Instrumental Variables and Identification

References

Davidson, Russell and James G. MacKinnon (2004). *Econometric Theory and Methods*. Oxford University Press.
Hayashi, Fumio (2000). *Econometrics*. Princeton University Press.

1 Two stage least squares: estimation, identification and testing

Consider the model

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + u_i, \; i = 1, 2, ..., n,$$

in which it is known that $\mathbb{E}[u_i] = 0$. *We also know that* $\mathbb{E}[z_{1i}u_i] = 0$ *and* $\mathbb{E}[z_{2i}u_i] = 0$. *You may assume that the observed variables have mean zero, the data on* $(y_i, x_{1i}, x_{2i}, z_{1i}, z_{2i})$ *are IID over* $i = 1, ..., n$, *and relevant conditional variances of* u_i *are homoskedastic.*

First, it is worth going over some definitions and derivations of instrumental variables estimators. Consider the model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \; \mathbb{E}[\mathbf{u}\mathbf{u}^\top] = \sigma^2\mathbf{I}, \tag{1}$$

where $\mathbf{X} = \begin{bmatrix}\mathbf{X}_1 & \mathbf{X}_2\end{bmatrix}$ is an $n \times k$ matrix, $\mathbb{E}[\mathbf{u}] = 0$, $\mathbb{E}[\mathbf{Z}^\top\mathbf{u}] = 0$, and that $\mathbf{y}$, and $\mathbf{X}$, and $\mathbf{Z}$ are IID.

Suppose that for each row observation, condition (1) is satisfied for some suitable information set Ω_t, and that we can form an $n \times l$ matrix $\mathbf{Z}$ with typical row $\mathbf{Z}_t$ such that all its elements belong to Ω_t. The l variables given by the l columns of $\mathbf{Z}$ are the instruments. For part 1.1.1, they will include $\mathbf{X}_1$ as a suitable instrument. In this case, $l > k$ and the model is said to be over-identified, because there is more than one way to formulate the moment condition

$$\mathbf{Z}^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0}. \tag{2}$$

Just identified models where $k = l$ allows the instrumental variable (IV) estimator to be unique, since there is only one k-dimensional linear space $\mathcal{S}(\mathbf{Z})$ that can be spanned by the $k = l$ instruments. We can treat an over-identified model as if it were just identified by choosing exactly k linear combinations of the l columns of $\mathbf{Z}$. Formally, we seek an $l \times k$ matrix $\mathbf{J}$ such that the $n \times k$ matrix $\mathbf{ZJ}$ is a valid instrument matrix and such that, by using $\mathbf{J}$, the asymptotic variance covariance matrix of the estimator is minimised relative to other relevant IV estimators. In this sense $\mathbf{ZJ}$ can be thought of as an ideal selection of correctly specified instruments.

There are three requirements that matrix $\mathbf{J}$ must satisfy: 1) it must have full column rank k, otherwise the space spanned by the columns of $\mathbf{ZJ}$ would have less rank than k, and the model would be underspecified; 2) $\mathbf{J}$ should be asymptotically deterministic;[1] and, 3) $\mathbf{J}$ should be chosen to minimise the asymptotic covariance matrix of the resulting IV estimator. The second condition is assumed to hold (given to us in the question), while the first condition is fairly trivial. The third condition requires some rigour.

To begin with, assume that $\mathbf{X}$ satisfies the following

$$\mathbf{X} = \bar{\mathbf{X}} + \mathbf{V}, \ \mathbb{E}[\mathbf{V}_t|\Omega_t] = \mathbf{0}, \tag{3}$$

where the t-th row of $\bar{\mathbf{X}}$ is $\bar{\mathbf{X}}_t = \mathbb{E}[\mathbf{X}_t|\Omega_t]$, and $\mathbf{X}_t$ is the t-th row of $\mathbf{X}$. Thus, $\bar{\mathbf{X}}$ is the expectation of $\mathbf{X}_t$ conditional on information set Ω_t. Using (3) and a LLN, we see that

$$\begin{aligned}\operatorname*{plim}_{n\to\infty}\frac{1}{n}\mathbf{X}^\top\mathbf{Z} &= \lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\mathbf{X}^\top\mathbf{Z}] \\ &= \lim_{n\to\infty}\frac{1}{n}\mathbb{E}[\bar{\mathbf{X}}^\top\mathbf{Z}] = \operatorname*{plim}_{n\to\infty}\frac{1}{n}\bar{\mathbf{X}}^\top\mathbf{Z}.\end{aligned} \tag{4}$$

If the model (1) is correctly specified with true parameter vector $\boldsymbol{\beta}_0$ and true error variance σ_0^2, the asymptotic covariance matrix of $\sqrt{n}(\hat{\boldsymbol{\beta}}_{IV} - \boldsymbol{\beta}_0)$ is given by[2]

$$\begin{aligned}\operatorname{Var}\left[\operatorname*{plim}_{n\to\infty}\sqrt{n}(\hat{\boldsymbol{\beta}}_{IV} - \boldsymbol{\beta}_0)\right] &= \sigma_0^2(\mathbf{S}_{\mathbf{Z}^\top\mathbf{X}})^{-1}\mathbf{S}_{\mathbf{Z}^\top\mathbf{Z}}(\mathbf{S}_{\mathbf{Z}^\top\mathbf{X}}^\top)^{-1} \\ &= \sigma_0^2\operatorname*{plim}_{n\to\infty}\left(n^{-1}\mathbf{X}^\top\mathbf{P}_\mathbf{Z}\mathbf{X}\right)^{-1},\end{aligned} \tag{5}$$

where $\mathbf{S}_{\mathbf{Z}^\top\mathbf{Z}} = \text{plim } n^{-1}\mathbf{Z}^\top\mathbf{Z}$, $\mathbf{P}_\mathbf{Z} = \mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top$ is the projection matrix which projects on to the linear space $\mathcal{S}(\mathbf{Z})$, and where instruments are chosen to minimise the above asymptotic covariance matrix. It follows from (4) and (5) that the asymptotic covariance matrix of the IV estimator computed using $\mathbf{ZJ}$ as the instrument matrix is

$$\sigma_0^2\operatorname*{plim}_{n\to\infty}\left(n^{-1}\bar{\mathbf{X}}^\top\mathbf{P}_{\mathbf{ZJ}}\bar{\mathbf{X}}\right)^{-1}. \tag{6}$$

