min_Tem  =Latitude longitude Foodgrain_Yld_FD  Min_Tem_FD Rain_FD ;
```

```
output out=endogeneity_test1 r= v ;
run;
proc reg data=endogeneity_test1;
model Foodgrain_Yld=  Min_Tem       Rain v ;
run;
/*the same is done for other independent variable Rain*/
proc reg data=endogeneity_test noprint;
model Rain     =Latitude longitude Foodgrain_Yld_FD  Min_Tem_FD Rain_FD  ;
output out=endogeneity_test2 r= v1 ;
run;
proc reg data=endogeneity_test2;
model Foodgrain_Yld=  Min_Tem       Rain v1 ;
run;
/**************************************************************************************************************/
/***************************Note: If other variables were present*****************************************/
/**************************************************************************************************************/
/*Note: The model we have been discussing is Foodgrain_Yld= Min_Tem  Rain. If our model were having an already
known exogenous variable i.e if the model were of the form Foodgrain_Yld= Min_Tem   Rain Farmer_age where
Farmer_age is the assumed exogenous variable. Then the code would be as follows*/

/*proc reg data=endogeneity_test noprint;
model Min_Tem= Farmer_age Latitude longitude Foodgrain_Yld_FD  Min_Tem_FD Rain_FD ;
output out=endogeneity_test1 r= v ;
run;
proc reg data=endogeneity_test1;
model Foodgrain_Yld=  Min_Tem       Rain Farmer_age v ;
run;*/
/**************************************************************************************************************/
/***************************Note: Using predictive value in place of residual*****************************/
/**************************************************************************************************************/
/* Predicted values can also be used in place of residual in the second stage regression. The result of
importance will be the same even if other accompanied outputs are different. The corresponding SAS code is as
follows.*/
/**************************************************************************************************************/
proc reg data=endogeneity_test noprint;
model Rain=Latitude longitude Foodgrain_Yld_FD  Min_Tem_FD Rain_FD ;
output out=endogeneity_test2 p=f p=f1;
run;
proc reg data=endogeneity_test2;
model Foodgrain_Yld=Min_Tem Rain f f1;
run;
/**************************************************************************************************************/
/***************************Section 3: Test for overidentification*********************************************/
/**************************************************************************************************************/
/*To determine if the regression model has more instruments than is necessary. The question we address here is,
“Are the extra instrument variables truly exogenous?” That is, are the extra instruments uncorrelated with the
error term? */
/*The first step is to estimate the true model by using 2SLS and to store the residuals. The following SAS
statements can be used. Note that the output has been suppressed because we are interested only in storing the
residuals from this analysis.*/
/**************************************************************************************************************/
proc syslin 2SLS noprint data=endogeneity_test out=step1_resid;
endogenous Min_Tem Rain;
instruments Latitude longitude Foodgrain_Yld_FD  Min_Tem_FD Rain_FD;
model Foodgrain_Yld=  Min_Tem       Rain;
output residual=out1_resid;
run;
/*Second step of overidentification. The next step is to regress the residuals from the 2SLS analysis on all
exogenous variables in the model. The following SAS statements can be used. */
proc reg data=step1_resid;
model out1_resid=Latitude longitude Foodgrain_Yld_FD  Min_Tem_FD Rain_FD ;
run;
/**************************************************************************************************************/
/******************Section 4: Further test for overidentification for appropriate instruments*****************/
/**************************************************************************************************************/
/*From the output of above code it is confirmed that the regression model has more instruments than is
necessary. Therefore the extra instrument variables may not be truly exogenous or the extra instruments are
correlated with the error term. In such case we revisit the assignment of instruments and after trial of
different combinations we reduce the number of instruments as follows.However selection or identification of
instruments is also an art, researchers should have a detail understanding of the variables with their cause and
effect relationship involved in the model. */
/**************************************************************************************************************/
proc syslin 2SLS noprint data=endogeneity_test out=step1_resid1;
endogenous Min_Tem Rain;
instruments  longitude   Min_Tem_FD Rain_FD;
```

