CHAPTER 12

Instrumental Variables Estimation

1. Instrumental Variables

Consider the following regression model with endogeneity: for all $i = 1, 2, ..., n$,

$$Y_i = X_i'\beta_0 + u_i$$

where $\mathbf{E}u_i = 0$, $\mathbf{E}u_i^2 = \sigma^2$, and $\mathbf{E}X_i u_i \neq 0$. The $k \times 1$ vector X_i is a column vector of endogeneous regressors. In this case, without further restrictions or additional data sets, we cannot identify β_0. Now, suppose that we have an observed random vector $Z_i \in \mathbf{R}^d$ for each individual i that satisfies the following restriction:

(Validity) $\mathbf{E}Z_i u_i = 0$, and
(Relevance) $\mathbf{E}Z_i X_i'$ is full column rank k.

When an observed random vector Z_i satisfies these two conditions, the random vector is called a vector of *instrumental variables*. When we have such a random vector, we can identify β_0 in the following way:

$$\beta_0 = \left[\left(\mathbf{E}\left[Z_i X_i'\right]\right)' \left(\mathbf{E}\left[Z_i X_i'\right]\right) \right]^{-1} \left(\mathbf{E}\left[Z_i X_i'\right]\right)' \mathbf{E}\left[Z_i Y_i\right].$$

Note that when $\mathbf{E}Z_i X_i'$ is not full column rank k, the matrix $\left(\mathbf{E}\left[Z_i X_i'\right]\right)' \left(\mathbf{E}\left[Z_i X_i'\right]\right)$ is not invertible. The relevance condition tells us that we need to have at least k number of instrumental variables. This rank condition is violated when Z_i and X_i are uncorrelated:

$$\mathbf{E}Z_i X_i' = \mathbf{E}Z_i \mathbf{E}X_i'.$$

The last matrix is of rank one. (Recall that $rk(AB) \leq \min\{rk(A), rk(B)\}$ for any conformable matrices A and B.) Suppose that a $m \times 1$ subvector X_{1i} of X_i is known to be uncorrelated with u_i, so that X_{1i} is in fact exogenous. Then, we can include this subvector X_{1i} in the instrument vector Z_i. Hence we need to find at least $k - m$ number of instrumental variables to identify the parameter β_0. The number $k - m$ is the number of endogenous regressors in X_i.

Let $Z_i \in \mathbf{R}^d$ and $X_i \in \mathbf{R}^k$. When $d > k$ so that the number of instrumental variables is larger than the dimension of X_i, we say that the model is *overidentified*. When $d = k$, we say that the model is *exactly identified*. In this case of exact identification, we identify β_0 as

$$\begin{aligned} \beta_0 &= \left[\left(\mathbf{E}\left[Z_i X_i'\right]\right)' \left(\mathbf{E}\left[Z_i X_i'\right]\right) \right]^{-1} \left(\mathbf{E}\left[Z_i X_i'\right]\right)' \mathbf{E}\left[Z_i Y_i\right] \\ &= \left(\mathbf{E}\left[Z_i X_i'\right]\right)^{-1} \left(\mathbf{E}\left[Z_i X_i'\right]'\right)^{-1} \left(\mathbf{E}\left[Z_i X_i'\right]\right)' \mathbf{E}\left[Z_i Y_i\right] \\ &= \left(\mathbf{E}\left[Z_i X_i'\right]\right)^{-1} \mathbf{E}\left[Z_i Y_i\right]. \end{aligned}$$

The most difficult part of this identification strategy is to find an instrumental variable. A typical strategy is to find an exogenous variable which influences the dependent variable Y_i only through affecting the regressor X_i. The requirement that it affects X_i fulfills the relevance condition, and affects Y_i without affecting u_i fulfills the validity condition. Such an exogeneous variable can be found through natural phenomenon such as temperature variations, rainfalls, or government tax system changes which affect individuals differently. There is no general method of finding a convincing instrumental variable. It is dealt with case by case.

2. Estimation

2.1. Method of Moment Estimation

The least squares estimator under exogeneity can be thought of as a method of moments estimator, with the exogeneity condition $\mathbf{E}X_i u_i = 0$ producing the moment condition:

$$\mathbf{E}X_i(Y_i - X_i'\beta_0) = 0.$$

Then, the least squares estimator is the minimizer of the Euclidean norm of the sample version of the expectation:

$$\hat{\beta} = \arg\min_{\beta \in \mathbf{R}^k} \left\| \frac{1}{n} \sum_{i=1}^{n} X_i(Y_i - X_i'\beta) \right\|^2.$$

This way of obtaining an estimator is called *method of moments estimation*. The moment condition from the exogeneity condition is not uniquely written. For any $k \times k$ matrix S such that $S'S$ is positive definite, one has the following moment condition:

$$S\mathbf{E}X_i(Y_i - X_i'\beta_0) = 0.$$

Therefore, alternatively, one may come up with an estimator by solving the following problem.

