Homework 4

ECON 7023: Econometrics II
Maghfira Ramadhani
March 18, 2022

Spring 2023

Chapter 6

Problem 6.8

The data in FERTIL1.RAW are a pooled cross section on more than a thousand U.S. women for the even years between 1972 and 1984, inclusive; the data set is similar to the one used by Sande (1992). These data can be used to study the relationship between women's education and fertility.

a. Use OLS to estimate a model relating number of children ever born to a woman (*kids*) to years of education, age, region, race, and type of environment reared in. You should use a quadratic in age and should include year dummies. What is the estimated relationship between fertility and education? Holding other factors fixed, has there been any notable secular change in fertility over the time period? Answer:

OLS Regression Output

```
. gen age2=age^2

. reg kids educ age age2 black east northcen west farm othrural town smcity y74-y84
```

Source	SS	df	MS			
Model	399.610888	17	23.5065228	Number of obs	=	1,129
Residual	2685.89841	1,111	2.41755033	F(17, 1111)	=	9.72
Total	3085.5093	1,128	2.73538059	Prob > F	=	0.0000
				R-squared	=	0.1295
				Adj R-squared	=	0.1162
				Root MSE	=	1.5548

kids	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
educ	-.1284268	.0183486	-7.00	0.000	-.1644286	-.092425
age	.5321346	.1383863	3.85	0.000	.2606065	.8036626
age2	-.005804	.0015643	-3.71	0.000	-.0088733	-.0027347
black	1.075658	.1735356	6.20	0.000	.7351631	1.416152
east	.217324	.1327878	1.64	0.102	-.0432192	.4778672
northcen	.363114	.1208969	3.00	0.003	.125902	.6003261
west	.1976032	.1669134	1.18	0.237	-.1298978	.5251041
farm	-.0525575	.14719	-0.36	0.721	-.3413592	.2362443
othrural	-.1628537	.175442	-0.93	0.353	-.5070887	.1813814
town	.0843532	.124531	0.68	0.498	-.1599893	.3286957
smcity	.2118791	.160296	1.32	0.187	-.1026379	.5263961
y74	.2681825	.172716	1.55	0.121	-.0707039	.6070689
y76	-.0973795	.1790456	-0.54	0.587	-.448685	.2539261
y78	-.0686665	.1816837	-0.38	0.706	-.4251483	.2878154
y80	-.0713053	.1827707	-0.39	0.697	-.42992	.2873093
y82	-.5224842	.1724361	-3.03	0.003	-.8608214	-.184147
y84	-.5451661	.1745162	-3.12	0.002	-.8875846	-.2027477
_cons	-7.742457	3.051767	-2.54	0.011	-13.73033	-1.754579

The OLS estimate shows that women with eight more years of education on average have about one fewer kid ($-0.128 \times 8 \approx -1$), holding all other variables the same. The estimate on years of education is statistically very significant. Observing the year dummies coefficient, in almost all periods except for the year 1974, fertility has been declining with a negative sign. However, the year dummy variables that are significant are the year dummy for 1982 and 1984, when women had about half a child less than a similar type of woman than the base year 1972.

b. Reestimate the model in part a, but use *motheduc* and *fatheduc* as instruments for *educ*. First, check that these instruments are sufficiently partially correlated with *educ*. Test whether *educ* is in fact exogenous in the fertility equation.
Answer:
From the reduced form regression result, we can see that *educ* is very significantly partially correlated with *feduc* and *meduc*. Also, the F-test result indicates the same thing with a p-value of zero.

Reduced Form Regression Output

```
. reg educ age age2 black east northcen west farm othrural town smcity y74-y84 meduc feduc

      Source |       SS           df       MS      Number of obs   =     1,129
-------------+----------------------------------   F(18, 1110)     =     24.82
       Model |  2256.26171        18  125.347873   Prob > F        =    0.0000
    Residual |  5606.85432     1,110  5.05122011   R-squared       =    0.2869
-------------+----------------------------------   Adj R-squared   =    0.2754
       Total |  7863.11603     1,128  6.97084755   Root MSE        =    2.2475

------------------------------------------------------------------------------
        educ | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         age |  -.2243687   .2000013    -1.12   0.262     -.616792    .1680546
        age2 |   .0025664   .0022605     1.14   0.256     -.001869    .0070018
       black |   .3667819   .2522869     1.45   0.146    -.1282311     .861795
        east |   .2488042   .1920135     1.30   0.195    -.1279462    .6255546
    northcen |   .0913945   .1757744     0.52   0.603    -.2534931    .4362821
        west |   .1010676   .2422408     0.42   0.677    -.3742339    .5763691
        farm |  -.3792615   .2143864    -1.77   0.077    -.7999099    .0413869
    othrural |   -.560814   .2551196    -2.20   0.028    -1.061385    -.060243
        town |   .0616337   .1807832     0.34   0.733    -.2930816     .416349
      smcity |   .0806634   .2317387     0.35   0.728    -.3740319    .5353588
         y74 |   .0060993    .249827     0.02   0.981    -.4840872    .4962858
         y76 |   .1239104   .2587922     0.48   0.632    -.3838667    .6316874
         y78 |   .2077861   .2627738     0.79   0.429    -.3078033    .7233755
         y80 |   .3828911   .2642433     1.45   0.148    -.1355816    .9013638
         y82 |   .5820401   .2492372     2.34   0.020     .0930108    1.071069
         y84 |   .4250429   .2529006     1.68   0.093    -.0711741      .92126
       meduc |   .1723015   .0221964     7.76   0.000     .1287499    .2158531
       feduc |   .2074188   .0254604     8.15   0.000     .1574629    .2573747
       _cons |   13.63334   4.396773     3.10   0.002     5.006421    22.26027
------------------------------------------------------------------------------
```

Testing Joint Significant of *meduc* and *feduc* in the Reduced Form

```
. test meduc feduc

 ( 1)  meduc = 0
 ( 2)  feduc = 0

       F(  2,  1110) =  155.79
            Prob > F =    0.0000
```

Endogeneity Test: Predict Residuals from Reduce Form and Include in the Original Regression

```
. predict vhat, resid

. reg kids educ age age2 black east northcen west farm othrural town smcity y74-y84 vhat

      Source |       SS           df       MS      Number of obs   =     1,129
-------------+----------------------------------   F(18, 1110)     =      9.21
       Model |  400.802376        18  22.2667987   Prob > F        =    0.0000
    Residual |  2684.70692     1,110  2.41865489   R-squared       =    0.1299
-------------+----------------------------------   Adj R-squared   =    0.1158
       Total |   3085.5093     1,128  2.73538059   Root MSE        =    1.5552

------------------------------------------------------------------------------
        kids | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        educ |  -.1527395   .0392012    -3.90   0.000    -.2296562   -.0758227
         age |   .5235536   .1389568     3.77   0.000     .2509059    .7962013
        age2 |   -.005716   .0015697    -3.64   0.000    -.0087959   -.0026362
       black |   1.072952    .173618     6.18   0.000     .7322958    1.413609
        east |   .2285554   .1337787     1.71   0.088    -.0339322     .491043
```

northcen	.3744188	.1219925	3.07	0.002	.1350569	.6137807
west	.2076398	.1675628	1.24	0.216	-.1211357	.5364153
farm	-.0770015	.1512869	-0.51	0.611	-.373842	.2198389
othrural	-.1952451	.1814491	-1.08	0.282	-.5512671	.1607769
town	.08181	.1246122	0.66	0.512	-.162692	.3263119
smcity	.2124996	.160335	1.33	0.185	-.1020943	.5270936
y74	.2721292	.172847	1.57	0.116	-.0670145	.6112729
y76	-.0945483	.1791319	-0.53	0.598	-.4460236	.2569269
y78	-.0572543	.1824512	-0.31	0.754	-.4152424	.3007337
y80	-.053248	.1846139	-0.29	0.773	-.4154795	.3089836
y82	-.4962149	.1764897	-2.81	0.005	-.842506	-.1499238
y84	-.5213604	.1778207	-2.93	0.003	-.8702631	-.1724578
vhat	.0311374	.0443634	0.70	0.483	-.0559081	.1181829
_cons	-7.241244	3.134883	-2.31	0.021	-13.39221	-1.09028

Check Estimation Result from 2SLS

```
. ivreg kids age age2 black east northcen west farm othrural town smcity y74-y84 (educ= meduc feduc)
Instrumental variables 2SLS regression
```

Source	SS	df	MS			
				Number of obs	=	1,129
				F(17, 1111)	=	7.72
Model	395.36632	17	23.2568424	Prob > F	=	0.0000
Residual	2690.14298	1,111	2.42137082	R-squared	=	0.1281
				Adj R-squared	=	0.1148
Total	3085.5093	1,128	2.73538059	Root MSE	=	1.5561

kids	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
educ	-.1527395	.0392232	-3.89	0.000	-.2296993	-.0757796
age	.5235536	.1390348	3.77	0.000	.2507532	.796354
age2	-.005716	.0015705	-3.64	0.000	-.0087976	-.0026345
black	1.072952	.1737155	6.18	0.000	.732105	1.4138
east	.2285554	.1338537	1.71	0.088	-.0340792	.4911901
northcen	.3744188	.122061	3.07	0.002	.1349228	.6139148
west	.2076398	.1676568	1.24	0.216	-.1213199	.5365995
farm	-.0770015	.1513718	-0.51	0.611	-.3740083	.2200053
othrural	-.1952451	.181551	-1.08	0.282	-.5514666	.1609764
town	.08181	.1246821	0.66	0.512	-.162829	.3264489
smcity	.2124996	.160425	1.32	0.186	-.1022706	.5272698
y74	.2721292	.172944	1.57	0.116	-.0672045	.6114629
y76	-.0945483	.1792324	-0.53	0.598	-.4462205	.2571239
y78	-.0572543	.1825536	-0.31	0.754	-.415443	.3009343
y80	-.053248	.1847175	-0.29	0.773	-.4156825	.3091865
y82	-.4962149	.1765888	-2.81	0.005	-.8427	-.1497297
y84	-.5213604	.1779205	-2.93	0.003	-.8704586	-.1722623
_cons	-7.241244	3.136642	-2.31	0.021	-13.39565	-1.086834

```
Instrumented: educ
 Instruments: age age2 black east northcen west farm othrural town smcity
              y74 y76 y78 y80 y82 y84 meduc feduc
```

Then using a residual-based test of *educ* against the null that *educ* is exogenous, the p-value of the residual from the reduced form in the original regression model is around one-half, showing little evidence that *educ* is endogenous in the equation. From the 2SLS estimate, the coefficient on *educ* is larger than the one from OLS, but since we find that there is not enough evidence of endogeneity, the difference can be due to the sampling problem.

c. Now allow the effect of education to change over time by including interaction terms such as $y74 \cdot educ$, $y76 \cdot educ$, and so on in the model. Use interactions of time dummies and parents' education as instruments for the interaction terms. Test that there has been no change in the relationship between fertility and education over time.
Answer:
Since there is no strong evidence of endogeneity of *educ*, I run the full model using OLS, and test the joint significance for all *time* and *educ* interaction terms. Also, I performed regression for the full model using 2SLS to compare.

