

```

1 //Question 5
2 clear all
3 cd "C:/Users/yw3982/Downloads"
4 use AJR2001.dta
5
6 //part a)
7 reg loggdp risk
8 **
9
10      Source |           SS           df           MS      Number of obs   =           64
11 -----+-----
12      Model |    36.2163143             1    36.2163143      F(1, 62)           =           68.17
13      Residual |    32.9382286           62     .531261752      Prob > F            =           0.0000
14 -----+-----
15      Total |    69.1545429           63     1.09769116      R-squared           =           0.5237
16                                     Adj R-squared        =           0.5160
17                                     Root MSE           =           .72888
18
19      loggdp |           Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
20 -----+-----
21      risk |     .516187     .0625186     8.26   0.000     .3912141     .6411599
22      _cons |     4.687415     .417441    11.23   0.000     3.852962     5.521867
23
24 **
25
26 reg risk logmort0
27 **
28      Source |           SS           df           MS      Number of obs   =           64
29 -----+-----
30      Model |    37.1754537             1    37.1754537      F(1, 62)           =           23.34
31      Residual |    98.7466676           62     1.59268819      Prob > F            =           0.0000
32 -----+-----
33      Total |   135.922121           63     2.15749399      R-squared           =           0.2735
34                                     Adj R-squared        =           0.2618
35                                     Root MSE           =           1.262
36
37      risk |           Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
38 -----+-----
39      logmort0 |   -.6132892     .1269412    -4.83   0.000    -.8670411    -.3595374
40      _cons |     9.365895     .6105941    15.34   0.000     8.145335    10.58646
41
42
43 **
44
45 ivregress 2sls loggdp (risk= logmort0)
46 **
47 Instrumental variables (2SLS) regression      Number of obs   =           64
48                                           Wald chi2(1)     =           36.60
49                                           Prob > chi2       =           0.0000
50                                           R-squared         =           0.1880
51                                           Root MSE         =           .93672
52
53
54      loggdp |           Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
55 -----+-----
56      risk |     .9294897     .1536318     6.05   0.000     .6283768     1.230603
57      _cons |     1.994296     1.007904     1.98   0.048     .0188405     3.969751
58
59 Instrumented:  risk
60 Instruments:  logmort0
61
62
63 **
64 ** Both OLS regression and 2SLS regression have point estimates that are different by 0.01
65 from the the reported value.
66
67
68 //part b)
69 **To find homoskedastic and heteroskedastic standard error, I run both robust and

```

non-robust regression below.

```
regress loggdp risk,vce(robust)
```

**

Linear regression

Number of obs	=	64
F(1, 62)	=	102.04
Prob > F	=	0.0000
R-squared	=	0.5237
Root MSE	=	.72888

		Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
loggdp							
risk		.516187	.051101	10.10	0.000	.4140376	.6183364
_cons		4.687415	.3244741	14.45	0.000	4.0388	5.336029

**

```
regress loggdp risk
```

**

Source	SS	df	MS	Number of obs	=	64
Model	36.2163143	1	36.2163143	F(1, 62)	=	68.17
Residual	32.9382286	62	.531261752	Prob > F	=	0.0000
Total	69.1545429	63	1.09769116	R-squared	=	0.5237
				Adj R-squared	=	0.5160
				Root MSE	=	.72888

		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
loggdp							
risk		.516187	.0625186	8.26	0.000	.3912141	.6411599
cons		4.687415	.417441	11.23	0.000	3.852962	5.521867

**

```
regress risk logmort0,vce(robust)
```

**

Linear regression

Number of obs	=	64
F(1, 62)	=	16.33
Prob > F	=	0.0001
R-squared	=	0.2735
Root MSE	=	1.262

		Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
risk							
logmort0		-.6132892	.1517849	-4.04	0.000	-.916703	-.3098755
_cons		9.365895	.7084137	13.22	0.000	7.949796	10.78199

**

```
regress risk logmort0
```

**

Source	SS	df	MS	Number of obs	=	64
Model	37.1754537	1	37.1754537	F(1, 62)	=	23.34
Residual	98.7466676	62	1.59268819	Prob > F	=	0.0000
Total	135.922121	63	2.15749399	R-squared	=	0.2735
				Adj R-squared	=	0.2618
				Root MSE	=	1.262

```

139      risk |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
140 -----+-----
141      logmort0 |   -.6132892   .1269412    -4.83   0.000    -1.8670411   -1.3595374
142      _cons |    9.365895   .6105941    15.34   0.000     8.145335    10.58646
143 -----+-----

```

```

144
145 **
146 ** For 2sls regression, I run three regressions with robust, unadjusted, and cluster
    respectively.