$$\hat{\beta}_S = \arg\min_{\beta \in \mathbf{R}^k} \left\| \frac{1}{n} \sum_{i=1}^{n} X_i(Y_i - X_i'\beta) \right\|_S^2,$$

where $||A||_S^2 = tr\left(A'S'SA\right)$. The resulting estimator takes the following form:

$$\begin{aligned} \hat{\beta}_S &= \arg\min_{\beta \in \mathbf{R}^k} \left(\frac{1}{n} \sum_{i=1}^{n} X_i(Y_i - X_i'\beta) \right)' S'S \left(\frac{1}{n} \sum_{i=1}^{n} X_i(Y_i - X_i'\beta) \right) \\ &= \arg\min_{\beta \in \mathbf{R}^k} \left(SX'y - SX'X\beta\right)' \left(SX'y - SX'X\beta\right), \end{aligned}$$

where $y = [Y_1, ..., Y_n]'$. Therefore, the solution is from projecting $SX'y$ onto $\mathcal{R}(SX'X)$. Using the least squares formula:

$$\begin{aligned} & \left[(SX'X)'\,(SX'X)\right]^{-1} (SX'X)'\,SX'y \\ =\ & \left[X'XS'SX'X\right]^{-1} X'XS'SX'y \\ =\ & (X'X)^{-1}\,(S'S)^{-1}\,(X'X)^{-1}X'XS'SX'y \\ =\ & (X'X)^{-1}X'y. \end{aligned}$$

Therefore, the use of S does not alter the least squares estimator.[1]

In general, the choice of S may affect the estimator and its asymptotic variance. Then one may ask what choice of S will yield the smallest asymptotic variance. Such a choice of S yields the *optimal weighting matrix* $S'S$. The optimal weighting matrix may depend on the distribution of (Y_i, X_i). However, we can replace the weighting matrix by a consistent estimator. After this replacement, the asymptotic distribution of the estimator remains unchanged.

---

[1]This is different from the GLS. Note that we can rewrite $\hat{\beta}_S$ as

$$\hat{\beta}_S = \arg\min_{\beta \in \mathbf{R}^k} (y - X\beta)'\,\Sigma\,(y - X\beta)$$

where $\Sigma = XS'SX'$. Suppose that we have $\mathbf{E}uu' = \Omega$, where $u = [u_1, \cdots, u_n]'$. Then, the GLS is given by

$$\hat{\beta}_{GLS} = \arg\min_{\beta \in \mathbf{R}^k} (y - X\beta)'\,\Omega^{-1}\,(y - X\beta)$$

or $\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$. However, there is no S such that

$$\Sigma = XS'SX' = \Omega^{-1}$$

because $XS'SX'$ is not invertible, having rank k not n.

Now, suppose that the regression model is in fact endogenous, but fortunately, we have instrumental variables Z_i. In this case, the validity condition of Z_i tells us that $\mathbf{E}Z_iu_i = 0$. Hence for any $s \times d$ matrix S such that $S'S$ is positive definite (this already requies that $s \geq d$), we can write

$$S\mathbf{E}Z_i(Y_i - X_i'\beta_0) = 0.$$

In fact, by the validity and relevance condition for Z_i, there exists a unique value of β_0 that satisfies the above equation. In other words, β_0 is identified by the moment equality restrictions.

The method of moments estimation suggests that we estimate β_0 in the following way:

$$\begin{aligned}
\hat{\beta}_S^{IV} &= \arg\min_{\beta \in \mathbf{R}^k} \left\| \frac{1}{n}\sum_{i=1}^n Z_i(Y_i - X_i'\beta) \right\|_S^2 \\
&= \arg\min_{\beta \in \mathbf{R}^k} \left(\frac{1}{n}\sum_{i=1}^n Z_i(Y_i - X_i'\beta) \right)' S'S \left(\frac{1}{n}\sum_{i=1}^n Z_i(Y_i - X_i'\beta) \right) \\
&= \arg\min_{\beta \in \mathbf{R}^k} \left(Z'(Y - X\beta)\right)' W \left(Z'(Y - X\beta)\right) \\
&= \arg\min_{\beta \in \mathbf{R}^k} \left(W^{1/2}Z'Y - W^{1/2}Z'X\beta\right)' \left(W^{1/2}Z'Y - W^{1/2}Z'X\beta\right)
\end{aligned}$$

where $W = S'S$ and

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}_{n\times 1}, \; Z = \begin{bmatrix} Z_1' \\ \vdots \\ Z_n' \end{bmatrix}_{n\times d} \text{ and } X = \begin{bmatrix} X_1' \\ \vdots \\ X_n' \end{bmatrix}_{n\times k}.$$