Estimation Result from OLS for the Full Model

```
. ivreg kids age age2 black east northcen west farm othrural town smcity y74 educ y74educ-y84educ
```

```
Instrumental variables 2SLS regression
```

Source	SS	df	MS			
				Number of obs	=	1,129
				F(18, 1110)	=	9.44
Model	409.673224	18	22.7596236	Prob > F	=	0.0000
Residual	2675.83608	1,110	2.41066313	R-squared	=	0.1328
				Adj R-squared	=	0.1187
Total	3085.5093	1,128	2.73538059	Root MSE	=	1.5526

kids	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
age	.510053	.1384743	3.68	0.000	.2383522	.7817539
age2	-.0055635	.001565	-3.55	0.000	-.0086342	-.0024928
black	1.06328	.1730631	6.14	0.000	.7237118	1.402847
east	.210392	.1326161	1.59	0.113	-.0498146	.4705986
northcen	.3558111	.1207703	2.95	0.003	.1188473	.5927749
west	.1844557	.166833	1.11	0.269	-.142888	.5117993
farm	-.054589	.146993	-0.37	0.710	-.3430044	.2338264
othrural	-.1670392	.1751397	-0.95	0.340	-.5106815	.176603
town	.0851755	.1244419	0.68	0.494	-.1589925	.3293434
smcity	.2153475	.1600898	1.35	0.179	-.0987653	.5294602
y74	-.200312	.6590335	-0.30	0.761	-1.493404	1.09278
educ	-.112692	.022643	-4.98	0.000	-.1571199	-.0682642
y74educ	.0344892	.0534433	0.65	0.519	-.0703721	.1393505
y76educ	-.0111314	.0143594	-0.78	0.438	-.039306	.0170433
y78educ	-.011372	.0143495	-0.79	0.428	-.0395271	.0167831
y80educ	-.0095471	.0142869	-0.67	0.504	-.0375795	.0184853
y82educ	-.0441485	.0134187	-3.29	0.001	-.0704773	-.0178196
y84educ	-.0472872	.0135507	-3.49	0.001	-.0738752	-.0206993
_cons	-7.387245	3.049727	-2.42	0.016	-13.37113	-1.403364

```
(no endogenous regressors)
```

Test Joint Significant of Interaction Variables

```
. test y74educ y76educ y78educ y80educ y82educ y84educ

 ( 1)  y74educ = 0
 ( 2)  y76educ = 0
 ( 3)  y78educ = 0
 ( 4)  y80educ = 0
 ( 5)  y82educ = 0
 ( 6)  y84educ = 0

       F(  6,  1110) =    4.11
            Prob > F =    0.0004
```

Estimation Result from 2SLS for the Full Model

```
. ivreg kids age age2 black east northcen west farm othrural town smcity y74 ///
> (educ y74educ-y84educ = meduc feduc y74meduc-y84feduc )

Instrumental variables 2SLS regression
```

Source	SS	df	MS			
				Number of obs	=	1,129
				F(18, 1110)	=	7.40
Model	406.162425	18	22.5645792	Prob > F	=	0.0000
Residual	2679.34688	1,110	2.41382601	R-squared	=	0.1316
				Adj R-squared	=	0.1176
Total	3085.5093	1,128	2.73538059	Root MSE	=	1.5536

kids	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
educ	-.1281456	.0442233	-2.90	0.004	-.2149162	-.041375
y74educ	.0392542	.1082955	0.36	0.717	-.1732328	.2517412
y76educ	-.0188478	.015178	-1.24	0.215	-.0486285	.010933
y78educ	-.0174725	.0150433	-1.16	0.246	-.0469891	.0120441
y80educ	-.0096295	.0151773	-0.63	0.526	-.0394089	.0201498
y82educ	-.0477111	.0144041	-3.31	0.001	-.0759734	-.0194487
y84educ	-.0477952	.0143867	-3.32	0.001	-.0760233	-.0195671
age	.4979232	.1398254	3.56	0.000	.2235712	.7722752
age2	-.0054354	.0015787	-3.44	0.001	-.008533	-.0023378
black	1.061408	.1737569	6.11	0.000	.7204788	1.402337
east	.2187422	.1337294	1.64	0.102	-.0436486	.4811331

```
    northcen |   .3640495   .1219907     2.98   0.003     .1246911    .6034078
        west |   .1850113   .1676614     1.10   0.270    -.1439576    .5139803
        farm |  -.0666629   .1511741    -0.44   0.659    -.3632822    .2299564
    othrural |  -.1829343   .1810601    -1.01   0.313     -.538193    .1723244
        town |   .0891165   .1246913     0.71   0.475    -.1555407    .3337737
      smcity |   .2171135   .1603366     1.35   0.176    -.0974836    .5317106
         y74 |  -.3023562   1.331661    -0.23   0.820    -2.915213    2.310501
       _cons |  -6.874234   3.170511    -2.17   0.030     -13.0951   -.6533642
------------------------------------------------------------------------------
Instrumented: educ y74educ y76educ y78educ y80educ y82educ y84educ
 Instruments: age age2 black east northcen west farm othrural town smcity
              y74 meduc feduc y74meduc y76meduc y78meduc y80meduc y82meduc
              y84meduc y74feduc y76feduc y78feduc y80feduc y82feduc y84feduc
```

Test Joint Significant of Interaction Variables

```
. test y74educ y76educ y78educ y80educ y82educ y84educ
 ( 1)  y74educ = 0
 ( 2)  y76educ = 0
 ( 3)  y78educ = 0
 ( 4)  y80educ = 0
 ( 5)  y82educ = 0
 ( 6)  y84educ = 0
       F(  6,  1110) =    3.45
            Prob > F =    0.0022
```

From the OLS model, the individual significance of the interaction terms is significant in the last two years that is 1984 and 1982 with negative coefficients. A similar result is obtained from the 2SLS estimate, with relatively close values to those from the OLS. From the joint significance of all interaction terms, there is enough evidence that there have been changes in the relationship between fertility and education over time.

Problem 6.9

Use the data in INJURY.RAW for this question.

a. Using the data for Kentucky, reestimate equation (6.54) adding as explanatory variables *male*, *married*, and a full set of industry- and injury-type dummy variables. How does the estimate on $afchnge \cdot highearn$ change when these other factors are controlled for? Is the estimate still statistically significant?
Answer:

OLS Regression Output

```
. reg ldurat afchnge highearn afhigh if ky
      Source |       SS           df       MS      Number of obs   =     5,626
-------------+----------------------------------   F(3, 5622)      =     39.54
       Model |  191.071442         3  63.6904807   Prob > F        =    0.0000
    Residual |   9055.9345     5,622  1.61080301   R-squared       =    0.0207
-------------+----------------------------------   Adj R-squared   =    0.0201
       Total |  9247.00594     5,625  1.64391217   Root MSE        =    1.2692
------------------------------------------------------------------------------
      ldurat | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
     afchnge |   .0076573   .0447173     0.17   0.864    -.0800058    .0953204
    highearn |   .2564785   .0474464     5.41   0.000     .1634652    .3494918
      afhigh |   .1906012   .0685089     2.78   0.005     .0562973    .3249051
       _cons |   1.125615   .0307368    36.62   0.000     1.065359    1.185871
------------------------------------------------------------------------------
```

OLS Regression Output with Added Explanatory Variables

```
. reg ldurat afchnge highearn afhigh male married head-construc if ky
      Source |       SS           df       MS      Number of obs   =     5,349
-------------+----------------------------------   F(14, 5334)     =     16.37
       Model |  358.441793        14  25.6029852   Prob > F        =    0.0000
    Residual |  8341.41206     5,334  1.56381928   R-squared       =    0.0412
-------------+----------------------------------   Adj R-squared   =    0.0387
             |
```

Total	8699.85385	5,348	1.62674904	Root MSE	= 1.2505

ldurat	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
afchnge	.0106274	.0449167	0.24	0.813	-.0774276	.0986824
highearn	.1757598	.0517462	3.40	0.001	.0743161	.2772035
afhigh	.2308768	.0695248	3.32	0.001	.0945798	.3671738
male	-.0979407	.0445498	-2.20	0.028	-.1852766	-.0106049
married	.1220995	.0391228	3.12	0.002	.0454027	.1987962
head	-.5139003	.1292776	-3.98	0.000	-.7673372	-.2604634
neck	.2699126	.1614899	1.67	0.095	-.0466737	.5864988
upextr	-.178539	.1011794	-1.76	0.078	-.376892	.0198141
trunk	.1264514	.1090163	1.16	0.246	-.0872651	.340168
lowback	-.0085967	.1015267	-0.08	0.933	-.2076305	.1904371
lowextr	-.1202911	.1023262	-1.18	0.240	-.3208922	.0803101
occdis	.2727118	.210769	1.29	0.196	-.1404816	.6859052
manuf	-.1606709	.0409038	-3.93	0.000	-.2408591	-.0804827
construc	.1101967	.0518063	2.13	0.033	.0086352	.2117581
_cons	1.245922	.1061677	11.74	0.000	1.03779	1.454054

Compared to without adding the new variables, the estimated coefficient on the interaction terms is now higher at 0.23 than at .19 from the previous result. It is also statistically more significant with a p-value of 0.001. Adding the new variables only slightly affects the standard error on the interaction terms.

b. What do you make of the small R-squared from part a? Does this mean the equation is useless?
Answer:
Comparing the R-squared that is quite small at 4.1% and adjusted R-square of 3.9%. It means that our model does not explain much of the variation. It means that we can not really use our model for making good predictions considering the variable that we have included so far. However, we can still get a good causal inference if the coefficient is statistically significant.

c. Estimate equation (6.54) using the data for Michigan. Compare the estimate on the interaction term for Michigan and Kentucky, as well as their statistical significance.
Answer:

OLS Regression Output

```
. reg ldurat afchnge highearn afhigh if mi
```

Source	SS	df	MS		
				Number of obs	= 1,524
				F(3, 1520)	= 6.05
Model	34.3850177	3	11.4616726	Prob > F	= 0.0004
Residual	2879.96981	1,520	1.89471698	R-squared	= 0.0118
				Adj R-squared	= 0.0098
Total	2914.35483	1,523	1.91356194	Root MSE	= 1.3765

ldurat	Coefficient	Std. err.	t	P>|t|	[95% conf. interval]	
afchnge	.0973808	.0847879	1.15	0.251	-.0689329	.2636945
highearn	.1691388	.1055676	1.60	0.109	-.0379348	.3762124
afhigh	.1919906	.1541699	1.25	0.213	-.1104176	.4943988
_cons	1.412737	.0567172	24.91	0.000	1.301485	1.523989

Comparing the estimate for Michigan and Kentucky we have similar estimates on the interaction terms consecutively .192 and .191, although the statistical significance for Michigan is lower probably due to the smaller sample size compared to those of Kentucky.