However, $\bar{\mathbf{X}}$ is not observed and in general it is not possible to find a matrix $\mathbf{J}$ such that $\bar{\mathbf{X}} = \mathbf{ZJ}$. But we can proceed by projecting $\bar{\mathbf{X}}$ on to the linear space $\mathcal{S}(\mathbf{Z})$, yielding the matrix of instruments:

$$\begin{aligned}\mathbf{ZJ} &= \mathbf{P}_\mathbf{Z}\bar{\mathbf{X}} = \mathbf{Z}(\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top\bar{\mathbf{X}} \\ \implies \mathbf{J} &= (\mathbf{Z}^\top\mathbf{Z})^{-1}\mathbf{Z}^\top\bar{\mathbf{X}}.\end{aligned} \tag{7}$$

[1] If not then $\operatorname*{plim}_{n\to\infty}\frac{1}{n}\mathbf{Z}^\top\mathbf{u} = \mathbf{0}$ would not hold and the IV estimator would be inconsistent.

[2] Here I assume the simple IV estimator

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

hence we get

$$(\hat{\boldsymbol{\beta}}_{\text{IV}} - \boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}}_{\text{IV}} - \boldsymbol{\beta}_0)^\top = (\mathbf{Z}^\top\mathbf{X})^{-1}\mathbf{Z}^\top\mathbf{u}\mathbf{u}^\top\mathbf{Z}(\mathbf{Z}^\top\mathbf{X})^{-1}.$$

To show that these instruments are optimal, begin by substituting $\mathbf{P_Z}\bar{\mathbf{X}}$ for $\mathbf{ZJ}$ in (6):

$$\sigma_0^2 \operatorname*{plim}_{n\to\infty} \left(n^{-1}\bar{\mathbf{X}}^\top \mathbf{P}_{\mathbf{P_Z}\bar{\mathbf{X}}}\bar{\mathbf{X}}\right)^{-1},$$

where

$$\bar{\mathbf{X}}\mathbf{P}_{\mathbf{P_Z}\bar{\mathbf{X}}}\bar{\mathbf{X}} = \bar{\mathbf{X}}^\top \mathbf{P_Z}\bar{\mathbf{X}}(\bar{\mathbf{X}}^\top \mathbf{P_Z}\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}^\top \mathbf{P_Z}\bar{\mathbf{X}} = \bar{\mathbf{X}}\mathbf{P_Z}\bar{\mathbf{X}}. \tag{8}$$

The precision matrix for the estimator that uses instruments $\mathbf{P_Z}\bar{\mathbf{X}}$ is proportional to $\bar{\mathbf{X}}^\top \mathbf{P_Z}\bar{\mathbf{X}}$. For the estimator which uses $\mathbf{ZJ}$ as instruments, the precision matrix is proportional to $\bar{\mathbf{X}}^\top \mathbf{P_{ZJ}}\bar{\mathbf{X}}$. The difference between the two precision matrices is proportional to

$$\bar{\mathbf{X}}^\top(\mathbf{P_Z}-\mathbf{P_{ZJ}})\bar{\mathbf{X}}.$$

The k-dimensional subspace $\mathcal{S}(\mathbf{ZJ})$, which is an image of the orthogonal projection $\mathbf{P_{ZJ}}$ is a subspace of the l-dimensional space $\mathcal{S}(\mathbf{Z})$, which is the image of $\mathbf{P_Z}$. Therefore, $\mathbf{P_Z} - \mathbf{P_{ZJ}}$ is itself an orthogonal projection matrix, implying that $\bar{X}^\top(\mathbf{P_Z}-\mathbf{P_{ZJ}})\bar{\mathbf{X}}$ is a positive semidefinite matrix. This proves that $\mathbf{ZJ}$ is the best choice of instruments.

As mentioned before, we do not know $\bar{\mathbf{X}}$ but if we know from (7) that

$$\operatorname*{plim}_{n\to\infty} \mathbf{J} = \operatorname*{plim}_{n\to\infty}(n^{-1}\mathbf{Z}^\top\mathbf{Z})^{-1}n^{-1}\mathbf{Z}^\top\bar{\mathbf{X}},$$

and from (4) that this is equal to

$$= \operatorname*{plim}_{n\to\infty}(n^{-1}\mathbf{Z}^\top\mathbf{Z})^{-1}n^{-1}\mathbf{Z}^\top\mathbf{X}.$$

This implies that we can use $\mathbf{P_Z X}$ instead of $\mathbf{P_Z}\bar{\mathbf{X}}$ without changing the asymptotic properties of the estimator. Thus, if we use $\mathbf{P_Z X}$ as the matrix of instrumental variables, the moment conditions (2) that define the estimator becomes

$$\mathbf{X}^\top \mathbf{P_Z}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0}, \tag{9}$$

which gives the generalised IV (GIV) estimator or two-stage least squares (2SLS) estimator

$$\hat{\boldsymbol{\beta}}_{\mathrm{GIV/2SLS}} = (\mathbf{X}^\top\mathbf{P_Z}\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{P_Z}\mathbf{y}. \tag{10}$$

1.1

First suppose it is known that $\mathbb{E}[x_{1i}u_i] = 0$, but it is not known whether $\mathbb{E}[x_{2i}u_i] = 0$ or $\mathbb{E}[x_{2i}u_i] \neq 0$.

1.1.1

Briefly explain how to calculate the two stage least squares (2SLS) estimator of the parameter vector $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$, treating x_{2i} as correlated with u_i.

