INSTRUMENTAL VARIABLES/ EXOGENEITY TESTS

$$y = X\beta + \epsilon,$$

y is $n\times1$, X is $n\times k$, β is $k\times1$, and ϵ is $n\times1$.

Suppose that *contamination* of X, where some of the X variables are correlated with ϵ, is suspected. This can occur if

- **ϵ contains omitted variables that are correlated with the included variables**
- **X contains measurement errors**
- **X contains endogenous variables that are determined jointly with y.**

***OLS Revisited*: Premultiply the regression equation by X' to get**

(1) $$X'y = X'X\beta + X'\epsilon.$$

One can interpret the OLS estimate b_{OLS} as the solution obtained from (1) by first approximating $X'\epsilon$ by zero, and then solving the resulting k equations in k unknowns,

(2) $$X'y = X'Xb_{OLS}.$$

Subtracting (1) from (2),

(3) $$X'X(b_{OLS} - \beta) = X'\epsilon,$$

The error in estimating β is linear in the error caused by approximating $X'\epsilon$ by zero.

If $X'X/n \to_p A$ positive definite and $X'\epsilon/n \to_p 0$, (3) implies $b_{OLS} \to_p \beta$.

If one has instead $X'\epsilon/n \to_p C \neq 0$, then b_{OLS} is <u>not</u> consistent for β, and instead $b_{OLS} \to_p \beta + A^{-1}C$.

***<u>Instrumental</u> <u>Variables</u>*: Suppose there is a n×j array of variables W, called *instruments*, that have two properties:**

(i) These variables are uncorrelated with ϵ; we say in this case that these instruments are *clean*.

(ii) The matrix of correlations between the variables in X and the variables in W is of maximum possible rank (= k); we say in this case that these instruments are *fully correlated*.

Call the instruments *proper* if they satisfy (i) and (ii). <u>The W array should include any variables from X that are themselves clean</u>.

To be fully correlated, W must include at least as many variables as are in X, so that $j \geq k$.

The number of instruments in W that are excluded from X must be at least as large as the number of contaminated variables that are included in X.

Instead of premultiplying the regression equation by X' as we did for OLS, premultiply it by $R'W'$, where R is a j×k weighting matrix that we get to choose. This gives

$$(4) \qquad R'W'y = R'W'X\beta + R'W'\varepsilon.$$

The idea of an instrumental variables (IV) estimator of β is to approximate $R'W'\varepsilon$ by zero, and solve

$$(5) \qquad R'W'y = R'W'X\, b_{IV}$$

for $b_{IV} = [R'W'X]^{-1}R'W'y$. Subtract (4) from (5) to get the IV analog of the OLS relationship (3),

$$(6) \qquad R'W'X(b_{IV} - \beta) = R'W'\varepsilon.$$

If $R'W'X/n$ converges in probability to a nonsingular matrix and $R'W'\varepsilon/n \to_p 0$, then $b_{IV} \to_p \beta$. Thus, in problems where OLS breaks down due to correlation of right-hand-side variables and the disturbances, you can use IV to get consistent estimates, provided you can find proper instruments.

The idea behind (5) is that W and ϵ are orthogonal in the population, a generalized moment condition. Then, (5) can be interpreted as the solution of a generalized method of moments problem, based on the sample moments $W'(y - X\beta)$. The properties of the IV estimator can be deduced as a special case of the general theory of GMM estimators.

If there are exactly as many instruments as there are explanatory variables, j = k, then the IV estimator is uniquely determined, $b_{IV} = (W'X)^{-1}W'y$, and R is irrelevant. However, if j > k, each R determines a different IV estimator. What is the best R? Premultiplying the regression equation by W' yields a system of j > k equations in k unknown β's, $W'y = W'X\beta + W'\epsilon$. Since there are more equations than unknowns, we cannot simply approximate all the $W'\epsilon$ terms by zero simultaneously, but will have to accommodate at least j-k non-zero residuals. But this is just like a regression problem, with j observations, k explanatory variables, and disturbances $\nu = W'\epsilon$. Suppose the disturbances ϵ have a covariance matrix $\sigma^2 I$, and hence the disturbances $\nu = W'\epsilon$ have a non-scalar covariance matrix $\sigma^2 I$. Then

$$(W'W)^{-1/2}W'y = (W'W)^{-1/2}W'X\beta + (W'W)^{-1/2}W'\epsilon$$

has a scalar covariance matrix, and when solved after approximating the last term by zero yields

