Hypothesis Testing for Instrumental Variable Estimators

Suppose we have the following linear regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \ \mathbf{u} \overset{d}{\sim} \mathrm{IID}(\mathbf{0}, \sigma^2 \mathbf{I}_n), \ \mathbb{E}[\mathbf{W}^\top \mathbf{u}] = \mathbf{0}, \tag{1}$$

and where

$$\mathbf{y} = \mathbf{X}_1 \beta_1 + \mathbf{X}_2 \beta_2 + \mathbf{u}, \ \mathbb{E}[\mathbf{u}|\mathbf{X}_2] = \mathbf{0},$$
$$\mathbf{W} = \begin{bmatrix} \mathbf{X}_2 & \mathbf{W}_1 \end{bmatrix},$$

where $\mathbf{X}_2$ would be our "included instruments" and $\mathbf{W}_1$ would be our "excluded instruments". Necessary and sufficient conditions for the GIV/2SLS estimator to estimate $\boldsymbol{\beta}$ consistently are the order condition (necessary) and the rank condition (sufficient). The order condition is that our model must not be underspecified, so $l \geq k$. For sufficiency, we first require that whenever $\mathbf{W}_t \in \Omega_t$,

$$\mathbb{E}[u_t | \mathbf{W}_t] = 0,$$

and $\mathbf{W}_t$ is assumed to be exogenous/predetermined with respect to the error term. For asymptotic identification, the above condition can be written as

$$\mathbf{S}_{\mathbf{W}^\top \mathbf{X}} = \operatorname*{plim}_{n \to \infty} \frac{1}{n} \mathbf{W}^\top \mathbf{X}$$

where $\mathbf{S}_{\mathbf{W}^\top \mathbf{X}}$ is deterministic and nonsingular. In addition we assume that $\mathbf{S}_{\mathbf{W}^\top \mathbf{W}}$ exists and is of full rank. If $\mathbf{S}_{\mathbf{W}^\top \mathbf{W}}$ does not have full rank, then at least one of the instruments is perfectly collinear with the others, asymptotically, and should be dropped. If $\mathbf{S}_{\mathbf{W}^\top \mathbf{X}}$ does not have full rank then the asymptotic version of the moment conditions ($\mathbf{X}^\top \mathbf{P}_{\mathbf{W}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$) has fewer than k linearly independent equations, and these conditions therefore have no unique solution. These asymptotic conditions are sufficient for consistency of the GIV/2SLS estimator. But the key necessary condition is

$$\operatorname*{plim}_{n \to \infty} \frac{1}{n} \mathbf{W}^\top \mathbf{u} = 0.$$

If this assumption did not hold, because any of the instruments was asymptotically correlated with the error terms, the first of the asymptotic sufficiency conditions would not hold either, and the GIV/2SLS estimator would not be consistent. It stands to reason, then, that testing the validity of our instruments is quite important. We derive testing procedures to do this below.

1 The Sargan-Hansen J-test statistic

Consider the model described previously in (1). Here, $\mathbf{X}$ is $n \times k_X$ and $\mathbf{W}$ is $n \times k_W$ and thus $(k_{W_1} - k_{X_1})$ is our overidentification degree. The IV estimator FOCs are

$$\mathbf{X}^\top \mathbf{P_W}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{IV}}) = 0 \ (k_X \text{ equations}), \tag{2}$$

$$\mathbb{E}[\mathbf{W}^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})] = 0 \ (k_W \text{ equations}), \tag{3}$$

but

$$\mathbf{W}^\top (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{IV}}) \neq 0,$$

thus we have (2) implying that $\mathbf{P_W X}$ are the effective instruments. Then let $\mathbf{W}^*$, which is an $n \times (k_W - k_X)$ matrix, be the ineffective yet valid instruments, such that

$$\mathbb{E}[(\mathbf{W}^*)^\top \mathbf{u}] = \mathbf{0}.$$

Note also that

$$\begin{aligned} \mathbf{P_W} &= \mathbf{P_{P_W X}} + \mathbf{P_{W^*}} \\ \implies \mathbf{P_{W^*}} &= \mathbf{P_W} - \mathbf{P_{P_W X}}, \end{aligned}$$

where $\mathbf{P_{W^*}}$ is of rank $k_W - k_X$. Now, we can formally begin the construct the Sargan-Hansen J-test statistic. Begin by assuming that

$$(\mathbf{W}^*)^\top \mathbf{u} \overset{d}{\sim} N(\mathbf{0}, \sigma^2((\mathbf{W}^*)^\top \mathbf{W})).$$