Calculate the GIV/2SLS estimator as given in (10) where $\mathbf{Z} = \begin{bmatrix}\mathbf{X}_1 & \mathbf{Z}_1 & \mathbf{Z}_2\end{bmatrix}$ since $\mathbf{X}_1$ is a valid instrument. Using the 2SLS approach, first regress $\mathbf{X}_2$ on $\mathbf{Z}$ with the following first stage regression

$$\mathbf{X}_2 = \mathbf{Z}\boldsymbol{\pi} + \mathbf{v},\ \mathbb{E}[\mathbf{Z}^\top\mathbf{v}] = \mathbf{0}. \tag{11}$$

The fitted values of $\mathbf{X}_2$ from the first regression, plus the actual values of $\mathbf{X}_1$ are collected to form the matrix $\mathbf{P_Z X}$. Then, the second stage regression

$$\mathbf{y} = \mathbf{P_Z X}\boldsymbol{\beta} + \boldsymbol{u},$$

is used to obtain the 2SLS estimator (10).

1.1.2

With reference to the first stage equation for x_{2i}, state a necessary and sufficient condition for the 2SLS estimator to estimate the parameter vector $\boldsymbol{\beta}$ consistently in this model.

Necessary and sufficient conditions for the 2SLS estimator to estimate $\boldsymbol{\beta}$ consistently are the order condition (necessary but not sufficient) and the rank condition (necessary and sufficient). The order condition is that our model must not be underspecified, so $l \geq k$. For sufficiency, we first require that whenever $\mathbf{Z}_t \in \Omega_t$,

$$\mathbb{E}[u_t|\mathbf{Z}_t] = 0, \tag{12}$$

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

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

where $\mathbf{S}_{\mathbf{Z}^\top \mathbf{X}}$ is deterministic and nonsingular. In addition we assume that $\mathbf{S}_{\mathbf{Z}^\top \mathbf{Z}}$ exists and is of full rank (defined just after (5)). If $\mathbf{S}_{\mathbf{Z}^\top \mathbf{Z}}$ does not have full rank, then at least one of the instruments is perfectly collinear with the others, asymptotically, and should be dropped. If $\mathbf{S}_{\mathbf{Z}^\top \mathbf{X}}$ does not have full rank then the asymptotic version of the moment conditions (9) has fewer than k linearly independent equations, and these conditions therefore have no unique solution. These asymptotic conditions are sufficient for consistency of the 2SLS estimator. But the key necessary condition is

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

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 2SLS estimator would not be consistent.

1.2

Now suppose it is known that $\mathbb{E}[X_{2i}u_i] \neq 0$ but it is not known whether $\mathbb{E}[X_{1i}u_i] = 0$ or not.

1.2.1

Briefly explain how to calculate the 2SLS estimator of the parameter vector $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$, treating both x_{1i} and x_{2i} as correlated with u_i.

The procedure here is similar to part 1.1.1, except that we can no longer use $\mathbf{X}_1$ as a valid instrument, and so $\mathbf{Z} = \begin{bmatrix}\mathbf{Z}_1 & \mathbf{Z}_2\end{bmatrix}$. Since $k = l$ our model is just specified, and so our moment conditions simplify down to

$$\mathbf{Z}^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0},$$

which we can solve directly to obtain the simple IV estimator

$$\hat{\boldsymbol{\beta}}_{\mathrm{IV}} = \left(\mathbf{Z}^\top\mathbf{X}\right)^{-1}\mathbf{Z}^\top\mathbf{y}. \tag{13}$$

1.2.2

State a necessary and sufficient condition for the 2SLS estimator to estimate the parameter vector $\boldsymbol{\beta}$ consistently in this case.

Again, like 1.1.2, we require an order condition (necessary but not sufficient) and rank condition (necessary and sufficient). The order condition requires that the model not be underspecified, so $l \geq k$. To satisfy our sufficiency condition, and for asymptotic identification, whenever $\mathbf{Z}_t \in \Omega_t$, $\mathbb{E}[u_t|\mathbf{Z}_t] = 0$ and $\mathbf{Z}_t$ is assumed to be predetermined with respect to the error term. Our rank condition ensures that

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

exists and is non-singular.

To illustrate how these conditions are important: If the model (1) is correctly specified with true parameter vector $\boldsymbol{\beta}_0$, then

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\mathrm{IV}} &= (\mathbf{Z}^\top\mathbf{X})^{-1}\mathbf{Z}^\top\mathbf{X}\boldsymbol{\beta}_0 + (\mathbf{Z}^\top\mathbf{X})^{-1}\mathbf{Z}^\top\mathbf{u} \\ &= \boldsymbol{\beta}_0 + (n^{-1}\mathbf{Z}^\top\mathbf{X})^{-1}n^{-1}\mathbf{Z}^\top\mathbf{u}.\end{aligned}$$

Similar to part 1.1.2, given the assumption of $\mathbf{S}_{\mathbf{Z}^\top\mathbf{X}}$, $\hat{\boldsymbol{\beta}}_{\mathrm{IV}}$ is consistent if and only if

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

In other words, we can only have consistency of the IV estimator if the error terms are asymptotically uncorrelated with the instruments.

1.2.3

Assuming that the condition stated in 1.2.1 is satisfied, suggest a procedure for testing whether x_{1i} is correlated with u_i in this case.

We can use a version of the Hausman test here. But before doing so, it's important to go over the variance and asymptotic distribution of the IV estimator. Like every estimator that we have covered, the IV estimator is asymptotically normally distributed with an asymptotic variance-covariance matrix that can be estimated consistently.

