**aAAEC 6311**

**LAB #1**

Objectives:

1. Learn to implement IV procedures using SAS

Card (1995) analyzed the effect of education (measured as yeas of schooling) on earnings (measured as hourly wage). Here we will use the same data from the Young Men Cohort of the 1976 National Longitudinal Survey. The model considered by Card (1995) was:

(1)

where denotes hourly wages, *s* denotes years of schooling, and **x**1is a vector of control variables.

In this model the schooling variable is considered endogenous due to the lack of data on ability. Thus, the error is correlated con *s*. The instrument used by Card (1995) is an indicator for whether a four year college is nearby.

Is the proximity to a college or university (*d*) a good instrument for schooling?

**Part 1. Basic Operations Using SAS**

* 1. Import and manipulate the data

Proc import

Data

* 1. Calculate basic summary statistics:

|  |
| --- |
| The SAS System |

The MEANS Procedure

| **Variable** | **Label** | **N** | **Mean** | **Std Dev** | **Minimum** | **Maximum** |
| --- | --- | --- | --- | --- | --- | --- |
| |  | | --- | | **wage76** | | **black** | | **age1415** | | **age1617** | | **grade76** | | **education** | | **exp76** | | **expsq76** | | **south76** | | **smsa76** | | **ca1** | | |  | | --- | | **wage76** | | **black** | | **age1415** | | **age1617** | | **grade76** | |  | | **exp76** | | **expsq76** | | **south76** | | **smsa76** | | **ca1** | | |  | | --- | | 3010 | | 3010 | | 3010 | | 3010 | | 3010 | | 3010 | | 3010 | | 3010 | | 3010 | | 3010 | | 3010 | | |  | | --- | | 1.6566642 | | 0.2335548 | | 0.2548173 | | 0.2408638 | | 13.2634551 | | 13.2634551 | | 8.8561462 | | 0.9557907 | | 0.4036545 | | 0.7129568 | | 0.3179402 | | |  | | --- | | 0.4437980 | | 0.4231624 | | 0.4358309 | | 0.4276788 | | 2.6769129 | | 2.6769129 | | 4.1416716 | | 0.8461831 | | 0.4907113 | | 0.4524571 | | 0.4657535 | | |  | | --- | | 0 | | 0 | | 0 | | 0 | | 1.0000000 | | 1.0000000 | | 0 | | 0 | | 0 | | 0 | | 0 | | |  | | --- | | 3.1796999 | | 1.0000000 | | 1.0000000 | | 1.0000000 | | 18.0000000 | | 18.0000000 | | 23.0000000 | | 5.2900000 | | 1.0000000 | | 1.0000000 | | 1.0000000 | |

**Part 2. LS and IV Estimation Using SAS**

2.1. Estimate the parameters of Equation (1) using LS:

|  |
| --- |
| The SAS System |

The REG Procedure

Model: MODEL1

Dependent Variable: wage76 wage76

|  |  |
| --- | --- |
| **Number of Observations Read** | 3010 |
| **Number of Observations Used** | 3010 |

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 6 | 172.16604 | 28.69434 | 204.93 | <.0001 |
| **Error** | 3003 | 420.47657 | 0.14002 |  |  |
| **Corrected Total** | 3009 | 592.64262 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 0.37419 | **R-Square** | 0.2905 |
| **Dependent Mean** | 1.65666 | **Adj R-Sq** | 0.2891 |
| **Coeff Var** | 22.58701 |  |  |

| **Parameter Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Label** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| **Intercept** | Intercept | **1** | 0.12848 | 0.06760 | 1.90 | 0.0575 |
| **education** |  | **1** | 0.07401 | 0.00351 | 21.11 | <.0001 |
| **exp76** | exp76 | **1** | 0.08360 | 0.00665 | 12.58 | <.0001 |
| **expsq76** | expsq76 | **1** | -0.22409 | 0.03178 | -7.05 | <.0001 |
| **black** | black | **1** | -0.18963 | 0.01763 | -10.76 | <.0001 |
| **south76** | south76 | **1** | -0.12486 | 0.01512 | -8.26 | <.0001 |
| **smsa76** | smsa76 | **1** | 0.16142 | 0.01557 | 10.37 | <.0001 |
|  |  |  |  |  |  |  |

2.1.1. Interpret

For every additional year of schooling, we would expect wages to increase by 7.4 percent, ceteris paribus.