Now, we now from that if $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I}_{k_2})$, then $\mathbf{z}^\top \mathbf{z} \sim \chi^2(k_2)$. But let's construct it:

$$\begin{aligned} \frac{\left((\mathbf{W}^*)^\top \mathbf{W}^*\right)^{-\frac{1}{2}}}{\sigma} (\mathbf{W}^*)^\top \mathbf{u} &\overset{d}{\sim} N(\mathbf{0}, \mathbf{I}_{k_W - k_X}) \\ \frac{\mathbf{u}^\top \mathbf{W}^* ((\mathbf{W}^*)^\top \mathbf{W}^*)^{-1} (\mathbf{W}^*)^\top \mathbf{u}}{\sigma^2} &\overset{d}{\sim} \chi^2(k_W - k_X) \\ J = \frac{\mathbf{u}^\top (\mathbf{P_W} - \mathbf{P_{P_W X}}) \mathbf{u}}{\sigma^2} &\overset{d}{\sim} \chi^2(k_W - k_X) \end{aligned} \tag{4}$$

We can, however, simplify the numerator on the LHS of (4):

$$\begin{aligned} (\mathbf{P_W} - \mathbf{P_{P_W X}})\mathbf{u} &= (\mathbf{P_W} - \mathbf{P_{P_W X}})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= (\mathbf{P_W} - \mathbf{P_{P_W X}})\mathbf{y} - \underbrace{(\mathbf{P_W} - \mathbf{P_{P_W X}})\mathbf{X}\boldsymbol{\beta}}_{=0} \\ &= \mathbf{P_W}(\mathbf{I} - \mathbf{X}(\mathbf{X}^\top \mathbf{P_W X})^{-1}\mathbf{X}^\top \mathbf{P_W})\mathbf{y} \\ &= \mathbf{P_W}(\mathbf{y} - \mathbf{X}(\mathbf{X}^\top \mathbf{P_W X})^{-1}\mathbf{X}^\top \mathbf{P_W y}) \\ &= \mathbf{P_W}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{IV}}). \end{aligned}$$

So (4) becomes:

$$J = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{IV}})^\top \mathbf{P}_\mathbf{W}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{IV}})}{\sigma^2} \sim \chi^2(k_W - k_X), \tag{5}$$

where

$$\sigma^2 = \frac{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{IV}})^\top (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{IV}})}{n}.$$

The Sargan test can be used test the null hypothesis that $\mathbb{E}[\mathbf{W}^\top \mathbf{u}] = \mathbf{0}$, against the alternative hypothesis that the instruments are invalid.

There is a caveat to the Sargan-Hansen test, however. Suppose that we have an overidentified model where the instruments of $\mathbf{W}$ are informative, but only one of them is valid. The Sargan-Hansen test is unable to distinguish between which instruments are valid and which are invalid. Thus even if the Sargen-Hansen test rejects the null hypothesis, it would be unable to pinpoint which instrument is invalid.

2 The Durbin-Wu-Hausman (DWH) test

In many cases, we do not know whether we actually need to use instrumental variables.[1] For example, we may suspect that some variables are measured with error, but we may not know whether the errors are large enough to cause enough inconsistency for us to worry about. Or we may suspect that certain explanatory variables are weakly endogenous. In such a case, it may or may not be perfectly reasonable to employ OLS estimation.