Begin by writing the sample error:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\text{IV}} - \boldsymbol{\beta}_0 &= (\mathbf{Z}^\top \mathbf{X})^{-1} \mathbf{Z}^\top \mathbf{u} \\ &= \left(\frac{1}{n} \mathbf{Z}^\top \mathbf{X} \right)^{-1} \frac{1}{n} \mathbf{Z}^\top \mathbf{u},\end{aligned}$$

where assumed that

$$\begin{aligned}\mathbf{S}_{\mathbf{Z}^\top \mathbf{X}}^{-1} &= \operatorname*{plim}_{n\to\infty} \left(\frac{1}{n} \mathbf{Z}^\top \mathbf{X} \right)^{-1}, \\ \frac{1}{n} \mathbf{Z}^\top \mathbf{u} &\xrightarrow{p} \mathbb{E}[\mathbf{Z}^\top \mathbf{u}] = \mathbf{0}.\end{aligned}$$

So,[3]

$$\begin{aligned}\operatorname*{plim}_{n\to\infty} \hat{\boldsymbol{\beta}}_{\text{IV}} - \boldsymbol{\beta}_0 &= \mathbf{S}_{\mathbf{Z}^\top \mathbf{X}}^{-1} \mathbf{0} \\ \implies \hat{\boldsymbol{\beta}}_{\text{IV}} &\xrightarrow{p} \boldsymbol{\beta}_0. \end{aligned} \tag{14}$$

Next, we show that

$$\sqrt{n} \left(\hat{\boldsymbol{\beta}}_{\text{IV}} - \boldsymbol{\beta}_0 \right) \xrightarrow{d} N \left(0, \widehat{\text{AVar}}(\hat{\boldsymbol{\beta}}_{\text{IV}}) \right). \tag{15}$$

Start by multiplying the sample error by $\sqrt{n}$:

$$\sqrt{n} \left(\hat{\boldsymbol{\beta}}_{\text{IV}} - \boldsymbol{\beta}_0 \right) = \mathbf{S}_{\mathbf{Z}^\top \mathbf{X}}^{-1} \frac{1}{\sqrt{n}} \mathbf{Z}^\top \mathbf{u}.$$

Now, we know by the CLT that $n^{-1/2} \mathbf{Z}^\top \mathbf{u} \xrightarrow{d} N(0, \sigma_0^2 \mathbf{S}_{\mathbf{Z}^\top \mathbf{Z}})$. So, by Slutsky's Theorem, $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\text{IV}} - \boldsymbol{\beta}_0)$ converges to a normal distribution with mean $\mathbf{0}$ and with asymptotic variance $\sigma_0^2 \mathbf{S}_{\mathbf{Z}^\top \mathbf{X}}^{-1} \mathbf{S}_{\mathbf{Z}^\top \mathbf{Z}} \mathbf{S}_{\mathbf{Z}^\top \mathbf{X}}^{-1}$. But we can write the asymptotic variance as:

$$\begin{aligned}\text{AVar}(\hat{\boldsymbol{\beta}}_{\text{IV}}) &= \sigma^2 \left(\mathbf{S}_{\mathbf{Z}^\top \mathbf{X}} \mathbf{S}_{\mathbf{Z}^\top \mathbf{Z}}^{-1} \mathbf{S}_{\mathbf{Z}^\top \mathbf{X}} \right)^{-1} \\ &= \sigma_0^2 \operatorname*{plim}_{n\to\infty} \left(\frac{\mathbf{X}^\top \mathbf{Z}}{n} \left(\frac{\mathbf{Z}^\top \mathbf{Z}}{n} \right)^{-1} \frac{\mathbf{Z}^\top \mathbf{X}}{n} \right)^{-1} \\ \widehat{\text{AVar}}(\hat{\boldsymbol{\beta}}_{\text{IV}}) &= \hat{\sigma}^2 \left(\frac{1}{n} \mathbf{X}^\top \mathbf{P}_{\mathbf{Z}} \mathbf{X} \right)^{-1}, \end{aligned} \tag{16}$$

giving our result. Note that we can do this due to the invertibility (rank condition met) of $\mathbf{Z}^\top \mathbf{X}$.

We can also replace $\mathbf{Z}$ by $\mathbf{P}_{\mathbf{Z}} \mathbf{X}$, and just like in (8), we find that

$$\mathbf{X}^\top \mathbf{P}_{\mathbf{P}_{\mathbf{Z}} \mathbf{X}} \mathbf{X} = \mathbf{X}^\top \mathbf{P}_{\mathbf{Z}} \mathbf{X} (\mathbf{X}^\top \mathbf{P}_{\mathbf{Z}} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{P}_{\mathbf{Z}} \mathbf{X} = \mathbf{X}^\top \mathbf{P}_{\mathbf{Z}} \mathbf{X},$$

[3] If we assume that $\mathbf{Z}^\top \mathbf{X}$ is ergodic stationarity, then by the Ergodic Theorem, we could state $\mathbf{S}_{\mathbf{Z}^\top \mathbf{X}} \xrightarrow{a.s.} \boldsymbol{\Sigma}_{\mathbf{Z}^\top \mathbf{X}}$.

from which it follows that (5) is unchanged if $\mathbf{Z}$ if replaced by $\mathbf{P_Z X}$ (showing the equality of IV and 2SLS/GIV in this case).

Now that we have our definitions clear, we can proceed with the Hausman test. We wish to test

$$\begin{aligned} H_0 &: \mathbb{E}[\mathbf{X}_1^\top \mathbf{u}] = 0 \\ H_1 &: \mathbb{E}[\mathbf{X}_1^\top \mathbf{u}] \neq 0, \end{aligned}$$

and so we should compare our instrumental variable estimators from 1.1.1 and 1.2.1. Under the null hypothesis, the estimator outlined in 1.1.1, $\hat{\boldsymbol{\beta}}_{\text{IV}}^a$, and the estimator outlined in 1.2.1, $\hat{\boldsymbol{\beta}}_{\text{IV}}^b$, say, both yield consistent estimators of $\boldsymbol{\beta}$. But $\hat{\boldsymbol{\beta}}_{\text{IV}}^b$ should be asymptotically less efficient than $\hat{\boldsymbol{\beta}}_{\text{IV}}^a$ under the null hypothesis. In other words, $\hat{\boldsymbol{\beta}}_{\text{IV}}^b - \hat{\boldsymbol{\beta}}_{\text{IV}}^a$ should be a positive semidefinite matrix.

Under the alternative hypothesis, the estimator $\hat{\boldsymbol{\beta}}_{\text{IV}}^b$ still yields a consistent estimator of $\boldsymbol{\beta}$, while $\hat{\boldsymbol{\beta}}_{\text{IV}}^a$ yields an inconsistent estimator.

