Econometrics E5040: Exercise 7: More on Instrumental Variables

1. Given the 2 variable model
$y = a + bx^t + u$ where $x = x^t + w$

It is suggested that a suitable instrument would be to use a value of plus or minus 1 according to whether the value of X is above or below the median X value, (assume an equal number of X observations). Write down the functional form of the parameter estimates.
Another instrumentation procedure is to rank the X values in ascending order. Write down the functional form for the estimate of the slope.
Are these good instruments?

- If instrument takes the value -1 if the X value is below its median value and 1 if >=, then putting X in ascending order gives

$$Z' = \begin{bmatrix} 1 & 1 & 1 & .. & 1 & 1 \\ -1 & -1 & -1 & 1 & 1 & 1 \end{bmatrix}$$

and
$$Z'X = \begin{bmatrix} 1 & 1 & 1 & .. & 1 & 1 \\ -1 & -1 & -1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X2 \\ & . \\ : & \\ & : \\ 1 & X_N \end{bmatrix} = \begin{bmatrix} N & \sum X \\ 0 & N/2(\bar{X}_2 - \bar{X}_1) \end{bmatrix}$$

Likewise
$$Z'y = \begin{bmatrix} \sum y \\ N/2(\bar{y}_2 - \bar{y}_1) \end{bmatrix}$$

Hence
$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} N & \sum X \\ 0 & N/2(\bar{X}_2 - \bar{X}_1) \end{bmatrix}^{-1} \begin{bmatrix} \sum y \\ N/2(\bar{y}_2 - \bar{y}_1) \end{bmatrix}$$

or
$$\begin{bmatrix} N & N\bar{X} \\ 0 & N/2(\bar{X}_2 - \bar{X}_1) \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} N\bar{y} \\ N/2(\bar{y}_2 - \bar{y}_1) \end{bmatrix}$$

which is a simultaneous equation system

$$N\hat{a} + N\bar{X}\hat{b} = N\bar{y} \qquad (1)$$

$$0 + N/2(\bar{X}_2 - \bar{X}_1)\hat{b} = N/2(\bar{y}_2 - \bar{y}_1) \qquad (2)$$

solving for b gives

$$\hat{b} = \frac{\bar{y}_2 - \bar{y}_1}{\bar{X}_2 - \bar{X}_1}$$

and so $\hat{a} = \bar{y} - \bar{X}\hat{b}$

- the slope estimate is therefore equivalent to partitioning the data by the median value and drawing a straight line through the mean points of each sub-sample.
This is unlikely to be highly correlated with all the variation in x

If instrumentation is done by ranking then

$$Z' = \begin{bmatrix} 1 & 1 & 1 & .. & 1 & 1 \\ 1 & 2 & 3 & .. & & N \end{bmatrix}$$

and $$Z'X = \begin{bmatrix} 1 & 1 & 1 & .. & 1 & 1 \\ 1 & 2 & 3 & .. & & N \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X2 \\ & . \\ : & \\ & : \\ 1 & X_N \end{bmatrix} = \begin{bmatrix} N & \sum X \\ \sum i & \sum iX \end{bmatrix}$$

$$Z'y = \begin{bmatrix} \sum y \\ \sum iy \end{bmatrix}$$

Hence $$\begin{bmatrix} N & \sum X \\ \sum i & \sum iX \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} \sum y \\ \sum iy \end{bmatrix}$$

and

$$N\hat{a} + N\bar{X}\hat{b} = N\bar{y} \qquad (1)$$
$$\sum i\,\hat{a} + \sum iX\,\hat{b} = \sum iy \qquad (2)$$

so $$\hat{b} = \frac{\sum iy - \sum i\bar{y}}{\sum iX - \sum i\bar{X}}$$

This is potentially a good instrument IF the measurement error is not so large as to change the rankings

2. Given the simultaneous equation system

$$y_1 = a_1 y_2 + u_1 \qquad (1)$$

$y_2 = a_2 y_1 + u_2$ (2)

with a1<0 a2>0 , determine the sign of the endogeneity bias in an OLS estimate of (1)

Consider a positive shock modelled by $u_1 > 0$
$\uparrow u_1 \rightarrow \uparrow y_1 \rightarrow \uparrow y_2$ if $a_2 > 0$