By projecting $Y_Z \equiv W^{1/2}Z'Y$ onto $\mathcal{R}(X_Z) \equiv \mathcal{R}(W^{1/2}Z'X)$, with $X_Z \equiv W^{1/2}Z'X$, we obtain the following:

$$\begin{aligned}
\hat{\beta}_S^{IV} &= (X_Z'X_Z)^{-1}X_Z'Y_Z \\
&= (X'ZWZ'X)^{-1}X'ZWZ'Y.
\end{aligned}$$

Now, let us study the asymptotic properties of the estimator $\hat{\beta}_S^{IV}$. We assume the following:

(A1) $\{(X_i', Z_i', u_i)\}_{i=1}^n$ is i.i.d. such that $\mathbf{E}||X_1||^4 < \infty$ and $\mathbf{E}||Z_1||^4 < \infty$.
(A2) $\mathbf{E}u_1 = 0$ and $\mathbf{E}[u_1^2|Z_1] = \sigma^2$, for some constants $\sigma^2 > 0$.
(A3) $\mathbf{E}Z_1u_1 = 0$ and $\mathbf{E}Z_1X_1'$ is full column rank k.
(A4) $\mathbf{E}Z_1Z_1'$ is invertible.

Observe that

$$\begin{aligned}
\hat{\beta}_S^{IV} - \beta_0 &= (X'ZWZ'X)^{-1}X'ZWZ'u \\
&= \left[\left(\frac{1}{n}\sum_{i=1}^n X_iZ_i'\right) W \left(\frac{1}{n}\sum_{i=1}^n Z_iX_i'\right)\right]^{-1} \\
&\quad \times \left(\frac{1}{n}\sum_{i=1}^n X_iZ_i'\right) W \left(\frac{1}{n}\sum_{i=1}^n Z_iu_i\right) \\
&\to_p \left[(\mathbf{E}X_iZ_i')\, W\, (\mathbf{E}Z_iX_i')\right]^{-1} (\mathbf{E}X_iZ_i')\, W\, (\mathbf{E}Z_iu_i) = 0.
\end{aligned}$$

Hence the estimator is consistent. Now, let us consider the asymptotic normality of the estimator. Write

$$\begin{aligned}
\sqrt{n}(\hat{\beta}_S^{IV} - \beta_0) &= \left[\left(\frac{1}{n}\sum_{i=1}^n X_iZ_i'\right) W \left(\frac{1}{n}\sum_{i=1}^n Z_iX_i'\right)\right]^{-1} \\
&\quad \times \left(\frac{1}{n}\sum_{i=1}^n X_iZ_i'\right) W \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_iu_i\right).
\end{aligned} \tag{2.1}$$

By the multivariate central limit theorem, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i u_i \to_d N(0, \sigma^2 \mathbf{E}[Z_i Z_i']).$$

(Note that we use homoskedasticity assumption $\mathbf{E}[u_1^2|Z_1] = \sigma^2$.) Therefore, using the weak law of large numbers and Slutsky's Theorem, we obtain that

$$\sqrt{n}(\hat{\beta}_S^{IV} - \beta_0) \to_d [(\mathbf{E}X_i Z_i') W (\mathbf{E}Z_i X_i')]^{-1} (\mathbf{E}X_i Z_i') W \times \zeta,$$

where $\zeta \sim N(0, \sigma^2 \mathbf{E}[Z_i Z_i'])$. Now, the asymptotic covariance matrix of $\hat{\beta}_S^{IV}$ is given by

$$\sigma^2 [(\mathbf{E}X_i Z_i') W (\mathbf{E}Z_i X_i')]^{-1} (\mathbf{E}X_i Z_i') W \mathbf{E}[Z_i Z_i'] W (\mathbf{E}Z_i X_i') [(\mathbf{E}X_i Z_i') W (\mathbf{E}Z_i X_i')]^{-1}. \tag{2.2}$$

2.2. Two Stage Least Squares Estimation

We have seen that β_0 can be estimated multiple different ways depending on the choice of the weighting matrix W as long as it is symmetric and positive definite. When the model is just identified, i.e., $d = k$, the asymptotic variance matrix is given by

$$\sigma^2 [\mathbf{E}Z_i X_i']^{-1} \mathbf{E}[Z_i Z_i'] [\mathbf{E}X_i Z_i']^{-1}. \tag{2.3}$$

In other words, the asymptotic variance does not depend on W. However, when the model is over-identified, the choice of W matters. Each estimator with each choice of W has different asymptotic distribution with different asymptotic variance matrix. Thus, in this case, one might wonder which choice of W is "best". We call W an *optimal weighting matrix* if W minimizes the asymptotic covariance matrix of $\hat{\beta}_S^{IV}$.