(8) $$b_{2SLS} = [X'W(W'W)^{-1}W'X]^{-1}X'W(W'W)^{-1}W'y.$$

This corresponds to using the weighting matrix $R = (W'W)^{-1}W'X$. But this formula provides another interpretation of (8). If you regress each variable in X on the instruments, the resulting OLS coefficients are $(W'W)^{-1}W'X$, the same as R. Then, the best linear combination of instruments WR equals the fitted value $X^* = W(W'W)^{-1}W'X$ of the explanatory variables from a OLS regression of X on W. Further, you have $X'W(W'W)^{-1}W'X = X'X^* = X^{*\prime}X^*$ and $X'W(W'W)^{-1}W'y = X^{*\prime}y$, so that the IV estimator (8) can also be written

(9) $$b_{2SLS} = (X^{*\prime}X)^{-1}X^{*\prime}y = (X^{*\prime}X^*)^{-1}X^{*\prime}y.$$

This provides a two-stage least squares (2SLS) interpretation of the IV estimator: First, a OLS regression of the explanatory variables X on the instruments W is used to obtain fitted values X^*, and second a OLS regression of y on X^* is used to obtain the IV estimator b_{2SLS}. Note that in the first stage, any variable in X that is also in W will achieve a perfect fit, so that this variable is carried over without modification in the second stage.

STATISTICAL PROPERTIES OF IV ESTIMATORS

IV estimators can behave badly in finite samples. In particular, they may fail to have moments. Their appeal relies on their behavior in large samples, although an important question is when samples are large enough so that the asymptotic approximation is reliable.

We show next that IV estimators are asymptotically normal Let $\sigma^2\Omega$ be the covariance matrix of ε, given W, and assume that it is finite and of full rank. Make the assumptions:

[1] $\mathrm{rank}(W) = j \geq k$
[2a] $W'W/n \to_p H$, a positive definite matrix
[2b] $W'\Omega W/n \to_p F$, a positive definite matrix
[3] $X'W/n \to_p G$, a matrix of rank k
[4] $W'\varepsilon/n \to_p 0$
[5] $n^{-1/2}W'\varepsilon \to_d N(0,\sigma^2F)$

Theorem: Assume that [1], [2b], [3] hold, and that an IV estimator is defined with a weighting matrix R_n that may depend on the sample n, but which converges to a matrix R of rank k. If [4] holds, then $b_{IV} \to_p \beta$. If both [4] and [5] hold, then

$$n^{1/2}(b_{IV} - \beta) \to_d N(0, \sigma^2(R'G')^{-1}R'FR(GR)^{-1}). \tag{10}$$

Suppose $R_n = (W'W)^{-1}W'X$ and [1]-[5] hold. Then the

IV estimator specializes to the 2SLS estimator b_{2SLS} given by (8) which satisfies $b_{2SLS} \to_p \beta$ and

$$\text{(11)}\quad n^{1/2}(b_{2SLS} - \beta) \to_d N(0, \sigma^2(GH^{-1}G')^{-1}(GH^{-1}FH^{-1}G')(GH^{-1}G')^{-1}).$$

If, further, $\Omega = I$,

$$\text{(12)}\qquad n^{1/2}(b_{2SLS} - \beta) \to_d N(0, \sigma^2(GH^{-1}G')^{-1}).$$

In order to use the large-sample properties of b_{IV} for hypothesis testing, it is necessary to find a consistent estimator for σ^2. The following estimator works: Define IV residuals

$$\begin{aligned} u = y - Xb_{IV} &= [I - X(R'W'X)^{-1}R'W']y \\ &= [I - X(R'W'X)^{-1}R'W']\varepsilon, \end{aligned}$$

the *Sum of Squared Residuals* SSR = $u'u$, and $s^2 = u'u/(n-k)$. If $\varepsilon'\varepsilon/n \to_p \sigma^2$, then s^2 is consistent for σ^2. To show this, simply write out the expression for $u'u/n$, and take the probability limit:

$$\begin{aligned} \text{(13)}\quad & \text{plim } u'u/n \\ &= \text{plim } \varepsilon'\varepsilon/n - 2 \text{ plim } [\varepsilon'W/n]R([X'W/n]R)^{-1}[X'\varepsilon/n] \\ &+ [\varepsilon'W/n]R([X'W/n]R)^{-1}[X'X/n](R'[W'X/n])^{-1}R'[W'\varepsilon/n] \\ &= \sigma^2 - 2\cdot 0'R'(GR)^{-1}C + 0'R'(GR)^{-1}A(R'G')^{-1}R'0 = \sigma^2. \end{aligned}$$