So if $a_2 > 0$ $Cov(u_1 y2) > 0$
Similarly if $a_2 < 0$ $Cov(u_1 y2) < 0$

and OLS on (1) gives (in mean deviation form)

$$\hat{a}_1 = \frac{\sum_i y_2 y_1}{\sum_i y_2^2} = \frac{\sum_i y_2(a_1 y_2 + u_1)}{\sum_i y_2^2} = a_1 + \frac{\sum_i y_2 u_1}{\sum_i y_2^2}$$

- by substitution

Hence OLS estimate of a_1 in (1) is upward biased (equals true value plus a positive constant if $Cov(u_1 y2) = \sum_i y_2 u_1 > 0$
(and downward biased if $a_2 < 0$)

3. Show that in the 2 variable model, the 2SLS estimator equals the ratio of the reduced form estimator to the 1st stage estimator

Given the 2sls IV estimator in the 2 variable (mean deviation) case with endogenous right hand side variable x and instrument z

can be written as $\hat{\beta}_{IV} = \frac{\sum_i z_i y_i}{\sum_i z_i x_i}$

Note that the 1st stage of 2SLS is a regression of x on z which gives

$$\hat{\gamma}_{1st} = \frac{\sum_i z_i x_i}{\sum_i z_i^2} \qquad (1)$$

and the reduced form regression of y on z gives

$$\hat{\delta}_{RF} = \frac{\sum_i z_i y_i}{\sum_i z_i^2} \qquad (2)$$

so (2)/(1) gives $\frac{\frac{\sum_i z_i y_i}{\sum_i z_i^2}}{\frac{\sum_i z_i x_i}{\sum_i z_i^2}} = \hat{\delta}_{RF} * \left(1/\hat{\gamma}_{1st}\right) =$

$$\frac{\sum_i z_i y_i}{\sum_i z_i x_i} = \hat{\beta}_{IV} \qquad (3)$$

Since the reduced form coefficient is therefore proportional to the IV coefficient this can be used
to assess the strength of the instrument (by rescaling the reduced form estimate by the reciprocal
of the first-stage coefficient).

The bigger the rescale, the weaker the instrument

If the first stage is weak the estimate of $\hat{\gamma}_{1st}$ will be close to zero (in terms of significance, not necessarily the point estimate)
and hence the estimate of $\hat{\beta}_{IV}$ large and uninformative

Note that if the reduced form coefficient is not significantly different from zero, (3) therefore suggests so is the IV estimate
Hence the effect of interest is either absent or the instruments are too weak to detect it.

4. Another way to interpret this estimator is that the numerator gives the difference in outcomes for those treated and not treated (sometimes called the intention to treat estimates)
divided by the difference in the endgenous variable for those treated and not treated

Consider a simple 2 variable model where the effect of interest is the impact of asignment to a treatment group on an outcome y

$$y = a + T\beta + u$$

and let there be a binary instrument z for T

The IV estimator is just $\hat{\beta}_{IV} = \dfrac{Cov(y_i z_i)}{Cov(T_i z_i)}$

Since $Cov(y_i z_i) = E(y_i - \bar{y})(z_i - \bar{z}) = E(y_i z_i) - \overline{yz}$

using Bayes law $= E(y_i/z_i = 1)E(z_i = 1) - \overline{yz}$

$= E(y_i/z_i = 1)E(z_i = 1) - \bar{y}\, E(z_i = 1)$

$= (E(y_i/z_i = 1) - \bar{y})E(z_i = 1)$

$= [E(y_i/z_i = 1) - \{E(y_i/z_i = 1)E(z_i = 1) + E(y_i/z_i = 0)E(z_i = 0)\}]E(z_i = 1)$

$= [E(y_i/z_i = 1)E(z_i = 0) - E(y_i/z_i = 0)E(z_i = 0)\}]E(z_i = 1)$

$= [E(y_i/z_i = 1) - E(y_i/z_i = 0)]E(z_i = 1)E(z_i = 0)$

Similarly the OLS regression of the endogenous variable on the instrument gives

$$\hat{\delta}_{RF} = \frac{Cov(D_i z_i)}{Cov(z_i z_i)}$$

where it can be shown that $Cov(D_i z_i) = [E(T_i/z_i = 1) - E(T_i/z_i = 0)]E(z_i = 1)E(z_i = 0)$