Theorem 1: *The optimal weighting matrix W is given by*

$$W = (\mathbf{E}[Z_i Z_i'])^{-1}.$$

Proof: When we take $W = (\mathbf{E}[Z_i Z_i'])^{-1}$, the asymptotic covariance matrix becomes

$$\sigma^2 \left\{(\mathbf{E}X_i Z_i') (\mathbf{E}[Z_i Z_i'])^{-1} (\mathbf{E}Z_i X_i')\right\}^{-1}.$$

One can check the following:

$$\begin{aligned} & (\mathbf{E}X_i Z_i') (\mathbf{E}[Z_i Z_i'])^{-1} (\mathbf{E}Z_i X_i') \\ & - (\mathbf{E}X_i Z_i') W (\mathbf{E}Z_i X_i') [(\mathbf{E}X_i Z_i') W \mathbf{E}[Z_i Z_i'] W (\mathbf{E}Z_i' X_i)]^{-1} (\mathbf{E}X_i Z_i') W (\mathbf{E}Z_i X_i') \\ = \; & (\mathbf{E}X_i Z_i') (\mathbf{E}[Z_i Z_i'])^{-1/2} \{I - P\} (\mathbf{E}[Z_i Z_i'])^{-1/2} (\mathbf{E}Z_i X_i') \geq 0 \end{aligned}$$

where

$$P = (\mathbf{E}[Z_i Z_i'])^{1/2} W (\mathbf{E}Z_i X_i') [(\mathbf{E}X_i Z_i') W \mathbf{E}[Z_i Z_i'] W (\mathbf{E}Z_i' X_i)]^{-1} (\mathbf{E}X_i Z_i') W (\mathbf{E}[Z_i Z_i'])^{1/2}.$$

Note that P is symmetric and idempotent, and hence it is positive semidefinite. The proof is complete. ■

Since we do not know the expectation $\mathbf{E}[Z_i Z_i']$, we consider using its sample version:

$$\left(\frac{1}{n}\sum_{i=1}^{n} Z_i Z_i'\right)^{-1} = \left(\frac{Z'Z}{n}\right)^{-1}.$$

Using $(Z'Z/n)^{-1}$ as our W, we obtain the IV estimator as:

$$\begin{aligned} \hat{\beta}^{IV} &= \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1} X'Z(Z'Z)^{-1}Z'Y \\ &= (X'P_Z X)^{-1} X'P_Z Y, \end{aligned}$$

where $P_Z = Z(Z'Z)^{-1}Z'$ is the projection matrix onto $\mathcal{R}(Z)$. And

$$\sqrt{n}(\hat{\beta}^{IV} - \beta_0) \to_d N\left(0, \sigma^2 \left\{(\mathbf{E}X_i Z_i') (\mathbf{E}[Z_i Z_i'])^{-1} (\mathbf{E}Z_i X_i')\right\}^{-1}\right).$$

Suppose that we are in the situation of overidentification. Then one may consider $\tilde{Z}_i = HZ_i$ as another instrumental variable for some $d \times d$ matrix H such that $H'H$ is positive definite and consider the moment condition

$$S\mathbf{E}\tilde{Z}_i(Y_i - X_i'\beta_0) = 0.$$

Theorem 1 tells us that the estimator $\hat{\beta}_S^{IV}$ using $\tilde{Z}_i$ instead of $\tilde{Z}_i$ is not better than $\hat{\beta}^{IV}$ that uses the optimal weighting matrix, $(Z'Z/n)^{-1}$. Hence one cannot improve the quality of the IV estimator over $\hat{\beta}^{IV}$ simply by reshuffling the instrumental variables.

Theorem 1 has another important implication. When $d > k$, there are multiple ways to identify the same β_0. For example, we can identify β_0 using only a subvector $Z_{1i} \in \mathbf{R}^m$, $k \leq m < d$, of Z_i such that $\mathbf{E}X_i Z_{1i}'$ is still full column rank k. Then, we can construct, say, $\hat{\beta}_1^{IV}$, using Z_{1i} in place of Z_i. In general, the IV estimator that uses Z_{1i} as the IV will be different from that uses Z_i as the IV. Let us compare the quality of the two estimators $\hat{\beta}^{IV}$ and $\hat{\beta}_1^{IV}$.

