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
s"
        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: lna

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	Standard Error	t Value	Pr > |t|
Intercept	1	5.20128	2.34569	2.22	0.0341
lnw	1	0.96630	0.11267	8.58	<.0001
lnb	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,

Given, $P_t = w_t / B_t$

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

Assume,

$\beta_2 = -\beta_3$, then replacing this assumption in above equn,

$$\ln(A_t) = \beta_1 + \beta_2 \ln(w_t) - \beta_2 \ln(B_t) + e_t$$

$$= \beta_1 + \beta_2 (\ln(w_t) - \ln(B_t)) + e_t$$

$$= \beta_1 + \beta_2 (\ln(w_t / B_t)) + e_t \quad \text{(log properties)}$$

$$= \beta_1 + \beta_2 \ln(P_t) + e_t \quad \therefore P_t = w_t / B_t$$

Proved

(b).

The SAS System

The REG Procedure
Model: MODEL1
Dependent Variable: lna

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	Standard Error	t Value	Pr > |t|
Intercept	1	5.20128	2.34569	2.22	0.0341
lnw	1	0.96630	0.11267	8.58	<.0001
lnb	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: lna

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 auto-correlated (no autocorrelation).

H_o: ρ = 0

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

H_a: ρ > 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;
```

The SAS System

The AUTOREG Procedure

Ordinary Least Squares Estimates			
SSE	3.0360454	DFE	31
MSE	0.09794	Root MSE	0.31295
SBC	24.9295416	AIC	20.3504601
MAE	0.24626269	AICC	21.1504601
MAPE	5.35024517	HQC	21.912058
Durbin-Watson	1.3097	Total R-Square	0.7078

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	5.2013	2.3457	2.22	0.0341
lnw	1	0.9663	0.1127	8.58	<.0001
lnb	1	-0.8360	0.3636	-2.30	0.0284

Estimates of Autocorrelations

Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
0	0.0893	1.000000	| |********************|
1	0.0292	0.327076	| |******* |

Preliminary MSE	0.0797

Estimates of Autoregressive Parameters

Lag	Coefficient	Standard Error	t Value
1	-0.327076	0.172532	-1.90

The SAS System

The AUTOREG Procedure

Yule-Walker Estimates			
SSE	2.68473544	DFE	30
MSE	0.08949	Root MSE	0.29915
SBC	24.3879436	AIC	18.2825015
MAE	0.23173962	AICC	19.6618118
MAPE	5.02459149	HQC	20.3646321
Durbin-Watson	1.9391	Transformed Regression R-Square	0.6338
		Total R-Square	0.7416

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Approx Pr > |t|
Intercept	1	5.6184	2.0133	2.79	0.0091
lnw	1	1.0127	0.1427	7.10	<.0001
lnb	1	-0.9328	0.2993	-3.12	0.0040

Test 1

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

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 lna				
Source	DF	Mean Square	F Value	Pr > F
Numerator	1	0.01482	0.15	0.6999
Denominator	31	0.09794		

Null hypothesis: H_o: $\beta_2 = \beta_3$.
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_o: $\beta_2 = \beta_3$.
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)+ ρ*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
t"
            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(l);
lnp = log(p);
lnr = log(r);
run;
```

Given:

Y = output; P = output price;

K = capital input; R = capital input price;

L = Labor input; W = labor input price.

Output = *f* (capital input, labor input) ………………….. (1)

Capital input = *f* (output price, output, capital input price) ………………….. (2)

Labor input = *f* (output price, output, labor input price) ………………….. (3)

Endogenous Variables: Y, K, and L. Exogenous variables: P, R, and W.

(a).