Hence the IV estimator

$$\hat{\beta}_{IV} = \frac{\sum_i z_i y_i}{\sum_i z_i x_i} == \frac{Cov(y_i z_i)}{Cov(T_i z_i)} = \frac{[E(y_i/z_i{=}1) - E(y_i/z_i{=}0)]E(z_i{=}1)E(z_i{=}0)}{E(T_i/z_i{=}1) - E(T_i/z_i{=}0)]E(z_i{=}1)E(z_i{=}0)}$$

$$= \frac{[E(y_i/z_i{=}1) - E(y_i/z_i{=}0)]}{E(T_i/z_i{=}1) - E(T_i/z_i{=}0)]}$$

Hence even if some of the non-treated group may get treated $E(T_i/z_i = 0) \neq 0$ (sometimes called non-compliance) and the IV estimator accounts for this

5. Given

$$C_t = b_1 + b_2 I_t + u_t$$
$$I_t = C_t + S_t + v_t \qquad\qquad u_t \sim N(0,1),$$

Derive the IV estimate of b_2 using a suitable instrument for I_t
How does it differ from the OLS equivalent?

- The OLS estimate of b_2 is given by

$$\hat{b}_2 = (X'X)^{-1}X'y$$

where X=[i : I]   y=C

Hence

$$\hat{b}_2^{ols} = \begin{bmatrix} N & \sum I \\ \sum I & \sum I^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum C \\ \sum CI \end{bmatrix} = \begin{bmatrix} 26 & 2 \\ 2 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 10 \\ 3 \end{bmatrix} = \begin{bmatrix} -0.2 \\ 0.52 \end{bmatrix}$$

Compare to the IV estimator $\hat{b}_2^{IV} = (Z'X)^{-1}Z'y$

In this case $Z' = \begin{bmatrix} 1 & 1 & .. & 1 \\ S_1 & S_2 & .. & S_T \end{bmatrix}$ $X' = \begin{bmatrix} 1 & 1 & .. & 1 \\ I_1 & I_2 & .. & I_T \end{bmatrix}$

S is a suitable instrument for I since it determines I but not C according to the simultaneous model above

$$\hat{b}_2^{iv} = \begin{bmatrix} N & \sum I \\ \sum S & \sum IS \end{bmatrix}^{-1} \begin{bmatrix} \sum C \\ \sum CS \end{bmatrix} = \begin{bmatrix} 26 & 2 \\ 1 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 10 \\ -1 \end{bmatrix} = \begin{bmatrix} 22/50 \\ -16/50 \end{bmatrix}$$

So can see that the OLS estimate is an (upward biased) estimate
(Note that the sample size in this sample is rather small, so the IV estimate is likely to suffer from small sample bias in practice)

6. Given the following information, find the IV (2SLS) estimates of the parameters and

their standard errors in the following 3 variable model where the first variable, x_1, is measured with error.

$y_i = b_1 x_1 + b_2 x_2 + u_i$

There is a potential instrument, x_3, available.

Let $X = [\, x_1 \; x_2 \; x_3]$ $Y = [y \; x_1]$ and the sample size is 100

$$X'X = \begin{bmatrix} 1 & 4 & 10 \\ 4 & 2 & 0 \\ 10 & 0 & 5 \end{bmatrix} = \begin{bmatrix} x_1'x_1 & x_1'x_2 & x_1'x_3 \\ x_2'x_1 & x_2'x_2 & x_2'x_3 \\ x_3'x_1 & x_3'x_2 & x_3'x_3 \end{bmatrix} \quad \text{and} \quad Y'Y = \begin{bmatrix} 20 & 10 \\ 10 & 1 \end{bmatrix}$$

$$\text{and } Y'X = \begin{bmatrix} 2 & 2 & 2 \\ 1 & 4 & 10 \end{bmatrix} = \begin{bmatrix} y'x_1 & y'x_2 & y'x_3 \\ x_1'x_1 & x_1'x_2 & x_1'x_3 \end{bmatrix}$$

x_1 is measured with error so need to instrument.

Let $x=[\, x_1 \; x_2]$ be the set of r.h.s. variables

Let $Z=[\, x_3 \; x_2]$ be the set of instruments

Since k=l then the IV estimator

$$\hat{b}_2^{IV} = \left[(X'Z)(Z'Z)^{-1}(Z'X) \right]^{-1} Z'y \quad \textit{reduces} \quad \textit{to} \quad \hat{b}_2^{IV} = (Z'X)^{-1}Z'y$$

$$\text{where now } Z'X = \begin{bmatrix} x_3' \\ x_2' \end{bmatrix} \begin{bmatrix} x_1 & x_2 \end{bmatrix} = \begin{bmatrix} x_3'x_1 & x_3'x_2 \\ x_2'x_1 & x_2'x_2 \end{bmatrix}$$