If the regressors are valid instruments, then they are also the optimal instruments. Consequently, the OLS estimator, which is consistent in this case, is preferable to an IV estimator computed with some other valid instrument matrix $\mathbf{W}$. In view of this, it would evidently be very useful to be able to test the null hypothesis that the error terms are uncorrelated with all the regressors against the alternative that they are correlated with some of the regressors, although not with the instruments $\mathbf{W}$. A test for this is the Durbin-Wu-Hausman test or DWH test.

The null and alternative hypotheses for the DWH test are as follows:

$$\begin{aligned} H_0: \quad \mathbf{y} = &\quad \mathbf{X}\boldsymbol{\beta} + \mathbf{u},\ \mathbf{u} \sim \text{IID}(\mathbf{0}, \sigma^2\mathbf{I}),\ \mathbb{E}(\mathbf{X}^\top \mathbf{u}) = \mathbf{0}, \\ H_1: \quad \mathbf{y} = &\quad \mathbf{X}\boldsymbol{\beta} + \mathbf{u},\ \mathbf{u} \sim \text{IID}(\mathbf{0}, \sigma^2\mathbf{I}),\ \mathbb{E}(\mathbf{W}^\top \mathbf{u}) = \mathbf{0}. \end{aligned}$$

[1] i.e., we may wish to test if the instruments we have are informative/effective.

Under the alternative hypothesis, the IV estimator is consistent, but the OLS estimator is not. Under the null, both are consistent. Thus $\text{plim}(\hat{\boldsymbol{\beta}}_{\text{IV}} - \hat{\boldsymbol{\beta}}_{\text{OLS}})$ is zero under the null and nonzero under the alternative.

The idea of the DWH test is to check whether the difference between the IV estimator and OLS estimator, $\hat{\boldsymbol{\beta}}_{\text{IV}} - \hat{\boldsymbol{\beta}}_{\text{OLS}}$, is significantly different from zero in the available sample. This difference, which is sometimes called the vector of contrasts, can be written as:

$$\hat{\boldsymbol{\beta}}_{\text{IV}} - \hat{\boldsymbol{\beta}}_{\text{OLS}} = (\mathbf{X}^\top \mathbf{P_W} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{P_W} \mathbf{y} - (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}.$$

This expression is not very useful as it stands But we can use a trick in econometrics. We pretend that the first factor of the IV estimator is common to both estimators, and take it out as a common factor. This gives:

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{\text{IV}} - \hat{\boldsymbol{\beta}}_{\text{OLS}} &= (\mathbf{X}^\top \mathbf{P_W} \mathbf{X})^{-1} (\mathbf{X}^\top \mathbf{P_W} \mathbf{y} - \mathbf{X}^\top \mathbf{P_W} \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}) \\
&= (\mathbf{X}^\top \mathbf{P_W} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{P_W} (\mathbf{I} - \mathbf{X} (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top) \mathbf{y} \\
&= (\mathbf{X}^\top \mathbf{P_W} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{P_W} \mathbf{M_X} \mathbf{y} \qquad (6)
\end{aligned}$$

The first factor in expression (6) is positive semidefinite, by the identification condition. Therefore, testing whether $\hat{\boldsymbol{\beta}}_{\text{IV}} - \hat{\boldsymbol{\beta}}_{\text{OLS}}$ is significantly different from zero is equivalent to testing whether the vector $\mathbf{X}^\top \mathbf{P_W} \mathbf{M_X} \mathbf{y}$ is significantly different from zero.

There are a few ways to derive the DWH test statistic. One way is via FWL theorem and eventually involves deriving a test based on the F distribution, as demonstrated in Davidson and MacKinnon.

