Evaluating IV Assumptions

More general Stata commands

Empirical Economics

Instrumental Variables Regression (Stata Seminar 4)



Teacher: Andrew Proctor andrew.proctor@phdstudent.hhs.se

October 6, 2017

Outline

 Instrumental Variables Regression Basics of IV/2SLS

2 Evaluating IV Assumptions Instrument Relevance Instrument Validity

3 More general Stata commands Presenting your results Miscellaneous

Instrumental Variables Regression •00000000	Evaluating IV Assumptions 00 00	More general Stata commands 000 0
Basics of IV/2SLS		

The endogeneity problem and IV solution

• Suppose we want to estimate:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

- But we know that x_i is endogenous (that is, Cov(x_i, u_i) ≠ 0) and we can't reasonably find control variables to remedy this problem. What can we do?
- One possibility is to look for an 'instrument' variable z_i that only affects our outcome y_i through it's effect on x_i . So that:

 z_i is a *relevant* instrument: $Cov(z_i, x_i) \neq 0$ ()

 z_i is a valid instrument (exogenous): $Cov(z_i, u_i) = 0$

Instrumental Variables Regression 00000000	Evaluating IV Assumptions	More general Stata commands 000 0
Basics of IV/2SLS		

The Instrumental Variables equations

• Our resulting model is then:

 $x_i = \pi_0 + \pi_1 z_i + v_i$ (first stage) $y_i = \beta_0 + \beta_1 x_i + u_i$ (structural equation)

• Another eq. of interest is the the relationship of y_i with z_i .

 $y_i = \gamma_0 + \gamma_1 z_i + \epsilon_i$ (reduced form)

• How do we estimate our parameter of interest (β₁) using these equations and our assumptions about the instrument?

Instrumental Variables Regression 00000000	Evaluating IV Assumptions 00 00	More general Stata commands 000 0
Basics of IV/2SLS		

Deriving the IV estimator

• With a single instrument, we have:

$$Cov(y_i, z_i) = Cov(\beta_0 + \beta_1 x_i + u_i, z_i) = \beta_1 Cov(x_i, z_i)$$
$$\implies \beta_1 = \frac{Cov(y_i, z_i)}{Cov(x_i, z_i)}$$

• Furthermore note, using the usual OLS formulas:

$$\gamma_1 = Cov(y_i, z_i)/V(z_i)$$

$$\pi_1 = Cov(x_i, z_i)/V(z_i)$$

• Hence we have:

$$\beta_1 = \frac{Cov(y_i, z_i)}{Cov(x_i, z_i)} = \frac{Cov(y_i, z_i)/V(z_i)}{Cov(x_i, z_i)/V(z_i)} = \frac{\gamma_1}{\pi_1}$$

Instrumental Variables Regression 00000000	Evaluating IV Assumptions 00 00	More general Stata commands 000 0
Basics of IV/2SLS		

Generalizing IV regression

- We began by assuming a model with the same number of instruments as explanatory variables (e.g. 1). This is called the *just-identified* case.
- Sometimes, however, you may have more potential instruments than endogenous variables (*over-identified case*).
 - The *just identified* IV regression is just a special case of **two-stage least squares**, which estimate effects by using *at least as many* instruments as there are endogenous regressors.
 - The first stage with k instruments case appears as:

 $x_i = \pi_0 + \pi_1 z_i + \pi_2 z_{2,i} + \dots + \pi_k z_{k,i} + v_i$

- Deriving the 2SLS Estimates (basic idea):
 - 2SLS is essentially derived by replacing x_i in the structural equation with its fitted values from the 1^{st} stage, then performing OLS (taking into account that \hat{x}_i is a statistic when estimating variance).

Instrumental Variables Regression 00000000	Evaluating IV Assumptions oo oo	More general Stata commands 000 0
Basics of IV/2SLS		

Including other exogenous variables in 2SLS

• Typically, you will probably want to include other variables in your model besides the endogenous variable(s) you are instrumenting. That is:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 w_i + u_i,$$

where x_i is an endogenous variable, and w_i is exogenous, but not an explicit instrument for x_i .