```

```

147 ivregress 2sls loggdp (risk = logmort0),vce(robust)

```

```

148 **
149
150 Instrumental variables (2SLS) regression          Number of obs   =           64
151                                                  Wald   chi2(1)   =           29.86
152                                                  Prob > chi2      =           0.0000
153                                                  R-squared       =           0.1880
154                                                  Root MSE      =           .93672
155

```

```

156 -----+-----
157      loggdp |      Coef.   Robust Std. Err.      z    P>|z|     [95% Conf. Interval]
158 -----+-----
159      risk |   .9294897   .1700872     5.46   0.000     .5961249     1.262855
160      _cons |   1.994296   1.134985     1.76   0.079    -1.2302345     4.218826
161 -----+-----

```

```

162 Instrumented:  risk
163 Instruments:   logmort0

```

```

164 **
165
166 ivregress 2sls loggdp (risk = logmort0), vce(cluster risk)
167 **

```

```

168
169 Instrumental variables (2SLS) regression          Number of obs   =           64
170                                                  Wald   chi2(1)   =           30.61
171                                                  Prob > chi2      =           0.0000
172                                                  R-squared       =           0.1880
173                                                  Root MSE      =           .93672
174

```

(Std. Err. adjusted for 48 clusters in risk)

```

175 -----+-----
176      loggdp |      Coef.   Robust Std. Err.      z    P>|z|     [95% Conf. Interval]
177 -----+-----
178      risk |   .9294897   .1680151     5.53   0.000     .6001861     1.258793
179      _cons |   1.994296   1.13931     1.75   0.080    -1.2387104     4.227302
180 -----+-----

```

```

181 Instrumented:  risk
182 Instruments:   logmort0

```

```

183 **
184
185 ivregress 2sls loggdp (risk = logmort0),vce(unadjusted)
186 **

```

```

187
188 Instrumental variables (2SLS) regression          Number of obs   =           64
189                                                  Wald   chi2(1)   =           36.60
190                                                  Prob > chi2      =           0.0000
191                                                  R-squared       =           0.1880
192                                                  Root MSE      =           .93672
193

```

```

194 -----+-----
195      loggdp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
196 -----+-----
197      risk |   .9294897   .1536318     6.05   0.000     .6283768     1.230603
198      _cons |   1.994296   1.007904     1.98   0.048     .0188405     3.969751
199 -----+-----