We could have used n-k instead of n in the denominator of this limit, as it makes no difference in large enough samples. The consistency of the estimator s^2 defined above holds for any IV estimator, and so holds in particular for the 2SLS estimators. Note that this consistent estimator of σ^2 substitutes the IV estimates of the coefficients into the <u>original</u> equation, and uses the original values of the X variables to form the residuals. When working with the 2SLS estimator, and calculating it by running the two OLS regression stages, you might be tempted to estimate σ^2 using a regression program printed values of SSR or the variance of the second stage regression, which is based on the residuals $\hat{u} = y - X^* b_{2SLS}$. This estimator is <u>not</u> consistent for σ^2.

Suppose EggN = σ^2I, so that 2SLS is best among IV estimators using instruments W. The sum of squared residuals SSR = uNu, where $u = y - Xb_{2SLS}$, can be used in hypothesis testing in the same way as in OLS estimation. For example, consider the hypothesis that $\beta_2 = 0$, where β_2 is a $r \times 1$ subvector of β. Let SSR_0 be the sum of squared residuals from the 2SLS regression of y on X with $\beta_2 = 0$ imposed, and SSR_1 be the sum of squared residuals from the unrestricted 2SLS regression of y on X. Then, $[(SSR_0 - SSR_1)/m]/[SSR_1/(n-k)]$ has an approximate F-distribution under the null with m and n-k degrees of freedom. There are several cautions to keep in mind when considering use of this test statistic. This is a large sample approximation, rather than an exact distribution, because it is derived from the asymptotic normality of the 2SLS estimator. Its actual size in small samples could differ substantially from its nominal (asymptotic) size. Also, the large sample distribution of the statistic assumed that the disturbances ϵ have a scalar covariance matrix.

What are the finite sample properties of IV estimators? Because you do not have the condition $E(\epsilon^*X) = 0$ holding in applications where IV is needed, you cannot get simple expressions for the moments of $b_{IV} = [R'W'X]^{-1}R'W'y = \beta + [R'W'X]^{-1}R'W'\epsilon$ by first taking expectations of ϵ conditioned on X and W. In particular, you cannot conclude that b_{IV} is unbiased, or that it has a covariance matrix given by its asymptotic covariance matrix. In fact, b_{IV} can have very bad small-sample properties. To illustrate, consider the case where the number of instruments equals the number of observations, j = n. (This can actually arise in dynamic models, where often all lagged values of the exogenous variables are legitimate instruments. It can also arise when the candidate instruments are not only uncorrelated with ϵ, but satisfy the stronger property that $E(\epsilon^*w) = 0$. In this case, all functions of w are also legitimate instruments.) In this case, W is a square matrix, and

$$b_{2SLS} = [X'W(W'W)^{-1}W'X]^{-1}X'W(W'W)^{-1}W'y$$
$$= [X'X]^{-1}X'y = b_{OLS}.$$

We know OLS is inconsistent when E(g*X) = 0 fails, so clearly the 2SLS estimator is also biased if we let the number of instruments grow linearly with sample size. This shows that for the IV asymptotic theory to be a good approximation, n must be much larger than j. One rule-of-thumb for IV is that n - j should exceed 40, and should grow linearly with n in order to have the large-sample approximations to the IV distribution work well.

Considerable technical analysis is required to characterize the finite-sample distributions of IV estimators analytically. However, simple numerical examples provide a picture of the situation. Consider a regression y = xβ + g where there is a single right-hand-side variable, and a single instrument w, and assume x, w, and g have the simple joint distribution given in the table below, where λ is the correlation of x and w, ρ is the correlation of x and g, and |λ| + |ρ| < 1. The interpretation of the second row of the table, for example, is that (x,w,ε) = (1,1,-1) and (x,w,ε) = (-1,-1,1) each occur with probability (1-ρ+λ)/8:

x	w	g	Prob
±1	±1	±1	(1+ρ+λ)/8
±1	±1	K1	(1-ρ+λ)/8
±1	K1	±1	(1+ρ-λ)/8
±1	K1	K1	(1-ρ-λ)/8

The random variables (x,w,ε) have mean zero, variance one, and $Ex\varepsilon = \rho$, $Exw = \lambda$, and $Ew\varepsilon = 0$. Their products have the joint distribution

xw	$x\varepsilon$	$w\varepsilon$	Prob
1	1	1	$(1+\rho+\lambda)/4$
1	-1	-1	$(1-\rho+\lambda)/4$
-1	1	-1	$(1+\rho-\lambda)/4$
-1	-1	1	$(1-\rho-\lambda)/4$