Problem 6.11

The following wage equation represents the populations of working people in 1978 and 1985:

$$\log(wage) = \beta_0 + \delta_0 y85 + \beta_1 educ + \delta_1 y85 \cdot educ + \beta_2 exper + \beta_3 exper^2 + \beta_4 union + \beta_5 female + \delta_5 y85 \cdot female + u,$$

where the explanatory variables are standard. The variable *union* is a dummy variable equal to one if the person belongs to a union and zero otherwise. The variable $y85$ is a dummy variable equal to one if the observation comes from 1985 and zero if it comes from 1978. In the file CPS78_85.RAW, there are 550 workers in the sample in 1978 and a different set of 534 people in 1985.

a. Estimate this equation and test whether the return to education has changed over the seven-year period.
   Answer:
   Table show the estimates of the regression model. The return to education increased by 1.85 percent between the years 1978 and 1985 interpreted from the coefficient on $y85 \cdot educ$, which is statistically significant at 5% level (two-sided).

   OLS Regression Output

```
. reg lwage y85 educ y85educ exper expersq union female y85fem

      Source |       SS           df       MS      Number of obs   =     1,084
-------------+----------------------------------   F(8, 1075)      =     99.80
       Model |  135.992074         8  16.9990092   Prob > F        =    0.0000
    Residual |  183.099094     1,075  .170324738   R-squared       =    0.4262
-------------+----------------------------------   Adj R-squared   =    0.4219
       Total |  319.091167     1,083   .29463635   Root MSE        =     .4127

------------------------------------------------------------------------------
       lwage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         y85 |   .1178062   .1237817     0.95   0.341     -.125075    .3606874
        educ |   .0747209   .0066764    11.19   0.000     .0616206    .0878212
     y85educ |   .0184605   .0093542     1.97   0.049      .000106     .036815
       exper |   .0295843   .0035673     8.29   0.000     .0225846     .036584
     expersq |  -.0003994   .0000775    -5.15   0.000    -.0005516   -.0002473
       union |   .2021319   .0302945     6.67   0.000     .1426888    .2615749
      female |  -.3167086   .0366215    -8.65   0.000    -.3885663    -.244851
      y85fem |    .085052    .051309     1.66   0.098    -.0156251     .185729
       _cons |   .4589329   .0934485     4.91   0.000     .2755707     .642295
------------------------------------------------------------------------------
```

b. What has happened to the gender gap over the period?
   Answer:
   The positive coefficient on $y85fem$ indicates that the estimated gender gap decreases around 8.5%. The coefficient on $y85fem$ is significant at the 10% level against the two-tail test. The gender wage difference is still large at approximately $(-31.67 + 8.5) \approx -23\%$ with women receiving lower wages.

c. Wages are measured in nominal dollars. What coefficients would change if we measure *wage* in 1978 dollars in both years? (Hint: Use the fact that for all 1985 observations, $\log(wage_i/P85) = \log(wage_i) - \log(P85)$, where $P85$ is the common deflator; $P85 = 1.65$ according to the Consumer Price Index.)
   Answer:
   If we use the 1978 dollars wage, the coefficient on $y85$ will change, mathematically

$$\beta'_{y85} = \frac{\partial(\log wage_i - \log P85)}{\partial y85} = \underbrace{\frac{\partial \log wage_i}{\partial y85}}_{\beta_{y85}} - \log P85 = .118 - \log(1.65) = -.118 - .501 \approx -.383.$$

d. Is there evidence that the variance of the error has changed over time?
   Answer:
   One idea is to estimate the variance using residual-squared. Then I regress it on the time dummy variable $y85$, if the coefficient is significant then there is some evidence that the variance error has changed over time. The coefficient is .418 and is statistically significant at 10% level. Thus, there is some evidence that the variance error has changed over time.

   OLS Regression of Residuals-Squared with Year Dummy

```
. predict u, resid
. gen u2=u^2
. reg u2 y85
```

```
      Source |       SS           df       MS      Number of obs   =     1,084
-------------+----------------------------------   F(1, 1082)      =      3.63
       Model |  .474853374         1  .474853374   Prob > F        =    0.0572
    Residual |  141.708808     1,082  .130969324   R-squared       =    0.0033
-------------+----------------------------------   Adj R-squared   =    0.0024
       Total |  142.183662     1,083  .131286853   Root MSE        =     .3619

------------------------------------------------------------------------------
          u2 | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         y85 |   .0418642   .0219861     1.90   0.057     -.001276    .0850043
       _cons |   .1482875   .0154313     9.61   0.000     .1180088    .1785662
------------------------------------------------------------------------------
```

e. With wages measured nominally, and holding other factors fixed, what is the estimated increase in nominal wage for a male with 12 years of education? Propose a regression to obtain a confidence interval for this estimate. (Hint: You must replace $y85 \cdot educ$ with something else.)
Answer:
The coefficient that we are interested are actually $\widehat{\theta} = \beta_{y85} + 12\beta_{y85educ}$. According to the introductory econometrics textbook, we can transform the regression model by replacing $y85 \cdot educ$ with $y85 \cdot (educ - 12)$ and then after performing the regression the coefficient that we are interested, $\widehat{\theta}$ will be the coefficient on $y85$. The result are showing the same results as follows.
Regression with Transformed Data to Test Linear Combination

```
. gen y85xeduc_12=y85*(educ-12)

. reg lwage y85 educ y85xeduc_12 exper expersq union female y85fem

      Source |       SS           df       MS      Number of obs   =     1,084
-------------+----------------------------------   F(8, 1075)      =     99.80
       Model |  135.992074         8  16.9990092   Prob > F        =    0.0000
    Residual |  183.099094     1,075  .170324738   R-squared       =    0.4262
-------------+----------------------------------   Adj R-squared   =    0.4219
       Total |  319.091167     1,083   .29463635   Root MSE        =     .4127

------------------------------------------------------------------------------
       lwage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         y85 |   .3393326   .0340099     9.98   0.000     .2725993    .4060659
        educ |   .0747209   .0066764    11.19   0.000     .0616206    .0878212
 y85xeduc_12 |   .0184605   .0093542     1.97   0.049      .000106     .036815
       exper |   .0295843   .0035673     8.29   0.000     .0225846     .036584
     expersq |  -.0003994   .0000775    -5.15   0.000    -.0005516   -.0002473
       union |   .2021319   .0302945     6.67   0.000     .1426888    .2615749
      female |  -.3167086   .0366215    -8.65   0.000    -.3885663    -.244851
      y85fem |    .085052    .051309     1.66   0.098    -.0156251     .185729
       _cons |   .4589329   .0934485     4.91   0.000     .2755707     .642295
------------------------------------------------------------------------------
```

Compare with Regular Stata Result

```
. quietly reg lwage y85 educ y85educ exper expersq union female y85fem

. lincom y85 +12*y85educ

 ( 1)  y85 + 12*y85educ = 0

------------------------------------------------------------------------------
       lwage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         (1) |   .3393326   .0340099     9.98   0.000     .2725993    .4060659
------------------------------------------------------------------------------
```

Chapter 7

Problem 7.2

In model (7.11), maintain Assumptions SOLS.1 and SOLS.2, and assume $B = \mathrm{E}(\mathbf{X}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{X}_i) = \mathrm{E}(\mathbf{X}_i'\boldsymbol{\Omega}\mathbf{X}_i)$, where $\boldsymbol{\Omega} \equiv \mathrm{E}(\mathbf{u}_i\mathbf{u}_i')$. (The last assumption is a different way of stating the homoskedasticity assumption for systems of equations; it always holds if assumption (7.53) holds.) Let $\widehat{\boldsymbol{\beta}}_{SOLS}$ denote the system OLS estimator.

a. Show that $\text{Avar}(\widehat{\boldsymbol{\beta}}_{SOLS}) = [\text{E}(\mathbf{X}_i'\mathbf{X}_i)]^{-1}[\text{E}(\mathbf{X}_i'\boldsymbol{\Omega}\mathbf{X}_i)][\text{E}(\mathbf{X}_i'\mathbf{X}_i)]^{-1}/N$.
   Answer:
   Recall model (7.11),

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{u}_i,$$

   where $\boldsymbol{\beta}$ is a $K \times 1$ vector, $\mathbf{X}_i$ is a $G \times K$ data matrix and $\mathbf{u}_i$ is a $G \times 1$ vector of error. Also, recall Assumptions SOLS.1: $\text{E}(\mathbf{X}_i'\mathbf{u}_i) = 0$, and SOLS.2: $\mathbf{A} = \text{E}(\mathbf{X}_i'\mathbf{X}_i)$ is nonsingular (has rank K). Under these two assumption we can write $\boldsymbol{\beta}$ as the following

$$\begin{aligned}\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{u}_i &\Leftrightarrow \mathbf{X}_i'\mathbf{y}_i = \mathbf{X}_i'\mathbf{X}_i\boldsymbol{\beta} + \mathbf{X}_i'\mathbf{u}_i \\ &\Leftrightarrow \text{E}(\mathbf{X}_i'\mathbf{y}_i) = \text{E}(\mathbf{X}_i'\mathbf{X}_i)\boldsymbol{\beta} + \text{E}(\mathbf{X}_i'\mathbf{u}_i) \\ &\Leftrightarrow \boldsymbol{\beta} = \text{E}(\mathbf{X}_i'\mathbf{X}_i)^{-1}\text{E}(\mathbf{X}_i'\mathbf{y}_i)\end{aligned}$$

   Now from the analogy principle we know that $\widehat{\boldsymbol{\beta}} = (N^{-1}\sum_{i=1}^N \mathbf{X}_i'\mathbf{X}_i)^{-1}(N^{-1}\sum_{i=1}^N \mathbf{X}_i'\mathbf{y}_i)$. Substitute the population model and with some algebraic manipulation, we can then write $\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ as the following

$$\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = (N^{-1}\sum_{i=1}^N \mathbf{X}_i'\mathbf{X}_i)^{-1}(N^{-1/2}\sum_{i=1}^N \mathbf{X}_i'\mathbf{u}_i).$$