```

```

200 Instrumented:  risk
201 Instruments:   logmort0

```

```

202 **
203

```

```

208  ** According to above stata output, the homoskedastic standard errors for all kinds of
    regression are closer to the reported output of the paper. Therefore, the authors use
    homoskedastic standard errors.
209
210  //part c
211  ** First get the predicted risk
212  regress risk logmort0
213  **
214      Source |           SS           df           MS       Number of obs   =           64
215  -----+-----
216      Model |   37.1754537           1   37.1754537       F(1, 62)           =          23.34
217      Residual |   98.7466676          62   1.59268819       Prob > F            =          0.0000
218  -----+-----
219      Total |  135.922121          63   2.15749399       R-squared           =          0.2735
220                                     Adj R-squared        =          0.2618
221                                     Root MSE            =          1.262
222
223      risk |           Coef.      Std. Err.      t    P>|t|     [95% Conf. Interval]
224  -----+-----
225      logmort0 |   -.6132892     .1269412    -4.83   0.000    -.8670411   -.3595374
226      _cons |    9.365895     .6105941    15.34   0.000     8.145335   10.58646
227
228  **
229  predict risk_hat
230
231  **Second, get the predicted loggdp based on logmort0
232  regress loggdp logmort0
233  **
234      Source |           SS           df           MS       Number of obs   =           64
235  -----+-----
236      Model |   32.1177728           1   32.1177728       F(1, 62)           =          53.77
237      Residual |   37.0367701          62   .59736726       Prob > F            =          0.0000
238  -----+-----
239      Total |   69.1545429          63   1.09769116       R-squared           =          0.4644
240                                     Adj R-squared        =          0.4558
241                                     Root MSE            =          .7729
242
243      loggdp |           Coef.      Std. Err.      t    P>|t|     [95% Conf. Interval]
244  -----+-----
245      logmort0 |   -.570046     .0777424    -7.33   0.000    -.7254509   -.4146412
246      _cons |   10.6998     .3739452    28.61   0.000     9.952292   11.4473
247
248  **
249  predict loggdp_hat
250
251
252
253  ** Third, run the regression of predicted loggdp on predicted risk
254  regress loggdp_hat risk_hat
255  **
256
257      Source |           SS           df           MS       Number of obs   =           64
258  -----+-----
259      Model |   32.117774           1   32.117774       F(1, 62)           > 99999.00
260      Residual |  3.2990e-12          62   5.3210e-14       Prob > F            =          0.0000
261  -----+-----
262      Total |   32.117774          63   .509805937       R-squared           =          1.0000
263                                     Adj R-squared        =          1.0000
264                                     Root MSE            =          2.3e-07
265
266      loggdp_hat |           Coef.      Std. Err.      t    P>|t|     [95% Conf. Interval]
267  -----+-----
268      risk_hat |   .9294897     3.78e-08    2.5e+07   0.000     .9294897     .9294898
269      _cons |   1.994295     2.48e-07    8.0e+06   0.000     1.994295     1.994296
270
271  **
272  ** The coefficients are the same. However, the standard error is different from the 2sls
    regression output.
273
274

```

```

275
276 //part d
277 ** the regression output of first stage and second stage done separately is the following:
278 regress risk logmort0
279 **
280
281 Source |      SS      df      MS      Number of obs      =      64
282 -----+-----
283 Model |  37.1754537      1  37.1754537      Prob > F      =      0.0000
284 Residual |  98.7466676     62  1.59268819      R-squared     =      0.2735
285 -----+-----
286 Total | 135.922121     63  2.15749399      Adj R-squared =      0.2618
287                               Root MSE      =      1.262
288
289 risk |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
290 -----+-----
291 logmort0 |  -.6132892   .1269412    -4.83   0.000   -1.8670411   -.3595374
292 _cons |   9.365895   .6105941   15.34   0.000    8.145335   10.58646
293 -----+-----
294
295 **
296 predict risk_hat
297 regress loggdp risk_hat
298 **
299 Source |      SS      df      MS      Number of obs      =      64
300 -----+-----
301 Model |  32.1177716      1  32.1177716      Prob > F      =      0.0000
302 Residual |  37.0367713     62  .597367279      R-squared     =      0.4644
303 -----+-----
304 Total |  69.1545429     63  1.09769116      Adj R-squared =      0.4558
305                               Root MSE      =      .7729
306
307 loggdp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
308 -----+-----
309 risk_hat |   .9294897   .126763     7.33   0.000    .676094    1.182885
310 _cons |   1.994296   .8316306     2.40   0.020    .3318896    3.656701
311 -----+-----
312
313 **
314 ** The coefficients are the same; however, the standard errors are different from 2sls
315 regression output.
316
317 //part e)
318 ** run the first stage regression and estimate the residual
319 regress risk logmort0
320 **
321 Source |      SS      df      MS      Number of obs      =      64
322 -----+-----
323 Model |  37.1754537      1  37.1754537      Prob > F      =      0.0000
324 Residual |  98.7466676     62  1.59268819      R-squared     =      0.2735
325 -----+-----
326 Total | 135.922121     63  2.15749399      Adj R-squared =      0.2618
327                               Root MSE      =      1.262
328
329 risk |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
330 -----+-----
331 logmort0 |  -.6132892   .1269412    -4.83   0.000   -1.8670411   -.3595374
332 _cons |   9.365895   .6105941   15.34   0.000    8.145335   10.58646
333 -----+-----
334
335 **
336 predict e_hat, residual
337
338 ** Now run the regression with the residual
339 regress loggdp risk e_hat
340 **
341 Source |      SS      df      MS      Number of obs      =      64
342 -----+-----
343 Model |  44.9573006      2  22.4786503      Prob > F      =      0.0000
344 Residual |  24.1972423     61  .396676104      R-squared     =      0.6501
345 -----+-----
346                               Adj R-squared =      0.6386