This implies $P(x\varepsilon=1) = (1+\rho)/2$. Then, in a sample of size n, $n((b_{OLS} - \beta) + 1)/2$ has an exact distribution that is binomial with n draws and probability $(1+\rho)/2$. Then $n^{1/2}(b_{OLS} - \beta)$ has mean $n^{1/2}\rho$ and variance $(1-\rho^2)$. Thus, $n\cdot MSE = n\cdot(Variance + Bias^{2)} = 1 + (n-1)\rho^2$.

Exercise: Draw 1000 samples of various sizes from the distribution above, calculate b_{IV}, calculate selected points of its CDF, and compare these with corresponding points of the CDF for b_{OLS}.

In practice, in problems where sample size minus the number of instruments exceeds 40, the asymptotic approximation to the distribution of the IV estimator is reasonably good, and one can use it to compare the OLS and IV estimates. To illustrate, continue the example of a regression in one variable, $y = x\beta + g$. Suppose as before that x and g have a correlation coefficient ρ … 0, so that OLS is biased, and suppose that there is a single proper instrument w that is uncorrelated with g and has a correlation λ … 0 with x. Then, the OLS estimator is asymptotically normal with mean $\beta + \rho\sigma_g/\sigma_x$ and variance $\sigma_g^2/n\sigma_x^2$. The 2SLS estimator is asymptotically normal with mean β and variance $\sigma_g^2/n\sigma_x^2\lambda^2$. The mean squares of the two estimators are then, approximately,

$$\mathrm{MSE}_{OLS} = (\rho^2 + 1/n)\sigma_g^2/\sigma_x^2$$
$$\mathrm{MSE}_{2SLS} = \sigma_g^2/n\sigma_x^2\lambda^2.$$

Then, 2SLS has a lower MSE than OLS when

$$1 < \rho^2\lambda^2 n/(1-\lambda^2) \;.\; (b_{2SLS}-b_{OLS})^2/(V(b_{2SLS})-V(b_{OLS})),$$

or approximately $n > (1 - \lambda^2)/\rho^2\lambda^2$. When $\lambda = 0.8$ and $\rho = 0.2$, this asymptotic approximation suggests that a sample size of about 14 is the tip point where b_{IV} should be better than b in terms of MSE. However, the asymptotic formula underestimates the probability of very large deviations

arising from a denominator in b_{IV} that is near zero, and as a consequence is too quick to reject b_{OLS}. The right-hand-side of this approximation to the ratio of the MSE is the Hausman test statistic for exogeneity, discussed below; for this one-variable case, one should reject the null hypothesis of exogeneity when the statistic exceeds one. Under the null, the statistic is approximately chi-square with one degree of freedom, so that this criterion corresponds to a type I error probability of 0.317.

RELATION OF IV TO OTHER ESTIMATORS

The 2SLS estimator can be interpreted as a member of the family of *Generalized Method of Moments* (GMM) estimators. You can verify by differentiating to get the first-order condition that the 2SLS estimator of the equation $y = X\beta + \varepsilon$ using the instruments W, where $E\varepsilon\varepsilon' = \sigma^2 I$, solves

$$(14) \qquad \mathrm{Min}_{\beta}\ (y-X\beta)'W(W'W)^{-1}W'(y-X\beta).$$

In this quadratic form objective function, $W'(y-X\beta)$ is the moment that has expectation zero in the population when β is the true parameter vector, and $(W'W)^{-1}$ is a "distance metric" in the center of the quadratic form. Define $P = W(W'W)^{-1}W'$, and note that P is idempotent, and thus is

a projection matrix. Then, the GMM criterion chooses β to minimize the length of the vector y- Xβ projected onto the subspace spanned by *P*. The properties of GMM hypothesis testing procedures follow readily from the observation that y-Xβ has mean zero and a scalar covariance matrix. In particular, $\mathrm{Min}_\beta\ (y-X\beta)'W(W'W)^{-1}W'(y-X\beta)/\sigma^2$ is asymptotically chi-squared distributed with degrees of freedom equal to the rank of *P*.