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: lny

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	Standard Error	t Value	Pr > |t|
Intercept	1	1.13757	0.14763	7.71	<.0001
lnk	1	0.00695	0.29386	0.02	0.9814
lnl	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;
```

Regression Prediction run

The MODEL Procedure

Model Summary	
Model Variables	3
Endogenous	2
Parameters	11
Equations	3
Number of Statements	3

Model Variables	lnk lnl lny
Parameters	a0 a1 a2 b0 b1 b2 b3 c0 c1 c2 c3
Equations	lny lnk lnl

The Equation to Estimate is	
lny =	F(a0(1), a1(lnk), a2(lnl))
Instruments	1 lnp lnr lnw

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

Regression Prediction run

The MODEL Procedure
2SLS Estimation Summary

Data Set Options	
DATA=	HW5D2

Minimization Summary	
Parameters Estimated	3
Method	Gauss
Iterations	1

Final Convergence Criteria	
R	0
PPC	0
RPC(a0)	10784.69
Object	0.999466
Trace(S)	0.562471
Objective Value	0.003853

Observations Processed	
Read	23
Solved	23

Regression Prediction run

The MODEL Procedure

Nonlinear 2SLS Summary of Residual Errors							
Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq
lny	3	20	11.2494	0.5625	0.7500	0.9070	0.8977

Nonlinear 2SLS Parameter Estimates				
Parameter	Estimate	Approx Std Err	t Value	Approx Pr > |t|
a0	1.089354	0.1820	5.99	<.0001
a1	0.247203	0.4275	0.58	0.5696
a2	0.707412	0.4262	1.66	0.1125

Number of Observations		Statistics for System	
Used	23	Objective	0.003853
Missing	0	Objective*N	0.0886

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).

```
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
s"
            DBMS=EXCEL REPLACE;
     RANGE="data";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

proc reg data = hw5d3;
model y = xper cap lab;
run;
```

Regression Prediction run

The REG Procedure
Model: MODEL1
Dependent Variable: y y

Number of Observations Read	75
Number of Observations Used	75

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	690.79130	230.26377	30.31	<.0001
Error	71	539.35083	7.59649		
Corrected Total	74	1230.14213			

Root MSE	2.75617	R-Square	0.5616
Dependent Mean	9.63442	Adj R-Sq	0.5430
Coeff Var	28.60756		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1.76226	1.05535	1.67	0.0994
xper	xper	1	0.14684	0.06343	2.31	0.0235
cap	cap	1	0.43796	0.11756	3.73	0.0004
lab	lab	1	0.23916	0.09980	2.40	0.0192

(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;
```

The REG Procedure
Model: MODEL1
Dependent Variable: xper xper

Number of Observations Read	75
Number of Observations Used	75

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	335.26440	111.75480	4.78	0.0043
Error	71	1658.65560	23.36135		
Corrected Total	74	1993.92000			

Root MSE	4.83336	R-Square	0.1681
Dependent Mean	13.88000	Adj R-Sq	0.1330
Coeff Var	34.82246		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	4.71600	2.57399	1.83	0.0711
cap	cap	1	0.40661	0.20702	1.96	0.0534
lab	lab	1	-0.11506	0.17868	-0.64	0.5217
age	age	1	0.16612	0.05303	3.13	0.0025

Regression Prediction run

The REG Procedure
Model: MODEL1
Dependent Variable: y y

Number of Observations Read	75
Number of Observations Used	75

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	710.20798	236.73599	32.33	<.0001
Error	71	519.93415	7.32302		
Corrected Total	74	1230.14213			

Root MSE	2.70611	R-Square	0.5773
Dependent Mean	9.63442	Adj R-Sq	0.5595
Coeff Var	28.08790		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	-2.48669	2.20705	-1.13	0.2637
pred	Predicted Value of xper	1	0.51210	0.17872	2.87	0.0055
cap	cap	1	0.33213	0.12521	2.65	0.0098
lab	lab	1	0.23998	0.09799	2.45	0.0168

- 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.

Regression Prediction run

The MODEL Procedure

Model Summary	
Model Variables	2
Endogenous	1
Parameters	4
Equations	1
Number of Statements	1

Model Variables	xper y
Parameters	a0 a1 a2 a3
Equations	y

The Equation to Estimate is	
y =	F(a0(1), a1(xper), a2(cap), a3(lab))
Instruments	1 age cap lab

NOTE: At 2SLS Iteration 1 convergence assumed because OBJECTIVE=2.193956E-27 is almost zero (<1E-12).

Regression Prediction run

The MODEL Procedure
2SLS Estimation Summary

Data Set Options	
DATA=	HW5D3

Minimization Summary	
Parameters Estimated	4
Method	Gauss
Iterations	1

Final Convergence Criteria	
R	1
PPC	0
RPC(a0)	24621.67
Object	1
Trace(S)	11.14411
Objective Value	2.19E-27

Observations Processed	
Read	75
Solved	75

The MODEL Procedure

Nonlinear 2SLS Summary of Residual Errors

Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq	Label
y	4	71	791.2	11.1441	3.3383	0.3568	0.3296	y

Nonlinear 2SLS Parameter Estimates

Parameter	Estimate	Approx Std Err	t Value	Approx Pr > |t|
a0	-2.48669	2.7226	-0.91	0.3642
a1	0.512102	0.2205	2.32	0.0231
a2	0.332134	0.1545	2.15	0.0349
a3	0.239975	0.1209	1.99	0.0510

