Part I: Autocorrelation test, GLS and MLE:

```
PROC IMPORT OUT= WORK.bm
     DATAFILE= "C:\Users\bmishra\Dropbox\Ph.D. Courseworks\Semest
er II, Spring 2019\Econometric Methods\Homeworks\Homework 5\HW5-DATA1.xl
     DBMS=EXCEL REPLACE;
  RANGE="SUGAR$";
  GETNAMES=YES;
  MIXED=NO;
  SCANTEXT=YES;
  USEDATE=YES;
  SCANTIME=YES;
RUN;
/* Data Management */
data bm; set bm;
a = a;
lna = log(a);
w = w;
lnw = log(w);
b = b;
lnb = log(b);
p = w/b;
lnp = log(p);
run;
```

(a).

```
/* Part I */
proc reg data = bm; /* OLS */
model lna = lnw lnb / dwprob; /* Q (a, b, d) */
p = lnayhatt;
r = lnaresid;
test lnw + lnb = 0; /* Q (d) */
run;
proc print;
run;
```

The SAS System The REG Procedure

Model: MODEL1
Dependent Variable: Ina

| Number of Observations Read | 35 |
|--|----|
| Number of Observations Used | 34 |
| Number of Observations with Missing Values | 1 |

| Analysis of Variance | | | | | | | |
|----------------------|----|-------------------|----------------|---------|--------|--|--|
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F | | |
| Model | 2 | 7.35324 | 3.67662 | 37.54 | <.0001 | | |
| Error | 31 | 3.03605 | 0.09794 | | | | |
| Corrected Total | 33 | 10.38929 | | | | | |

| Root MSE | 0.31295 | R-Square | 0.7078 |
|----------------|---------|----------|--------|
| Dependent Mean | 4.70727 | Adj R-Sq | 0.6889 |
| Coeff Var | 6.64820 | | |

| Parameter Estimates | | | | | | | | |
|---------------------|----|-----------------------|---------|---------|---------|--|--|--|
| Variable | DF | Parameter Estimate | | t Value | Pr > t | | | |
| Intercept | 1 | 5.20128 | 2.34569 | 2.22 | 0.0341 | | | |
| Inw | 1 | 0.96630 | 0.11267 | 8.58 | <.0001 | | | |
| Inb | 1 | -0.83605 | 0.36357 | -2.30 | 0.0284 | | | |

Interpretations:

 β_2 : with the one unit percentage increase in the price of w measured in \$/ton and keeping other variable constant the percentage of acreage in wheat in thousands of hectares increase by 0.96630 unit an average.

 β_3 : with the one unit percentage increase in the price of jute (b) measured in \$/ton and keeping other variable constant, the percentage of acreage in wheat decrease by 0.836025 unit on an average.

| Here, | | |
|------------|---|----------------|
| Giv | en, P = Wt/P | |
| , 0 | (A) = R + R ln(W) + B, ln(Re) + e. | |
| Assume, | Len, $\beta_t = Wt/\beta_t$ $\ln(A_t) = \beta_1 + \beta_2 \ln(W_t) + \beta_3 \ln(B_t) + \epsilon_t$ | |
| 0. | - 1 | |
| Above equa | 5- fz, they replacing this accomplety in | |
| lu | (Mz) = B, + B2 ln(W) - B2 ln (B+) + ex | |
| | | |
| | = R, + B2 (ln(w+) - ln(B+)) +e+ | |
| | = B. + Bz (In (wt/B+)+ ex (log | propul tes) |
| | = Bi+ Bz luffe) + Ce 00 /2= | WL, |
| | | |
| | Prove | |

(b).

The SAS System

The REG Procedure Model: MODEL1 Dependent Variable: Ina

| Number of Observations Read | 35 |
|--|----|
| Number of Observations Used | 34 |
| Number of Observations with Missing Values | 1 |

| Analysis of Variance | | | | | | | |
|----------------------|----|-------------------|----------------|---------|--------|--|--|
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F | | |
| Model | 2 | 7.35324 | 3.67662 | 37.54 | <.0001 | | |
| Error | 31 | 3.03605 | 0.09794 | | | | |
| Corrected Total | 33 | 10.38929 | | | | | |

| Root MSE | 0.31295 | R-Square | 0.7078 |
|----------------|---------|----------|--------|
| Dependent Mean | 4.70727 | Adj R-Sq | 0.6889 |
| Coeff Var | 6.64820 | | |