Informally, we expect the two estimators to produce similar estimates of the parameter vector $\boldsymbol{\beta}$ if the null hypothesis is valid and both estimators are consistent, and to produce different estimates of $\boldsymbol{\beta}$ if the null hypothesis is rejected and only $\hat{\boldsymbol{\beta}}_{\text{IV}}^b$ is consistent. One version of the Hausman which can be constructed is

$$H = (\hat{\boldsymbol{\beta}}_{\text{IV}}^b - \hat{\boldsymbol{\beta}}_{\text{IV}}^a)^\top \left[\widehat{\text{AVar}}\left(\hat{\boldsymbol{\beta}}_{\text{IV}}^b\right) - \widehat{\text{AVar}}\left(\hat{\boldsymbol{\beta}}_{\text{IV}}^a\right)\right]^{-1} (\hat{\boldsymbol{\beta}}_{\text{IV}}^b - \hat{\boldsymbol{\beta}}_{\text{IV}}^a) \overset{a}{\sim} \chi^2(2). \tag{17}$$

Alternatively, we could construct a Hausman test based on the corresponding estimates of individual scalar parameters. For example

$$\begin{aligned} H_1 &= (\hat{\beta}_1^b - \hat{\beta}_1^a)\left(\hat{v}_{11}^b - \hat{v}_{11}^a\right)^{-1}(\hat{\beta}_1^b - \hat{\beta}_1^a) \\ &= \frac{(\hat{\beta}_1^b - \hat{\beta}_1^a)^2}{\hat{v}_{11}^b - \hat{v}_{11}^a} \overset{a}{\sim} \chi^2(1), \end{aligned} \tag{18}$$

under $H_0 : \mathbb{E}[\mathbf{X}_1^\top \mathbf{u}] = 0$. This is is also equivalent to

$$\sqrt{H_1} = \frac{\hat{\beta}_1^b - \hat{\beta}_1^a}{\sqrt{\hat{v}_{11}^b - \hat{v}_{11}^a}} \overset{a}{\sim} N(0,1), \tag{19}$$

under the null hypothesis, where $\hat{v}_{11}^i$ is the first element on the main diagonal of $\hat{\text{Var}}\left[\hat{\boldsymbol{\beta}}_{\text{IV}}^i\right]$, for $i = (a, b)$. The scalar version of the Hausman test would be useful in a setting where $\hat{\text{Var}}\left[\hat{\boldsymbol{\beta}}_{\text{IV}}^b\right] - \hat{\text{Var}}\left[\hat{\boldsymbol{\beta}}_{\text{IV}}^a\right]$ is singular, and may be more powerful than H in a setting where neglecting correlation between x_{1i} and u_i mostly affects the 2SLS estimates of the scalar parameter β_1.

1.3

For an application where the sample size n is large, a researcher reports OLS and 2SLS results as follows:

Table 1: **OLS and 2SLS Results**

OLS estimates	$y_i = \underset{(4)}{14}x_{1i} + \underset{(3)}{9}x_{2i}$
2SLS estimates (a)	$y_i = \underset{(6)}{16}x_{1i} - \underset{(5)}{12}x_{2i}$
2SLS estimates (b)	$y_i = \underset{(10)}{20}x_{1i} - \underset{(7)}{14}x_{2i}$

Asymptotic standard errors are reported in brackets below the estimated coefficients, 2SLS estimates (a) are computed treating only x_{2i} as correlated with u_i, while 2SLS estimates (b) are computed treating both x_{1i} and x_{2i} as correlated with u_i. For 2SLS estimates (a), the Sargan test of the validity of the over-identifying restriction is reported to be 0.6503 (the 95th percentile of the $\chi^2(1)$ distribution is 3.841).

1.3.1

What do these results suggest about the correlation between x_{1i} and u_i, and the correlation between x_{2i} and u_i?

So here we have a situation where OLS is consistent if and only BOTH $\mathbb{E}[\mathbf{X}_1^\top \mathbf{u}] = \mathbf{0}$ and $\mathbb{E}[\mathbf{X}_2^\top \mathbf{u}] = \mathbf{0}$, 2SLS (a) is consistent if and only if $\mathbb{E}[\mathbf{X}_1^\top \mathbf{u}] = \mathbf{0}$ but we allow for $\mathbb{E}[\mathbf{X}_2^\top \mathbf{u}] \neq \mathbf{0}$, and 2SLS (b) is consistent if $\mathbb{E}[\mathbf{X}_1^\top \mathbf{u}] \neq \mathbf{0}$ or $\mathbb{E}[\mathbf{X}_2^\top \mathbf{u}] \neq \mathbf{0}$, or both $\mathbb{E}[\mathbf{X}_1^\top \mathbf{u}] \neq \mathbf{0}$ and $\mathbb{E}[\mathbf{X}_2^\top \mathbf{u}] \neq \mathbf{0}$.

Recall the formula for our confidence interval (say, a 95% confidence interval):

$$\Pr\left(\hat{\beta}_i - t_{0.025}SE_i \leq \beta_i \leq \hat{\beta}_i - t_{0.975}SE_i\right) = 0.95,$$

and we can clearly say that the confidence interval for the OLS estimate of β_2 is approximately (3,15), which our 2SLS estimates are clearly outside of. Conversely, the 95% confidence intervals for 2SLS (a) is approximately (-22,-2) and for 2SLS (b) it is (-28,0). So there does seem to be some overlap between the 2SLS regressions.

We could run a scalar Hausman test based on 19 for the coefficient of $\mathbf{X}_2$ between the OLS estimate and 2SLS (a) estimate, where the null hypothesis is that $\mathbb{E}[\mathbf{X}_2^\top \mathbf{u}] = \mathbf{0}$ and the alternative that $\mathbb{E}[\mathbf{X}_2^\top \mathbf{u}] \neq \mathbf{0}$. The test statistic is thus

$$\sqrt{H_1} = \frac{-12 - 9}{\sqrt{5^2 - 3^2}} = \frac{-21}{\sqrt{16}} = -5.25 < -1.96,$$

and so we would reject the null hypothesis as our test statistic is well in the rejection region, suggesting that $\mathbf{X}_2$ is in fact correlated with $\mathbf{u}$.