Number of Observations		Statistics for System	
Used	75	Objective	2.194E-27
Missing	0	Objective*N	1.645E-25

Regression Prediction run

The MODEL Procedure

Model Summary	
Model Variables	2
Endogenous	1
Parameters	4
Equations	1
Number of Statements	1

Model Variables	xper y
Parameters	a0 a1 a2 a3
Equations	y

The Equation to Estimate is	
y =	F(a0(1), a1(xper), a2(cap), a3(lab))
Instruments	1 age cap lab

NOTE: At 2SLS Iteration 1 convergence assumed because OBJECTIVE=2.193956E-27 is almost zero (<1E-12).

Regression Prediction run

The MODEL Procedure
2SLS Estimation Summary

Data Set Options	
DATA=	HW5D3

Minimization Summary	
Parameters Estimated	4
Method	Gauss
Iterations	1

Final Convergence Criteria	
R	1
PPC	0
RPC(a0)	24621.67
Object	1
Trace(S)	11.14411
Objective Value	2.19E-27

Observations Processed	
Read	75
Solved	75

The MODEL Procedure

Nonlinear 2SLS Summary of Residual Errors								
Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq	Label
y	4	71	791.2	11.1441	3.3383	0.3568	0.3296	y

Nonlinear 2SLS Parameter Estimates				
Parameter	Estimate	Approx Std Err	t Value	Approx Pr > |t|
a0	-2.48669	2.7226	-0.91	0.3642
a1	0.512102	0.2205	2.32	0.0231
a2	0.332134	0.1545	2.15	0.0349
a3	0.239975	0.1209	1.99	0.0510

Number of Observations		Statistics for System	
Used	75	Objective	2.194E-27
Missing	0	Objective*N	1.645E-25

I ran GMM test.

Regression Prediction run

The MODEL Procedure
GMM Estimation Summary

Data Set Options	
DATA=	HW5D3

Minimization Summary	
Parameters Estimated	4
Kernel Used	PARZEN
l(n)	2.371441
Method	Gauss
Iterations	0

Final Convergence Criteria	
R	1
PPC	0
RPC	.
Object	.
Trace(S)	11.14411
Objective Value	3.46E-28

Observations Processed	
Read	75
Solved	75

The MODEL Procedure

Nonlinear GMM Summary of Residual Errors								
Equation	DF Model	DF Error	SSE	MSE	Root MSE	R-Square	Adj R-Sq	Label
y	4	71	791.2	11.1441	3.3383	0.3568	0.3296	y

Nonlinear GMM Parameter Estimates				
Parameter	Estimate	Approx Std Err	t Value	Approx Pr > |t|
a0	-2.48669	2.3302	-1.07	0.2895
a1	0.512102	0.1919	2.67	0.0094
a2	0.332134	0.1585	2.10	0.0397
a3	0.239975	0.1179	2.03	0.0456

Number of Observations		Statistics for System	
Used	75	Objective	3.464E-28
Missing	0	Objective*N	2.598E-26

GMM Test Statistics			
Test	DF	Statistic	Prob
Overidentifying Restrictions	0	0.00	.

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
s"
            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
t"
            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(l);
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
s"
            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 $\ln(A_t) = \beta_1 + \beta_2 \ln(P_t) + e_t$ where A_t 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

(a) 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 $\beta_2 = -\beta_3$.

(b) 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.

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

(d) 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)?

(e) 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)

Y K L P R W

Consider the following production functions as:

$$\ln(Y_t) = \beta_1 + \beta_2 \ln(K_t) + \beta_3 \ln(L_t) + e_{1t} \quad (1)$$
$$\ln(K_t) = \alpha_1 + \alpha_2 \ln(P_t) + \alpha_3 \ln(Y_t) + \alpha_4 \ln(R_t) + e_{2t} \quad (2)$$
$$\ln(L_t) = \lambda_1 + \lambda_2 \ln(P_t) + \lambda_3 \ln(Y_t) + \lambda_4 \ln(W_t) + e_{3t} \quad (3),$$

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
K = Capital input
L = labor input

P = Price output
R = Price of capital input
W = price of labor input

(a) Is the production function (1) identified?
(b) Find OLS estimates of the production function, and comment the problem of OLS.
(c) 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*

(a) 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.
(c) Run Hausman test to check on the validity of the instrumental variables.