| Parameter Estimates | | | | | | | | |
|---------------------|----|-----------------------|---------|---------|---------|--|--|--|
| Variable | DF | Parameter Estimate | | t Value | Pr > t | | | |
| Intercept | 1 | 5.20128 | 2.34569 | 2.22 | 0.0341 | | | |
| Inw | 1 | 0.96630 | 0.11267 | 8.58 | <.0001 | | | |
| Inb | 1 | -0.83605 | 0.36357 | -2.30 | 0.0284 | | | |

d = 1.310. Number of observations (n) = 34 and # of variables (k) (no intercept) = 2. At 5% Significance point dL = 1.333 and dU = 1.580

Since d < dL, we reject null hypothesis at 5% significance level. There might be evidence of autocorrelation.

The SAS System

The REG Procedure Model: MODEL1 Dependent Variable: Ina

| Durbin-Watson D | 1.310 |
|---------------------------|--------|
| Pr < DW | 0.0131 |
| Pr > DW | 0.9869 |
| Number of Observations | 34 |
| 1st Order Autocorrelation | 0.327 |

Note: Pr<DW (DW and d in lecture note are same) is the p-value for testing positive autocorrelation, and Pr>DW is the p-value for testing negative autocorrelation.

Ist order autocorrelation (ρ) = 0.327

Null Hypothesis: The residuals from an OLS are not autocorrelated (no autocorrelation).

 H_0 : $\rho = 0$

Alternative Hypothesis: The residuals follows an AR1 process (auto-correlated).

 H_a : $\rho > 0$ (because d < 2)

Conclusion: we reject null hypothesis (P = 0.0131) at 5% significance level.. There might be autocorrelation.

(c).

```
proc autoreg data = bm;
model lna = lnw lnb / nlag = 1; /* Q (c) */
test lnw + lnb = 0; /* Q (d) */
output out = autoreg1 r = auregehatt;
run;
```



Figure: OLS (Left) & GLS (Right)

- The SSE, MSE SBC, MAE, MAPE, DFE, Root MSE, AIC, AICC, HQC are larger in OLS compared to GLS and higher Total R-squared value in GLS.
- Looking at parameters estimates table, parameter estimates are smaller, standard errors are larger from and t-values are smaller in OLS compared to GLS except for lnb. Both models have parameters with similar significance trend at 95% CI (all significant) but OLS has some non-significant variables at 99% CI.

(d).

Estimated result from (b):

| The REG Procedure Model: MODEL1 Test 1 Results for Dependent Variable Ina | | | | | | |
|--|----|---------|---------|--------|--|--|
| Source | | Mean | F Value | | | |
| Numerator | 1 | 0.01482 | 0.15 | 0.6999 | | |
| Denominator | 31 | 0.09794 | | | | |

Null hypothesis: H_0 : $\beta_2 = \beta_3$. Alternative Hypothesis H_a : $\beta_2 \neq \beta_3$.

Conclusion: We fail to reject null hypothesis at 95% CI. (P = 0.6999)

Estimated result from (c):

| Test 1 | | | | | | |
|-------------|----|----------------|---------|--------|--|--|
| Source | DF | Mean Square | F Value | Pr > F | | |
| Numerator | 1 | 0.006580 | 0.07 | 0.7881 | | |
| Denominator | 30 | 0.089491 | | | | |

Null hypothesis: H_0 : $\beta_2 = \beta_3$. Alternative Hypothesis H_a : $\beta_2 \neq \beta_3$.

Conclusion: We fail to reject null hypothesis at 95% CI. (P = 0.7881)

The hypothesis test results do not differ between (b) and (c).

(e).

Predicting for first year using GLS result:

First Year: $Ln(A) = 5.6184 + (LN(500)*1.0127)-(LN(500)*0.9328) + \rho*Err.$

Second Year: Ln(A) =5.6184 + (LN(500)*1.0127)-(LN(500)*0.9328)+ ρ *Err.