• All exogenous variables in the structural equation should also always appear in the first-stage.

$$x_i = \pi_0 + \pi_1 z_i + \pi_2 w_i + v_i$$

Instrumental Variables Regression 000000000	Evaluating IV Assumptions 00 00	More general Stata commands 000 0
Basics of IV/2SLS		

Including other exogenous variables in 2SLS ctd

- In econometrics, W_i are referred to as the "included instruments" (because they are included in the structural equation) and Z_i are referred to as the "excluded instruments" (because they do not appear in the structural equation).
- The requirement remains the same that you need as many excluded instruments as there are endogenous explanatory variables.

Instrumental Variables Regression	Evaluating IV Assumptions	More general Stata commands 000 0
Parise of IV/2818		

Performing IV regression using the -ivreg2- package

- There are multiple regression commands/packages to perform IV/2SLS in Stata, but the most comprehensive is **ivreg2**.
- Since this package was not originally written by Stata Corp (but instead was contributed by the Stata user community), you may need to install it.
 - To install a package, in the package enter ssc install [NAME OF PACKAGE]

Instrumental Variables Regression	Evaluating IV Assumptions 00 00	More general Stata commands 000 0
Pasies of IV/2818		

Performing IV regression using the -ivreg2- package ctd

• Basic syntax of ivreg2:

ivreg2 [DEPVAR] [EXOGENOUS VARS] ([ENDOG VAR] = [EXCL. INSTRUMENTS])

- Common options for ivreg2:
 - You can specify heteroskedastic-robust or clustered standard errors, in the same manner as **regress** or **xtregress**.
 - Include first-stage results by indicating first.
 - Include reduced-form results by indicating rf.
 - Omit results for identification tests by indicating noid.