It is possible to give the 2SLS estimator a *pseudo-MLE* interpretation. Premultiply the regression equation by W' to obtain $W'y = W'X\beta + W'\epsilon$. Now treat $W'\epsilon$ *as if* it were normally distributed with mean zero and j×j covariance matrix $\lambda^2 W'W$, conditioned on $W'X$. Then, the log likelihood of the sample would be

$$L = -\,(j/2)\log 2\pi - (j/2)(½)\log \lambda^2 - (½)\log \det(W'W)$$
$$-\,(1/2\lambda^2)(W'y-W'X\beta)'(W'W)^{-1}(W'y-W'X\beta).$$

The first-order condition for maximization of this pseudo-likelihood is the same as the condition defining the 2SLS estimator.

TESTING EXOGENEITY

Sometimes one is unsure whether some potential instruments are clean. If they are, then there is an asymptotic efficiency gain from including them as instruments. However, if they are not, estimates will be inconsistent. Because of this tradeoff, it is useful to have a specification test that permits one to judge whether suspect instruments are clean or not. To set the problem, consider a regression $y = X\beta + g$, an array of proper instruments Z, and an array of instruments W that includes Z plus other variables that may be either clean or contaminated.

Several superficially different problems can be recast in this framework:

(1) The regression may be one in which some right-hand-side variables are known to be exogenous and others are suspect, Z is an array that contains the known exogenous variables and other clean instruments, and W contains Z and the variables in X that were excluded from Z because of the possibility that they might be dirty. In this case, 2SLS using W reduces to OLS, and the problem is to test whether the regression can be estimated consistently by OLS.

(2) The regression may contain known endogenous and known exogenous variables, Z is an array that contains the known exogenous variables and other clean instruments, and W is an array that contains Z and additional suspect instruments from outside the equation. In this case, one has a consistent 2SLS estimator using instruments Z, and a 2SLS estimator using instruments W that is more efficient under the hypothesis that W is exogenous, but inconsistent otherwise. The question is whether to use the more inclusive array of instruments.

(3) The regression may contain known endogenous, known exogenous, and suspect right-hand-side variables, Z is an array that contains the known exogenous variables plus other instruments from outside the equation, and W is an array that contains Z plus the suspect variables from the equation. The question is whether it is necessary to instrument for the suspect variables, or whether they are clean and can themselves be used as instruments.

In the regression $y = X\beta + \varepsilon$, you can play it safe and use only the Z instruments. This gives $b_Q = (X'QX)^{-1}X'Qy$, where $Q = Z(Z'Z)^{-1}Z'$. Alternately, you use W, including the suspect instruments, taking a chance with inconsistency to gain efficiency. This gives

$$b_P = (X'PX)^{-1}X'Py, \text{ where } P = W(W'W)^{-1}W'.$$

If the suspect instruments are clean and both estimators are consistent, then b_Q and b_P should be close together, as they are estimates of the same β; further, b_P is efficient relative to b_Q, implying that the covariance matrix of $(b_Q - b_P)$ equals the covariance matrix of b_Q minus the covariance matrix of b_P. However, if the suspect instruments are contaminated, b_P is inconsistent, and $(b_Q - b_P)$ has a nonzero probability limit. This suggests a test statistic of the form

$$(15) \qquad (b_Q - b_P)'[V(b_Q) - V(b_P)]^{-}(b_Q - b_P),$$

where $[\cdot]^{-}$ denotes a generalized inverse, could be used to test if W is clean. This form is the exogeneity test originally proposed by Hausman. Under the null hypothesis that W is clean, this statistic will be asymptotically chi-square with degrees of freedom equal to the rank of the covariance matrix in the center of the quadratic form.

Another formulation of an exogeneity test is more convenient to compute, and can be shown (in one manifestation) to be equivalent to the Hausman test statistic. This alternative formulation has the form of an omitted variable test, with appropriately constructed auxiliary variables.

First do an OLS regression of X on Z and retrieve fitted values $X^* = QX$, where $Q = Z(Z'Z)^{-1}Z'$. (This is necessary only for variables in X that are not in Z, since otherwise this step just returns the original variable.) Second, using W as instruments, do a 2SLS regression of y on X, and retrieve the sum of squared residuals SSR_1. Third, do a 2SLS regression of y on X and a subset of m columns of X^* that are linearly independent of X, and retrieve the sum of squared residuals SSR_2. Finally, form the statistic $[(SSR_1 - SSR_2)/m]/[SSR_2/(n-k)]$. Under the null hypothesis that W is clean, this statistic has an approximate F-distribution with m and n-k degrees of freedom, and can be interpreted as a test for whether the m auxiliary variables from X^* should be omitted from the regression. When a subset of X^* of maximum possible rank is chosen, this statistic turns out to be asymptotically equivalent to the Hausman test statistic. Note that if W contains X, then the 2SLS in the second and third steps reduces to OLS.