Theorem 2: *The asymptotic covariance matrix of $\hat{\beta}_1^{IV}$ is larger than or equal to that of $\hat{\beta}^{IV}$.*

Proof: The asymptotic covariance matrix of $\hat{\beta}_1^{IV}$ is given by

$$\sigma^2 \left\{ (\mathbf{E}X_i Z_{1i}') \left(\mathbf{E}\left[Z_{1i}Z_{1i}'\right]\right)^{-1} (\mathbf{E}Z_{1i}X_{1i}') \right\}^{-1}.$$

Let A be a $m \times d$ selection matrix such that $AZ_i = Z_{1i}$. Then, we can rewrite the inverse of the asymptotic covariance matrix as

$$(\mathbf{E}X_i Z_i') A' \left(A\mathbf{E}\left[Z_i Z_i'\right] A'\right)^{-1} A (\mathbf{E}Z_i X_i').$$

Similarly as in the proof of Theorem 1,

$$\begin{aligned} & (\mathbf{E}X_i Z_i') \left(\mathbf{E}\left[Z_i Z_i'\right]\right)^{-1} (\mathbf{E}Z_i X_i') \\ & - (\mathbf{E}X_i Z_i') A' \left(A\mathbf{E}\left[Z_i Z_i'\right] A'\right)^{-1} A (\mathbf{E}Z_i X_i') \\ = \; & (\mathbf{E}X_i Z_i') \left(\mathbf{E}\left[Z_i Z_i'\right]\right)^{-1/2} \left\{ I - \left(\mathbf{E}\left[Z_i Z_i'\right]\right)^{1/2} A' \left(A\mathbf{E}\left[Z_i Z_i'\right] A'\right)^{-1} A \left(\mathbf{E}\left[Z_i Z_i'\right]\right)^{1/2} \right\} \\ & \times \left(\mathbf{E}\left[Z_i Z_i'\right]\right)^{-1/2} (\mathbf{E}Z_i X_i'). \end{aligned}$$

(Note that $(A\mathbf{E}\left[Z_i Z_i'\right] A')$ will not be invertible if $m > d$.) Now, observe that

$$\left(\mathbf{E}\left[Z_i Z_i'\right]\right)^{1/2} A' \left(A\mathbf{E}\left[Z_i Z_i'\right] A'\right)^{-1} A \left(\mathbf{E}\left[Z_i Z_i'\right]\right)^{1/2}$$

is symmetric and idempotent. Hence the above matrix is positive semidefinite. The proof is complete. ■

This implies that the quality of the estimator becomes worse, when we remove some of the intrumental variables. Since $P_Z = P_Z^2$, let $X^* = P_Z X$. Then, we can rewrite the IV estimator as

$$\hat{\beta}^{IV} = \left(X^{*\prime} X^*\right)^{-1} X^{*\prime} Y.$$

Therefore, the IV estimator is obtained, in the first stage, by projecting X onto $\mathcal{R}(Z)$, which means that we regress first X on Z and obtain the predicted value $P_Z X$, and then in the second stage, we project Y onto $\mathcal{R}(P_Z X)$, which means that we regress Y on $P_Z X$. The resulting estimator of β_0 is $\hat{\beta}^{IV}$. For this reason, the estimator $\hat{\beta}^{IV}$ is often called a *two stage least squares estimator* (2SLS).

More specifically, consider the following simultaneous equation model:

$$\begin{aligned} Y_i &= t_i\beta_1 + W_i'\beta_2 + u_i, \\ t_i &= Z_{1i}'\pi_1 + W_i'\pi_2 + v_i, \end{aligned} \tag{2.4}$$

where t_i is a scalar random variable and W_i is a $k_1 \times 1$ vector and $k = k_1 + 1$. Let Z_{1i} be a $(d - k_1) \times 1$ vector such that $d - k_1 \geq 1$. We assume that Z_{1i} is a vector of instrumental variables that are not included in the endogenous regresssion model. We set

$$Z_i = \begin{bmatrix} Z_{1i} \\ W_i \end{bmatrix}.$$

Furthermore, we assume that

$$\begin{aligned} \mathbf{E}t_iu_i &\neq 0, \\ \mathbf{E}W_iu_i &= 0, \\ \mathbf{E}Z_iu_i &= 0, \text{ and} \\ \mathbf{E}Z_iv_i &= 0. \end{aligned}$$

Therefore, t_i is a one-dimensional endogenous variable and W_i is a vector of exogenous variables and Z_i is a vector of instrumental variables. Then each entry, say, $\hat{t}_i$, of P_Zt is the predicted value of t_i from regressing t_i on Z_i, (this is the first stage regression) and also $P_ZW = W$, where

$$t = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_n \end{bmatrix} \text{ and } W = \begin{bmatrix} W_1' \\ W_2' \\ \vdots \\ W_n' \end{bmatrix},$$

because W belongs to $\mathcal{R}(Z)$. We take $X_i = [t_i, W_i']$. Then the 2SLS estimator of $\beta = (\beta_1, \beta_2')'$ is obtained by regressing Y_i on the $\hat{t}_i$ and W_i. (This is the second stage regression.)