Given the values in X'X above, it follows that

$$Z'X = \begin{bmatrix} 10 & 0 \\ 4 & 2 \end{bmatrix}$$

$$\text{Similarly } Z'y = \begin{bmatrix} x_3' \\ x_2' \end{bmatrix} y = \begin{bmatrix} x_3'y \\ x_2'y \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$\text{So } \hat{b}_2^{IV} = (Z'X)^{-1}Z'y = \begin{bmatrix} 10 & 0 \\ 4 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 2/20 & 0 \\ -4/20 & 10/20 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

$$\Rightarrow \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} 0.2 \\ 0.6 \end{bmatrix}$$

To find $\text{Var}\left(\hat{b}_2^{IV}\right)$

use $\text{Var}\left(\hat{b}_2^{IV}\right) = s_{IV}^2 (X' P_Z X)^{-1} = s_{IV}^2 \left[Z'X\right]^{-1} [Z'Z][X'Z]^{-1} \qquad if \quad l = k$

with $s_{IV}^2 = \dfrac{\left(y - X\hat{\beta}_{IV}\right)\left(y - X\hat{\beta}_{IV}\right)'}{N} = \dfrac{\left[y'y - 2y'X\hat{\beta}_{IV} + \hat{\beta}_{IV}' X'X\hat{\beta}_{IV}\right]}{N}$

From above Y'Y= $\begin{bmatrix} y' \\ x_1' \end{bmatrix} \begin{bmatrix} y & x_1 \end{bmatrix} = \begin{bmatrix} y'y & y'x_1 \\ x_1'y & x_1'x_1 \end{bmatrix} = \begin{bmatrix} 20 & 10 \\ 10 & 1 \end{bmatrix}$

so that y'y=20

Also X'X= $\begin{bmatrix} x_1'x_1 & x_1'x_2 \\ x_1'x_2 & x_2'x_2 \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ 4 & 2 \end{bmatrix}$

so $s_{IV}^2 = \dfrac{1}{100}\left(20 - 2\begin{bmatrix} 2 & 2 \end{bmatrix}\begin{bmatrix} 0.2 \\ 0.6 \end{bmatrix} + \begin{bmatrix} 0.2 & 0.6 \end{bmatrix}\begin{bmatrix} 1 & 4 \\ 4 & 2 \end{bmatrix}\begin{bmatrix} 0.2 \\ 0.6 \end{bmatrix}\right)$

= 1/100[20 - 3.2 +1.7]
=0.185

Hence $\text{Var}\left(\hat{b}_2^{IV}\right) = 0.185 \begin{bmatrix} x_3'x_1 & x_3'x_2 \\ x_2'x_1 & x_2'x_2 \end{bmatrix} \begin{bmatrix} x_3'x_3 & x_3'x_2 \\ x_2'x_3 & x_2'x_2 \end{bmatrix} \begin{bmatrix} x_1'x_3 & x_1'x_2 \\ x_2'x_3 & x_2'x_2 \end{bmatrix}$

$=0.185 \begin{bmatrix} 10 & 0 \\ 4 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 5 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 10 & 4 \\ 0 & 2 \end{bmatrix}^{-1}$

$=0.185 \begin{bmatrix} 0.05 & -0.1 \\ -0.1 & 0.7 \end{bmatrix}$

so $\text{Var}(\hat{b}_1^{IV}) = 0.185 * 0.05$ $=0.00925$

and $\therefore$ $s.e.\left(\hat{b}_1^{IV}\right) = 0.096$

so $\text{Var}\left(\hat{b}_2^{IV}\right) = 0.185 * 0.7$ $= 0.1295$

$s.e.\left(\hat{b}_2^{IV}\right) = 0.36$

(Note that the implied t values $t_{\beta_1} = \frac{0.2}{0.096} = 2.08$ $\quad \beta_1 = \frac{0.6}{0.36} = 1.67$

indicate that only β_1 is significantly different from zero at the 5% level)

7. Show that the efficiency ratio of OLS relative to IV in the 2 variable model is equal to the square of the correlation coefficient between X and the instrument Z.