We next show that this test is indeed an exogeneity test. Consider the 2SLS regression

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{X}_1^{*}\gamma + \eta,$$

where X_1^* is a subset of $X^* = QX$ such that $[X,X_1^*]$ is of full rank. The 2SLS estimates of the parameters in this model, using W as instruments, satisfy

$$\begin{bmatrix} b_P \\ c_P \end{bmatrix} = \begin{bmatrix} X'PX & X'QX_1 \\ X_1'QX & X_1'QX_1 \end{bmatrix}^{-1} \begin{bmatrix} X'Py \\ X_1'Qy \end{bmatrix} = \begin{bmatrix} \beta \\ 0 \end{bmatrix} - \begin{bmatrix} X'PX & X'QX_1 \\ X_1'QX & X_1'QX_1 \end{bmatrix}^{-1} \begin{bmatrix} X'P\varepsilon \\ X_1'Q\varepsilon \end{bmatrix}.$$

But $X'Q\varepsilon/n \to_p \text{plim}(X'Z/n)\cdot(\text{plim}(Z'Z/n))^{-1}\cdot\text{plim}(Z'\varepsilon/n) = 0$ by assumptions [1]-[4] when Z is clean. Similarly, $X'P\varepsilon/n \to_p GH^{-1}\cdot\text{plim}(W'\varepsilon/n) = 0$ when W is clean, but $X'P\varepsilon/n \to_p GH^{-1}\cdot\text{plim}(W'\varepsilon/n) \neq 0$ when W is contaminated. Define

$$\begin{bmatrix} X'PX/n & X'QX_1/n \\ X_1'QX/n & X_1'QX_1/n \end{bmatrix}^{-1} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

From the formula for a partitioned inverse,

$$A_{11} = (X'[P - QX_1(X_1'QX_1)^{-1}X_1'Q]X/n)^{-1}$$

$$A_{22} = (X_1'Q[I - X(X'PX)^{-1}X']QX_1/n)^{-1}$$

$$A_{21} = -(X_1'QX_1)^{-1}X_1'QX\cdot A_{11}$$

$$= -A_{22}(X_1'QX)(X'PX)^{-1} = A_{12}'$$

Hence,

$$(16) \qquad c_P = A_{22}\cdot\{X_1'Q\varepsilon/n - (X_1'QX)(X'PX)^{-1}\cdot X'P\varepsilon/n\}.$$

If W is clean and satisfies assumptions [4] and [5], then $c_P \to_p 0$ and $n^{1/2}c_P$ is asymptotically normal. On the other hand, if W is contaminated, then c_P has a non-zero

probability limit. Then, a test for $\gamma = 0$ using c_P is a test of exogeneity.

The test above can be reinterpreted as a Hausman test involving differences of b_P and b_Q. Recall that $b_Q = \beta + (X'QX)^{-1}X'Q\varepsilon$ and $b_P = \beta + (X'PX)^{-1}X'P\varepsilon$. Then

(17) $$(X'QX)(b_Q - b_P) = \{X'Q\varepsilon/n - (X'QX)(X'PX)^{-1}{''}X'P\varepsilon/n\}.$$

Then in particular for a linearly independent subvector X_1 of X,

$$\begin{gathered} A_{22}(X_1'QX)(b_Q - b_P) \\ = A_{22}\{X_1'Q\varepsilon/n - (X_1'QX)(X'PX)^{-1}{''}X'P\varepsilon/n\} = c_P. \end{gathered}$$

Thus, c_P is a linear transformation of $(b_Q - b_P)$. Then, testing whether c_P is near zero is equivalent to testing whether a linear transformation of $(b_Q - b_P)$ is near zero. When X_1 is of maximum rank, this equivalence establishes that the Hausman test in its original form is the same as the test for c_P.