Under the null hypothesis, we estimate via OLS, and we know that the OLS residuals are given by the vector $\mathbf{M_X y}$. Therefore, we wish to test whether the k columns of the matrix $\mathbf{P_W X}$ are orthogonal to this vector of residuals. Let us partition the matrix of regressors X, so that $\mathbf{X} = \begin{bmatrix} \mathbf{Z} & \mathbf{Y} \end{bmatrix}$, where the k_1 columns of $\mathbf{Z}$ are included in the matrix of instruments $\mathbf{W}$, and the $k_2 = k - k_1$ columns of $\mathbf{Y}$ are treated as potentially endogenous. We know that the OLS residuals are orthogonal to all the columns of $\mathbf{X}$, in particular to those of $\mathbf{Z}$. For these regressors, there is therefore nothing to test. The relation

$$\mathbf{Z}^\top \mathbf{P_W} \mathbf{M_X} \mathbf{y} = \mathbf{Z}^\top \mathbf{M_X} \mathbf{y} = \mathbf{0}$$

holds because $\mathbf{P_W Z} = \mathbf{Z}$ and $\mathbf{M_X Z} = \mathbf{O}$. The test is thus concerned only with the k_2 elements of $\mathbf{Y}^\top \mathbf{P_W} \mathbf{M_X} \mathbf{y}$, which are not in general identically zero, but should not differ from it significantly under the null.

The easiest way to test whether $\mathbf{Y}^\top \mathbf{P_W M_X y}$ is significantly different from zero is to use an F test for the k_2 restrictions $\boldsymbol{\delta} = \mathbf{0}$ in the OLS regression:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{P_W Y}\boldsymbol{\delta} + \mathbf{u}. \tag{7}$$

Using the FWL theorem, we can attain the OLS estimate of $\boldsymbol{\delta}$:

$$\hat{\boldsymbol{\delta}} = (\mathbf{Y}^\top \mathbf{P_W M_X P_W Y})^{-1}\mathbf{Y}^\top \mathbf{P_W M_X y}.$$

The OLS estimate of $\boldsymbol{\delta}$ from (7), is the same as those from the FWL regression of $\mathbf{M_X y}$ on $\mathbf{M_X P_W Y}$, via the FWL theorem. We know that the inverted component in the above equation is positive definite, so testing whether or not $\boldsymbol{\delta} = \mathbf{0}$ is equivalent to testing whether $\mathbf{Y}^\top \mathbf{P_W M_X y} = \mathbf{0}$, as desired. Also, note that under the null

$$\mathbf{X}^\top \mathbf{P_W M_X} \underbrace{\mathbf{y}}_{\mathbf{X}\boldsymbol{\beta}+\mathbf{u}} = \mathbf{X}^\top \mathbf{P_W M_X u}.$$

The F test based in (7) has k_2 and $n-k-k_2$ degrees of freedom. Under the null, if we assume that $\mathbf{X}$ and $\boldsymbol{W}$ are not merely predetermined but also exogenous, and that the error terms $\mathbf{u}$ are multivariate normal ($\mathbf{u} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$), the F statistic does indeed have the $F(k_2, n-k-k_2)$ distribution. Thus

$$\mathbf{X}^\top \mathbf{P_W M_X u} \overset{d}{\sim} N\left(\mathbf{0}, \sigma^2(\mathbf{X}^\top \mathbf{P_W M_X P_W X})\right),$$

so

$$(\mathbf{X}^\top \mathbf{P_W X})^{-1}\mathbf{X}^\top \mathbf{P_W M_X u} \overset{d}{\sim} N\left(\mathbf{0}, \sigma^2 \underbrace{(\mathbf{X}^\top \mathbf{P_W X})^{-1}(\mathbf{X}^\top \mathbf{P_W M_X P_W X})(\mathbf{X}^\top \mathbf{P_W X})^{-1}}_{(\mathbf{X}^\top \mathbf{P_W X})^{-1}[\mathbf{X}^\top \mathbf{P_W X} - \mathbf{X}^\top \mathbf{P_W X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{P_W X}](\mathbf{X}^\top \mathbf{P_W X})^{-1}}\right),$$

so

$$\begin{aligned}(\hat{\boldsymbol{\beta}}_{\text{IV}} - \hat{\boldsymbol{\beta}}_{\text{OLS}}) &\overset{d}{\sim} N\left(\mathbf{0}, \sigma^2[(\mathbf{X}^\top \mathbf{P_W X})^{-1} - (\mathbf{X}^\top \mathbf{X})^{-1}]\right) \\ &\overset{d}{\sim} N(\mathbf{0}, \text{Var}(\hat{\boldsymbol{\beta}}_{\text{IV}}) - \text{Var}(\hat{\boldsymbol{\beta}}_{\text{OLS}}))\end{aligned}$$