- in the 2 variable model $\quad$ y = a + bx+u

$\text{Var}(\hat{b}^{ols}) = \frac{\sigma^2}{\sum x^2}$ $\quad$ $\text{Var}(\hat{b}^{IV}) = \frac{\sigma^2 \sum z^2}{\sum zx^2}$

So the relative efficiency of the two estimators is given by

$$\frac{Var(\hat{b}^{ols})}{Var(\hat{b}^{IV})} = \frac{\sigma^2 / \sum x^2}{\sigma^2 \sum z^2 / \sum zx^2}$$

$$= \frac{\sum zx^2}{\sum z^2 \sum x^2} = \frac{Cov(Z,X)^2}{Var(Z)Var(X)}$$

= the square of the correlation coefficient between X and Z
(Hence since z and x are never colinear then the relative efficiency of IV is always <1)

9. To show that $s_{IV}^2 = \frac{\left(y - X\hat{\beta}_{IV}\right)\left(y - X\hat{\beta}_{IV}\right)'}{N}$ is a consistent estimator of the unobserved true residual variance

Let the vector of IV residuals $\hat{u}^{IV} = y - X\,\hat{b}^{IV} = y - X[Z'X]^{-1}Z'y \qquad if \quad l = k$

sub. in for y from observed model

$= X\beta + u - X(Z'X)^{-1}Z'(X\beta + u)$
$= [I - X(Z'X)^{-1}Z']u$
$= M_Z u$

Since $s^2_{IV} = \frac{\left(y - X\hat{\beta}_{IV}\right)\left(y - X\hat{\beta}_{IV}\right)'}{N} = \frac{\hat{u}'_{IV}\hat{u}_{IV}}{N} = \frac{u'M'_Z uM_Z}{N}$

$$= \frac{u'[I - X(Z'X)^{-1}Z']'[I - X(Z'X)^{-1}Z']u}{N}$$

$$= \frac{u'u}{N} + \left(\frac{u'Z}{N}\right)\left(\frac{X'Z}{N}\right)^{-1}\left(\frac{X'X}{N}\right)\left(\frac{Z'X}{N}\right)^{-1}\left(\frac{Z'u}{N}\right) - 2\left(\frac{u'X}{N}\right)\left(\frac{X'Z}{N}\right)^{-1}\left(\frac{Z'u}{N}\right)$$

(since l=k then Z'X is a square matrix and its inverse exists)

using assumptions about relationship between instrumental variables and residuals ie
$\text{plim}\left(\frac{\sum z_i u_i}{N}\right) = 0$

then second term on right hand side is zero (in the limit)

Hence $s^2_{IV} \overset{p}{\to} \sigma^2$

10. Prove that when a set of variables, X_1 are regressed on an instrument set containing the same exogenous variables in addition to other variables then the predicted values from this regression equal X_1

- the 2SLS estimator is given by
$$\hat{\beta}_{2sls} = \left(\hat{X}'\hat{X}\right)^{-1}\hat{X}'y = \left[X'Z(Z'Z)^{-1}Z'X\right]^{-1}X'Z(Z'Z)^{-1}Z'y$$

where Z=[$X_1 : Z_2$] is N by l and X=[$X_1 : X_2$] is N by k
and X_2 is the sub-set of endogenous variables
So 2sls obtains a set of N predicted values for each of the k right hand side variables and then regresses y on these predicted X values
Consider the term $(Z'Z)^{-1}Z'X$
this gives an k by k matrix of estimated OLS coefficients from a regression of X on Z [Eg if L=3 & k=3 then $(Z'Z)^{-1}Z'X$ is a 3 by 3 matrix of coefficients]
and the ith column gives the set of estimates from the regression of the instruments Z on variable $X_i = \hat{\gamma}_i$

Hence the ith column of $Z(Z'Z)^{-1}Z'X_i = Z\hat{\gamma}_i = \hat{X}_i$
is the set of N predicted values for the variable X_i when it is regressed on all the variables in Z

and since Z contains the variable X_i (unless it is endogenous) then this is equivalent to running a regression of one variable on itself

(In 2 variable case $\hat{\gamma}_i = (X_i'X_i)^{-1}X_i'X_i = I_k$ and so $\hat{X}_i = Z\hat{\gamma}_i = X_iI_k$)

More generally let $\hat{\beta}_{2sls} = \left[X'Z(Z'Z)^{-1}Z'X\right]^{-1}X'Z(Z'Z)^{-1}Z'y = (X'P_ZX)^{-1}X'P_Zy$

where $X'P_ZX = \begin{bmatrix} X_1'P_ZX_1 & X_1'P_ZX_2 \\ X_2'P_ZX_1 & X_2'P_ZX_2 \end{bmatrix}$