 ρ = -0.327 & Err. = 0.172532

First Year:

Ln(A) = 5.6184 + (LN(500)*1.0127) - (LN(500)*0.9328) + (-0.327*0.172532) = 6.0675

Second Year:

 $Ln(A) = 5.6184 + (LN(500)*1.0127) - (LN(500)*0.9328) + (-0.327)^2*0.172532 = 6.3223$

Part II: Two-stage Least Square Estimation:

```
PROC IMPORT OUT= WORK.HW5D2
            DATAFILE= "C:\Users\bmishra\Dropbox\Ph.D. Courseworks\Semest
er II, Spring 2019\Econometric Methods\Homeworks\Homework 5\HW5-DATA2.tx
            DBMS=TAB REPLACE;
     GETNAMES=YES;
     DATAROW=2;
RUN:
/* Part II */
data hw5d2; set hw5d2;
lny = log(y);
lnk = log(k);
lnw = log(w);
lnl = log(1);
lnp = log(p);
lnr = log(r);
run;
```

Given:

In our case, we have M=3 simultaneous equations to jointly determine the values of M=3 endogenous variables. At least M-1=2 variables must be absent from an equation for estimation of its parameters to be possible. (For an equation to be identified in a system the total number of variables excluded from the equation but included in other equations must be at least equal to one less than the number of equations in the system.)

For equation 1, at least one variable in the system must be absent from equation 1. We have total of three variables (output price, capital input price and labor input price) absent from the system in this equation. This is over identified.

For equation 2, at least two variables should be absent from the equation 2. We have two missing variables (labor input and labor input price) from the system in this equation. This equation is identified.

For Equation 3, at least two variables should be absent from the equation 2 to be identified. We have two missing variables (capital input and capital input price) from the system in this equation. This equation is also identified.

(b).

```
/* b */
proc reg data = hw5d2;
model lny = lnk lnl;
run;
```

Regression Prediction run

The REG Procedure Model: MODEL1 Dependent Variable: Iny

Number of Observations Read 23 Number of Observations Used 23

| Analysis of Variance | | | | | | | | |
|----------------------|----|-------------------|----------------|---------|--------|--|--|--|
| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F | | | |
| Model | 2 | 112.89975 | 56.44987 | 140.08 | <.0001 | | | |
| Error | 20 | 8.05949 | 0.40297 | | | | | |
| Corrected Total | 22 | 120.95924 | | | | | | |

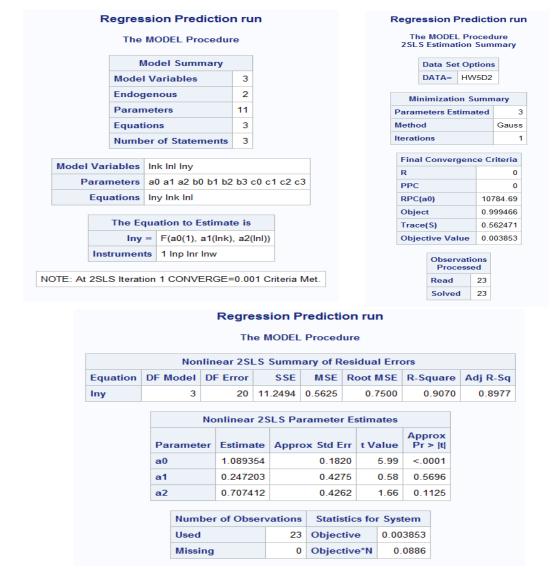
| Root MSE | 0.63480 | R-Square | 0.9334 |
|----------------|----------|----------|--------|
| Dependent Mean | 1.71760 | Adj R-Sq | 0.9267 |
| Coeff Var | 36.95860 | | |

| Parameter Estimates | | | | | | | | |
|---------------------|----|-----------------------|---------|---------|---------|--|--|--|
| Variable | DF | Parameter Estimate | | t Value | Pr > t | | | |
| Intercept | 1 | 1.13757 | 0.14763 | 7.71 | <.0001 | | | |
| Ink | 1 | 0.00695 | 0.29386 | 0.02 | 0.9814 | | | |
| Inl | 1 | 0.81368 | 0.29550 | 2.75 | 0.0123 | | | |

Theoretically, output (Y) is dependent upon capital input and labor input. Further capital input is dependent upon output price, output, capital input price and labor input is dependent upon output price, output, and labor input price. However, with the OLS regression above we estimated equation only based on capital input and labor input but their dependence upon other six variables in the system are not accounted. This is not accounted while estimated output equation and thus the prediction is inaccurate/not consistent and also biased.

(c).