   Since $\text{E}(\mathbf{X}_i'\mathbf{u}_i) = 0$ from CLT we have $N^{-1/2}\sum_{i=1}^N \mathbf{X}_i'\mathbf{u}_i \overset{a}{\sim} \mathbb{N}(\mathbf{0}, \mathbf{B})$. Further, we can then write $N^{-1/2}\sum_{i=1}^N \mathbf{X}_i'\mathbf{u}_i = O_p(1)$, and by LLN and Slutsky's theorem we also have $(\mathbf{XX}/N)^{-1} = \mathbf{A}^{-1} + o_p(1)$. Then we can rewrite,

$$\begin{aligned}\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) &= \mathbf{A}^{-1}(N^{-1/2}\sum_{i=1}^N \mathbf{X}_i'\mathbf{u}_i) + [(\mathbf{X}'\mathbf{X}/N)^{-1} - \mathbf{A}^{-1}](N^{-1/2}\sum_{i=1}^N \mathbf{X}_i'\mathbf{u}_i) \\ &= \mathbf{A}^{-1}(N^{-1/2}\sum_{i=1}^N \mathbf{X}_i'\mathbf{u}_i) + o_p(1)\cdot O_p(1) = \mathbf{A}^{-1}(N^{-1/2}\sum_{i=1}^N \mathbf{X}_i'\mathbf{u}_i) + o_p(1).\end{aligned}$$

   Thus we can derive now by CLT, that $\sqrt{N}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \overset{a}{\sim} \mathbb{N}(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}) = \mathbb{N}(\mathbf{0}, \mathbf{A}^{-1}\text{E}(\mathbf{X}_i'\boldsymbol{\Omega}\mathbf{X}_i)\mathbf{A}^{-1})$. We can write it similarly as $\text{Avar}(\widehat{\boldsymbol{\beta}}_{SOLS}) = [\text{E}(\mathbf{X}_i'\mathbf{X}_i)]^{-1}[\text{E}(\mathbf{X}_i'\boldsymbol{\Omega}\mathbf{X}_i)][\text{E}(\mathbf{X}_i'\mathbf{X}_i)]^{-1}/N$.

b. How would you estimate the asymptotic variance in part a?
   Answer:
   To estimate $\text{Avar}(\widehat{\boldsymbol{\beta}}_{SOLS}) = [\text{E}(\mathbf{X}_i'\mathbf{X}_i)]^{-1}[\text{E}(\mathbf{X}_i'\boldsymbol{\Omega}\mathbf{X}_i)][\text{E}(\mathbf{X}_i'\mathbf{X}_i)]^{-1}/N$, we can use analogy principle. From LLN we know that, $\widehat{\mathbf{A}} = \mathbf{X}'\mathbf{X}/N = N^{-1}\sum_{i=1}^N \mathbf{X}_i'\mathbf{X}_i \xrightarrow{p} \text{E}(\mathbf{X}_i'\mathbf{X}_i)$. Then to estimate the $\mathbf{u}_i$ we use the SOLS residuals $\widehat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i\widehat{\boldsymbol{\beta}}$ which is a $G \times 1$ vector of residuals. Then we have $\widehat{\boldsymbol{\Omega}} = N^{-1}\sum_{i=1}^N \widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i' \xrightarrow{p} \boldsymbol{\Omega}$. Then we can estimate $\text{Avar}(\widehat{\boldsymbol{\beta}}_{SOLS})$ by

$$\widehat{\mathbf{V}} = \widehat{\mathbf{A}}^{-1}\left(\sum_{i=1}^N \mathbf{X}_i'\widehat{\boldsymbol{\Omega}}\mathbf{X}_i\right)\widehat{\mathbf{A}}^{-1} = \left(\sum_{i=1}^N \mathbf{X}_i'\mathbf{X}_i\right)^{-1}\left(\sum_{i=1}^N \mathbf{X}_i'\widehat{\boldsymbol{\Omega}}\mathbf{X}_i\right)\left(\sum_{i=1}^N \mathbf{X}_i'\mathbf{X}_i\right)^{-1}.$$

c. Now add Assumptions SGLS.1-SGLS.3. Show that $\text{Avar}(\widehat{\boldsymbol{\beta}}_{SOLS}) - \text{Avar}(\widehat{\boldsymbol{\beta}}_{FGLS})$ is positive semidefinite. (Hint: Show that $[\text{Avar}(\widehat{\boldsymbol{\beta}}_{FGLS})]^{-1} - [\text{Avar}(\widehat{\boldsymbol{\beta}}_{SOLS})]^{-1}$ is p.s.d.)
   Answer:
   Recall Assumptions SGLS.1: $\text{E}(\mathbf{X}_i \otimes \mathbf{u}_i) = \mathbf{0}$, SGLS.2: $\boldsymbol{\Omega}$ is p.s.d. and $\text{E}(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i)$ is nonsingular, and SGLS.3: $\text{E}(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{u}_i\mathbf{u}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i) = \text{E}(\mathbf{X}_i'\boldsymbol{\Omega}^{-1}\mathbf{X}_i)$, where $\boldsymbol{\Omega} \equiv \text{E}(\mathbf{u}_i\mathbf{u}_i')$.
   We need to show that $\text{Avar}(\widehat{\boldsymbol{\beta}}_{SOLS}) - \text{Avar}(\widehat{\boldsymbol{\beta}}_{FGLS})$ is p.s.d which is equivalent to showing that $[\text{Avar}(\widehat{\boldsymbol{\beta}}_{FGLS})]^{-1} - [\text{Avar}(\widehat{\boldsymbol{\beta}}_{SOLS})]^{-1}$ is p.s.d. Recall that under Assumptions SGLS.1-SGLS.3 from Theorem 7.4 in textbook we have $\text{Avar}(\widehat{\boldsymbol{\beta}}_{FGLS}) = \text{E}(\mathbf{X}_i'\boldsymbol{\Omega}\mathbf{X}_i)^{-1}/N$. Disregard the N and we have

$$= \text{Avar}(\widehat{\boldsymbol{\beta}}_{FGLS})]^{-1} - [\text{Avar}(\widehat{\boldsymbol{\beta}}_{SOLS})]^{-1} = E(\mathbf{X}_i'\boldsymbol{\Omega}\mathbf{X}_i) - \text{E}(\mathbf{X}_i'\mathbf{X}_i)[\text{E}(\mathbf{X}_i'\boldsymbol{\Omega}\mathbf{X}_i)]^{-1}\text{E}(\mathbf{X}_i'\mathbf{X}_i)$$

need to be p.s.d. Now, we need some algebraic manipulation to help, make $E(\mathbf{X}_i'\mathbf{\Omega}\mathbf{X}_i) = E(\mathbf{Z}_i'\mathbf{Z}_i)$, by construction we have $\mathbf{Z}_i = \mathbf{\Omega}^{-1/2}\mathbf{X}_i$. Similarly, make $\mathrm{E}(\mathbf{X}_i'\mathbf{\Omega}\mathbf{X}_i) = \mathrm{E}(\mathbf{W}_i'\mathbf{W}_i)$, by construction we have $\mathbf{W}_i = \mathbf{\Omega}^{1/2}\mathbf{X}_i$. Now, we are left to show that

$$\mathrm{E}(\mathbf{X}_i'\mathbf{\Omega}\mathbf{X}_i) - \mathrm{E}(\mathbf{X}_i'\mathbf{X}_i)[\mathrm{E}(\mathbf{X}_i'\mathbf{\Omega}\mathbf{X}_i)]^{-1}\mathrm{E}(\mathbf{X}_i'\mathbf{X}_i) = \mathrm{E}(\mathbf{Z}_i'\mathbf{Z}_i) - \mathrm{E}(\mathbf{Z}_i'\mathbf{W}_i)\mathrm{E}(\mathbf{W}_i'\mathbf{W}_i)\mathrm{E}(\mathbf{W}_i'\mathbf{Z}_i)$$

The latter form look familiar with the matrix form when we show efficiency of 2SLS. Now define linear projection of $\mathbf{Z}_i$ on $\mathbf{W}_i$: $\mathbf{Z}_i = \mathbf{W}_i\mathbf{\Pi} + \mathbf{R}_i$, with $\mathbf{\Pi} = \mathrm{E}(\mathbf{W}_i'\mathbf{W}_i)^{-1}\mathrm{E}(\mathbf{W}_i'\mathbf{Z}_i)$, and $\mathbf{R}_i$ is $G \times K$ matrix of population residual from the projection. By algebraic manipulation we can show that

$$\mathrm{E}(\mathbf{R}_i'\mathbf{R}_i) = \mathrm{E}(\mathbf{Z}_i'\mathbf{Z}_i) - \mathrm{E}(\mathbf{Z}_i'\mathbf{W}_i)\mathrm{E}(\mathbf{W}_i'\mathbf{W}_i)\mathrm{E}(\mathbf{W}_i'\mathbf{Z}_i),$$

which is a positive semi-definite since it is a quadratic form of a matrix, with identity as the meat in the sandwich form. Thus we show that under these assumptions and the rank condition satisfied FGLS is more efficient than OLS. □

d. If, in addition to the previous assumptions, $\mathbf{\Omega} = \sigma^2\mathbf{I}_G$, show that SOLS and FGLS have the same asymptotic variance.
Answer:
Recall that

$$\begin{aligned}
\mathrm{Avar}(\widehat{\boldsymbol{\beta}}_{FGLS}) &= \mathrm{E}(\mathbf{X}_i'\mathbf{\Omega}\mathbf{X}_i)^{-1}/N, \\
\mathrm{Avar}(\widehat{\boldsymbol{\beta}}_{SOLS}) &= [\mathrm{E}(\mathbf{X}_i'\mathbf{X}_i)]^{-1}[\mathrm{E}(\mathbf{X}_i'\mathbf{\Omega}\mathbf{X}_i)][\mathrm{E}(\mathbf{X}_i'\mathbf{X}_i)]^{-1}/N,
\end{aligned}$$

and substitute the given assumption. We have

$$\begin{aligned}
\mathrm{Avar}(\widehat{\boldsymbol{\beta}}_{FGLS}) &= \mathrm{E}(\mathbf{X}_i'\mathbf{\Omega}\mathbf{X}_i)^{-1}/N = \mathrm{E}(\mathbf{X}_i'\sigma^2\mathbf{I}_G\mathbf{X}_i)^{-1}/N = \sigma^2\mathrm{E}(\mathbf{X}_i'\mathbf{X}_i)^{-1}/N, \\
\mathrm{Avar}(\widehat{\boldsymbol{\beta}}_{SOLS}) &= [\mathrm{E}(\mathbf{X}_i'\mathbf{X}_i)]^{-1}[\mathrm{E}(\mathbf{X}_i'\mathbf{\Omega}\mathbf{X}_i)][\mathrm{E}(\mathbf{X}_i'\mathbf{X}_i)]^{-1}/N \\
&= [\mathrm{E}(\mathbf{X}_i'\mathbf{X}_i)]^{-1}[\mathrm{E}(\mathbf{X}_i'\sigma^2\mathbf{I}_G\mathbf{X}_i)][\mathrm{E}(\mathbf{X}_i'\mathbf{X}_i)]^{-1}/N \\
&= \sigma^2\mathrm{E}(\mathbf{X}_i'\mathbf{X}_i)^{-1}/N.
\end{aligned}$$

We showed that they are the same. □

e. Evaluate the following statement: "Under the assumption of part c, FGLS is never asymptotically worse that SOLS, even if $\mathbf{\Omega} = \sigma^2\mathbf{I}_G$."
Answer:
The statement is true provided of what we showed in part c and part d, and provided that any other condition such as rank conditions holds.