```

```

344      Total |    69.1545429      63  1.09769116  Root MSE      =    .62982
345
346

```

```

347      -----+-----
348      loggdp |      Coef.      Std. Err.      t    P>|t|      [95% Conf. Interval]
349      -----+-----
350      risk   |    .9294897    .1032975     9.00  0.000    .7229335    1.136046
351      e_hat  |   -.5689     .1211919    -4.69  0.000   -.8112382   -.3265617
352      _cons  |    1.994296    .6776848     2.94  0.005    .6391811    3.34941
353      -----+-----

```

```

354  **
355  ** The estimated coefficient is the same as the 2SLS estimate. However, the standard error
356  is different.
357

```

```

358  //part f)
359  regress loggdp risk latitude africa
360  **

```

```

361
362      Source |      SS      df      MS      Number of obs      =      64
363      -----+-----
364      Model  |  45.8984283      3  15.2994761      F(3, 60)      =      39.47
365      Residual |  23.2561146     60   .387601911      Prob > F      =      0.0000
366      -----+-----
367      Total  |  69.1545429     63  1.09769116      R-squared     =      0.6637
368                                     Adj R-squared  =      0.6469
369                                     Root MSE     =      .62258

```

```

370      -----+-----
371      loggdp |      Coef.      Std. Err.      t    P>|t|      [95% Conf. Interval]
372      -----+-----
373      risk   |    .3765228    .0608421     6.19  0.000    .2548205    .4982252
374      latitude |    1.382463    .6440401     2.15  0.036    .0941905    2.670735
375      africa  |   -.7232696    .1712967    -4.22  0.000   -1.065914   -.3806251
376      _cons  |    5.652234    .4152858    13.61  0.000    4.821539    6.482929
377      -----+-----

```

```

378  **
379
380  ** The p-values for latitude and africa are both smaller than 0.05. Therefore, they are
381  statistically significant. To be safe, run the F-statistics

```

```

382
383  test latitude africa
384  **
385  ( 1)  latitude = 0
386  ( 2)  africa = 0
387
388      F( 2, 60) = 10.84
389      Prob > F = 0.0001
390

```

```

391  **
392  di "The critical value is " invFtail(2,60,0.05)
393  **The critical value is 3.1504113
394  ** The F statistics is larger than the critical value. Therefore, they are also jointly
395  significant. Now, let me compare the R square before and after adding these two variables.
396  regress loggdp risk
397  **

```

```

397      Source |      SS      df      MS      Number of obs      =      64
398      -----+-----
399      Model  |  36.2163143      1  36.2163143      F(1, 62)      =      68.17
400      Residual |  32.9382286     62   .531261752      Prob > F      =      0.0000
401      -----+-----
402      Total  |  69.1545429     63  1.09769116      R-squared     =      0.5237
403                                     Adj R-squared  =      0.5160
404                                     Root MSE     =      .72888

```

```

405      -----+-----
406      loggdp |      Coef.      Std. Err.      t    P>|t|      [95% Conf. Interval]
407      -----+-----
408      risk   |    .516187    .0625186     8.26  0.000    .3912141    .6411599
409      cons   |    4.687415    .417441    11.23  0.000    3.852962    5.521867
410      -----+-----