Next, for $\mathbf{X}_1$, we test the null hypothesis that $\mathbb{E}[\mathbf{X}_1^\top \mathbf{u}] = \mathbf{0}$ against the alternative that $\mathbb{E}[\mathbf{X}_1^\top \mathbf{u}] \neq \mathbf{0}$ using a Hausman test, comparing the 2SLS (a) estimates and the 2SLS (b) estimates as described in 1.2.3. Doing this for β_1 yields the following test statistic

$$\sqrt{H_1} = \frac{20 - 16}{\sqrt{10^2 - 6^2}} = \frac{4}{\sqrt{64}} = 0.5 < 1.96,$$

which does not reject the null hypothesis at any reasonable significance level, suggesting that $\mathbf{X}_1$ is uncorrelated with $\mathbf{u}$.

The Sargan test for 2SLS (a) tests the following:

$$H_0 : \mathbb{E}[\mathbf{Z}^\top \mathbf{u}] = \mathbf{0},$$
$$H_1 : \mathbb{E}[\mathbf{Z}^\top \mathbf{u}] \neq \mathbf{0},$$

where $\mathbf{Z} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{Z}_1 & \mathbf{Z}_2 \end{bmatrix}$ hence it is a test of the effectiveness the overspecified model. The reported value of 0.6503 is well below the 95th percentile of the $\chi^2(1)$ which is approximately 3.841, hence we cannot reject the null hypothesis. This is consistent with out finding about the absence of any correlation between $\mathbf{X}_1$ and $\mathbf{u}$.

1.3.2

What conclusion, if any, would you draw about the sign β_2?

β_2 is most likely negative, as only the OLS estimate suggests a positive value, and we just showed the OLS estimator to be inconsistent as it fails to account for the correlation between $\mathbf{X}_2$ and $\mathbf{u}$. Furthermore, a t-test to see if $\beta_2 = 0$ using the 2SLS estimates would strong reject a null suggesting that $\beta_2 = 0$.

1.3.3

What additional information could be reported here, that would increase your confidence in your answers to parts 1.3.1 and 1.3.2?

- Results from the first stage linear projection for the two 2SLS estimators to assess validity of the identification conditions.
- Strong indications that the instruments are informative (conditional on $\mathbf{X}_1$) for $\mathbf{X}_2$ would increase confidence in conclusions drawn on the basis of the 2SLS (a) estimates.
- Strong indications that at least one instrument is informative for both $\mathbf{X}_1$ and $\mathbf{X}_2$, and that the predicted values $\hat{\mathbf{X}}_1$ and $\hat{\mathbf{X}}_2$ are not highly collinear would increase confidence in conclusions drawn on the basis of the 2SLS (b) estimates.
- Formal rank tests could also be presented.
- For the 2SLS (a) specification, with one endogenous variable, this simplifies to an F-test for the exclusion of the (outside) instruments from the one (non-trivial) first stage linear projection. Given that we have IID data and conditional homoskedasticity, the value of this F-test statistic could be compared with critical values tabulated by Stock and Yogo (2005) to assess the strength of these instruments.

2 Identification

The 5 random variables $(y_i, x_{1i}, x_{2i}, z_{1i}, z_{2i})$ are generated by the processes

$$\begin{aligned}
y_i &= \beta_1 x_{1i} + \beta_2 x_{2i} + u_i, \\
x_{1i} &= \gamma_{11} z_{1i} + \gamma_{12} z_{2i} + \theta_1 u_i + v_{1i}, \\
x_{2i} &= \delta x_{1i} + \gamma_{21} z_{1i} + \gamma_{22} z_{2i} + \theta_2 u_i + v_{2i}, \\
z_{1i} &= \phi_1 u_i + w_{1i}, \\
z_{2i} &= \phi_2 u_i + w_{2i},
\end{aligned}$$

where the random vector $(u_i, v_{1i}, v_{2i}, w_{1i}, w_{2i})^\top$ is IID with mean zero and diagonal variance-covariance matrix.

2.1

First, suppose it is known that $\beta_2 = \gamma_{12} = 0$. Data is observed on (y_i, x_{1i}, z_{1i}) only. The DGP in this case simplifies down to

$$\begin{aligned}
y_i &= \beta_1 x_{1i} + u_i, \\
x_{1i} &= \gamma_{11} z_{1i} + \theta_1 u_i + v_{1i}, \\
z_{1i} &= \phi_1 u_i + w_{1i},
\end{aligned}$$

2.1.1

Evaluate $\mathbb{E}[z_{1i} u_i]$, $\mathbb{E}[x_{1i} u_i]$, *and* $\mathbb{E}[x_{1i} z_i]$.

The DGP in this case simplifies down to

$$\begin{aligned}
y_i &= \beta_1 x_{1i} + u_i, \\
x_{1i} &= \gamma_{11} z_{1i} + \theta_1 u_i + v_{1i}, \\
z_{1i} &= \phi_1 u_i + w_{1i},
\end{aligned}$$

where, after direct substitution, we have

$$\begin{aligned}
\mathbb{E}[z_{1i} u_i] &= \mathrm{Cov}[z_{1i}, u_i] = \mathbb{E}[\phi_1 u_i^2 + w_{1i} u_i] = \phi_1 \mathbb{E}[u_i^2] = \phi_1 \mathrm{Var}[u_i], \\
\mathbb{E}[x_{1i} u_i] &= \mathrm{Cov}[x_{1i}, u_i] = (\gamma_{11} \phi_1 + \theta_1) \mathrm{Var}[u_i], \\
\mathbb{E}[x_{1i} z_i] &= \mathrm{Cov}[x_{1i}, z_i] = (\gamma_{11} \phi_1^2 + \theta_1 \phi_1) \mathrm{Var}[u_i] + \gamma_{11} \mathrm{Var}[w_{1i}].
\end{aligned}$$

Note, the covariance between the error terms is 0 since the random vector $[u_i, v_{1i}, v_{2i}, w_{1i}, w_{2i}]^\top$ has a diagonal variance-covariance matrix.