Instrumental	Variables	Regression
00000000		

Evaluating IV Assumptions

Nore general Stata commands

. ivreg2 IMTLYNEE X22* ACE 0.AGEQ (ADEQ = CAEEQ C TR1* QTR22* QT Naring - collimenties detected Yaring - collimenties XE29 QTR129 QTR329 QTR329 QTR329 IV (SLS) estimation								1	
Warning - collinearities detected Ver dropped: Statistics consistent for homoskedasticity only Statistics consistent for homoskedasticity only <tr< td=""><td></td><td>ivreg2 LW</td><td>KLYWGE YR2*</td><td>AGEQ c.A</td><td>GEQ#c.AG</td><td>Q (EDUC = QTR</td><td>1* QTR22* Q1</td><td></td><td></td></tr<>		ivreg2 LW	KLYWGE YR2*	AGEQ c.A	GEQ#c.AG	Q (EDUC = QTR	1* QTR22* Q1		
Vers dropped: YES QTRISS QTRISS QTRISS IV (1515) estimation Estimates efficient for homoskedasticity only Statistics consistent for homoskedasticity only Statistics consistent for homoskedasticity only Total (contered) 55 = 104853.0198 Contal (uncentered) 55 = 648731.774 Uncentered 25 = 648731.774 Total (uncentered) 55 = 93812.46542 Residual 55 = 93812.46542 Residual 55 = 93812.46542 Residual 55 = 0.0895 TRED115313 TRED1053764 LURLIVEE Coef. Std. Err. z Fill TRED1065764 .0645056 TRED1065764 .056847 .056847 .056842 .050854 TRED1065764 .0508543 .0508543 .0508543 .0508541 .050774 .0508543 .0508543 .0508543 .0508543 .0508543 .0508543 .0508543 .0508543 <td>Warning - coll</td> <td>inearities de</td> <td>tected</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>	Warning - coll	inearities de	tected						
IV (2515) estimation Estimates efficient for homoskedasticity only Statistic consistent for homoskedasticity only Total (centered) 55 = 104853.0198 Coref. Std. Err. z P>(z) [95% Conf. Interval] EUCC 1.199233 0.333314 0.90 0.000 0.0403598 TEDC -1.113319 0.0708764 -1.37 0.116 -2.39461 0.0273812 TED -1.0687764 0.064942 -1.44 0.010 -2.234664 0.0207076 TED -0.0698173 0.037398 0.0423431 -1.68 0.061 -1.163297 0.0034717 TED -0.0698173 0.037398 0.0148 0.022 -1.123066 0.0023113 TED -0.0698173 0.037398 0.0148 0.022 -0.033 0.034843 0.2000718 TED -0.0423750 0.012736 0.014 0.972 -0.013766 TED -0.0423750 0.012737 0.014 0.922 -0.033 0.034843 0	Vars dropped:	YR29 QT	R129 QTR328	QTR329					
Image: Statistics consistent for homoskedasticity only Statistics consistent for homoskedasticity only Statistics consistent for homoskedasticity only F(12,247180) = 0.0000 Total (uncentered) S3 = 104853.0198 Centered R2 = 0.0000 Total (uncentered) S3 = 6674371.774 Uncentered R2 = 0.0959 Residual S3 = 93812.46542 EUCC	IV (2SLS) esti	mation							
Estimates efficient for homoskedasticity only Statistic consistent for homoskedasticity only Number of obs = 247199 r[12:347186] = 0.0000 Total (centered) 55 = 104853.0198 Centered R2 = 0.1033 Total (uncentered) 55 = 647371.774 Uncentered R2 = 0.1039 Total (uncentered) 55 = 93812.46542 Root MSE = .616 LWRLYWOE Ceff. Std. Err. z F)z [5% Conf. Interval] LUC11299233 .033514 3.90 0.000 .044555 VRE0 115313 .0708794 -1.57 0.116 2248641 .0270716 TR21 1055764 .065942 1217941 .0064556 .135290e TR22 025877 .0256524 123464 .0270776 TR22 025877 .0256873 .031389 119341 .0064395 TR24 079844 .0423343 18 .0610 123741 .0041396 TR24 079844 .02177 16 .022 .02337 .02119 TR25 .0029174 16 .002 .12397 .0202119									
Statistics consistent for homoskedasticity only Number of obs = 247199 F(12,247186) = 8.68 Prob > F = 0.0000 Total (centered) 55 = 104053.0198 Centered R2 = 0.1033 Total (uncentered) 55 = 6674371.774 Residual 55 = 93812.46542 Denetered R2 = 0.9859 Residual 55 = 0.9859 Residual 55 = 0000 Cent. x F>(x) [95% Conf. Interval] LWKLYWOE Ccef. Std. Err. x F>(x) [95% Conf. Interval] EUCC .1299233 .0333314 3.90 0.000 .0643556 TR21 -1065764 .064942 -1.64 0.101 230461 .0207971 TR23 0262975 .055624 168 0.061 163297 .0043917 TR23 0269475 .0325393 168 0.062 121374 .0043917 TR23 0269475 .022391 1689 0.062 .1213297 .0033917 TR25 0695473 .0327392 169 .0029 .1232466 .0029171 TR25 .0069473 .012477 .1.07 .02446 .005051 .0154917	Estimates effi	cient for hom	oskedasticit	v only					