The set of predicted values for the exogenous variables X$_1$ based on the instrument set Z is
$P_ZX_1 = Z(Z'Z)^{-1}Z'X_1$

$$=\begin{bmatrix} X_1 & Z_2 \end{bmatrix}\begin{bmatrix} X_1'X_1 & X_1'Z_2 \\ Z_2'X_1 & Z_2'Z_2 \end{bmatrix}^{-1}\begin{bmatrix} X_1' \\ Z_2' \end{bmatrix}X_1 \qquad (1)$$

Using rules on inverses of partitioned matrices (see for example Johnston & DiNardo appendix)

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \begin{bmatrix} b_{11} & -b_{11}a_{12}a_{22}^{-1} \\ -a_{22}^{-1}a_{21}b_{11} & a_{22}^{-1} + a_{22}^{-1}a_{21}b_{11}a_{12}a_{22}^{-1} \end{bmatrix}$$

where b$_{11} = (a_{11} - a_{12}a_{22}^{-1}a_{21})^{-1}$

This means that the b$_{11} = \left[X_1'X_1 - X_1'Z_2(Z_2'Z_2)^{-1}Z_2'X_1\right] = (X_1'M_{Z_2}X_1)^{-1}$

where $M_{Z_2} = I - Z_2(Z_2'Z_2)^{-1}Z_2'$
So the 1st row of $(Z'Z)^{-1}Z'X_1$ in (1), which is the portion of the matrix that isolates the effect of X$_1$
on X$_1$in the regression of X$_1$ on Z=[X$_1$: Z_2] is given by

$$\left(b_{11} * X_1'Z_2\right) + \left(-b_{11}a_{12}a_{22}^{-1} * Z_2'X_1\right)$$

$$= (X_1'M_{Z_2}X_1)^{-1} * X_1'X_1 - \left((X_1'M_{Z_2}X_1)^{-1}X_1'Z_2(Z_2'Z_2)^{-1} * Z_2'X_1\right)$$

$$= (X_1'M_{Z_2}X_1)^{-1}\left[X_1'X_1 - X_1'Z_2(Z_2'Z_2)^{-1}Z_2'X_1\right]$$

$$=(X_1'M_{Z_2}X_1)^{-1}X_1'\left[I - Z_2(Z_2'Z_2)^{-1}Z_2'\right]X_1$$
$$=(X_1'M_{Z_2}X_1)^{-1}(X_1'M_{Z_2}X_1)$$
$$=I_k \qquad (2)$$

Hence the coefficient in a regression of X_1 on itself is an identity matrix with dimensions equal to the number of columns in X_1

Now using the alternative formulation for the inverse of a partitioned matrix

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}^{-1} = \begin{bmatrix} a_{11}^{-1} + a_{11}^{-1}a_{12}b_{22}a_{21}a_{11}^{-1} & -a_{11}^{-1}a_{12}b_{22} \\ -b_{22}a_{21}a_{11}^{-1} & b_{22} \end{bmatrix}$$

where

$b_{22} = (a_{22} - a_{21}a_{11}^{-1}a_{12})^{-1} = \left[Z_2'Z_2 - Z_2'X_1(X_1'X_1)^{-1}X_1'Z_2\right] = (Z_2'M_{X_1}Z_2)^{-1}$

where $M_{X_1} = I - X_1(X_1'X_1)^{-1}X_1'$

Hence can show that the second row of $(Z'Z)^{-1}Z'X_1$ is given by

$$\left(-b_{22}a_{21}a_{11}^{-1} * X_1'X_1\right) + \left(b_{22} * Z_2'X_1\right) = \left(-(Z_2'M_{X_1}Z_2)^{-1}Z_2'X_1(X_1'X_1)^{-1}X_1'X_1\right) +$$

$$= -(Z_2'M_{X_1}Z_2)^{-1}Z_2'X_1 + \left((Z_2'M_{X_1}Z_2)^{-1}Z_2'X_1\right)$$

$= 0$ (3)

(so the regression of the instruments on X_1 net of X_1 gives a set of zero coefficients)

Together (2) & (3) imply that $\hat{\gamma}_1 = (Z'Z)^{-1}Z'X_1 = \begin{bmatrix} I_k \\ 0 \end{bmatrix}$

so the set of predicted values $Z\hat{\gamma}_1 = Z(Z'Z)^{-1}Z'X_1 = P_ZX_1 = [X_1 : Z_2]\begin{bmatrix} I_k \\ 0 \end{bmatrix} = X_1$

- including the original exogenous variables from the model among the set of instruments is equivalent to including these original X_1 variables in the second stage of 2sls

11. Show why the Hausman test for exogeneity can be derived as a test of the joint explanatory power of the residuals, rhat, (from regressions of each of the endogenous variables on the exogenous variables in the system) included as additional variables in the original structural form.