One should be careful in implementing this using a built-in regression command in the software. The standard errors generated from the second stage regression is not correct in general, because the resulting standard error is based on the following estimator of σ^2 :

$$\frac{1}{n}\left(Y - P_ZX\hat{\beta}^{IV}\right)'\left(Y - P_ZX\hat{\beta}^{IV}\right).$$

However, the correct estimator of σ^2 is $\frac{1}{n}\left(Y - X\hat{\beta}^{IV}\right)'\left(Y - X\hat{\beta}^{IV}\right)$.

2.3. Overidentifying Restrictions Test

One may wonder if there is any way to formally check whether Z_i is a plausible instrumental variable. In this situation, we need to check whether Z_i satisfies the validity condition and the relevance condition. For the validity condition, the check amounts to testing the moment condition:

$$\begin{aligned} H_0 &: \mathbf{E}Z_i(Y_i - X_i'\beta_0) = 0. \\ H_1 &: \mathbf{E}Z_i(Y_i - X_i'\beta_0) \neq 0. \end{aligned} \tag{2.5}$$

A natural starting point in constructing a test is first to see whether the sample version of the moment:

$$\frac{1}{n}\sum_{i=1}^{n} Z_i\hat{u}_i$$

is close to zero or not, where

$$\hat{u}_i = Y_i - X_i'\hat{\beta}^{IV},$$

and $\hat{\beta}^{IV}$ is a two-stage least squares estimator.

One should be careful here. If $k = d$, i.e., the model is exactly identified, the above sample version is zero always by the definition of $\hat{\beta}^{IV}$, regardless of whether the null hypothesis holds or not, and hence we cannot use it as a basis for our test. However, when the model is overidentified, this idea of testing the moment condition works. The idea follows that of the Wald test. First define

$$J = n\left(\frac{1}{n}\sum_{i=1}^{n} Z_i\hat{u}_i\right)'\hat{C}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i\hat{u}_i\right),$$

where

$$\hat{C} = \left(\frac{1}{n}\sum_{i=1}^{n} \hat{u}_i^2 Z_iZ_i'\right)^{-1}.$$

Let us consider the following conditions.

(C1) $\mathbf{E}[\|X_i\|^4] + \mathbf{E}[\|Z_i\|^4] < \infty$.
(C2) $\mathbf{E}[u_i^2|Z_i] = \sigma^2$.
(C3) $\mathbf{E}[Z_i Z_i']$ is positive definite.

Theorem 1: *Suppose that Conditions (C1)-(C3) hold. Then, under the null hypothesis,*

$$J_d \to_d \chi^2_{d-k}.$$

Proof: First, note that

$$\frac{1}{n}\sum_{i=1}^n \hat{u}_i^2 Z_i Z_i' = \frac{1}{n}\sum_{i=1}^n u_i^2 Z_i Z_i' + \frac{1}{n}\sum_{i=1}^n (\hat{u}_i^2 - u_i^2) Z_i Z_i'.$$

Let the (m,ℓ)-th entry of $Z_i Z_i'$ be denoted by $H_{m,\ell,i}$. We show that for each $m, \ell = 1, ..., d$,

$$\frac{1}{n}\sum_{i=1}^n (\hat{u}_i^2 - u_i^2) H_{m,\ell,i} = o_P(1).$$

Then this proves that

$$\frac{1}{n}\sum_{i=1}^n (\hat{u}_i^2 - u_i^2) Z_i Z_i' = o_P(1).$$

Write

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^n (\hat{u}_i^2 - u_i^2) H_{m,\ell,i} \\ = \; & \frac{1}{n}\sum_{i=1}^n (\hat{u}_i - u_i)(\hat{u}_i + u_i) H_{m,\ell,i} \\ = \; & 2\frac{1}{n}\sum_{i=1}^n (\hat{u}_i - u_i) u_i H_{m,\ell,i} + \frac{1}{n}\sum_{i=1}^n (\hat{u}_i - u_i)^2 H_{m,\ell,i}. \end{aligned}$$

Since

$$\hat{u}_i - u_i = -X_i'(\hat{\beta}^{IV} - \beta),$$

we write

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^n (\hat{u}_i - u_i)(\hat{u}_i + u_i) H'_{m,\ell,i} \\ = \; & -2\left(\frac{1}{n}\sum_{i=1}^n H_{m,\ell,i} u_i X_i'\right)(\hat{\beta}^{IV} - \beta) + (\hat{\beta}^{IV} - \beta)'\left(\frac{1}{n}\sum_{i=1}^n X_i X_i' H_{m,\ell,i}\right)(\hat{\beta}^{IV} - \beta)'. \end{aligned}$$

Since $\hat{\beta}^{IV}$ is a consistent estimator of β, by the Law of the Large Numbers and Slutsky's lemma, the above terms are $o_P(1)$. This implies that