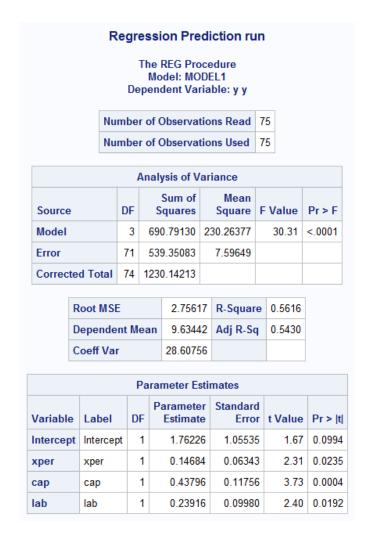
```
/* Model Procedure: Creates One Production Model */
proc model data = hw5d2;
instruments lnp lnr lnw;
endogenous lnk lnl;
production: lny = a0 + a1*lnk + a2*lnl;
demandk: lnk = b0 + b1*lnp + b2*lny + b3*lnr;
demandl: lnl = c0 + c1*lnp + c2*lny + c3*lnw;
fit lny / 2sls;
run;
```



Two stage least square estimates are available in nonlinear 2SLS parameter estimates table. Parameters are still biased but they are consistent (not consistent before). Parameters are in same directions in both models but parameter of lnl (labor input) become non-significant (it was significant before). lnk is also not significant.

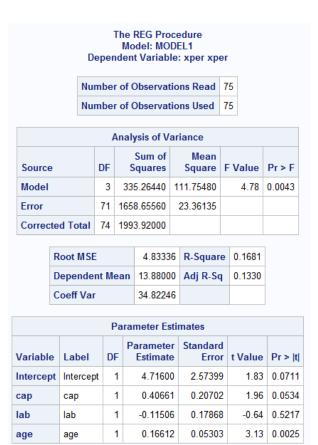
Part III: Instrumental Variable/ Two-stage Least Square Estimation:

(a).



(b).

```
/* b */
proc reg data = hw5d3;
model xper = cap lab age;
output out = aut r = resid p = pred;
run;
proc reg data = aut;
model y = pred cap lab;
run;
```





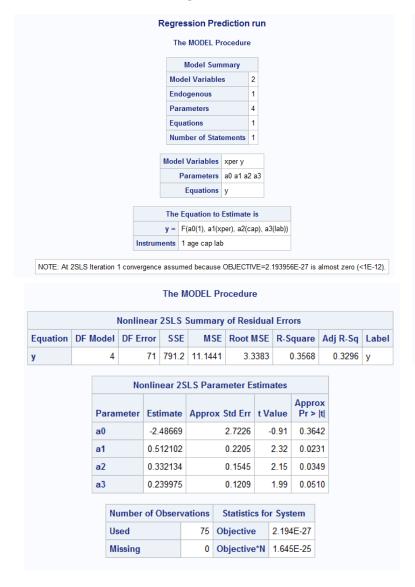
- After checking instrument, lab is not significant (P = 0.5217). This mode is identified.
- After 2SLS, all variables are significant except intercept, the sign of intercept and lab also changed. The estimates also varies.

(c).

```
/* c */
proc model data = hw5d3;
instruments age cap lab;
endogenous xper;
BeerProdn: y = a0 + a1*xper + a2*cap + a3*lab;
fit y / 2sls hausman;
run;

proc model data = hw5d3;
instruments age cap lab;
endogenous xper;
BeerProdn: y = a0 + a1*xper + a2*cap + a3*lab;
fit y / 2sls gmm;
run;
```

I ran Hausman test and did not get the test result.



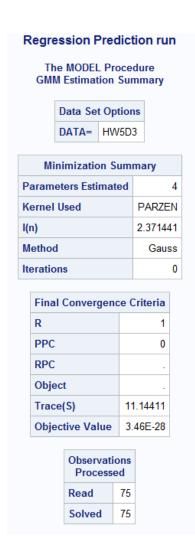


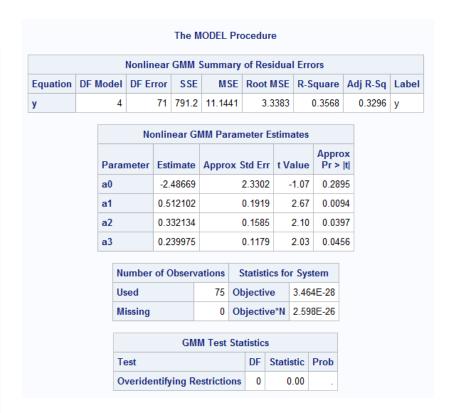


4

Gauss

I ran GMM test.