2.1.2

Consider the linear model $y_i = \alpha x_{1i} + \epsilon_i$. *State restrictions on the parameters of the process generating* (x_{1i}, z_{1i}) *under which the parameter* β_1 *in the DGP process could be* *estimated consistently by i) the OLS estimate of α in this linear model, and; ii) the 2SLS estimator of α in this linear model, using z_{1i}, as an instrumental variable.*

First note that the model can be obtained directly from the DGP by setting $\alpha = \beta_1$ and $\epsilon_i = u_i$. For consistent OLS estimation of α we require the orthogonality condition $\mathbb{E}[x_{1i}u_i] = 0$. Based on our answer previously, we thus require

$$\begin{aligned}\mathbb{E}[x_{1i}u_i] &= (\gamma_{11}\phi_1 + \theta_1)\text{Var}[u_i] = 0\\ &\implies \gamma_{11}\phi_1 + \theta_1 = 0.\end{aligned}$$

For the 2SLS/IV case, we simply require $\mathbb{E}[z_{1i}u_i] = 0$ (validity requirement), which implies that we need $\phi_1 = 0$, furthermore we need $\gamma_{11} \neq 0$ (informative requirement).

2.2

Now suppose it is known that $\gamma_{12} = \gamma_{22} = 0$. Data is observed on $(y_i, x_{1i}, x_{2i}, z_{1i})$ only. The DGP simplifies down to

$$\begin{aligned}y_i &= \beta_1 x_{1i} + \beta_2 x_{2i} + u_i,\\ x_{1i} &= \gamma_{11} z_{1i} + \theta_1 u_i + v_{1i},\\ x_{2i} &= \delta x_{1i} + \gamma_{21} z_{1i} + \theta_2 u_i + v_{2i},\\ z_{1i} &= \phi_1 u_i + w_{1i}.\end{aligned}$$

2.2.1

Evaluate $\mathbb{E}[z_{1i}u_i]$, $\mathbb{E}[x_{1i}u_i]$, and $\mathbb{E}[x_{2i}u_i]$.

After direct substitution we have

$$\begin{aligned}\mathbb{E}[z_{1i}u_i] &= \text{Cov}[z_{1i}, u_i] = \phi_1 \text{Var}[u_i],\\ \mathbb{E}[x_{1i}u_i] &= \text{Cov}[x_{1i}, u_i] = (\gamma_{11}\phi_1 + \theta_1)\text{Var}[u_i],\\ \mathbb{E}[x_{2i}u_i] &= \text{Cov}[x_{2i}, u_i] = (\delta(\gamma_{11}\phi_1 + \theta_1) + \gamma_{21}\phi_1 + \theta_2)\text{Var}[u_i].\end{aligned}$$

2.2.2

Suppose that $\phi_1 = \theta_1 = 0$ and that $\theta_2 \neq 0$. Explain what restriction on the parameters of the DGP is then required for z_{1i} to be an informative instrument for x_{2i}, such that the parameters (β_1, β_2) can be estimated consistently using 2SLS from the linear model $y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$.

The assumption that $\phi_1 = \theta_1 = 0$ ensures validity (it also means that x_1 can be used as an instrument). In order for z_{1i} to be informative on x_{2i} we require that $\gamma_{21} \neq 0$. Consider the first stage linear projection:

$$x_{2i} = \pi_x x_{1i} + \pi_z z_{1i} + e_i,$$

and in the DGP we have

$$x_{2i} = \delta x_{1i} + \gamma_{21} z_{1i} + \theta_2 u_i + v_{2i},$$

where we showed that x_1 and z_1 are valid instruments in estimating x_2. Given the assumptions in this question we also have $z_{1i} = w_{1i}$, and so

$$x_{1i} = \gamma_{11} z_{1i} + v_{1i} = \gamma_{11} w_{1i} + v_{1i}.$$

So substituting our values into the first stage linear progression gives

$$x_{2i} = \underbrace{\pi_x}_{\delta} \underbrace{x_{1i}}_{\gamma_{11} w_{1i} + v_{1i}} + \underbrace{\pi_z}_{\gamma_{21}} \underbrace{z_{1i}}_{w_{1i}} + \underbrace{e_i}_{\theta_2 u_i + v_{2i}},$$

justifying our requirement of $\gamma_{21} \neq 0$ for informativeness.

2.3

Now suppose it is known that $\delta = 0$. Data is observed on $(y_i, x_{1i}, x_{2i}, z_{1i}, z_{2i})$.

2.3.1

Evaluate $\mathbb{E}[z_{1i} u_i]$, $\mathbb{E}[z_{2i} u_i]$, $\mathbb{E}[x_{1i} u_i]$, *and* $\mathbb{E}[x_{2i} u_i]$.

Our DGP is now

$$\begin{aligned}
y_i &= \beta_1 x_{1i} + \beta_2 x_{2i} + u_i, \\
x_{1i} &= \gamma_{11} z_{1i} + \gamma_{12} z_{2i} + \theta_1 u_i + v_{1i}, \\
x_{2i} &= \gamma_{21} z_{1i} + \gamma_{22} z_{2i} + \theta_2 u_i + v_{2i}, \\
z_{1i} &= \phi_1 u_i + w_{1i}, \\
z_{2i} &= \phi_2 u_i + w_{2i},
\end{aligned}$$

and so after direct substitution, we have

$$\begin{aligned}
\mathbb{E}[z_{1i} u_i] &= \text{Cov}[z_{1i}, u_i] = \phi_1 \text{Var}[u_i], \\
\mathbb{E}[z_{2i} u_i] &= \text{Cov}[z_{2i}, u_i] = \phi_2 \text{Var}[u_i], \\
\mathbb{E}[x_{1i} u_i] &= \text{Cov}[x_{1i}, u_i] = (\gamma_{11} \phi_1 + \gamma_{12} \phi_2 + \theta_1) \text{Var}[u_i], \\
\mathbb{E}[x_{2i} u_i] &= \text{Cov}[x_{2i}, u_i] = (\gamma_{21} \phi_1 + \gamma_{22} \phi_2 + \theta_2) \text{Var}[u_i].
\end{aligned}$$