$y = X_1 \beta + X_2\delta + \gamma$ rhat + residuals (A)

(This exposition draws on Johnston &DiNardo ch. 10. Interested readers might also want to consult this)

- remember that the Frisch-Waugh theorem says that the estimates for B can be obtained by an alternative regression that first nets out the influence of other variables

Consider then the original model

$y = X_1 B + X_2\delta + \epsilon$

where Z=[$X_1 : Z_2$] is N by l and X=[$X_1 : X_2$] is N by k

(so that X_2 is endogenous)

Now consider a vector of contrasts between the OLS estimator and any other linear estimator given by

$$\hat{\beta}_A = \left(X'AX\right)^{-1}X'Ay$$

-eg think of the IV estimator $\hat{\beta}_{IV} = \left(X'P_ZX\right)^{-1}X'P_Zy$

It follows that the difference $\hat{\beta}_A - \hat{\beta}_{OLS} = \left[\left(X'AX\right)^{-1}X'Ay - \left(X'X\right)^{-1}X'y\right]$

$$= \left(X'AX\right)^{-1}X'AM_Xy \qquad (1)$$

where $M_X = [I - X(X'X)^{-1}X']$

This applies to any linear estimator. In general will be interested in the asymptotic properties of this difference

Consider $\text{plim}\left(\frac{X'AX}{N}\right)^{-1}$

since this is a k by k matrix of (effectively squared) values it will stay finite as the sample size gets infinitely large

So consistency depends on the behaviour of

$$\text{plim}\left(\frac{X'AM_Xy}{N}\right) \qquad (2)$$

if it equals zero then the coefficient estimates produced by the two procedures are asymptotically equivalent

In the case of 2SLS then A=P_Z=Z(Z'Z)$^{-1}Z'$

and we are interested in whether $\text{plim}\left(\frac{X'P_ZM_Xy}{N}\right) = 0$

Given X=[$X_1 : X_2$]

question 6 establishes that $P_ZX = \hat{X} = \left[\hat{X}_1 : \hat{X}_2\right]$

so $X'P_Z = \begin{bmatrix} \hat{X}_1' \\ \hat{X}_2' \end{bmatrix}$

If X_1 is exogenous and Z contains X_1 question 6 also shows that $\hat{X}_1 = X_1$

so $X'P_Z = \begin{bmatrix} X_1' \\ \hat{X}_2' \end{bmatrix}$

and therefore $X'P_Z M_X$ in (3) = $\begin{bmatrix} X_1' \\ \hat{X}_2' \end{bmatrix} M_X = \begin{bmatrix} 0 \\ \hat{X}_2' M_X \end{bmatrix}$

since the residuals from a regression of X_1 on Z which contains X_1 will be zero

this means that $\text{plim}\left(\frac{X'P_Z M_X y}{N}\right) = 0$ iff $\text{plim}\left(\frac{\hat{X}_2' M_X y}{N}\right) = 0$

Since $\hat{X}_2' M_X y$ is equivalent to the effect of $\hat{X}_2$ on y after netting out the influence of X (the variables contained in M_X)

on $\hat{X}_2$

then can base a test for consistency on the following model

$$y = XB + \hat{X}_2 \delta + u$$

and use an F test for the joint significance of the δ vector
(if estimated F > critical F value, reject the null that the two estimation procedures are asymptotically equivalent)

This is one way to test for endogeneity of variables using the Hausman logic of using a vector of contrasts

Now consider a different way of examining the issue.