Number of obs = 247199 F(12,24716) = 0.0000 Total (uncentered) 55 = 104853.0198 Centered R2 = 0.0000 Total (uncentered) 55 = 6674371.774 Uncentered R2 = 0.1033 Residual 53 = 93812.46542 LMKLYWGE Coef. Std. Err. z $P > z $ [59% Conf. Interval] EUC .1999233 VIEL 115319 VIEL 166716 VIEL 1053764 VIEL 1054764 VIEL 056473 0564737 .0337384 056 119 0564737 .037398 056 119 VIEL 0124737	Statistics con	sistent for h	omoskedastic	ity only					
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $					37-	mber of obs -	2471.00		
Introduction Introduction Introduction Total (contered) S5 = 104853.0198 Brob F = 0.0000 Total (uncentered) S5 = 6674371.774 Chintered R = 0.0899 Residual S5 = 93812.46542 Root MSE = .616 LMKLYWOE Coef. Std. Err. z P> z [95% Conf. Interval] EUCC .1299233 .033314 3.00 0.000 .064359 TWE1 -1057764 .04894 -1.47 0.116 .2073812 TWE1 -1057764 .064942 -1.44 0.010 .02473812 TWE2 065764 .0485341 -1.88 0.061 -1893747 .0044717 TWE2 0696473 .032398 -1.63207 .0034717 TWE2 0642795 .010245 .0034717 TWE2 0642745 .012746 .0202 .1212066 .0024317 TWE2 .003651 .0154974 TWE2 0629475 .0127457 .0107 0.024 .0222061 .0034717 TWE2 0629475 .0124797 .0107 0.024 .0203055 .0155497 T					240	imper of obs =	24/199		
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $						(12,24/100) =	0.00		
10441 (centered) 55 = 10483.0198 Centered R2 = 0.1053 Total (uncentered S2 = 664731.774 Uncentered R2 = 0.9859 Residual 55 = 93812.46542 Root H5Z = .616 LMKLYWGE Coef. Std. Err. z P>[z] [59% Conf. Interval] EEOC .1099233 .033314 3.00 0.000 .064559 YE20 .115319 070704 .137 0.116 .0077812 YE21 1055764 .0694942 -1.64 0.101 239461 .0077812 YE21 1055764 .055624 -1.84 0.061 199747 .0041576 YE23 0298895 .0453541 -1.88 0.062 123737 .0045764 YE25 0698473 .037388 -1.69 0.092 123297 .0034717 YE25 0642745 .0271461 055649 .056409 .015409 YE27 0107343 .0147477 -1.07 .0284 .003955 .0155409 YE26 0.014 .0272 .029708 .022301 .025619 YE27 .0107345 .012477 .01707 .022302 .003453 YE26 0.1042379 .0073048 2.02					P.	- 1 < 20	0.0000		
104.1 (uncentered) 0.9893 0.987471.774 Uncentered NC = 0.9893 Residual SS = 93812.46542 Root MSE = .616 IMMELYWGE Coef. Std. Err. z F) z [5% Conf. Interval] EDUC .1299233 .0333314 3.90 0.000 .044555 .1352908 TRE0 .1115313 .070754 -1.57 0.116 .2294451 .0273912 TRE1 1065764 .064942 -1.64 0.101 2294451 .0064594 TRE3 0228795 .0425341 -1.88 0.061 182977 .0044594 TRE3 0298995 .0425341 -1.88 0.061 .182397 .0044594 TRE3 0298995 .0425343 -1.68 0.006 .182397 .0034717 TRE5 0423755 .0127467 .016 .022946 .0034917 TRE3 .0029617 .012748 .0246 .003955 .0158499 TRE3 .0029617 .012747 .0.07 .014 .02291 .0034717 TRE5 0.002777 .0.04 .972 .003	Iotal (centere	1) 55 =	104853.0198		0	entered R2 =	0.1053		
Residual 33 = 93812.46542 Root MSE = 0.616 LMKLYWGE Coef. Std. Err. z P> z [55% Conf. Interval] EDCC .1299233 .0333314 3.90 0.000 .0645558 .1952902 THE 0 113316 .0703744 -1.57 0.118 200451 .0273812 THE 0 1026278 .055662 213141 .0064564 THE 1 0262895 .0855641 218 .0064564 THE 2 078940 .0262318 188 .061 1893747 .0034717 THE 2 0664713 .0327988 1.89 .0529 123297 .0034717 THE 2 06497345 .014797 1.07 .0284 0509551 .0154979 THE 2 06497345 .014797 1.07 .0284 0509551 .015492 THE 2 0.0167345 .012477 007 .0142373 202501 THE 2 0.006851 .012477 .0079 .022301 <	Total (uncente	rea) SS =	6674371.774		01	icentered R2 =	0.9859		
LWKLYWGE Ccef. Std. Err. z P> z [95% Conf. Interval] EUCC .1299233 .0333514 3.90 0.000 .0644556 .1352908 YRE0 115319 .0708794 -1.57 0.116 2204451 .0270776 YRE1 1065764 0.664942 164 0.101 238644 .0207076 YRE2 1062564 0.055624 168 0.061 2117341 .0064925 YRE2 0262895 .0393381 189 0.061 1637341 .00641956 TRE3 079894 .042343 189 0.060 16397347 .0034137 TRE3 0479844 .042343 189 0.060 1639747 .0034137 TRE3 0479845 .037386 .1027 .0041356 .1004213 TRE3 0427353 .0271461 -1.56 0.012 .1035450 .10082113 TRE3 .0427753 .0107166 .0044 .9770 .0280050	Residual SS	-	93812.46542		Re	ot MSE =	.616		