```

```

411 **
412
413 ** The stata output shows that both R-squared and adjusted R-squared increase by over 0.1
    after adding latitude and africa. Therefore, the model with these two control variables is
    doing a better job at explaining the variation of loggdp. Therefore, latitude and africa
    are predictive of the level of GDP.
414
415 //part g)
416 ** The 2sls regression output with control variables is the following:
417 ivregress 2sls loggdp (risk= logmort0) latitude africa
418 **
419
420 Instrumental variables (2SLS) regression                Number of obs   =           64
421                                                         Wald chi2(3)     =           57.55
422                                                         Prob > chi2      =           0.0000
423                                                         R-squared       =           0.3922
424                                                         Root MSE      =           .81039
425
426 -----
427      loggdp |           Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
428 -----+-----
429      risk   |    .799968     .2497417     3.20   0.001     .3104831    1.289453
430  latitude |   -.0553109    1.161701    -0.05   0.962    -2.332203    2.221581
431  africa   |   -.3479258    .3062581    -1.14   0.256    -.9481806    .252329
432   _cons   |    2.99507     1.581523     1.89   0.058    -.1046587    6.094798
433 -----
434 Instrumented:  risk
435 Instruments:   latitude africa logmort0
436
437 **
438 ** In this case, both africa and latitude have the p values that are larger than 0.05.
    Therefore, they are not statistically significant.
439
440 test latitude africa
441
442   ( 1)  latitude = 0
443   ( 2)  africa = 0
444
445           chi2( 2) =      1.61
446           Prob > chi2 =      0.4471
447
448 ** The p value is still larger than 0.05. Therefore, latitude and africa are not jointly
    significant as well. When we run the OLS regression, these two regressors are both
    individually and jointly significant. However, when we run the 2sls regression, Neither are
    individually and jointly significant.
449
450 //part h)
451 ** First, I need to generate the mortality
452 gen mort0 = exp(logmort0)
453 ** the reduced form of risk is presented as:
454 regress risk mort0
455 **
456      Source |           SS           df           MS       Number of obs   =           64
457 -----+-----
458      Model |    8.69746685           1    8.69746685       F(1, 62)         =           4.24
459      Residual |   127.224655           62    2.05201056       Prob > F          =           0.0437
460 -----+-----
461      Total |   135.922121           63    2.15749399       R-squared         =           0.0640
462                                     Adj R-squared      =           0.0489
463                                     Root MSE         =           1.4325
464
465 -----
466      risk |           Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
467 -----+-----
468      mort0 |   -.0007862     .0003819    -2.06   0.044    -.0015495    -.0000228
469   _cons   |    6.709419     .2021894    33.18   0.000     6.305248    7.11359
470 -----
471 **
472 ** The authors prefer logmort0 over mort0 because the logmort0 is more economically and
    statistically significant than mort0. The coefficient of logmort0 deviates more from 0 than
    the coefficient of mort0 does. The p value of logmort0 is 0 which is much smaller than that

```

```
for mort0.
```

```
//part i)
```

```
gen logmort02 = logmort0^2
```

```
regress risk logmort0 logmort02
```

```
**
```

Source	SS	df	MS	Number of obs	=	64
Model	51.1839123	2	25.5919562	F(2, 61)	=	18.42
Residual	84.738209	61	1.38915097	Prob > F	=	0.0000
Total	135.922121	63	2.15749399	R-squared	=	0.3766
				Adj R-squared	=	0.3561
				Root MSE	=	1.1786

risk	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
logmort0	-2.645684	.6508988	-4.06	0.000	-3.947237 -1.344132
logmort02	.2101141	.066166	3.18	0.002	.077807 .3424211
_cons	13.94859	1.551696	8.99	0.000	10.84579 17.0514

```
**
** The square of logmort0 is statistically significant. Therefore, the relationship between
risk and logmort0 is nonlinear. Since the regression incorporates quadratic form, we can
intepret the marginal effect of the logmort0.
```

```
** We need to run the regression again using specific quadratic form notation so that stata
can identify it when estimating marginal effect.
```

```
regress risk logmort0 c.logmort0#c.logmort0
```

```
**
```

Source	SS	df	MS	Number of obs	=	64
Model	51.1839124	2	25.5919562	F(2, 61)	=	18.42
Residual	84.738209	61	1.38915097	Prob > F	=	0.0000
Total	135.922121	63	2.15749399	R-squared	=	0.3766
				Adj R-squared	=	0.3561
				Root MSE	=	1.1786

risk	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
logmort0	-2.645684	.6508988	-4.06	0.000	-3.947237 -1.344132
c.logmort0#c.logmort0	.2101141	.066166	3.18	0.002	.077807 .3424211
_cons	13.94859	1.551696	8.99	0.000	10.84579 17.0514

```
sum logmort0,detail
```

```
**
```

```
Original Log Settler Mortality
```