Compare model (A) with an augmented model that includes the *additional* instruments in Z
$y = XB + Z_2\gamma + \eta$ (B)

The Frisch-Waugh theorem tells us that

$$\hat{B} = \left(X'M_{Z2}'M_{Z2}X\right)^{-1}X'M_{Z2}'M_Z y = \left(X'M_{Z2}X\right)^{-1}X'M_Z y = \left(X'M_{Z2}'M_{Z2}X\right)^{-1}X'M_Z y$$

where the idempotent matrix $M_{Z2} = [I - Z_2(Z_2'Z_2)^{-1}Z_2']$ and $M_{Z2}'M_{Z2} = M_{Z2}$

and $M_{Z2}y$ is the vector of residuals from a regression of y on Z_2
$M_{Z2}X$ is the matrix of residuals from a regression of X on Z_2

So an alternative way to get an estimate of B is to

1) regress X on Z_2 and save the residuals
2) regress these residuals on y

The regression of X on Z_2 gives coefficient estimates $(Z_2^{\prime}Z_2)^{-1}Z_2^{\prime}X$

and predicted values $\hat{X} = P_{Z_2}X = Z_2(Z_2^{\prime}Z2)^{-1}Z_2^{\prime}X$

The residuals from this regression $M_{Z_2}X = [I - Z_2(Z_2^{\prime}Z2)^{-1}Z_2^{\prime}]X$

so the second step can be formulated as

$$y = M_{Z_2}XB + \upsilon$$

and the associated OLS estimate of B from this model is as in (3)

$$\hat{B} = \left(X^{\prime}M_{Z2}^{\prime}M_{Z_2}X\right)^{-1}X^{\prime}M_{Z2}^{\prime}y = \left(X^{\prime}M_{Z_2}X\right)^{-1}X^{\prime}M_{Z}y$$

In (1) this means that now A=M_{Z2}

and so using the logic above, the vector of contrasts is influenced by whether

$$\text{plim}\left(\frac{X^{\prime}AM_{X}y}{N}\right) = \text{plim}\left(\frac{X^{\prime}M_{Z2}M_{X}y}{N}\right) = 0$$

as above in the 2sls case X'A reduces to $\hat{X}_2$

similarly $X^{\prime}M_{Z2}$ reduces to the set of residuals from a from a regression of X on Z_2

which is just the residuals from a regression of the variables in the original (structural) model on the set of instruments
and because the residuals from the sub-set of exogenous variables in X will be zero this is equivalent to
an alternative regression to test endogeneity

$$y = XB + u\gamma_{X_2.Z}\gamma + \upsilon \qquad (4)$$

and test for the significance of $\hat{\gamma}$

If $\hat{\gamma} \neq 0$ then reject null that the two estimation procedures produce the same coefficient estimates

[Note that the estimates (but not the standard errors) of B in (4) will be identical to those from the 2SLS model. This gives a way to compare OLS and 2SLS by simply removing the residual term and comparing the difference

Proof:

Let X=$[X_1 : X_2]$ and $\hat{w} = u_{X_2.Z}$

(so that $\hat{w}$ is the residual from an ols regression of the endogenous variable X_2 in the original regression on all the instruments Z=[$X_1 : Z_2$]
$X_2 = Z\delta + w$)

Now consider the model y $= XB + w\gamma + \upsilon = Q\rho + \upsilon$ (5)

where Q=[X:W] $\rho = \begin{bmatrix} B \\ \gamma \end{bmatrix}$

so OLS on (5) gives $\hat{\rho}' = [\hat{B}' : \hat{\gamma}']$

But the Frisch-Waugh theorem says that $\hat{B}$ can also be obtained from a two-step regression of
i) regress X=$[X_1 : X_2]$ on w and save the residuals
ii) regress y on these residuals

note that a regression of the exogenous variables X_1 on w will be orthogonal to X_1 because the algebra of OLS tells us that the residuals from a regression are uncorrelated (orthogonal) to the right hand side variables so $\sum_i \hat{w} X_1 = \sum_i \hat{w} Z_2 = 0$

This means that an OLS regression of $\hat{w}$ on X_1 will give an estimated coefficient of zero (see also question 6) and hence a residual = X_1

Also because by construction $X_2 = \hat{X}_2 + \hat{w}$ then any regression of X_2 on $\hat{w}$ must give an OLS estimate of 1 (see also question 6) and a residual of $\hat{X}_2$

so the residuals from the 1st step of the Frisch-Waugh analysis = (X_1 $\hat{X}_2$)

and regressing these residuals on y in the 2nd step of Frisch-Waugh is therefore exactly the same as the second step of 2SLS

but since the Frisch-Waugh is also equivalent to netting out the effect of the variable in a single regression then

y $= X_1B_1 + X_2B_2 + \hat{w}\gamma + \upsilon$

will give the same coefficient estimates (but not standard errors) as the 2SLS estimator

This can be useful to compare the difference between OLS from

y $= X_1 B_1 + X_2 B_2 + e$

and IV estimates (though stata's output from the Hausman test will also do this)