Line Line Line Line EUCC .1299233 .0333514 3.90 0.000 0.044355 1.952308 YE20 1115333 .0708764 -1.87 0.116 238644 .0373812 YE21 1065764 0.64944 -1.64 0.101 238644 .0373812 YE21 10626376 .0556624 -1.84 0.065 2117341 .0064584 YE23 0268973 .0453341 -1.88 0.065 2117341 .0064584 YE24 079844 .0423343 -1.88 0.061 163297 .0041356 YE25 0793948 .0237389 -1.69 0.092 1232946 .0092117 YE25 0423755 .027146 .10284 053095 .0158499 YE25 .0014 .0727 0770348 .024801 .022051 YE25 0 (cmitted) .024 .022051 .024513 .202051 YE25 0 (cmitted)	LWKLYWGE	Coef.	Std. Err.	z	P>1z1	[95% Conf.	Intervall		
ELOC .1298233 .0333314 3.00 0.000 .0643536 .1392308 TRED 1115313 .0708754 70 0.116 203451 204513 TRE1 065764 .064942 64 0.010 220451 204545 2037812 TRE2 026376 655662 115314 60 2117341 0064594 TRE3 026376 695662 188 0.061 1829747 0044594 TRE3 026379 425343 488 0.061 1829747 0043717 TRE5 0423755 075944 01266 18297 003971 TRE5 0423755 071461 05 19 050955 0158409 TRE7 0423755 012477 07 07 02464 0108299 TRE7 0013746 012477 07 07 014 022051 TRE2 0013746 0				_					
TRED 1115319 .0708794 -1.57 0.116 204451 .0273812 TRED 1065764 0.64842 -1.64 0.101 2338044 .0270776 TRED 1065764 0.056542 -1.84 0.065 2117341 .0041356 TRED 0928935 0.0565 11199747 .0041356 TRED 0927935 0.03904 090050 020166 TRED 0279735 0.027451 .05604 .000155 TRED 0279873 0.037904 109 .0500 .000155 TRED 0297375 0.027461 .050 .05005 .015840 TRED 0297375 0.027461 .05005 .015840 TRED .010736 0.014 0.977 .029707 .020501 TRED 0.010776 0.044 0.9770 .020501 TRED 0 (mittred) .0004843 .2800738 c.A6E026 .142277 0.703048 2.02 0.043 .0240739 c.A6E026 .0007866 -1.75 0.0408 .022032 .000463 000851 .0007866 -1.75 0.0408 .02020501 00085 .0007866	EDUC	.1299233	.0333514	3.90	0.000	.0645558	.1952908		
TR21 1055764 .0645942 -1.64 0.101 2338044 .0227076 TR23 1026376 .0556624 184 0.065 2117341 .0064394 TR23 079694 .0423343 -1.08 0.061 1899747 .0004395 TR24 079694 .0423343 -1.68 0.062 163297 .0034717 TR25 054375 .0337398 -1.69 0.092 1632046 .0092411 TR25 0423753 .0174077 -1.07 0.2244 0550095 .0156429 TR27 0103743 .0174077 -1.07 0.224 0530095 .0156409 TR26 .0003451 .0102736 .0044 .922 .020501 .022501 TR27 .0112737 .01027376 .0024 .020201 .020501 .022501 TR29 0 (mittred) .002 .022902 .000163 _c.A62045.A6202 0013786 .0077866 -1.75 0.901 0229202 .000163	YR20	1115319	.0708754	-1.57	0.116	2504451	.0273812		
YR21 1025778 .0556242 -1.04 0.065 2117341 .0064364 YR23 0226895 0.689341 -1.0889 0.061 1859747 .0041356 YR24 079694 .042343 -1.88 0.061 182977 .0034117 YR25 056913 .037396 -1.66 0.092 .1213044 .0052115 YR25 .056913 .0271467 -1.56 0.192 .123046 .0022115 YR26 0423735 .0271467 -1.56 0.192 .1022995 YR25 .003631 .0102736 0.04 0.972 .012859 YR28 .003641 .0102736 0.044843 .200738 C.AGEOFC.AECD 0013766 .0703048 2.02 0.043 .0248643 .200738 c.AGEOFC.AECD 0013766 .007966 -1.75 0.000 022802 .00163 00006 007966 -1.75 0.000 022802 .00163	YR21	1065764	.064942	-1.64	0.101	2338604	.0207076		
YR23 022895 .0425341 -1.088 0.061 1899747 .0041555 YR24 079894 0.4023543 -1.08 0.060 1632247 .0034717 YR25 0625173 .0327399 -1.68 0.092 1632464 .009215 YR26 0423755 .027461 -1.5 0.119 0558095 .0158409 YR27 0107343 .017477 -1.07 0.284 0550095 .0155409 YR27 0.003651 .0102736 0.04 0.927 0157706 .002601 YR28 0 (cmitted) .02 .043 .0044843 .2800739 c.AG204ci.AGEQ 0013766 .007966 -1.75 0.080 0022922 .000163 00004 .121397 1.6500966 0.07 0.941 -3.112731 3.355525	YR22	1026378	.0556624	-1.84	0.065	2117341	.0064584		