$$\implies (\hat{\boldsymbol{\beta}}_{\text{IV}} - \hat{\boldsymbol{\beta}}_{\text{OLS}})^\top \left[\text{Var}(\hat{\boldsymbol{\beta}}_{\text{IV}}) - \text{Var}(\hat{\boldsymbol{\beta}}_{\text{OLS}})\right]^{-1} (\hat{\boldsymbol{\beta}}_{\text{IV}} - \hat{\boldsymbol{\beta}}_{\text{OLS}}) \sim \chi^2$$

Under the null, its asymptotic distribution is $F(k_2, \infty)$, and k_2 times the statistic is asymptotically distributed as $\chi^2(k_2)$. If the null hypothesis is rejected, we are faced with the same sort of ambiguity of interpretation as for the test of Sargan-Hansen test.

Note, we can justify the degrees of freedom of our test statistic by the following: We proposed that the DWH test has k_2 degrees of freedom, the number of possibly endogenous variables on the RHS of our null hypothesis. This is smaller than k, the dimension of vector $\boldsymbol{\beta}$. Consider:

$$\mathbf{X}^\top \mathbf{P}_\mathbf{W} \mathbf{M}_\mathbf{X} \mathbf{u},$$

where, as before, we have

$$\mathbf{X} = \begin{bmatrix} \mathbf{Z} & \mathbf{Y} \end{bmatrix}, \; \mathbf{W} = \begin{bmatrix} \mathbf{Z} & \mathbf{W}_1 \end{bmatrix},$$

then:

$$\begin{aligned} \mathbf{P}_\mathbf{W} &= \mathbf{P}_\mathbf{Z} + \mathbf{P}_{\mathbf{M}_\mathbf{Z} \mathbf{W}_1} \\ \implies \mathbf{P}_\mathbf{W} \mathbf{M}_\mathbf{X} &= \mathbf{P}_{\mathbf{M}_\mathbf{Z} \mathbf{W}_1} \mathbf{M}_\mathbf{X}, \end{aligned}$$

since $\mathbf{M}_\mathbf{X} \mathbf{P}_\mathbf{Z} = \mathbf{O}$. Thus, we can write

$$\mathbf{X}^\top \mathbf{P}_\mathbf{W} \mathbf{M}_\mathbf{X} \mathbf{u} = \mathbf{X}^\top \mathbf{P}_{\mathbf{M}_\mathbf{Z} \mathbf{W}_1} \mathbf{M}_\mathbf{X} \mathbf{u},$$

where $\mathbf{X}^\top \mathbf{P}_{\mathbf{M}_\mathbf{Z} \mathbf{W}_1}$ on the RHS of the above equation looks like

$$\begin{bmatrix} \mathbf{Z}^\top \\ \mathbf{Y}^\top \end{bmatrix} \mathbf{P}_{\mathbf{M}_{\mathbf{X}_2} \mathbf{W}_1} = \begin{bmatrix} \mathbf{X}_1^\top \mathbf{P}_{\mathbf{M}_\mathbf{Z} \mathbf{W}_1} \\ \mathbf{O} \end{bmatrix}.$$

Therefore:

$$\mathbf{X}^\top \mathbf{P}_\mathbf{W} \mathbf{M}_\mathbf{X} \mathbf{u} = \begin{bmatrix} \mathbf{X}_1^\top \mathbf{P}_{\mathbf{M}_\mathbf{Z} \mathbf{W}_1} \mathbf{M}_\mathbf{X} \mathbf{u} \\ \mathbf{O} \end{bmatrix}_{k \times 1},$$

hence

$$\chi^2(k_\mathbf{Y}),$$

where $k_\mathbf{Y} = k_2$.