EXOGENICITY TESTS ARE GMM TESTS FOR OVER-IDENTIFICATION

***The Hausman Exogeneity Test*. Consider the regression model $y = X\beta + \varepsilon$, and suppose one wants to test the exogeneity of p variables X_1 in X. Suppose R is an array of instruments, including X_2; then $Z = P_R X_1$ are instruments for X_1. Let $W = [Z\ X]$ be all the variables that are orthogonal to ε in the population under the null hypothesis that X and ε are uncorrelated. As in the omitted variables problem, consider the test statistic for over-identifying restrictions, $2nQ_n = \min_b u'P_W u/\sigma^2$, where $u = y - Xb$. Decompose $P_W = P_X + (P_W - P_X)$. Then $u'(P_W - P_X)u = y'(P_W - P_X)y$ and the minimizing b sets $u'P_X u = 0$, so that $2nQ_n = y'(P_W - P_X)y/\sigma^2$. Since $P_W - P_X = P_{Q_X W}$, one also has $2nQ_n = y' P_{Q_X W}\, y$. This statistic is the same as the test statistic for the hypothesis that the coefficients of Z are zero in a regression of y on X and Z; thus the test for over-identifying restrictions is an omitted variables test. One can also write $2nQ_n = \|\hat{y}_W - \hat{y}_X\|^2/\sigma^2$, so that a computationally convenient equivalent test is based on the difference between the fitted values of y from a regression on X and Z and a regression on X alone. Finally, we will show that the statistic can be written**

$$2nQ_n$$
$$= (b_{1,2SLS} - b_{1,OLS})[V(b_{1,2SLS}) - V(b_{1,OLS})]^{-1}(b_{1,2SLS} - b_{1,OLS}).$$

In this form, the statistic is the Hausman test for exogenicity in the form developed by Hausman and Taylor, and the result establishes that the Hausman test for exogeneity is equivalent to a GMM test for over-identifying restrictions.

Several steps are needed to demonstrate this equivalence. Note that $\mathbf{b_{2SLS} = (X'P_MX)^{-1}X'P_My}$, where M = [Z $\mathbf{X_2}$]. Write

$$\mathbf{b_{2SLS} - b_{OLS} = (X'P_MX)^{-1}X'P_My - (X'X)^{-1}X'y}$$
$$\mathbf{= (X'P_MX)^{-1}[X'P_M - X'P_MX(X'X)^{-1}X']y}$$
$$\mathbf{= (X'P_MX)^{-1}X'P_MQ_Xy.}$$

Since $\mathbf{X_2}$ is in M, $\mathbf{P_MX_2 = X_2}$, implying $\mathbf{X'P_MQ_X} = \begin{bmatrix} X_1'P_MQ_X \\ X_2'P_MQ_X \end{bmatrix}$ $= \begin{bmatrix} X_1'P_MQ_X \\ X_2'Q_X \end{bmatrix} = \begin{bmatrix} X_1'P_MQ_X \\ 0 \end{bmatrix}$. Also, $\mathbf{X'P_MX} = \begin{bmatrix} X_1'P_MX_1 & X_1'P_MX_2 \\ X_2'P_MX_1 & X_2'P_MX_2 \end{bmatrix} = \begin{bmatrix} X_1'P_MX_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}$. Then $\begin{bmatrix} X_1'P_MQ_Xy \\ 0 \end{bmatrix} = \mathbf{(X'P_MX)(b_{2SLS} - b_{OLS})}$ / $\begin{bmatrix} X_1'P_MX_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix} \begin{bmatrix} b_{1,2SLS} - b_{1,OLS} \\ b_{2,2SLS} - b_{2,OLS} \end{bmatrix}$. From the second block of equations, one obtains the result that the second subvector is a linear combination of the first subvector. This implies that a test statistic that is a function of the full vector of differences of 2SLS and OLS estimates can be written equivalently as a function of the first subvector

of differences. From the first block of equations, substituting in the solution for the second subvector of differences expressed in terms of the first, one obtains

$$[X_1'P_MX_1 - X_1'X_2(X_2'X_2)^{-1}X_2'X_1](b_{1,2SLS} - b_{1,OLS})$$
$$= X_1'P_MQ_Xy$$

The matrix on the left-hand-side can be rewritten as $X_1'P_M Q_{X_2} P_MX_1$, so that

$$b_{1,2SLS} - b_{1,OLS} = (X_1'P_M Q_{X_2} P_MX_1)^{-1}X_1'P_MQ_Xy.$$

Next, we calculate the covariance matrix of b_{2SLS} - b_{OLS}, and show that it is equal to the difference of $V(b_{2SLS}) = \sigma^2(X'P_MX)^{-1}$ and $V(b_{OLS}) = \sigma^2(X'X)^{-1}$. From the formula $b_{2SLS} - b_{OLS} = (X'P_MX)^{-1}X'P_MQ_Xy$, one has $V(b_{2SLS} - b_{OLS}) = \sigma^2(X'P_MX)^{-1}X'P_MQ_XP_MX(X'P_MX)^{-1}$.
On the other hand,

$$V(b_{2SLS}) - V(b_{OLS}) = \sigma^2(X'P_MX)^{-1}\{X'P_MX -$$
$$X'P_MX(X'X)^{-1}X'P_MX\}(X'P_MX)^{-1}$$
$$= \sigma^2(X'P_MX)^{-1}\{X'P_M[I - X(X'X)^{-1}X']P_MX\}(X'P_MX)^{-1}$$
$$= \sigma^2(X'P_MX)^{-1}X'P_MQ_XP_MX(X'P_MX)^{-1}.$$