Problem 7.7

Consider the panel data model

$$\begin{aligned}
& y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + u_{it}, && t = 1, 2, \ldots, T, \\
& \mathrm{E}(u_{it}|\mathbf{x}_{it}, u_{i,t-1}, \mathbf{x}_{i,t-1}, \ldots,) = 0, && \\
& \mathrm{E}(u_{it}^2|\mathbf{x}_{it}) = \mathrm{E}(u_{it}^2) = \sigma_t^2, && t = 1, 2, \ldots, T.
\end{aligned}$$

(Note that $\mathrm{E}(u_{it}^2|\mathbf{x}_{it})$ does not depend on $\mathbf{x}_{it}$, but it is allowed to be a different constant in each time period.)

a. Show that $\mathbf{\Omega} = \mathrm{E}(\mathbf{u}_i\mathbf{u}_i')$ is a diagonal matrix.
Answer:
From the last given condition $\mathrm{E}(u_{it}^2) = \sigma_t^2$. Now take the element with different t, $\mathrm{E}(u_{it}u_{is})$ with $s \neq t$. From the second given condition $\mathrm{E}(u_{it}|\mathbf{x}_{it}, u_{i,t-1}, \mathbf{x}_{i,t-1}, \ldots,) = 0$, then we have $\mathrm{E}(u_{it}|u_{is}) = 0$. By LIE we have $\mathrm{E}(u_{it}u_{is}) = \mathrm{E}(u_{it}u_{is}|u_{is}) = \mathrm{E}(u_{is}\mathrm{E}(u_{it}|u_{is})) = 0$ with $s \neq t$. Thus, $\mathbf{\Omega} = \mathrm{E}(\mathbf{u}_i\mathbf{u}_i')$ is a diagonal matrix. □

b. Write down the GLS estimator assuming that $\mathbf{\Omega}$ is known.

Answer:

Recall the GLS estimator from equation (7.45) in textbook but we don't need to estimate $\widehat{\mathbf{\Omega}}$ because $\mathbf{\Omega}$ is know is known. We have

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &= \left(\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{\Omega}^{-1} \mathbf{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \mathbf{X}_i' \mathbf{\Omega}^{-1} \mathbf{y}_i\right) \\
&= \left(\sum_{i=1}^{N} \mathbf{X}_i' \mathrm{E}(\mathbf{u}_i\mathbf{u}_i')^{-1} \mathbf{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \mathbf{X}_i' \mathrm{E}(\mathbf{u}_i\mathbf{u}_i')^{-1} \mathbf{y}_i\right) \\
&= \left(\sum_{i=1}^{N} \sum_{t=1}^{T} (\sigma_t^2)^{-1} \mathbf{x}_{it}' \mathbf{x}_{it}\right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{N} (\sigma_t^2)^{-1} \mathbf{x}_{it}' \mathbf{y}_{it}\right) \\
&= \left(\sum_{i=1}^{N} \sum_{t=1}^{T} {\sigma_t}^{-2} \mathbf{x}_{it}' \mathbf{x}_{it}\right)^{-1} \left(\sum_{i=1}^{N} \sum_{t=1}^{N} {\sigma_t}^{-2} \mathbf{x}_{it}' \mathbf{y}_{it}\right).
\end{aligned}$$

The σ_t^{-2}, the inverse of variance (taken from the diagonal element of $\mathbf{\Omega}$). □

c. Argue that Assumption SGLS.1 does not necessarily hold under the assumptions made. (Setting $\mathbf{x}_{it} = y_{i,t-1}$ might help in answering this part.) Nevertheless, show that the GLS estimator from part b *is* consistent for $\boldsymbol{\beta}$ by showing that $\mathrm{E}(\mathbf{X}_i'\mathbf{\Omega}^{-1}\mathbf{u}_i) = 0$. (This proof shows that Assumption SGLS.1 is sufficient, but not necessary, for consistency. Sometimes $\mathrm{E}(\mathbf{X}_i'\mathbf{\Omega}^{-1}\mathbf{u}_i) = 0$ even though Assumption SGLS.1 does not hold.)

Answer:

From the hint, if $\mathbf{x}_{it} = y_{i,t-1}$ then we have $y_{it} = f(y_{i,t-1})$ or written differently, $y_{it} = \delta_0 + \delta_1 y_{i,t-1} + u_{it}$, or said differently we will have $x_{i,t+1} = y_{it}$ is correlated with u_{it}. If this correlation exist, then SGLS.1 does not hold. However, the sufficient condition for consistency of the GLS estimator is $\mathrm{E}(\mathbf{X}_i'\mathbf{\Omega}^{-1}\mathbf{u}_i) = 0$. Since $\mathbf{\Omega}$ is known and a diagonal matrix then we have

$$\mathbf{X}_i'\mathbf{\Omega}^{-1}\mathbf{u}_i = \sum_{t=1}^{T} \mathbf{x}_{it}' {\sigma_t}^{-2} u_i t \Leftrightarrow \mathrm{E}(\mathbf{X}_i'\mathbf{\Omega}^{-1}\mathbf{u}_i) - \sum_{t=1}^{T} {\sigma_t}^{-2} \mathrm{E}(\mathbf{x}_{it}' u_{it}) = \mathbf{0}.$$

It follows from the second given assumption, $\mathrm{E}(u_{it}|\mathbf{x}_{it}, u_{i,t-1}, \mathbf{x}_{i,t-1}, \ldots,) = 0$, by LIE that, $\mathrm{E}(\mathbf{x}_{it}' u_{it}) = 0$. Thus the GLS estimator is consistent in this case without necessarily having SGLS.1 hold.

d. Show that Assumptions SGLS.3 holds under the given assumptions.

Answer:

Recall SGLS.3: $\mathrm{E}(\mathbf{X}_i'\mathbf{\Omega}^{-1}\mathbf{u}_i\mathbf{u}_i'\mathbf{\Omega}^{-1}\mathbf{X}_i) = \mathrm{E}(\mathbf{X}_i'\mathbf{\Omega}^{-1}\mathbf{X}_i)$, where $\mathbf{\Omega} \equiv \mathrm{E}(\mathbf{u}_i\mathbf{u}_i')$. Since $\mathbf{\Omega}^{-1}$ is diagonal and known, we have that $\mathbf{X}_i'\mathbf{\Omega}^{-1} = (\sigma_1^{-2}\mathbf{x}_{i1}', \ldots, \sigma_T^{-2}\mathbf{x}_{iT}')'$. Thus, we can write the following

$$\begin{aligned}
\mathrm{E}(\mathbf{X}_i'\mathbf{\Omega}^{-1}\mathbf{u}_i\mathbf{u}_i'\mathbf{\Omega}^{-1}\mathbf{X}_i) &= \mathrm{E}((\sigma_1^{-2}\mathbf{x}_{i1}', \ldots, \sigma_T^{-2}\mathbf{x}_{iT}')'\mathbf{u}_i\mathbf{u}_i'(\sigma_1^{-2}\mathbf{x}_{i1}', \ldots, \sigma_T^{-2}\mathbf{x}_{iT}')) \\
&= \sum_{t=1}^{T} \sum_{s=1}^{T} \sigma_s^{-2}\sigma_t^{-2} \mathrm{E}(u_{is}u_{it}\mathbf{x}_{is}'\mathbf{x}_{it}).
\end{aligned}$$

From the second given assumption, $\mathrm{E}(u_{it}|\mathbf{x}_{it}, u_{i,t-1}, \mathbf{x}_{i,t-1}, \ldots,) = 0$, when $s \neq t$, by LIE we have $\mathrm{E}(u_{is}u_{it}\mathbf{x}_{is}'\mathbf{x}_{it}) = \mathrm{E}(u_{it}\mathrm{E}(u_{is}\mathbf{x}_{is}'\mathbf{x}_{it}|\mathbf{x}_{it}, \mathbf{x}_{is}, u_{is})) = 0$. And then for $s = t$, for each t, also by LIE, we have

$$\mathrm{E}(u_{it}^2\mathbf{x}_{it}'\mathbf{x}_{it}) = \mathrm{E}(\mathbf{x}_{it}'\mathbf{x}_{it}\mathrm{E}(u_{it}^2|\mathbf{x}_{it})) = \sigma_t^2\mathrm{E}(\mathbf{x}_{it}'\mathbf{x}_{it}).$$

Finally, we can show that $\mathrm{E}(\mathbf{X}_i'\mathbf{\Omega}^{-1}\mathbf{u}_i\mathbf{u}_i'\mathbf{\Omega}^{-1}\mathbf{X}_i) = \sum_{t=1}^{T} \sigma_t^2 \mathrm{E}(\mathbf{x}_{it}'\mathbf{x}_{it}) = \mathrm{E}(\mathbf{X}_i'\mathbf{\Omega}^{-1}\mathbf{X}_i)$, thus SGLS.3 holds. □

e. Explain how to consistently estimate each σ_t^2 (as $N \to \infty$).
Answer:
To estimate each σ_t^2 we can run pooled OLS across all i and t and save each residual $\widehat{u}_i t$. Then, we compute the sample variance across t using $\widehat{\sigma}_t^2 = (N-K)^{-1}\sum_{i=1}^N \widehat{u}_i t^2$. In this case, we might not need to adjust for degree of freedom as $N \to \infty$. We can implement these using the foreach iteration in Stata and save the value in a new variable.

f. Argue that, under the assumptions made, valid inference is obtained by weighting each observation $(y_{it}, \mathbf{x}_{it})$ by $1/\widehat{\sigma}_t$ and then running pooled OLS.
Answer:
Recall