```
model Foodgrain_Yld=   Min_Tem        Rain;
output residual=out1_resid1;
run;
/*Second step of overidentification. The next step is to regress the residuals from the 2SLS analysis on all
exogenous variables in the model. The following SAS statements can be used. */
proc reg data=step1_resid1;
model out1_resid1= longitude   Min_Tem_FD Rain_FD ;
run;
/*****************************************************************************************************/
/************************************Section 5: Test for weak instrument****************************************/
/*****************************************************************************************************/
/*We now turn our attention to the problem of weak instruments—that is, the case when the selected instrumental
variables used in estimation have a poor correlation with the endogenous variable. We will discuss a general
method for determining if weak instruments have been used in the model. */
/*****************************************************************************************************/
proc reg data=endogeneity_test;
model Min_Tem Rain= longitude   Min_Tem_FD Rain_FD;
run;
/*the above code is same as the codes below*/
proc reg data=endogeneity_test;
model Min_Tem = longitude   Min_Tem_FD Rain_FD;
model  Rain= longitude   Min_Tem_FD Rain_FD;
run;
/*****************************************************************************************************/
/********************Section 6: Test for endogeneity with idendified instruments*******************/
/*****************************************************************************************************/
/*Thus we have identified the appropriate instruments viz Longitude, Min_Tem_FD and Rain_FD out of the pool of
instruments by testing overidentifying restrictions and they are not found weak. Now we can retest endogeneity
of the regressors Min_Tem and Rain using the identified instrumental variables (Longitude, Min_Tem_FD and
Rain_FD). The SAS codes are as follows.*/
/*****************************************************************************************************/
proc reg data=endogeneity_test noprint;
model Min_Tem = longitude   Min_Tem_FD Rain_FD ;
output out=endogeneity_test1 r= v ;
run;
proc reg data=endogeneity_test1;
model Foodgrain_Yld=   Min_Tem        Rain v ;
run;
/*the same is done for other independent variable Rain*/
proc reg data=endogeneity_test noprint;
model Rain     = longitude   Min_Tem_FD Rain_FD   ;
output out=endogeneity_test2 r= v1 ;
run;
proc reg data=endogeneity_test2;
model Foodgrain_Yld=   Min_Tem        Rain v1 ;
run;
/*****************************************************************************************************/
/*******************************************Section 7: 2SLS estimate******************************************/
/*****************************************************************************************************/
/*2SLS estimate can be achieved using the following SAS codes. In the first proc reg the predicted value of
endogenous regressors Min_Tem and Rain designated by f and f1 respectively are saved in the temporary dataset
endogeneity_test1. In the second proc reg endogenous regressors Min_Tem and Rain are replaced by f and f1 as
regressors in the original model.*/
/*****************************************************************************************************/
proc reg data=endogeneity_test noprint;
model Min_Tem Rain     = longitude   Min_Tem_FD Rain_FD   ;
output out=endogeneity_test1 p= f p= f1 ;
run;
proc reg data=endogeneity_test1;
model Foodgrain_Yld=    f f1 ;
run;
/*Above 2SLS estimates can be compared with the OLS estimate result given below.*/
proc reg data=endogeneity_test;
model Foodgrain_Yld=Min_Tem Rain ;
run;
/*****************************************************************************************************/
/*Section 8:Hausman’s test: To determine if there are significant differences between OLS and 2SLS estimators*/
/*****************************************************************************************************/
/*In the previous sections of Hausman’s test for endogeneity and estimation of 2SLS, various prog reg statements
are used. But in SAS a dedicated package of codes of Hausman’s test can be used to accomplish the same as
illustrated below. Here, Hausman’s specification test is used to determine if there are significant differences
between the OLS and the IV estimators. Under the null hypothesis of no endogeneity, both the OLS and the IV
estimators are consistent. Under the alternative hypothesis of endogeneity, only the IV estimator is consistent.
Accordingly endogeneity can also be concluded. If significant difference is found between OLS and IV estimates,
then the suggested estimate is IV and in turn it suggest evidence of at least one endogenous explanatory
```

```
variable in the model. On the other hand, if OLS is suggested by the test, then all the regressors are
exogenous.*/

/*This Hausman's test can be performed in SAS by using the proc model procedure as shown in the following codes.
Notice that we specify the endogenous variable in the "endo" statement, the instruments in the "instruments"
statement, and then write down the linear model to be estimated. beta1, beta2 and beta3 are the parameters to be
estimated. This is followed by the "fit" statement using the dependent variable with the option that the Hausman
test be used to compare the OLS and the instrumental variable estimator.*/
/*************************************************************************************************************/
proc model data=endogeneity_test;
endo Min_Tem Rain;
instruments  longitude   Min_Tem_FD Rain_FD ;
Foodgrain_Yld=beta1+beta2*Min_Tem+beta3*Rain;
fit Foodgrain_Yld/ols 2sls  hausman;
run;
/*************************************************************************************************************/
/*Note that the OLS and 2SLS estimates given Section 7 and the OLS and 2SLS estimates given by Section 8 are
supposed to give correspondingly the same results.However the results are deviated a little from each other
though the underlying interpretation is the same. This deviation may be due to the method-specific computational
variations of the different procedures in estimating the parameters.  */
```