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n \hat{u}_i^2 Z_i Z_i' &= \frac{1}{n}\sum_{i=1}^n u_i^2 Z_i Z_i' + o_P(1) = \mathbf{E}[u_i^2 Z_i Z_i'] + o_P(1) \\ &= \sigma^2 \mathbf{E}[Z_i Z_i'] + o_P(1), \end{aligned}$$

the second to the last equality coming from the Law of the Large Numbers, and the last equality coming from the conditional homoskedasticity assumption (C2). Letting

$$C = (\sigma^2 \mathbf{E}[Z_i Z_i'])^{-1},$$

we conclude that

$$\hat{C} = C + o_P(1). \tag{2.6}$$

We write

$$\begin{aligned} & \hat{C}^{1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i\hat{u}_i \\ = \;& \hat{C}^{1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i u_i + \hat{C}^{-1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i(\hat{u}_i - u_i) = A_{1n} + A_{2n}, \text{ say.} \end{aligned}$$

By the Central Limit Theorem,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i u_i \to_d N(0, \mathbf{E}[u_i^2 Z_i Z_i']),$$

and hence it is $O_P(1)$. Therefore,

$$\hat{C}^{1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i u_i = C^{1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i u_i + o_P(1).$$

Let us turn to A_{2n}. Write it as

$$\begin{aligned} & -\hat{C}^{1/2}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i X_i'\right)(\hat{\beta}^{IV} - \beta) \\ = \;& -\hat{C}^{1/2}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i X_i'\right)\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_i Z_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n} Z_i Z_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} Z_i X_i'\right)\right]^{-1} \\ & \times\left(\frac{1}{n}\sum_{i=1}^{n} X_i Z_i'\right)\left(\frac{1}{n}\sum_{i=1}^{n} Z_i Z_i'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i u_i\right). \end{aligned}$$

(See (2.1).) Applying the Law of of the Large Numbers, Slutsky's Lemma and (2.6), we find that

$$\begin{aligned} A_{2n} = \;& -C^{1/2}\mathbf{E}[Z_i X_i']\left[\mathbf{E}[X_i Z_i']\left(\sigma^2\mathbf{E}[Z_i Z_i']\right)^{-1}\mathbf{E}[Z_i X_i']\right]^{-1} \\ & \times\mathbf{E}[X_i Z_i']\left(\sigma^2\mathbf{E}[Z_i Z_i']\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i u_i + o_P(1). \end{aligned}$$

We conclude that

$$A_{1n} + A_{2n} = PC^{1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i u_i + o_P(1), \tag{2.7}$$

where

$$P = I - C^{1/2}\mathbf{E}[Z_i X_i']\left[\mathbf{E}[X_i Z_i']\left(\sigma^2\mathbf{E}[Z_i Z_i']\right)^{-1}\mathbf{E}[Z_i X_i']\right]^{-1}\mathbf{E}[X_i Z_i']C^{1/2}.$$

It is not hard to see that P is symmetric and idempotent and hence is a projection matrix. By the Central Limit Theorem, we find that

$$C^{1/2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i u_i \to_d N(0, I_d).$$

Since

$$J = (A_{1n} + A_{2n})'(A_{1n} + A_{2n}) \to_d \chi^2_{rk(P)}.$$

The rank of P is $d - k$, delivering the desired result. ■

Inspecting the proof shows that the premuliplication by projection matrix P in (2.7) arises due to the term that comes from $\hat{u}_i - u_i$. If we knew u_i and used it in place of $\hat{u}_i$ in constructing J test statistic, the test statistic would have limiting distribution χ^2_d. However, we used part of the data to estimate β to obtain $\hat{u}_i$, and this leads to the reduction in the degree of freedom in the χ^2_{d-k} distribution.

Thus, the overidentifying restrictions test rejects the null hypothesis if J is larger than the critical value from the distribution χ^2_{d-k} distribution. The test is also called a *J test*.

3. Relevance of IV

3.1. Testing for the Relevance of IV

The relevance of the IV can be tested by considering the following regression model:

$$t_i = Z'_{1i}\pi_1 + W'_i\pi_2 + v_i.$$

When all the parameters in π_1 are significantly different from zero, we take this as evidence in favor of the relevance condition. The null hypothesis is taken to be

$$\begin{aligned} H_0 &: \pi_1 = 0 \text{ against} \\ H_1 &: \pi_1 \neq 0. \end{aligned}$$

The test can be done by using the usual F-test. The failure of the rejection of the null hypothesis indicates the problem with the relevance of the IV. However, one needs to be careful that the F-test that we are using is not the one that testing $H_0 : \pi_1 = \pi_2 = 0$. If we reject the null hypothesis of this latter form, this does not say that the instrumental variable is relevant with statistical significance.