2.2. Estimate the parameters of Equation (1) using 2SLS in the proc syslin procedure:

|  |
| --- |
| The SAS System |

The SYSLIN Procedure

Two-Stage Least Squares Estimation

|  |  |
| --- | --- |
| **Model** | wage76 |
| **Dependent Variable** | wage76 |
| **Label** | wage76 |

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 6 | 110.8566 | 18.47610 | 120.83 | <.0001 |
| **Error** | 3003 | 459.1862 | 0.152909 |  |  |
| **Corrected Total** | 3009 | 592.6426 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 0.39104 | **R-Square** | 0.19447 |
| **Dependent Mean** | 1.65666 | **Adj R-Sq** | 0.19286 |
| **Coeff Var** | 23.60382 |  |  |

| **Parameter Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** | **Variable Label** |
| **Intercept** | 1 | -0.85249 | 0.829348 | -1.03 | 0.3041 | Intercept |
| **education** | 1 | 0.132295 | 0.049234 | 2.69 | 0.0072 |  |
| **exp76** | 1 | 0.107501 | 0.021301 | 5.05 | <.0001 | exp76 |
| **expsq76** | 1 | -0.22841 | 0.033414 | -6.84 | <.0001 | expsq76 |
| **black** | 1 | -0.13080 | 0.052873 | -2.47 | 0.0134 | black |
| **south76** | 1 | -0.10490 | 0.023073 | -4.55 | <.0001 | south76 |
| **smsa76** | 1 | 0.131321 | 0.030130 | 4.36 | <.0001 | smsa76 |

2.2.1 Compare the LS and 2SLS (from proc syslin) estimates of and their corresponding standard errors.

|  |  |  |
| --- | --- | --- |
|  | **LS** | **2SLS** |
| **Parameter** | 0.07401 | 0.132295 |
| **Standard error** | 0.00351 | 0.049234 |

Interpret:

We see that using LS, we estimate that for every additional year of schooling, we would expect wages to increase by 7.4 percent, ceteris paribus. However, when we use 2SLS, we estimate that for every additional year of schooling, we would expect wages to increase by 13.2295 percent, ceteris paribus. The spread of our data is narrower with LS with a smaller standard error (0.00351) while 2SLS has a larger standard error (0.049234).

2.3. Estimate the parameters of Equation (1) using 2SLS in the proc model procedure:

|  |
| --- |
| The SAS System |

The MODEL Procedure

| **Nonlinear 2SLS Summary of Residual Errors** | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Equation** | **DF Model** | **DF Error** | **SSE** | **MSE** | **Root MSE** | **R-Square** | **Adj R-Sq** | **Label** |
| **wage76** | 7 | 3003 | 459.2 | 0.1529 | 0.3910 | 0.2252 | 0.2236 | wage76 |

| **Nonlinear 2SLS Parameter Estimates** | | | | |
| --- | --- | --- | --- | --- |
| **Parameter** | **Estimate** | **Approx Std Err** | **t Value** | **Approx Pr > |t|** |
| **bo** | -0.85249 | 0.8293 | -1.03 | 0.3041 |
| **b1** | 0.132295 | 0.0492 | 2.69 | 0.0072 |
| **b2** | 0.107501 | 0.0213 | 5.05 | <.0001 |
| **b3** | -0.22841 | 0.0334 | -6.84 | <.0001 |
| **b4** | -0.1308 | 0.0529 | -2.47 | 0.0134 |
| **b5** | -0.1049 | 0.0231 | -4.55 | <.0001 |
| **b6** | 0.131321 | 0.0301 | 4.36 | <.0001 |

| **Number of Observations** | | **Statistics for System** | |
| --- | --- | --- | --- |
| **Used** | 3010 | **Objective** | 1.19E-27 |
| **Missing** | 0 | **Objective\*N** | 3.582E-24 |

2.3.1 Compare the 2SLS proc syslin and 2SLS proc model estimates of .

|  |  |  |
| --- | --- | --- |
|  | **2SLS (proc syslin)** | **2SLS (proc model)** |
| **Parameter** | 0.132295 | 0.132295 |
| **Standard error** | 0.0492 | 0.0492 |

Interpret:

Using 2SLS proc sylin and proc model yields us the same exact results for the parameter estimate and standard errors for the variable education.