Thus, $V(b_{2SLS} - b_{OLS}) = V(b_{2SLS}) - V(b_{OLS})$. This is a consequence of the fact that under the null hypothesis OLS is efficient among the class of linear estimators including 2SLS. Expanding the center of this expression, and using the results $P_MX_2 = X_2$ and hence $Q_XP_MX_2 = 0$, one has

$$X'P_MQ_XP_MX = \begin{bmatrix} X_1'P_MQ_XP_MX_1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Hence, $V(b_{2SLS}) - V(b_{OLS})$ is of rank p; this also follows by noting that $b_{2,2SLS} - b_{2,OLS}$ could be written as a linear transformation of $b_{1,2SLS} - b_{1,OLS}$.

Next, use the formula for partitioned inverses to show for N = M or N = I that the northwest corner of $\begin{bmatrix} X_1'P_NX_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix}^{-1}$ is $(X_1'P_NQ_{X_2}P_NX_1)^{-1}$. Then,

$$V(b_{1,2SLS} - b_{1,OLS}) =$$
$$\sigma^2(X_1'P_M Q_{X_2} P_MX_1)^{-1}X_1'P_MQ_XP_MX_1(X_1'P_M Q_{X_2} P_MX_1)^{-1}.$$

Using the expressions above, the quadratic form can be written

$$(b_{1,2SLS} - b_{1,OLS})V(b_{1,2SLS} - b_{1,OLS})^{-1}(b_{1,2SLS} - b_{1,OLS})$$
$$= y'Q_XP_MX_1(X_1'P_MQ_XP_MX_1)^{-1}X_1'P_MQ_Xy/\sigma^2.$$

Finally, one has, from the test for over-identifying restrictions,

$$2nQ_n = y'(P_W - P_X)y/\sigma^2 = y'P_{Q_XW}y\ /\sigma^2$$
$$\equiv y'Q_XP_MX_1(X_1'P_MQ_XP_MX_1)^{-1}X_1'P_MQ_Xy/\sigma^2,$$

so that the two statistics coincide.

A Generalized Exogenicity Test: **Consider the regression $y = X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + \epsilon$, and the null hypothesis that X_1 is exogenous, where X_2 is known to be exogenous, and X_3 is known to be endogenous. Suppose N is an array of instruments, including X_2, that are sufficient to identify the coefficients when the hypothesis is false. Let $W = [N\ X_1]$ be the full set of instruments available when the null hypothesis is true. Then the best instruments under the null hypothesis are $X_o = P_W X \equiv [X_1\ X_2\ X_3^*]$, and the best instruments under the alternative are $X_u = P_N X \equiv [X_1^*\ \ X_2\ \ X_3^*]$. The test statistic for over-identifying restrictions is $2nQ_n = y'(P_{X_o} - P_{X_u})y/\sigma^2$, as in the previous cases. This can be written $2nQ_n = (SSR_{X_o} - SSR_{X_u})/\sigma^2$, with the numerator the difference in sum of squared residuals from a OLS regression of y on X_u and a OLS regression of y on X_o. Also, $2nQ_n = \|\hat{y}_{X_o} - \hat{y}_{X_u}\|^2/\sigma^2$, the difference between the fitted values of y from a regression on X_u and a regression on X_o. Finally,**

$$2nQ_n = (b_{2SLS_o} - b_{2SLS_u})'[V(b_{2SLS_u}) - V(b_{2SLS_o})]^- (b_{2SLS_o} - b_{2SLS_u}),$$

an extension of the Hausman-Taylor exogeneity test to the problem where some variables are suspect and others are known to be exogenous. One can show that the quadratic form in the center of this quadratic form has rank equal to the rank of X_1, and that the test statistic can be written equivalently as a quadratic form in the subvector of differences of the 2SLS estimates for the X_1 coefficients,

with the ordinary inverse of the corresponding submatrix of differences of variances in the center of the quadratic form.