YR14 079894 .0425343 -1.68 0.060 165297 .003417 YR15 056913 .033739 -1.68 0.060 123046 .0092115 YR16 0423755 .0271461 -1.56 0.19 0956903 .0108299 YR17 0187343 .014297 .0103756 .0.04 1.9 .020561 YR12 .0003651 .0102736 0.04 0.972 0197708 .020501 YR12 .0003651 .0102736 0.04 0.972 0197708 .020501 YR12 .0103796 .00348 2.02 0.043 .0044843 .2800738 C.A0E026c.A0EQ 0013766 .0007866 -1.75 0.080 0029202 .000163 0001 .121397 1.650096 0.07 9.91 -3.112731 3.355525	YR23	0928895	.0495341	-1.88	0.061	1899747	.0041956		
WI25 0659173 .0337398 -1.68 0.092 123046 .009215 WI26 0423755 .0274461 155 0.119 0958095 .0108299 WI27 0197343 .0174977 0<70	YR24	079894	.0425343	-1.88	0.060	1632597	.0034717		
YR26 042375 .0271461 -1.56 0.119 0955090 .0108299 YR27 0187343 .0174877 .107 0.284 .055009 .0108249 YR28 .0003651 .012736 0.04 0.972 0197708 .020501 YR29 .0003651 .0102736 0.04 0.972 0197708 .020501 YR29 .0703048 2.02 0.043 .00448643 .2800738 c.AGE20fc.AGE2 0013786 .0007866 -1.75 0.080 0029202 .000163 oon .121397 1.650096 0.07 0.941 -3.112731 3.35525	YR25	0569173	.0337398	-1.69	0.092	123046	.0092115		
YK27 -0.017743 0.1014977 -1.07 0.284 -0.053095 0.0155409 YK28 0.0003651 0.012736 0.04 0.972 -0.0197708 0.020501 YK25 0 (mitted) .0013746 0.49 0.972 -0.0197708 .200501 XACC 1.42279 .0703048 2.02 0.043 .0044443 .2800739 c.AGEQ#c.AGEQ 0013766 -1.75 0.080 0022920 .000163 00ms .121397 1.650096 0.07 0.941 -3.112731 3.355525	YR26	0423755	.0271461	-1.56	0.119	0955809	.0108299		
YR28 .0003651 .0102736 0.04 0.972 0197708 .020501 YR29 0 0 0 .0012730 2.02 0.043 .2800738 AGEQ .42279 .0703049 2.02 0.043 .0044943 .2800738 c.AGEQ#c.AGEQ 013786 .0007866 -1.75 0.060 0029202 .000163 oons .121397 1.650096 0.07 0.941 -3.112731 3.35525	YR27	0187343	.0174877	-1.07	0.284	0530095	.0155409		
YE25 0 (emitted) AGEO .142279 .0703048 2.02 0.043 .0044843 .2800738 c.AGEQ/et.AGEO 0013766 0.00766 -1.75 0.080 0022202 .000163 oons 1.21397 1.650096 0.07 0.941 -3.112731 3.355525	YR28	.0003651	.0102736	0.04	0.972	0197708	.020501		
AGEQ .142279 .0703048 2.02 0.043 .0044843 .2800738 c.AGEQ#c.AGEQ 0013786 .0007866 -1.75 0.080 0022202 .000163 oons .121397 1.650096 0.07 0.941 -3.112731 3.355525	YR29	0	(omitted)						
c.AGEQ+c.AGEQ0013786 .0007866 -1.75 0.0800029202 .000163 cons .121397 1.650096 0.07 0.941 -3.112731 3.355525	AGEQ	.142279	.0703048	2.02	0.043	.0044843	.2800738		
_cons .121397 1.650096 0.07 0.941 -3.112731 3.355525	c.AGEQ#c.AGEQ	0013786	.0007866	-1.75	0.080	0029202	.000163		
	_cons	.121397	1.650096	0.07	0.941	-3.112731	3.355525		
LISETOED REATINEIC LOVETIGEDITIEICETION TERE OF ALL INVERTMENTALS VO. 300	bargan boabist	(010224000		50 01 ui	Chi-	sq(26) P-val =	0.5018		
Chi-sq(26) P-val = 0.5018	Collinearities	detected amo	ng instrumen	ts: 1 in	strument	(s) dropped			
Callerative detected among instruments: 1 instruments: 23.505	Instrumented:	EDUC						1	
Sargem Exclusion (overlassing) 20.303 Chi-sqi(2) -val = 0.5018 Chi-sqi(2) -val = 0.5018 Collinearities detected emong instruments: l instrument(s) dropped Instrumente(s) dropped	Included instr	aments: YR20	YR21 YR22 YR	23 YR24	YR25 YR2	S YR27 YR28 AG	ΕQ		
Collinearities detected among instruments: 1 instrument(s) dropped Collinearities detected among instruments: 1 instrument(s) dropped Instrumented: Included instruments: YR20 YR21 YR22 YR24 YR25 YR26 YR27 YR28 AGE0		c.AGE	Q#c.AGEQ					1	
Collinearities detected among instruments: 1 instruments): 20.303 Chi-sq(26) P-val = 0.5018 Collinearities detected among instruments: 1 instrument(s) dropped Instrumented: EDOC Included instruments: VRIO YRIS YRIS YRIS YRIS YRIS YRIS AGEO c.REDie-AGEO	Excluded instr	aments: QTR1	QTR120 QTR12	1 QTR122	QTR123 (TR124 QTR125	QTR126	1	
Sargem statistic (overigentification test of all instruments): 20.303 Chilinearities detected among instruments: 1 instrument(g) 0.5018 Collinearities detected among instruments: 1 instrument(s) 1 instrument(s) Instrumented: EDOC		07010	7 070120 070	220 OTD2	21 OTP22	OTRA22 OTR22	4 OTD225	1	