**Part 3. Estimate the parameters of Equation (1) using the 2 step 2SLS procedure described in class (Run two separate regressions).**

First stage regression:

|  |
| --- |
| The SAS System |

The REG Procedure

Model: MODEL1

Dependent Variable: education

|  |  |
| --- | --- |
| **Number of Observations Read** | 3010 |
| **Number of Observations Used** | 3010 |

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 6 | 10230 | 1705.08072 | 451.87 | <.0001 |
| **Error** | 3003 | 11332 | 3.77343 |  |  |
| **Corrected Total** | 3009 | 21562 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 1.94253 | **R-Square** | 0.4745 |
| **Dependent Mean** | 13.26346 | **Adj R-Sq** | 0.4734 |
| **Coeff Var** | 14.64574 |  |  |

| **Parameter Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Label** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| **Intercept** | Intercept | **1** | 16.99650 | 0.17608 | 96.53 | <.0001 |
| **ca1** | ca1 | **1** | -0.33732 | 0.08250 | -4.09 | <.0001 |
| **exp76** | exp76 | **1** | -0.41001 | 0.03369 | -12.17 | <.0001 |
| **expsq76** | expsq76 | **1** | 0.07323 | 0.16499 | 0.44 | 0.6572 |
| **black** | black | **1** | -1.00614 | 0.08965 | -11.22 | <.0001 |
| **south76** | south76 | **1** | -0.29146 | 0.07922 | -3.68 | 0.0002 |
| **smsa76** | smsa76 | **1** | 0.40388 | 0.08489 | 4.76 | <.0001 |

Second stage regression:

|  |
| --- |
| The SAS System |

The REG Procedure

Model: MODEL1

Dependent Variable: wage76 wage76

|  |  |
| --- | --- |
| **Number of Observations Read** | 3010 |
| **Number of Observations Used** | 3010 |

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 6 | 110.85657 | 18.47610 | 115.16 | <.0001 |
| **Error** | 3003 | 481.78604 | 0.16043 |  |  |
| **Corrected Total** | 3009 | 592.64262 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 0.40054 | **R-Square** | 0.1871 |
| **Dependent Mean** | 1.65666 | **Adj R-Sq** | 0.1854 |
| **Coeff Var** | 24.17770 |  |  |

| **Parameter Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Label** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| **Intercept** | Intercept | **1** | -0.85249 | 0.84951 | -1.00 | 0.3157 |
| **eduhat** | Predicted Value of education | **1** | 0.13229 | 0.05043 | 2.62 | 0.0088 |
| **exp76** | exp76 | **1** | 0.10750 | 0.02182 | 4.93 | <.0001 |
| **expsq76** | expsq76 | **1** | -0.22841 | 0.03423 | -6.67 | <.0001 |
| **black** | black | **1** | -0.13080 | 0.05416 | -2.42 | 0.0158 |
| **south76** | south76 | **1** | -0.10490 | 0.02363 | -4.44 | <.0001 |
| **smsa76** | smsa76 | **1** | 0.13132 | 0.03086 | 4.26 | <.0001 |

3.1. How do the parameter estimates obtained in the second stage compare to those obtained using proc syslin and proc model?

The parameter estimates for education obtained in the second stage was 0.13229, which is identical to the values obtained from the proc sylin and proc model.

3.2. How do the standard errors of the coefficients obtained in the second stage compare to those obtained using proc syslin and proc model? If they are different, which ones would you prefer?