GMM Test Statistics shows DF = 0. Hausman test might not work.

```
SAS Code:
Homework 5 Part I:
PROC IMPORT OUT= WORK.bm
            DATAFILE= "C:\Users\bmishra\Dropbox\Ph.D. Courseworks\Semest
er II, Spring 2019\Econometric Methods\Homeworks\Homework 5\HW5-DATA1.xl
            DBMS=EXCEL REPLACE;
    RANGE="SUGAR$";
     GETNAMES=YES;
    MIXED=NO;
     SCANTEXT=YES;
    USEDATE=YES;
     SCANTIME=YES;
RUN;
proc print;
run;
/* Data Management */
data bm; set bm;
lna = log(a);
lnw = log(w);
lnb = log(b);
run;
proc print;
run;
/* Part I */
proc reg data = bm; /* OLS */
model lna = lnb lnw / dwprob; /* Q (a, b, d) */
test lnb + lnw = 0; /* Q (d) */
run;
proc autoreg data = bm;
model lna = lnw lnb / nlag = 1; /* Q (c) */
test lnw + lnb = 0; /* Q (d) */
output out = autoreg1 r = auregehatt;
run;
proc autoreg data = bm;
model lna = lnw lnb / nlag = 1; /* Q (c) */
test lnw + lnb = 0; /* Q (d) */
run;
Homework 5 Part II:
PROC IMPORT OUT= WORK.HW5D2
            DATAFILE= "C:\Users\bmishra\Dropbox\Ph.D. Courseworks\Semest
er II, Spring 2019\Econometric Methods\Homeworks\Homework 5\HW5-DATA2.tx
            DBMS=TAB REPLACE;
     GETNAMES=YES;
    DATAROW=2;
RUN:
proc print;
run;
/* Part II */
```

```
data hw5d2; set hw5d2;
lny = log(y);
lnk = log(k);
lnw = log(w);
lnl = log(1);
lnp = log(p);
lnr = log(r);
run;
/* b */
proc reg data = hw5d2;
model lny = lnk lnl;
run;
/* c: Model Procedure */
proc model data = hw5d2;
instruments lnp lnr lnw;
endogenous lnk lnl;
production: lny = a0 + a1*lnk + a2*lnl;
demandk: lnk = b0 + b1*lnp + b2*lny + b3*lnr;
demandl: lnl = c0 + c1*lnp + c2*lny + c3*lnw;
fit lny / 2sls;
run;
/* c: Syslin Procedure */
proc syslin data = hw5d2;
instruments lnp lnr lnw;
endogenous lnk lnl lny;
production: model lny = lnk lnl / overid;
demandk: model lnk = lnp lny lnr / overid;
demandl: model lnl = lnp lny lnw / overid;
run;
Homework 5 Part III:
PROC IMPORT OUT= WORK.HW5D3
            DATAFILE= "C:\Users\bmishra\Dropbox\Ph.D. Courseworks\Semest
er II, Spring 2019\Econometric Methods\Homeworks\Homework 5\HW5-DATA3.xl
            DBMS=EXCEL REPLACE;
    RANGE="data";
    GETNAMES=YES;
    MIXED=NO;
    SCANTEXT=YES;
    USEDATE=YES;
    SCANTIME=YES;
RUN;
proc print;
run;
/* Part III */
/* a */
proc reg data = hw5d3;
model y = xper cap lab;
run;
/* b */
proc reg data = hw5d3;
model xper = cap lab age;
output out = aut r = resid p = pred;
run;
```

```
proc reg data = aut;
model y = pred cap lab;
run;

/* c */
proc model data = hw5d3;
instruments age cap lab;
endogenous xper;
BeerProdn: y = a0 + a1*xper + a2*cap + a3*lab;
fit y / 2sls hausman;
run;

proc model data = hw5d3;
instruments age cap lab;
endogenous xper;
BeerProdn: y = a0 + a1*xper + a2*cap + a3*lab;
fit y / 2sls gmm;
run;
```

AGEC5213: ECONOMETRIC METHODS **Spring 2019**

PROBLEM SET NO. 5- due on April 29, 2018

Part I. Autocorrelation test, GLS and MLE (10 points)

Consider a wheat acreage model $\ln(A_1) = \beta_1 + \beta_2 \ln(P_1) + e_1$ where A_1 was area of wheat (in thousands of hectares) sown in Oklahoma in year t and P_t was the price of wheat. This specification was, in fact, a simplification of a slightly more general specification; the price of wheat was measured relative to the price of Barley. Specifically, let

 W_t = price of in year t (dollar/ton), and

 B_t = price of jute in year t (dollar/ton).