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^N\sum_{t=1}^T {\sigma_t}^{-2}\mathbf{x}_{it}'\mathbf{x}_{it}\right)^{-1}\left(\sum_{i=1}^N\sum_{t=1}^N {\sigma_t}^{-2}\mathbf{x}_{it}'\mathbf{y}_{it}\right)$$

$$\Rightarrow \sqrt{N}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) = \left(N^{-1}\sum_{i=1}^N\sum_{t=1}^T {\sigma_t}^{-2}\mathbf{x}_{it}'\mathbf{x}_{it}\right)^{-1}\left(N^{-1/2}\sum_{i=1}^N\sum_{t=1}^N {\sigma_t}^{-2}\mathbf{x}_{it}'u_{it}\right) + o_p(1).$$

To have the same inference, we need to show that if $\widehat{\sigma}_t^2 \xrightarrow{p} \sigma_t^2$, thus transforming the data by weighting it by $1/\widehat{\sigma}_t$ will not change the asymptotic variance of the GLS. For the first term, we can use the consistency of sample variance estimation for each t, we have

$$\begin{aligned}\widehat{\sigma}_t^2 \xrightarrow{p} \sigma_t^2,\ \forall t \Leftrightarrow \sum_{i=1}^N\sum_{t=1}^T \widehat{\sigma}_t^{-2}\mathbf{x}_{it}'\mathbf{x}_{it} &\xrightarrow{p} \sum_{i=1}^N\sum_{t=1}^T \sigma_t^{-2}\mathbf{x}_{it}'\mathbf{x}_{it} \\ \Leftrightarrow \sum_{i=1}^N\sum_{t=1}^T \widehat{\sigma}_t^{-2}\mathbf{x}_{it}'\mathbf{x}_{it} &= \sum_{i=1}^N\sum_{t=1}^T \sigma_t^{-2}\mathbf{x}_{it}'\mathbf{x}_{it} + o_p(1).\end{aligned}$$

Now consider the second term in the distribution, from Slutsky's theorem, we have $\widehat{\sigma}_t^{-2} \xrightarrow{p} \sigma_t^{-2}$. Also from CLT, we have $N^{-1/2}\sum_{i=1}^N \mathbf{x}_{it}'u_{it} = O_p(1)$. We have

$$\begin{aligned}N^{-1/2}\sum_{i=1}^N\sum_{t=1}^N \widehat{\sigma}_t^{-2}\mathbf{x}_{it}'u_{it} - N^{-1/2}\sum_{i=1}^N\sum_{t=1}^N {\sigma_t}^{-2}\mathbf{x}_{it}'u_{it} &= \sum_{t=1}^T\left(N^{-1/2}\sum_{i=1}^N \mathbf{x}_{it}'u_{it}\right)(\widehat{\sigma}_t^{-2}-\sigma_t^{-2}) \\ &= O_p(1)\cdot \o_p(1) = o_p(1).\end{aligned}$$

Finally we will have

$$\begin{aligned}\sqrt{N}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}) &= \left(N^{-1}\sum_{i=1}^N\sum_{t=1}^T {\sigma_t}^{-2}\mathbf{x}_{it}'\mathbf{x}_{it}\right)^{-1}\left(N^{-1/2}\sum_{i=1}^N\sum_{t=1}^N {\sigma_t}^{-2}\mathbf{x}_{it}'u_{it}\right) + o_p(1) \\ &= \left(N^{-1}\sum_{i=1}^N\sum_{t=1}^T {\widehat{\sigma}_t}^{-2}\mathbf{x}_{it}'\mathbf{x}_{it} - o_p(1)\right)^{-1}\left(N^{-1/2}\sum_{i=1}^N\sum_{t=1}^N \widehat{\sigma}_t^{-2}\mathbf{x}_{it}'u_{it} - o_p(1)\right) + o_p(1) \\ &= \left(N^{-1}\sum_{i=1}^N\sum_{t=1}^T {\widehat{\sigma}_t}^{-2}\mathbf{x}_{it}'\mathbf{x}_{it}\right)^{-1}\left(N^{-1/2}\sum_{i=1}^N\sum_{t=1}^N \widehat{\sigma}_t^{-2}\mathbf{x}_{it}'u_{it}\right) + o_p(1).\end{aligned}$$

Now we showed that by transforming the data $(y_{it}, \mathbf{x}_{it})$ to $(y_{it}/\widehat{\sigma}_t, \mathbf{x}_{it}/\widehat{\sigma}_t)$ have the same asymptotic distribution. Note that the way we get the residuals should be different and refer to answer in point e. □

g. What happen if we assume that $\sigma_t^2 = \sigma^2$ for all $t = 1, \ldots, T$?
Answer:
If we assume $\sigma_t^2 = \sigma^2$ for all $t = 1, \ldots, T$ then we can use the standard OLS regression pooled across i and t because now the variance is independently distributed across i and t.

Chapter 8

Problem 8.1

a. Show that GMM estimator that solves the problem (8.27) satisfies the first order-condition

$$\left(\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{X}_i\right)'\widehat{\mathbf{W}}\left(\sum_{i=1}^{N}\mathbf{Z}_i'(\mathbf{y}_i-\mathbf{X}_i\widehat{\boldsymbol{\beta}})\right)=0$$

Answer:
Recall the minimization problem

$$\min_{\mathbf{b}} Q(\mathbf{b})=\min_{\mathbf{b}}\left[\sum_{i=1}^{N}\mathbf{Z}_i'(\mathbf{y}_i-\mathbf{X}_i\mathbf{b})\right]'\widehat{\mathbf{W}}\left[\sum_{i=1}^{N}\mathbf{Z}_i'(\mathbf{y}_i-\mathbf{X}_i\mathbf{b})\right].$$

Since $\widehat{\boldsymbol{\beta}}=\arg\min_{\mathbf{b}} Q(\mathbf{b})$, take the first order condition with respect to $\mathbf{b}$, we have

$$\frac{\partial Q(\mathbf{b})'}{\partial \mathbf{b}}=-2\left(\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{X}_i\right)'\widehat{\mathbf{W}}\left(\sum_{i=1}^{N}\mathbf{Z}_i'(\mathbf{y}_i-\mathbf{X}_i\mathbf{b})\right)=\mathbf{0}.$$

At the solution of the F.O.C, we get $\widehat{\boldsymbol{\beta}}$ by solving

$$\left(\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{X}_i\right)'\widehat{\mathbf{W}}\left(\sum_{i=1}^{N}\mathbf{Z}_i'(\mathbf{y}_i-\mathbf{X}_i\widehat{\boldsymbol{\beta}})\right)=\mathbf{0}.$$

b. Use this expression to obtain (8.28)
Answer:
We can write result from a in full matrix

$$\begin{aligned}
&\left(\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{X}_i\right)'\widehat{\mathbf{W}}\left(\sum_{i=1}^{N}\mathbf{Z}_i'(\mathbf{y}_i-\mathbf{X}_i\widehat{\boldsymbol{\beta}})\right)=\mathbf{0}\\
&\Leftrightarrow (\mathbf{X}'\mathbf{Z})\widehat{\mathbf{W}}(\mathbf{Z}'\mathbf{Y}-\mathbf{Z}'\mathbf{X}\widehat{\boldsymbol{\beta}})=\mathbf{0}\\
&\Leftrightarrow (\mathbf{X}'\mathbf{Z}\widehat{\mathbf{W}}\mathbf{Z}'\mathbf{Y})=(\mathbf{X}'\mathbf{Z}\widehat{\mathbf{W}}\mathbf{Z}'\mathbf{X})\widehat{\boldsymbol{\beta}}\\
&\Leftrightarrow \widehat{\boldsymbol{\beta}}=(\mathbf{X}'\mathbf{Z}\widehat{\mathbf{W}}\mathbf{Z}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Z}\widehat{\mathbf{W}}\mathbf{Z}'\mathbf{Y}).
\end{aligned}$$

□

Problem 8.5

Verify that the difference $(\mathbf{C}'\boldsymbol{\Lambda}^{-1}\mathbf{C})-(\mathbf{C}'\mathbf{W}\mathbf{C})(\mathbf{C}'\mathbf{W}\boldsymbol{\Lambda}\mathbf{W}\mathbf{C})\mathbf{C})^{-1}(\mathbf{C}'\mathbf{W}\mathbf{C})$ in expression (8.34) is positive semidefinite for any symmetric positive definite matrices $\mathbf{W}$ and $\boldsymbol{\Lambda}$. (Hint: Show that the difference can be expressed as $\mathbf{C}'\boldsymbol{\Lambda}^{-1/2}[\mathbf{I}_L-\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}']\boldsymbol{\Lambda}^{-1/2}\mathbf{C}$ where $\mathbf{D}\equiv\boldsymbol{\Lambda}^{1/2}\mathbf{W}\mathbf{C}$. Then, note that for any $L\times K$ matrix $\mathbf{D},\mathbf{I}_L-\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$ is a symmetric, idempotent matrix, and therefore positive semidefinite.)
Answer:
Following the hint, let $\mathbf{D}=\boldsymbol{\Lambda}^{1/2}\mathbf{W}\mathbf{C}$, then we have

$$\begin{aligned}
\mathbf{C}'\boldsymbol{\Lambda}^{-1/2}[\mathbf{I}_L-\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}']\boldsymbol{\Lambda}^{-1/2}\mathbf{C}&=\mathbf{C}'\boldsymbol{\Lambda}^{-1/2}[\mathbf{I}_L-(\boldsymbol{\Lambda}^{1/2}\mathbf{W}\mathbf{C})((\boldsymbol{\Lambda}^{1/2}\mathbf{W}\mathbf{C})'(\boldsymbol{\Lambda}^{1/2}\mathbf{W}\mathbf{C}))^{-1}(\boldsymbol{\Lambda}^{1/2}\mathbf{W}\mathbf{C})']\boldsymbol{\Lambda}^{-1/2}\mathbf{C}\\
&=(\mathbf{C}'\boldsymbol{\Lambda}^{-1}\mathbf{C})\\
&\qquad-\mathbf{C}'\boldsymbol{\Lambda}^{-1/2}(\boldsymbol{\Lambda}^{1/2}\mathbf{W}\mathbf{C})((\boldsymbol{\Lambda}^{1/2}\mathbf{W}\mathbf{C})'(\boldsymbol{\Lambda}^{1/2}\mathbf{W}\mathbf{C}))^{-1}(\boldsymbol{\Lambda}^{1/2}\mathbf{W}\mathbf{C})'\boldsymbol{\Lambda}^{-1/2}\mathbf{C}\\
&=(\mathbf{C}'\boldsymbol{\Lambda}^{-1}\mathbf{C})\\
&\qquad-(\mathbf{C}'\mathbf{W}\mathbf{C})((\mathbf{C}'\mathbf{W}\boldsymbol{\Lambda}^{1/2\prime})(\boldsymbol{\Lambda}^{1/2\prime}\mathbf{W}\mathbf{C}))^{-1}(\boldsymbol{\Lambda}^{1/2}(\mathbf{C}'\mathbf{W}\mathbf{C})\\
&=(\mathbf{C}'\boldsymbol{\Lambda}^{-1}\mathbf{C})-(\mathbf{C}'\mathbf{W}\mathbf{C})(\mathbf{C}'\mathbf{W}\boldsymbol{\Lambda}\mathbf{W}\mathbf{C})\mathbf{C})^{-1}(\mathbf{C}'\mathbf{W}\mathbf{C}).
\end{aligned}$$