2.3.2

Suppose that $\phi_1 = \phi_2 = 0$ *and that* $\theta_1 \neq 0$ *and* $\theta_2 \neq 0$. *Explain whether or not the parameter vector* $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$ *is identified in the cases with*:

1. $\gamma_{21} = \gamma_{22} = 0$
2. $\gamma_{12} = \gamma_{22} = 0$
3. $\gamma_{11} = 3\gamma_{21}$ and $\gamma_{12} = 3\gamma_{22}$

Under these assumptions our expectation terms (validity conditions) are

$$\begin{aligned}
\mathbb{E}[z_{1i}u_i] &= 0, \\
\mathbb{E}[z_{2i}u_i] &= 0, \\
\mathbb{E}[x_{1i}u_i] &= \theta_1 \text{Var}[u_i], \\
\mathbb{E}[x_{2i}u_i] &= \theta_2 \text{Var}[u_i],
\end{aligned}$$

thus, only z_1 and z_2 are valid instruments in the estimation of x_1 and x_2. Consider the first stage linear projection

$$\begin{aligned}
\mathbf{X}^\top &= \mathbf{Z}^\top \mathbf{\Pi} + \mathbf{E}^\top, \\
\Leftrightarrow \begin{bmatrix} x_{1i} & x_{2i} \end{bmatrix} &= \begin{bmatrix} z_{1i} & z_{2i} \end{bmatrix} \begin{bmatrix} \pi_{x_1z_1} & \pi_{x_1z_2} \\ \pi_{x_2z_1} & \pi_{x_2z_2} \end{bmatrix} + \begin{bmatrix} e_{1i} & e_{2i} \end{bmatrix},
\end{aligned}$$

$$\begin{aligned}
\implies x_{1i} &= \pi_{x_1z_1} z_{1i} + \pi_{x_1z_2} z_{2i} + e_{1i}, \\
x_{2i} &= \pi_{x_2z_1} z_{1i} + \pi_{x_2z_2} z_{2i} + e_{2i}.
\end{aligned}$$

For case 1), comparing the first stage linear projection to our DGP, we have

$$\begin{aligned}
y_i &= \beta_1 x_{1i} + \beta_2 x_{2i} + u_i, \\
x_{1i} &= \gamma_{11} z_{1i} + \gamma_{12} z_{2i} + \theta_1 u_i + v_{1i}, \\
x_{2i} &= \delta x_{1i} + \theta_2 u_i + v_{2i}, \\
z_{1i} &= w_{1i}, \\
z_{2i} &= w_{2i},
\end{aligned}$$

thus because we have no valid instruments for x_2 in our DGP, $\boldsymbol{\beta}$ will be unidentified.

For case 2), we have the following DGP

$$\begin{aligned}
y_i &= \beta_1 x_{1i} + \beta_2 x_{2i} + u_i, \\
x_{1i} &= \gamma_{11} z_{1i} + \theta_1 u_i + v_{1i}, \\
x_{2i} &= \delta x_{1i} + \gamma_{21} z_{1i} + \theta_2 u_i + v_{2i}, \\
z_{1i} &= w_{1i}, \\
z_{2i} &= w_{2i},
\end{aligned}$$

where we only have one informative instrument, z_1, for two endogenous variables. Therefore $\boldsymbol{\beta}$ will be unidentified.

Finally, for case 3) we have

$$\begin{aligned}
y_i &= \beta_1 x_{1i} + \beta_2 x_{2i} + u_i, \\
x_{1i} &= 3\gamma_{21} z_{1i} + 3\gamma_{22} z_{2i} + \theta_1 u_i + v_{1i}, \\
x_{2i} &= \delta x_{1i} + \gamma_{21} z_{1i} + \gamma_{22} z_{2i} + \theta_2 u_i + v_{2i}, \\
z_{1i} &= w_{1i}, \\
z_{2i} &= w_{2i},
\end{aligned}$$

which then implies our system for the first stage linear projection is akin to

$$\begin{bmatrix} x_{1i} & x_{2i} \end{bmatrix} = \begin{bmatrix} z_{1i} & z_{2i} \end{bmatrix} \begin{bmatrix} 3\pi_{x_2 z_1} & 3\pi_{x_2 z_2} \\ \pi_{x_2 z_1} & \pi_{x_2 z_2} \end{bmatrix} + \begin{bmatrix} e_{1i} & e_{2i} \end{bmatrix},$$

where we clearly have a violation of our rank condition. In other words, the instruments are linearly dependent, and so in the limit $\hat{x}_1$ and $\hat{x}_2$ are perfectly correlated with each other, thus $\boldsymbol{\beta}$ is not identified.

2.3.3

Again supposing that $\phi_1 = \phi_2 = 0$ and that $\theta_1 \neq 0$ and $\theta_2 \neq 0$, state a necessary and sufficient condition for the parameter vector $\boldsymbol{\beta} = (\beta_1, \beta_2)^\top$ to be identified, and express this in terms of the parameters of the DGP.

As previously stated in part 1.2.2, necessary and sufficient conditions for $\boldsymbol{\beta}$ to be identified are the order condition and the rank condition. The order condition requires that our model not be underspecified, so we require at least as many instruments as we have regressors. Here, we have z_{1i} and z_{2i} as our instruments, and our regressors are x_{1i} and x_{2i} (so, $l = k$, and our model is just specified). Secondly, we would require the rank of our matrix of instruments to be of full rank so that $\mathbf{S}_{\mathbf{Z}^\top \mathbf{X}}$ and $\mathbf{S}_{\mathbf{Z}^\top \mathbf{Z}}$ exists and is nonsingular. Thus, we would require

$$\operatorname{rank} \begin{bmatrix} \gamma_{11} & \gamma_{21} \\ \gamma_{12} & \gamma_{22} \end{bmatrix} = 2.$$

This rank condition was not met in cases 1), 2), or 3) in the previous question.