3.2. Weak Instruments

Certainly the relevance condition that $\mathbf{E}[XZ']$ is full row rank k is necessary for the identification of β_0. When the rank of $\mathbf{E}[X_iZ'_i]$ almost fails to be k (e.g. having the smallest nonzero eigenvalue is close to zero), we say that the instrument Z_i is a *weak instrument*. This happens especially when the correlations between an entry of Z_i and all the elements of X_i are close to zero. In this case, the asymptotic normal distribution of the IV estimator may not be a good approximation of its fininte sample distribution. First, let us consider an extreme situation where there is no correlation between the endogenous regressor and the instrumental variable. Observe that

$$\begin{aligned} \hat{\beta}^{IV} - \beta_0 &= \left[\left(\frac{1}{n}\sum_{i=1}^n X_iZ'_i\right)\left(\frac{1}{n}\sum_{i=1}^n Z_iZ'_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n Z_iX'_i\right)\right]^{-1} \\ &\quad \times \left(\frac{1}{n}\sum_{i=1}^n X_iZ'_i\right)\left(\frac{1}{n}\sum_{i=1}^n Z_iZ'_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n Z_iu_i\right) \\ &= \left[\left(\frac{1}{n}\sum_{i=1}^n X_iZ'_i\right)\left(\frac{1}{n}\sum_{i=1}^n Z_iZ'_i\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_iX'_i\right)\right]^{-1} \\ &\quad \times \left(\frac{1}{n}\sum_{i=1}^n X_iZ'_i\right)\left(\frac{1}{n}\sum_{i=1}^n Z_iZ'_i\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_iu_i\right), \end{aligned}$$

by multiplying by $\sqrt{n}$ and dividing by $\sqrt{n}$. Since $\mathbf{E}Z_iX'_i = 0$ and $\mathbf{E}Z_iu_i = 0$, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i[X'_i, u_i] \to_d N\left(0, \mathbf{E}\left[[X'_i, u_i]'Z_iZ'_i[X'_i, u_i]\right]\right),$$

and

$$\left(\frac{1}{n}\sum_{i=1}^n X_iZ'_i\right)\left(\frac{1}{n}\sum_{i=1}^n Z_iZ'_i\right)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n Z_iX'_i \to_d N(0, V)$$

where $[X'_i, u_i]$ is a $1 \times (k+1)$ row vector, and V is a positive definite matrix. Therefore, $\hat{\beta}^{IV} - \beta_0$ converges in distribution to a random quantity that is non-normal in general, and the estimator $\hat{\beta}^{IV}$ is not consistent. This means that the limiting distribution of $\sqrt{n}(\hat{\beta}^{IV} - \beta_0)$ behaves *discontinuously* as we move from a distribution with $\mathbf{E}[XZ']$ is full column rank to a distribution with $\mathbf{E}[XZ'] = 0$. This discontinuity happens purely because we have sent $n \to \infty$. In finite samples, the exact distribution of $\hat{\beta}^{IV}$ may change continuously as we move one distribution toward a distribution with $\mathbf{E}[XZ'] = 0$. Therefore, when the underlying distribution is such that $\mathbf{E}[XZ']$ is full column rank but close to zero, the asymptotic distribution of $\hat{\beta}^{IV}$ will be one that is obtained under the usual relevance condition for the IV. However, the finite sample distribution of $\hat{\beta}^{IV}$ is expected to be closer to one under $\mathbf{E}[XZ'] = 0$. The standard errors based on the asymptotic normality of $\hat{\beta}^{IV}$ cannot be a reliable one. This problem is called a *weak instrument problem*. Usually, the weak intrument problem is detected by checking the relevance condition using the observations. When it is suspected that there is a weak instrument, one has to rely on a different asymptotic theory that reflects this situation.

How do I know if my regression may suffer from the weak IV problem? Consider the situation where there is only one endogenous regressor denoted by t_i as in the model of (2.4). The first idea will be that one performs an F-test in the regression of t_i on the IVs in the regression equation:

$$t_i = Z_{1i}'\pi_1 + W_i'\pi_2 + v_i.$$

The null hypothesis is taken to be

$$H_0 : \pi_1 = 0 \text{ against } H_1 : \pi_1 \neq 0.$$

The rule-of-thumb critical value for this F-test at 5% is 10. This rule is different from the usual F-test, because this rule is based on the null hypothesis that is weaker than the conventional null hypothesis that the regression coefficients in the first stage regression are all zero. The former null hypothesis covers not only the latter null hypothesis but also the case where the coefficients are very close to zero.

If you have only one endogenous regressor and at least one instrument that is strongly relevant, there is no weak instrument problem. However, if all the instruments are weak, you need to use a different standard error formula that is robust to the weakness of the instruments. The inference under weak instruments lies beyond the scope of this course, and interested students are referred to advanced textbook that contains topics on weak instruments.