Using the standard error from the first stage and comparing this to the proc sylin and proc model, our value from the first stage (0.0825) compared to the proc sylin and proc model (0.0492) is greater. Therefore, all else equal, we would prefer to use the model with smaller standard errors, that being the proc sylin and proc model.

3.3. Use proc model to perform the Hausman test (the degrees of freedom reported by SAS are incorrect).

/\*IV(2SLS) in proc model with Hausman test\*/

**proc** **model** data=CARD1;

endogenous education;

instruments ca1 exp76 expsq76 black south76 smsa76;

parms bo b1 b2 b3 b4 b5 b6;

wage76 = bo+ b1\*education + b2\*exp76 + b3\*expsq76 + b4\*black + b5\*south76 + b6\*smsa76;

fit wage76 / ols **2**sls hausman ;

**run**;

/\*Be careful: Automatic Hausman test has incorrect degrees of freedom\*/

**data** chisq;

df = **1**;

p\_val = **1** - probchi(**1.41**,df);

chi1\_90 = cinv(**.90**,df);

chi1\_95 = cinv(**.95**,df);

chi1\_99 = cinv(**.99**,df);

**run**;

3.4. What are the conclusions of the test?

| **Hausman's Specification Test Results** | | | | |
| --- | --- | --- | --- | --- |
| **Efficient under H0** | **Consistent under H1** | **DF** | **Statistic** | **Pr > ChiSq** |
| **OLS** | 2SLS | 7 | 1.41 | 0.9853 |

|  |
| --- |
| **The SAS System** |

| **df** | **p\_val** | **chi1\_90** | **chi1\_95** | **chi1\_99** |
| --- | --- | --- | --- | --- |
| 1 | 0.23506 | 2.70554 | 3.84146 | 6.63490 |

H­0: plim(biv-bLS)=0

H­A: plim(biv-bLS)≠0

Since 1.41<2.70554 at even a 90% level, we fail to reject the null hypothesis and conclude there is no evidence of endogeneity.

**Part 4. Overidentification tests.**

4.1. Use proc syslin

Report and interpret the result of the over-identification test

The over-identification test can only be used if the number of instrumental variables is greater than the number of endogenous variables. Here, we only have one instrumental variable proximity to nearby 4-year college (ca1) as an instrument for one endogenous variable education.

Part 5. Detection of “weakness” in instrumental variables.

5.1. Conduct appropriate analyses to evaluate the weakness of the instruments. Provide conclusions.

|  |
| --- |
| The SAS System |

The REG Procedure

Model: MODEL1

Dependent Variable: education

|  |  |
| --- | --- |
| **Number of Observations Read** | 3010 |
| **Number of Observations Used** | 3010 |

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 6 | 10230 | 1705.08072 | 451.87 | <.0001 |
| **Error** | 3003 | 11332 | 3.77343 |  |  |
| **Corrected Total** | 3009 | 21562 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 1.94253 | **R-Square** | 0.4745 |
| **Dependent Mean** | 13.26346 | **Adj R-Sq** | 0.4734 |
| **Coeff Var** | 14.64574 |  |  |

| **Parameter Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Label** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| **Intercept** | Intercept | **1** | 16.99650 | 0.17608 | 96.53 | <.0001 |
| **ca1** | ca1 | **1** | -0.33732 | 0.08250 | -4.09 | <.0001 |
| **exp76** | exp76 | **1** | -0.41001 | 0.03369 | -12.17 | <.0001 |
| **expsq76** | expsq76 | **1** | 0.07323 | 0.16499 | 0.44 | 0.6572 |
| **black** | black | **1** | -1.00614 | 0.08965 | -11.22 | <.0001 |
| **south76** | south76 | **1** | -0.29146 | 0.07922 | -3.68 | 0.0002 |
| **smsa76** | smsa76 | **1** | 0.40388 | 0.08489 | 4.76 | <.0001 |

H0: ca1 is not a weak instrument

HA: ca1 is a weak instrument

Since we only have one instrumental variable, we can square the t-value to get the F-value for ca1: F-valueca1 = (t-valueca1)2 = (-4.09)2 = 16.7281. Because the F-valuecal (16.7281) > 10, our instrumental variable ca1 on the proximity of nearby 4-year colleges is not a weak instrumental variable.