The price that was utilized was $P_t = W_t/B_t$. Since barley is the crop which competes with wheat for the use of the land, it is the price of wheat relative to the price of barley that affects farmers' decisions about how much wheat to plant. Data on A_t , W_t and B_t appear in the file, HW5-DATA1.xls



Consider the model

 $ln(A_t) = \beta_1 + \beta_2 ln(W_t) + \beta_3 ln(B_t) + e_t$

Give interpretations for β_2 and β_3 . Show that this model is equivalent to the original model if $\beta_2 = -\beta_3$.



Estimate the model in part (a) using least squares and test for the presence of autocorrelated errors using the DW test at the 5% significance level.



Re-estimate the model assuming the existence of AR(1) errors (GLS). Are there any noticeable changes in the results?



Using the estimation results from (b) and (c), test H_0 : $\beta_2 = -\beta_3$ against H_1 : $\beta_2 \neq -\beta_3$ at the 5% significance level. Do hypothesis test results differ between (b) and (c)?



Predict ln(A) for the next two years assuming that $W_{T+1} = W_{T+2} = 500$ and $B_{T+1} = B_{T+2} = 500$.

Part II. Two-Stage Least Square Estimation (5 points)

YKLPRW

Consider the following production functions as:

$$\ln(Y_t) = \beta_1 + \beta_2 \ln(K_t) + \beta_3 \ln(L_t) + e_{1t}$$
 (1)

$$\ln(K_1) = \alpha_1 + \alpha_2 \ln(P_1) + \alpha_3 \ln(Y_1) + \alpha_4 \ln(R_1) + e_{21}$$
 (2)

$$\ln(L_{t}) = \lambda_{1} + \lambda_{2} \ln(P_{t}) + \lambda_{3} \ln(Y_{t}) + \lambda_{4} \ln(W_{t}) + e_{3t}$$
 (3),

where Y_t , K_t and L_t are output, capital input, and labour input, and P_t , R_t and W_t are corresponding prices. Given that Y_t , K_t and L_t are simultaneously determined by equations (1), (2) and (3), and P_t , R_t and W_t are determined from outside, it seems appropriate to treat the three equations as a simultaneous system where Y_t , K_t and L_t are the endogenous variables and P_t , R_t and W_t are exogenous variables. Data for this example (23 observations) are in the file HW5-DATA2

Y = output

| P = Psice output

| K = Capital input | R = Price of capital input.

| L = labor input | W = psice of labor input.



Is the production function (1) identified?

Find OLS estimates of the production function, and comment the problem of OLS.

Find two stage least squares estimates of β_2 and β_3 . Comment on the results.

Part III. Instrumental Variable/Two-Stage Least Square Estimation (5 points)

Consider a production function for local beer production as:

$$Y_{t} = \beta_{1} + \beta_{2}mgt_{t} + \beta_{3}cap_{t} + \beta_{4}lab_{t} + e_{1t},$$

where y_t is an index of beer output for t-th brewery, taking into account both quantity and quality, mgt_t , is a variable reflecting the efficiency of management, cap_t and lab_t are indices of capital input and labor inputs. Because Kelly cannot get data an management efficiency, mgt_t , she collects data on the number of years of experience on the brewery business $(xpert_t)$ of each brewery manager and uses that variable in place of mgt_t . Data for this example (75 observations) are in the file HW5-DATA3

Estimate the equation using OLS.

(b) Estimate the instrumental variable/2SLS with instruments age, cap, lab, trend, and trend2, and compare the results with OLS results.

Run Hausman test to check on the validity of the instrumental variables.

YKLPRW

P = Psice of capital input. N = Psice of capital input.

y = eadsult K = Capilal Impult L = dabut Impult