It turns out it is true. Then since $\mathbf{C}'\mathbf{\Lambda}^{-1/2}[\mathbf{I}_L - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}']\mathbf{\Lambda}^{-1/2}\mathbf{C}$ is a matrix quadratic form, it is p.s.d. if the meat matrix in the sandwich form is p.s.d. We know that $\mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$ is a projection matrix, call $P_{\mathbf{D}}$, then $(\mathbf{I}_L - P_{\mathbf{D}})$ is an idempotent matrix so it must be p.s.d and we showed that the difference is p.s.d. $\square$

Chapter 9

Problem 9.8

a. Extend Problem 5.4b using CARD.RAW to allow $educ^2$ to appear in the $\log(wage)$ equation, without using $nearc2$ as an instrument. Specifically, use interactions of $nearc4$ with some or all of the other exogenous variables in the $\log(wage)$ equation as instruments for $educ^2$. Compute a heteroskedasticity-robust test to be sure that at least one of these additional instruments appears in the linear projection of $educ^2$ onto your entire list of instruments. Test whether $educ^2$ needs to be in the $\log(wage)$ equation.
Answer:

Generate Interaction Variables

```
. gen educ2=educ^2
. gen nearc4exper=nearc4*exper
. gen nearc4expersq=nearc4*expersq
. gen nearc4black=nearc4*black
```

Reduced Form Estimates for Extension of Problem 5.4b

```
. reg educ2 exper expersq black south smsa reg661-reg668 smsa66 nearc4 ///
> nearc4exper nearc4expersq nearc4black, robust
Linear regression                                Number of obs     =      3,010
                                                 F(18, 2991)       =     233.34
                                                 Prob > F          =     0.0000
                                                 R-squared         =     0.4505
                                                 Root MSE          =     52.172
```

educ2	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
exper	-18.01791	1.229128	-14.66	0.000	-20.42793	-15.60789
expersq	.3700966	.058167	6.36	0.000	.2560452	.4841479
black	-21.04009	3.569591	-5.89	0.000	-28.03919	-14.04098
south	-.5738389	3.973465	-0.14	0.885	-8.36484	7.217162
smsa	10.38892	3.036816	3.42	0.001	4.434463	16.34338
reg661	-6.175308	5.574484	-1.11	0.268	-17.10552	4.754903
reg662	-6.092379	4.254714	-1.43	0.152	-14.43484	2.250083
reg663	-6.193772	4.010618	-1.54	0.123	-14.05762	1.670077
reg664	-3.413348	5.069994	-0.67	0.501	-13.35438	6.527681
reg665	-12.31649	5.439968	-2.26	0.024	-22.98295	-1.650031
reg666	-13.27102	5.693005	-2.33	0.020	-24.43362	-2.10842
reg667	-10.83381	5.814901	-1.86	0.063	-22.23542	.567801
reg668	8.427749	6.627727	1.27	0.204	-4.567616	21.42312
smsa66	-.4621454	3.058084	-0.15	0.880	-6.458307	5.534016
nearc4	-12.25914	7.012394	-1.75	0.081	-26.00874	1.490464
nearc4exper	4.192304	1.55785	2.69	0.007	1.137738	7.24687
nearc4expersq	-.1623635	.0753242	-2.16	0.031	-.310056	-.014671
nearc4black	-4.789202	4.247869	-1.13	0.260	-13.11824	3.53984
_cons	307.212	6.617862	46.42	0.000	294.2359	320.188

Test Joint Significant of Instrument

```
. test nearc4exper nearc4expersq nearc4black
 ( 1)  nearc4exper = 0
 ( 2)  nearc4expersq = 0
 ( 3)  nearc4black = 0
       F(  3,  2991) =    3.72
            Prob > F =    0.0110
```

2SLS Regression Result

```
. ivreg lwage exper expersq black south smsa reg661-reg668 smsa66 ///
> (educ educ2 = nearc4 nearc4exper nearc4expersq nearc4black)

Instrumental variables 2SLS regression

      Source |       SS           df       MS      Number of obs   =     3,010
-------------+----------------------------------   F(16, 2993)     =     45.92
       Model |  116.731381        16  7.29571132   Prob > F        =    0.0000
    Residual |  475.910264     2,993  .159007773   R-squared       =    0.1970
-------------+----------------------------------   Adj R-squared   =    0.1927
       Total |  592.641645     3,009  .196956346   Root MSE        =    .39876

------------------------------------------------------------------------------
       lwage | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
        educ |   .3161298   .1457578     2.17   0.030     .0303342    .6019254
       educ2 |  -.0066592   .0058401    -1.14   0.254    -.0181103    .0047918
       exper |   .0840117   .0361077     2.33   0.020     .0132132    .1548101
     expersq |  -.0007825   .0014221    -0.55   0.582    -.0035709    .0020058
       black |  -.1360751   .0455727    -2.99   0.003    -.2254322   -.0467181
       south |   -.141488   .0279775    -5.06   0.000    -.1963451   -.0866308
        smsa |   .1072011   .0290324     3.69   0.000     .0502755    .1641267
      reg661 |  -.1098848   .0428194    -2.57   0.010    -.1938432   -.0259264
      reg662 |   .0036271   .0325364     0.11   0.911    -.0601688    .0674231
      reg663 |   .0428246   .0315082     1.36   0.174    -.0189554    .1046045
      reg664 |  -.0639842   .0391843    -1.63   0.103    -.1408151    .0128468
      reg665 |   .0480365   .0445934     1.08   0.281    -.0394003    .1354734
      reg666 |   .0672512   .0498043     1.35   0.177    -.0304028    .1649052
      reg667 |   .0347783   .0471451     0.74   0.461    -.0576617    .1272183
      reg668 |  -.1933844   .0512395    -3.77   0.000    -.2938526   -.0929161
      smsa66 |   .0089666   .0222745     0.40   0.687    -.0347083    .0526414
       _cons |   2.610889   .9706341     2.69   0.007     .7077116    4.514067
------------------------------------------------------------------------------
Instrumented: educ educ2
 Instruments: exper expersq black south smsa reg661 reg662 reg663 reg664
              reg665 reg666 reg667 reg668 smsa66 nearc4 nearc4exper
              nearc4expersq nearc4black
```

After performing heteroskedasticity-robust Wald test for the joint significance of the instrument, it can be reported that the three interaction terms are partially correlated with $educ^2$. The p-value is .011. In the 2SLS estimates the coefficient of $educ^2$ is not significant, thus we can leave it out of the equation.

b. Start again with the model estimated in Problem 5.4b, but suppose we add the interaction $black \cdot educ$. Explain why $black \cdot z_j$ is a potential IV for $black \cdot educ$, where z_j is any exogenous variable in the system (including $nearc4$).
Answer:
Reduced Form from 5.4b

```
. gen blackeduc=black*educ

. reg educ exper expersq black south smsa reg661-reg668 smsa66 nearc4 blackeduc

      Source |       SS           df       MS      Number of obs   =     3,010
-------------+----------------------------------   F(16, 2993)     =    225.46
       Model |   11784.607        16  736.537939   Prob > F        =    0.0000
    Residual |  9777.47304     2,993  3.26678017   R-squared       =    0.5465
-------------+----------------------------------   Adj R-squared   =    0.5441
       Total |  21562.0801     3,009  7.16586243   Root MSE        =    1.8074

------------------------------------------------------------------------------
        educ | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       exper |   -.448666   .0314333   -14.27   0.000     -.510299   -.3870329
     expersq |   .0059278   .0015552     3.81   0.000     .0028784    .0089773
       black |  -8.299265   .3548985   -23.38   0.000    -8.995135   -7.603395
       south |   .0826884   .1262945     0.65   0.513    -.1649443    .3303212
        smsa |   .2728048   .0978085     2.79   0.005     .0810262    .4645835
      reg661 |  -.3028809   .1886186    -1.61   0.108    -.6727162    .0669544
      reg662 |  -.2851939   .1372326    -2.08   0.038    -.5542737    -.016114
      reg663 |  -.3059376   .1328891    -2.30   0.021    -.5665007   -.0453744
      reg664 |  -.1897754   .1732839    -1.10   0.274    -.5295429    .1499922
      reg665 |  -.6319416   .1754165    -3.60   0.000    -.9758906   -.2879925
      reg666 |  -.6838073   .1954178    -3.50   0.000    -1.066974   -.3006405
      reg667 |  -.6105922   .1917077    -3.19   0.001    -.9864845   -.2346999
```

reg668	.2442232	.2251193	1.08	0.278	-.1971811	.6856274
smsa66	-.0099628	.0985277	-0.10	0.919	-.2031517	.1832261
nearc4	.2459321	.0819096	3.00	0.003	.0853273	.4065369
blackeduc	.6077667	.0283914	21.41	0.000	.5520979	.6634354
_cons	16.91173	.1966621	85.99	0.000	16.52613	17.29734

```
. reg educ exper expersq black south smsa reg661-reg668 smsa66 nearc4 nearc2 blackeduc
```

Source	SS	df	MS		
				Number of obs =	3,010
				F(17, 2992) =	212.73
Model	11799.6365	17	694.096263	Prob > F =	0.0000
Residual	9762.44359	2,992	3.26284879	R-squared =	0.5472
				Adj R-squared =	0.5447
Total	21562.0801	3,009	7.16586243	Root MSE =	1.8063

educ	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
exper	-.448436	.0314146	-14.27	0.000	-.5100323	-.3868396
expersq	.0059122	.0015543	3.80	0.000	.0028646	.0089599
black	-8.326544	.3549125	-23.46	0.000	-9.022441	-7.630647
south	.0951701	.1263524	0.75	0.451	-.1525762	.3429164
smsa	.2715175	.0977514	2.78	0.006	.0798507	.4631843
reg661	-.2508726	.1900563	-1.32	0.187	-.6235269	.1217816
reg662	-.2601786	.1376444	-1.89	0.059	-.5300659	.0097087
reg663	-.2456871	.1357437	-1.81	0.070	-.5118474	.0204733
reg664	-.1203049	.1761786	-0.68	0.495	-.4657484	.2251386
reg665	-.5746888	.1773289	-3.24	0.001	-.9223876	-.2269899
reg666	-.6704957	.1953986	-3.43	0.001	-1.053625	-.2873665
reg667	-.5486189	.1937561	-2.83	0.005	-.9285276	-.1687102
reg668	.3301183	.2285158	1.44	0.149	-.1179456	.7781822
smsa66	-.0419958	.0995931	-0.42	0.673	-.2372738	.1532822
nearc4	.2466392	.081861	3.01	0.003	.0861297	.4071487
nearc2	.1547519	.0721046	2.15	0.032	.0133723	.2961316
blackeduc	.6090166	.0283803	21.46	0.000	.5533697	.6646636
_cons	16.81692	.2014477	83.48	0.000	16.42193	17.21191