Part 6. Use of the control function approach to control for endogeneity

6.1. Compare the 2SLS and control function approach estimates of and their corresponding standard errors.

|  |
| --- |
| The SAS System |

The MODEL Procedure

| **Nonlinear OLS Summary of Residual Errors** | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Equation** | **DF Model** | **DF Error** | **SSE** | **MSE** | **Root MSE** | **R-Square** | **Adj R-Sq** | **Label** |
| **wage76** | 8 | 3002 | 420.3 | 0.1400 | 0.3742 | 0.2909 | 0.2892 | wage76 |

| **Nonlinear OLS Parameter Estimates** | | | | |
| --- | --- | --- | --- | --- |
| **Parameter** | **Estimate** | **Approx Std Err** | **t Value** | **Approx Pr > |t|** |
| **bo** | -0.85249 | 0.7935 | -1.07 | 0.2828 |
| **b1** | 0.132295 | 0.0471 | 2.81 | 0.0050 |
| **b2** | 0.107501 | 0.0204 | 5.27 | <.0001 |
| **b3** | -0.22841 | 0.0320 | -7.14 | <.0001 |
| **b4** | -0.1308 | 0.0506 | -2.59 | 0.0098 |
| **b5** | -0.1049 | 0.0221 | -4.75 | <.0001 |
| **b6** | 0.131321 | 0.0288 | 4.56 | <.0001 |
| **b7** | -0.05861 | 0.0472 | -1.24 | 0.2148 |

| **Number of Observations** | | **Statistics for System** | |
| --- | --- | --- | --- |
| **Used** | 3010 | **Objective** | 0.1396 |
| **Missing** | 0 | **Objective\*N** | 420.2611 |

|  |  |  |
| --- | --- | --- |
|  | **Control function** | **2SLS** |
| **Parameter** | 0.132295 | 0.132295 |
| **Standard error** | 0.0471 | 0.0492 |

Interpret:

The estimated parameter of education on wages is the same, however, the estimated standard error in the control function approach is smaller than the 2SLS approach. Theoretically this makes sense, since the control function approach attempts to “clean up” possible correlation between our endogenous variables and the unobservables in our error term.

6.2. Report and interpret the results of the Hausman test for endogeneity.

|  |
| --- |
| The SAS System |

The REG Procedure

Model: MODEL1

Dependent Variable: wage76 wage76

|  |  |
| --- | --- |
| **Number of Observations Read** | 3010 |
| **Number of Observations Used** | 3010 |

| **Analysis of Variance** | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 7 | 172.38154 | 24.62593 | 175.91 | <.0001 |
| **Error** | 3002 | 420.26108 | 0.13999 |  |  |
| **Corrected Total** | 3009 | 592.64262 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 0.37416 | **R-Square** | 0.2909 |
| **Dependent Mean** | 1.65666 | **Adj R-Sq** | 0.2892 |
| **Coeff Var** | 22.58498 |  |  |

| **Parameter Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Label** | **DF** | **Parameter Estimate** | **Standard Error** | **t Value** | **Pr > |t|** |
| **Intercept** | Intercept | **1** | -0.85249 | 0.79355 | -1.07 | 0.2828 |
| **education** |  | **1** | 0.13229 | 0.04711 | 2.81 | 0.0050 |
| **exp76** | exp76 | **1** | 0.10750 | 0.02038 | 5.27 | <.0001 |
| **expsq76** | expsq76 | **1** | -0.22841 | 0.03197 | -7.14 | <.0001 |
| **black** | black | **1** | -0.13080 | 0.05059 | -2.59 | 0.0098 |
| **south76** | south76 | **1** | -0.10490 | 0.02208 | -4.75 | <.0001 |
| **smsa76** | smsa76 | **1** | 0.13132 | 0.02883 | 4.56 | <.0001 |
| **ehat1** | Residual | **1** | -0.05861 | 0.04724 | -1.24 | 0.2148 |

H0: rho = 0

HA: rho ≠ 0

Since the t-value for rho is -1.24 and the p-value is 0.2148, it is not significant at a α=0.05 level. We fail to reject the null hypothesis and conclude that there is no evidence of endogeneity.