Instrumental Variables Regression	Evaluating IV Assumptions •0 •0	More general Stata commands 000 0
Instrument Relevance		

Evaluating instrument relevance

• Recall that we in order for a variable, *z_i* to be a relevant instrument for endogenous variable, *x_i*, we require:

 $Cov(z_i, x_i) \neq 0$

- Or, in terms of $\beta_{IV}=rac{\gamma_1}{\pi_1}$, we require: $\pi_1
 eq 0$
- Generally speaking, we can test this condition with a t-test (or F-test in the case of multiple instruments) of the excluded instruments in the first-stage.
- ivreg2 by default reports more advanced tests for the relevance condition, which it calls "underidentification tests"
 - Because if instruments are irrelevant, then the number of relevant instruments is less than the number of endogenous explanatory variables and so the model is *underidentified*.

Weak Instruments

- Two-stage least squares, however, in fact has a problem not only if $\pi = 0$ but even if $\pi \approx 0$. Why?
 - Although IV regression is consistent, it does suffer from finite sample bias. This bias is inversely related to the correlation of x_i with z_i .
 - As the number of instruments increase for a given endogenous variable (each with $\pi \approx 0$), the bias of 2SLS becomes as large as OLS.
- Returning again to the F-test for the excluded instruments, a good "rule of thumb" is that (for a single endogenous explanatory variable) the F-Stat should be at least 10.
 - $F^{crit} = 10$ in this case corresponds to a relative bias of $\sim 10\%$ compared to the bias of OLS.
 - **ivreg2** presents weak identification test statistics and the critical values corresponding to different levels of relative bias.