If the error term in the reduced form is independent of z_j, then any function of $\mathbf{z}$ will not be correlated with the error of the population model as well, including any interaction taking the form of $black \cdot z_j$. Thus, if any z_j is correlated with *educ* it is very likely that it is correlated with $black \cdot z_j$ as well. Thus it will be a good potential IV.

c. In Example 6.2 we used $black \cdot nearc4$ as the IV for $black \cdot educ$. Now use 2SLS with $black \cdot \widehat{educ}$ as the IV for $black \cdot educ$, where $\widehat{educ}$ are the fitted values from the first-stage regression of *educ* on all exogenous variables (including *nearc4*). What do you find?
Answer:
2SLS Regression Result

```
. ivreg lwage exper expersq black south smsa reg661-reg668 smsa66 (educ blackeduc = nearc4 nearc4black), robust
```

Instrumental variables 2SLS regression

Number of obs =	3,010
F(16, 2993) =	52.35
Prob > F =	0.0000
R-squared =	0.2435
Root MSE =	.38702

lwage	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
educ	.1273557	.0561622	2.27	0.023	.0172352	.2374762
blackeduc	.0109036	.0399278	0.27	0.785	-.0673851	.0891923
exper	.1059116	.0249463	4.25	0.000	.0569979	.1548253
expersq	-.0022406	.0004902	-4.57	0.000	-.0032017	-.0012794
black	-.282765	.5012131	-0.56	0.573	-1.265522	.6999922
south	-.1424762	.0298942	-4.77	0.000	-.2010914	-.083861
smsa	.1111555	.0310592	3.58	0.000	.050256	.1720551
reg661	-.1103479	.0418554	-2.64	0.008	-.1924161	-.0282797
reg662	-.0081783	.0339196	-0.24	0.809	-.0746863	.0583298
reg663	.0382413	.0335008	1.14	0.254	-.0274456	.1039283
reg664	-.0600379	.0398032	-1.51	0.132	-.1380824	.0180066

reg665	.0337805	.0519109	0.65	0.515	-.0680042	.1355652
reg666	.0498975	.0559569	0.89	0.373	-.0598204	.1596155
reg667	.0216942	.0528376	0.41	0.681	-.0819075	.1252959
reg668	-.1908353	.0506182	-3.77	0.000	-.2900853	-.0915853
smsa66	.0180009	.0205709	0.88	0.382	-.0223337	.0583356
_cons	3.84499	.9545666	4.03	0.000	1.973317	5.716663

```
Instrumented: educ blackeduc
 Instruments: exper expersq black south smsa reg661 reg662 reg663 reg664
              reg665 reg666 reg667 reg668 smsa66 nearc4 nearc4black
```

Generate Fitted Value for educ from First Stage Regression

```
. reg educ exper expersq black south smsa reg661-reg668 smsa66 nearc4, robust

Linear regression                               Number of obs     =      3,010
                                                F(15, 2994)       =     244.92
                                                Prob > F          =     0.0000
                                                R-squared         =     0.4771
                                                Root MSE          =     1.9405
```

educ	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
exper	-.4125334	.0320751	-12.86	0.000	-.4754249	-.3496418
expersq	.0008686	.0017076	0.51	0.611	-.0024795	.0042167
black	-.9355287	.0925281	-10.11	0.000	-1.116954	-.7541037
south	-.0516126	.1419604	-0.36	0.716	-.3299623	.2267371
smsa	.4021825	.1112278	3.62	0.000	.1840918	.6202731
reg661	-.210271	.1993703	-1.05	0.292	-.6011876	.1806456
reg662	-.2889073	.1513383	-1.91	0.056	-.5856449	.0078302
reg663	-.2382099	.1431446	-1.66	0.096	-.5188817	.0424618
reg664	-.093089	.1799452	-0.52	0.605	-.4459179	.2597398
reg665	-.4828875	.1950961	-2.48	0.013	-.8654234	-.1003516
reg666	-.5130857	.2090161	-2.45	0.014	-.9229154	-.103256
reg667	-.4270887	.2110058	-2.02	0.043	-.8408198	-.0133576
reg668	.3136204	.2337813	1.34	0.180	-.144768	.7720087
smsa66	.0254805	.1106315	0.23	0.818	-.1914409	.2424019
nearc4	.3198989	.0850763	3.76	0.000	.153085	.4867128
_cons	16.84852	.1865621	90.31	0.000	16.48272	17.21433

```
. predict double educhat
(option xb assumed; fitted values)

. gen blackeduchat=black*educhat
```

2SLS Regression Result

```
. ivreg lwage exper expersq black south smsa reg661-reg668 smsa66 (educ blackeduc = nearc4 blackeduchat), robust

Instrumental variables 2SLS regression          Number of obs     =      3,010
                                                F(16, 2993)       =      52.52
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2501
                                                Root MSE          =     .38535
```

lwage	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
educ	.1178141	.0554036	2.13	0.034	.0091811	.226447
blackeduc	.035984	.0105707	3.40	0.001	.0152573	.0567106
exper	.1004843	.0241951	4.15	0.000	.0530436	.147925
expersq	-.0020235	.0003597	-5.63	0.000	-.0027288	-.0013183
black	-.5955669	.1587782	-3.75	0.000	-.9068923	-.2842415
south	-.1374265	.0294259	-4.67	0.000	-.1951236	-.0797294
smsa	.1096541	.0306748	3.57	0.000	.0495083	.1697998
reg661	-.1161759	.0409317	-2.84	0.005	-.196433	-.0359189
reg662	-.0107817	.0335743	-0.32	0.748	-.0766127	.0550494
reg663	.0331736	.0326007	1.02	0.309	-.0307484	.0970955
reg664	-.064916	.0388398	-1.67	0.095	-.1410715	.0112395
reg665	.023022	.0505787	0.46	0.649	-.0761506	.1221946
reg666	.0379568	.0534653	0.71	0.478	-.0668757	.1427892
reg667	.0100466	.0513629	0.20	0.845	-.0906637	.1107568

reg668	-.1907066	.0502527	-3.79	0.000	-.2892399	-.0921733
smsa66	.0167814	.0203639	0.82	0.410	-.0231472	.0567101
_cons	4.00836	.9416251	4.26	0.000	2.162062	5.854658

```
Instrumented: educ blackeduc
 Instruments: exper expersq black south smsa reg661 reg662 reg663 reg664
              reg665 reg666 reg667 reg668 smsa66 nearc4 blackeduchat
```

Comparing the standard error for the two procedures, it turns out that using $black \cdot \widehat{educ}$ as the IV yields a smaller standard error than using $black \cdot nearc4$.

d. If $\mathrm{E}(educ|\mathbf{z})$ is linear and $\mathrm{Var}(u_1|\mathbf{z}) = \sigma_1^2$, where $\mathbf{z}$ is the set of all exogenous variables and u_1 is the error in the $\log(wage)$ equation, explain why the estimator using $black \cdot \widehat{educ}$ as the IV is asymptotically more efficient than the estimator using $black \cdot nearc4$ as the IV.
Answer:
Suppose $\mathrm{E}(educ|\mathbf{z})$ is linear, write as $\mathrm{E}(educ|\mathbf{z}) = \mathbf{z}\boldsymbol{\pi}_2$. We also have $\mathrm{Var}(u_1|\mathbf{z}) = \sigma_1^2$. We can find the optimal choice of instruments from Theorem 8.5 in the textbook. The theorem says if we have $\mathrm{E}(u_{ig}|\mathbf{w}_i) = 0,\ g = 1, \ldots, G$ for some vector $\mathbf{w}_i$, suppose $\boldsymbol{\Omega}(\mathbf{w}_i) = \mathrm{E}(\mathbf{u}_i'\mathbf{u}_i|\mathbf{w}_i)$ and that rank $\mathrm{E}(\mathbf{Z}_i^{*\prime}\mathbf{X}_i) = K$, then the optimal instruments is $\mathbf{Z}_i^* = \boldsymbol{\Omega}(\mathbf{w}_i)^{-1}\mathrm{E}(\mathbf{X}_i|\mathbf{w}_i)$. Applying the theorem we have the optimal IV for $educ$ and $black \cdot educ$ are the following

$$\mathbf{Z}^*_{educ} = \mathrm{E}(\mathbf{u}_1'\mathbf{u}_1|\mathbf{z})^{-1}\mathrm{E}(educ|\mathbf{z}) = \mathrm{Var}(u_1|\mathbf{z})^{-1}\mathbf{z}\boldsymbol{\pi}_2 = \sigma_1^{-2}\mathbf{z}\boldsymbol{\pi}_2.$$
$$\mathbf{Z}^*_{black \cdot educ} = \mathrm{E}(\mathbf{u}_1'\mathbf{u}_1|\mathbf{z})^{-1}\mathrm{E}(black \cdot educ|\mathbf{z}) = \mathrm{E}(\mathbf{u}_1'\mathbf{u}_1|\mathbf{z})^{-1} black \cdot \mathrm{E}(educ|\mathbf{z}) = \sigma_1^{-2} black \cdot \mathbf{z}\boldsymbol{\pi}_2.$$

Now since the variance is constant it is just a scalar for the remaining term. We can take the optimal IV as $\mathbf{z}\boldsymbol{\pi}_2$ and $black \cdot \mathbf{z}\boldsymbol{\pi}_2$. Note that if we get $\mathbf{z}\boldsymbol{\pi}_2$ from the linear projection of $educ$ on $\mathbf{z}$ and implement using $\widehat{\boldsymbol{\pi}}_2$.It turns out that $\mathbf{z}\widehat{\boldsymbol{\pi}}_2$ is the fitted value of $educ$ i.e. $\widehat{educ}$ from the regressing $educ$ on all exogenous variable $\mathbf{z}$. Thus our optimal instrument now are $\mathbf{z}$, $\widehat{educ}$, and $black \cdot \widehat{educ}$. Thus, using 2SLS with these instruments will give us an asymptotically more efficient estimator than any other option of IV.