Percentiles	Smallest		
1%	2.145931	2.145931	
5%	2.70805	2.145931	
10%	2.791165	2.701361	Obs 64
25%	4.232656	2.70805	Sum of Wgt. 64
50%	4.35863		Mean 4.646749
			Std. Dev. 1.252543
75%	5.480639	6.504288	
90%	6.284209	7.293018	Variance 1.568863
95%	6.504288	7.602901	Skewness .3052336
99%	7.986165	7.986165	Kurtosis 3.171949

```
** Now take 5 percentile, 25 percentile, 50 percentile, 75 percentile, 95 percentile ,and
99 percentile to estimate their conditional marginal effect.
```

```
margins, dydx(logmort0) at(logmort0 = (2.70805 4.232656 4.35863 5.480639 6.504288 7.986165))
```

```
**
```

Conditional marginal effects	Number of obs	=	64
------------------------------	---------------	---	----

```
Model VCE : OLS
```



```

537
538 Expression   : Linear prediction, predict()
539 dy/dx w.r.t. : logmort0
540
541 1._at        : logmort0          =      2.70805
542
543 2._at        : logmort0          =      4.232656
544
545 3._at        : logmort0          =      4.35863
546
547 4._at        : logmort0          =      5.480639
548
549 5._at        : logmort0          =      6.504288
550
551 6._at        : logmort0          =      7.986165
552
553 -----
554 |                               Delta-method
555 |               dy/dx   Std. Err.      t    P>|t|     [95% Conf. Interval]
556 -----+-----
557 logmort0 |
558   _at |
559     1 |   -1.507685   .3055837   -4.93   0.000   -2.118738   -.8966331
560     2 |   -.8670031   .1429619   -6.06   0.000   -1.152873   -.5811331
561     3 |   -.8140652   .1343585   -6.06   0.000   -1.082732   -.5453987
562     4 |   -.3425655   .146023   -2.35   0.022   -.6345566   -.0505745
563     5 |    .0876007   .2505381    0.35   0.728   -.4133812    .5885825
564     6 |    .710327   .4333453    1.64   0.106   -.1562005    1.576855
565 -----
566 **
567 ** This output suggests that the marginal effect is increasing. At first, risk decreases
when logmort0 rises. Laster, risk rises as logmort0 rises. The curve relating to
relationship between logmort0 and risk should be convex.
568 ** The output of the 2sls regression with risk instrumented by both logmort0 and logmort02
is the following:
569 ivregress 2sls loggdp (risk = logmort0 logmort02)
570 **
571 Instrumental variables (2SLS) regression               Number of obs   =           64
572                                                         Wald chi2(1)     =           46.68
573                                                         Prob > chi2       =           0.0000
574                                                         R-squared         =           0.3948
575                                                         Root MSE         =           .80865
576
577 -----
578 loggdp |               Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
579 -----+-----
580   risk |   .7722554   .1130303    6.83   0.000    .55072    .9937908
581   _cons |   3.018849   .7434204    4.06   0.000    1.561772    4.475926
582 -----
583 Instrumented:   risk
584 Instruments:    logmort0 logmort02
585
586 **
587 ** The coefficient of risk in the 2sls regression with risk instrumented only by logmort0
is .9294897 while the risk coefficient falls to .7722554 after risk is instrumented by
both logmort0 and its square. The constant rises from 1.994296 to 3.018849 as the square of
logmort0 is introduced to instrumental variables in the 2sls regression. While there is no
change in p value of risk before and after logmort02 is added to instrumental variable, the
p value of the constant drops from 0.048 to 0.
588
589 // part j)
590 ** The result of the endogeneity test is the following:
591 estat endogenous
592 **
593 Tests of endogeneity
594 Ho: variables are exogenous
595
596 Durbin (score) chi2(1)          =   10.4601   (p = 0.0012)
597 Wu-Hausman F(1,61)             =   11.9176   (p = 0.0010)
598 **