Instrumental Variables Regression	Evaluating IV Assumptions ○○ ●○	More general Stata commands 000 0
Instrument Validity		

Evaluating instrument validity

• In addition to instrument relevance, we also require that the instrument is *valid* (that is, exogenous).

 $Cov(z_i, u_i) = 0$

- In general, it is not possible to test instrumental validity.
- But if you have more than one instrument for a single endogenous variable, you can perform a type of falsification test of your instrumental validity assumptions:
 - The **test of overidentifying restrictions** assumes (under the null) that all instruments are valid.
 - If only some of the proposed instruments are valid, the test will tend to reject the null.
 - If none of the instruments are valid, however, the test is not helpful.
- ivreg2 reports the results of the test of overidentifying restrictions by default.

LWKLYWG	E Coef.	Std. Err.	z	P> z	[95% Conf.	Interval
EDU	C .0517289	.0157576	3.28	0.001	.0208445	.0826133
c.AGEQ#c.AGE	Q0014031	.000026	-53.96	0.000	0014541	0013521
Underidentif	ication test (A	Anderson cano	n. corr.	LM stat: Chi-s	istic): sq(30) P-val =	176.566 0.0000
Weak identif	ication test ((Cragg-Donald	Wald F st	tatistic)	:	5.886
Stock-Yogo w	eak ID test cr	itical values	: 5% max	kimal IV	relative bias	21.42
			10% mag	kimal IV	relative bias	11.32
			20% max	kimal IV	relative bias	6.09
			30% max	kimal IV	relative bias	4.29
			10% max	kimal IV	size	86.17
			15% max	kimal IV	size	44.78
			20% max	kimal IV	size	30.72
			- 25≩ mai	kimal IV	size	23.65

Instrumental Variables Regression	Evaluating IV Assumptions 00 00	More general Stata commands ●○○ ○
Presenting your results		

Intro to the -estout- package

- A key component of effective data analysis is professional presentation of results.
- Likely the best Stata package for preparing professional-looking descriptive statistic and regression output tables is the **-estout- package**.
- We will briefly go over the basics of using this package to produce regression output tables.

Basics of the -estout- package

- The first step to creating table with -estout- is to store your the results of your table.
 - To do this, simply put **eststo:** immediately before your regression command (on the same line).
- After saving any regressions that you want to appear in a table, use the **esttab** command to save generate the table (and save it to disk).
 - The syntax is esttab using "[YOURFILENAME]", options
- To clear the saved results, use the command eststo clear.

Instrumental Variables Regression	Evaluating IV Assumptions	More general Stata commands ○○● ○

Specifying Table Output with **esttab**

- There are several common options:
 - Most of the time, you will want to add a title for your table by specifying the option title("[YOUR TITLE HERE]")
 - Titles for each regression model saved in a table can be specified by mtitles("[TITLE1]" "[TITLE2]" ...)
 - You should also specify the format that you want your table saved as. Formats include **scml csv rtf html** & **tex**.
 - A great option is **onecell**, which specifies that estimates and standard errors should appear in the same cells of your table.
 - **label** displays variable labels instead of names (generally preferable).
 - If you have long variable names/labels, use **wrap** and **noabbrev** to word-wrap and not abbreviate these labels.
 - If your table is big, you can make it more compact by specifying the option **compress**.
 - Finally, use the option replace to overwrite the table if it has been previously saved.

Presenting

Local Macros and Scalars

Sometimes you may want to store information not as a variable in a dataset, but just as a short-hand for values or text that you use frequently. A good way to do this is with scalars and locals.

- To save a specific number (not as a variable), you can define it as a sclar. To do this, the syntax is scalar [SCALARNAME]
 = [VALUE/EXPRESSION]. Scalars can then be multiplied or used in expressions just like a normal variable.
- Other times, you may want to refer to a long expression (of any sort, a list of variables, values, part of a regression code etc) in a shorthand to avoid having to rewrite values repeatedly.
 - To do this, define a local via the syntax: local [NAMEFORLOCAL] [EXPRESSION...]
 - To then refer to the local, put it in quotes just like the index values in a loop, like this: `[NAMEFORLOCAL]'.