**SAS code for the lab**

/\*Data from Card, D. 1995. "Using Geographic Variation in College Proximity to Estimate The Return to Schooling." NBER Working Paper No. 4483.

http://econweb.tamu.edu/gan/econometrics1/w4483.pdf \*/

/\*Importing data from Excel file\*/

**proc** **import**

datafile="C:\ccarpio\Documents\Teaching TTU\AAEC6311\data\CARD.xls"

out = CARD replace;

sheet = "sheet1";

getnames=yes;

**run**;

/\*Variables to be used in estimation\*/

/\*

smsa76 %If If lived in SMSA in 1976 (r0437515=1,2)

ca1 %If If any 4-year college nearby (r0004000!=4)

south76 %1f If lived in South in 1976 (r0437511=1)

wage76 %If '76 log(Wage)

exp76 %If '76 experience, (10 + age66) - grade76 - 6)

expsq76 %If '76 experience, exp76 ^2/100 \*/

/\*Cleaning up the data\*/

**data** CARD1;

set CARD;

if wage76="." then delete;

if grade76="." then delete;

/\*Delete observations that do not have interview month\*/

if intmo66=-**999** then delete;

education = grade76;

**run**;

/\*Basic summary statistics\*/

**proc** **means** data=CARD1;

var wage76 black age1415 age1617 grade76 education exp76 expsq76 black south76 smsa76 ca1;

**run**;

/\*Basic OLS regression\*/

**proc** **reg** data=CARD1;

model wage76 = education exp76 expsq76 black south76 smsa76;

output out=CARD2 predicted=yhat residual=ehat; /\*Save residuals and predicted values in original dataset\*/

**run**;

**quit**;

/\*IV(2SLS) in proc syslin\*/

**proc** **syslin** data=CARD1 **2**sls; /\*Note: you could use ols here too\*/

endogenous education;

instruments ca1 exp76 expsq76 black south76 smsa76;

model wage76 = education exp76 expsq76 black south76 smsa76/ overid;

**run**;

/\*IV(2SLS) in proc model\*/

**proc** **model** data=CARD1;

endogenous wage76 education;

instruments ca1 exp76 expsq76 black south76 smsa76;

parms bo b1 b2 b3 b4 b5 b6;

wage76 = bo+ b1\*education + b2\*exp76 + b3\*expsq76 + b4\*black + b5\*south76 + b6\*smsa76;

fit wage76 / **2**sls ; /\*Note: you could use ols here too\*/

**run**;

**STATA Appendix**

Importing and managing data:

**import excel "C:\ccarpio\Documents\Teaching TTU\AAEC6311\data\CARD.xls", sheet("Sheet1") firstrow**

**drop if if intmo66=-999**

Summary statistics:

summarize wage76 black smsa76 south76 exp76 expsq76 grade76 ca1

Variable | Obs Mean Std. Dev. Min Max

-------------+--------------------------------------------------------

wage76 | 3029 1.657025 .4441405 0 3.1797

black | 5139 .274567 .4463393 0 1

smsa76 | 5139 .4915353 .499977 0 1

south76 | 3637 .4003299 .4900326 0 1

exp76 | 3613 8.949903 4.215668 0 25

-------------+--------------------------------------------------------

expsq76 | 3613 .978677 .8789053 0 6.25

grade76 | 3613 13.2253 2.749741 0 18

ca1 | 5139 .3107608 .4628501 0 1

Basic OLS:

**regress wage76 black smsa76 south76 exp76 expsq76**

**Source | SS df MS Number of obs = 3010**

**-------------+------------------------------ F( 5, 3004) = 136.55**

**Model | 109.752506 5 21.9505012 Prob > F = 0.0000**

**Residual | 482.89011 3004 .160749038 R-squared = 0.1852**

**-------------+------------------------------ Adj R-squared = 0.1838**

**Total | 592.642616 3009 .196956669 Root MSE = .40094**

**------------------------------------------------------------------------------**

**wage76 | Coef. Std. Err. t P>|t| [95% Conf. Interval]**

**-------------+----------------------------------------------------------------**

**black | -.2643379 .0185019 -14.29 0.000 -.3006156 -.2280602**

**smsa76 | .1996459 .0165732 12.05 0.000 .1671499 .2321419**

**south76 | -.1502084 .0161476 -9.30 0.000 -.1818698 -.1185469**

**exp76 | .0532429 .0069544 7.66 0.000 .0396071 .0668787**

**expsq76 | -.2186054 .0340546 -6.42 0.000 -.2853781 -.1518327**

**\_cons | 1.374109 .0353634 38.86 0.000 1.30477 1.443448**

**------------------------------------------------------------------------------**

2SLS:

**ivregress 2sls wage76 black smsa76 south76 exp76 expsq76 (grade76 = ca1), first**

First-stage regressions

-----------------------

Number of obs = 3010

F( 6, 3003) = 451.87

Prob > F = 0.0000

R-squared = 0.4745

Adj R-squared = 0.4734

Root MSE = 1.9425

------------------------------------------------------------------------------

grade76 | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

black | -1.006138 .0896454 -11.22 0.000 -1.181911 -.8303656

smsa76 | .4038769 .0848872 4.76 0.000 .2374339 .5703199

south76 | -.291464 .0792247 -3.68 0.000 -.4468042 -.1361238

exp76 | -.4100081 .0336939 -12.17 0.000 -.4760735 -.3439427

expsq76 | .0732287 .1649946 0.44 0.657 -.2502852 .3967426

ca1 | -.3373208 .0825004 -4.09 0.000 -.4990839 -.1755577

\_cons | 16.9965 .1760758 96.53 0.000 16.65125 17.34174

------------------------------------------------------------------------------

Instrumental variables (2SLS) regression Number of obs = 3010

Wald chi2(6) = 726.67

Prob > chi2 = 0.0000

R-squared = 0.2252

Root MSE = .39058

------------------------------------------------------------------------------

wage76 | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

grade76 | .1322948 .0491764 2.69 0.007 .0359109 .2286788

black | -.130795 .0528112 -2.48 0.013 -.2343031 -.0272869

smsa76 | .1313207 .030095 4.36 0.000 .0723355 .1903059

south76 | -.1048971 .0230465 -4.55 0.000 -.1500673 -.0597269

exp76 | .1075006 .021276 5.05 0.000 .0658004 .1492008

expsq76 | -.2284089 .0333747 -6.84 0.000 -.2938221 -.1629957

\_cons | -.8524886 .8283829 -1.03 0.303 -2.476089 .771112

------------------------------------------------------------------------------

Instrumented: grade76

Instruments: black smsa76 south76 exp76 expsq76 ca1

**estat endogenous**

Tests of endogeneity

Ho: variables are exogenous

Durbin (score) chi2(1) = 1.54263 (p = 0.2142)

Wu-Hausman F(1,3002) = 1.53932 (p = 0.2148)

**estat firststage**

First-stage regression summary statistics

--------------------------------------------------------------------------

| Adjusted Partial

Variable | R-sq. R-sq. R-sq. F(1,3003) Prob > F

-------------+------------------------------------------------------------

grade76 | 0.4745 0.4734 0.0055 16.7176 0.0000

--------------------------------------------------------------------------

Minimum eigenvalue statistic = 16.7176

Critical Values # of endogenous regressors: 1

Ho: Instruments are weak # of excluded instruments: 1

---------------------------------------------------------------------

| 5% 10% 20% 30%

2SLS relative bias | (not available)

-----------------------------------+---------------------------------

| 10% 15% 20% 25%

2SLS Size of nominal 5% Wald test | 16.38 8.96 6.66 5.53

LIML Size of nominal 5% Wald test | 16.38 8.96 6.66 5.53

---------------------------------------------------------------------

**estat overid**

no overidentifying restrictions

Further references about post-estimation tests

http://www.stata.com/manuals13/rivregresspostestimation.pdf