```

```

599  ** Since p value of the Hausman test is smaller than 0.05, we reject the null hypothesis
600  that variables are exogeneous at 5% significance level. Therefore, variables are
601  endogeneous.
602
603  // part k)
604
605  ** Do the first stage F test:
606  reg risk logmort0 logmort02
607  **
608
609  -----+-----
610  Source |          SS      df      MS      Number of obs   =          64
611  -----+-----
612  Model |  51.1839123      2    25.5919562   F(2, 61)         =         18.42
613  Residual |  84.738209     61    1.38915097   Prob > F          =         0.0000
614  Total |  135.922121     63    2.15749399   R-squared         =         0.3766
615  -----+-----
616  Adj R-squared   =         0.3561
617  Root MSE       =         1.1786
618
619  -----+-----
620  risk |          Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
621  -----+-----
622  logmort0 | -2.645684   .6508988    -4.06   0.000   -3.947237   -1.344132
623  logmort02 |  .2101141   .066166     3.18   0.002    .077807    .3424211
624  _cons |  13.94859   1.551696     8.99   0.000   10.84579    17.0514
625  -----+-----
626
627  **
628  test logmort0 logmort02
629  **
630  ( 1)  logmort0 = 0
631  ( 2)  logmort02 = 0
632
633  F( 2, 61) = 18.42
634  Prob > F = 0.0000
635
636  **
637  ** F-statistics is greater than 10 (18.42>10). Instrumental variables therefore are relevant.
638
639  //part l)
640
641  ** Do the J-test:
642  ivregress 2sls loggdp (risk = logmort0 logmort02)
643  **
644  Instrumental variables (2SLS) regression
645  Number of obs   =          64
646  Wald chi2(1)    =          46.68
647  Prob > chi2      =         0.0000
648  R-squared       =         0.3948
649  Root MSE       =         .80865
650
651  -----+-----
652  loggdp |          Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
653  -----+-----
654  risk |  .7722554   .1130303     6.83   0.000    .55072    .9937908
655  _cons |  3.018849   .7434204     4.06   0.000   1.561772   4.475926
656  -----+-----
657  Instrumented:  risk
658  Instruments:   logmort0 logmort02
659
660  **
661  drop e_hat
662  predict e_hat, residual
663  regress e_hat logmort0 logmort02
664  **
665
666  Source |          SS      df      MS      Number of obs   =          64
667  -----+-----
668  Model |  3.35810086      2    1.67905043   F(2, 61)         =         2.66
669  Residual |  38.4926828     61    .631027586   Prob > F          =         0.0780
670  Total |  41.8507836     63    .664298153   R-squared         =         0.0802
671  -----+-----
672  Adj R-squared   =         0.0501
673  Root MSE       =         .79437
674  -----+-----

```

```

667           e_hat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
668 -----+-----
669      logmort0 |      .7516194   .4386953     1.71   0.092    - .125606   1.628845
670      logmort02 |     -.0876735   .0445948    -1.97   0.054    - .1768463   .0014993
671      _cons |     -1.464118   1.045818    -1.40   0.167    -3.555361   .6271237
672 -----+-----
673
674 **
675
676 test logmort0 logmort02
677 **
678   ( 1)  logmort0 = 0
679   ( 2)  logmort02 = 0
680
681       F(  2,    61) =    2.66
682       Prob > F =    0.0780
683
684 **
685 di "The J-test = 2 * 2.66 = " 2* 2.66
686 **The J-test = 2 * 2.66 = 5.32
687
688 ** In this case, l = 2, k =1, so l-k = 1
689
690 di "The critical value of chi square at 1 degree of freedom at 10% significance level is "
691 invchi2tail(1,0.01)
692 **The critical value of chi square at 1 degree of freedom at 10% significance level is
693 2.7055435
694 ** 5.32 > 2.7055435. we reject the null hypothesis that IVs are exogeneous at 10%
695 significance level.
696
697 di "The critial value of chi square at 1 degree of freedom at 5% significance level is "
698 invchi2tail(1,0.05)
699 **The critical value of chi square at 1 degree of freedom at 5% significace level is 3.8414588
700 ** 5.32 > 3.8414588. We reject the null hypothesis that IVs are exogeneous (all
701 coefficients of IVs are zero) at 5% significance level.
702
703 di "The critical value of chi square at 1 degree of freedom at 1% significance level is "
704 invchi2tail(1,0.01)
705 **The critical value of chi square at 1 degree of freedom at 1% significance level is
706 6.6348966
707 ** 5.32 < 6.6348966. We cannot reject the null hypothesis that the IVs are exogeneous at 1%
708 significance level.
709
710 ** In conclusion, the instruments are exogeneous at 1% significance level; however, the
711 instrumental variables are not exogeneous at 5% and 10% significance level.
712
713

```