

Introduction to Econometrics II: Recitation 3*

Seung-hun Lee[†]

February 5th, 2020

1 Instrumental Variable

1.1 Limited Information Maximum Likelihood Approach

1.1.1 Setup and Motivation

Assume a data generating process

$$y_i = \beta'_1 x_{1i} + \beta'_2 x_{2i} + e_i, \quad (x_{1i} \in \mathbb{R}^{k_1}, x_{2i} \in \mathbb{R}^{k_2})$$

where $E(x_{1i}e_i) = 0$, but $E(x_{2i}e_i) \neq 0$. A **limited information maximum likelihood (LIML) estimator** derives the maximum likelihood estimator for the joint distribution of (y_i, x_{2i}) using structural equation of y_i and the reduced form equation for x_{2i} . This differs from the full information maximum likelihood (FIML) in the sense that the FIML requires structural equation for x_{2i} as well.

Briefly speaking, the reason we prefer LIML is as follows. When the number of the instruments are fixed, 2SLS and LIML have the same asymptotic distribution, with the former not requiring the assumption of normality (Greene, 2012). However, when there is a problem of weak instrument variable or too many instrumental variables, it can be shown that 2SLS becomes biased towards OLS. Others have shown that LIML estimators perform better in the presence of weak IVs and/or too many IVs.¹

*This is based on the lecture notes of Professors Jushan Bai and Bernard Salanie. I was also greatly helped by previous recitation notes from Paul Sungwook Koh and Dong Woo Hahm. The remaining errors are mine.

[†]Contact me at sl4436@columbia.edu if you spot any errors or have suggestions on improving this note.

¹I refer further explanation to Jorn-Steffen Pischke's notes, found at <http://econ.lse.ac.uk/staff/spischke/ec533/Weak%20IV.pdf>

1.1.2 Derivation

The structural equation for y_i is introduced in the previous page. Assume that there is a $z_i = \begin{pmatrix} x_{1i} \\ z_{2i} \end{pmatrix} \in \mathbb{R}^l$ where $E(z_i e_i) = 0$. I now define a reduced form equation for x_{2i} , written as

$$x_{2i} = \Gamma'_{12} x_{1i} + \Gamma'_{22} z_{2i} + u_{2i}, \quad (\Gamma'_{12} \in \mathbb{R}^{k_2 \times k_1}, \Gamma'_{22} \in \mathbb{R}^{k_2 \times (l-k_1)}, z_{2i} \in \mathbb{R}^{l-k_1})$$

By putting the structural and reduced form equation together in a matrix form, we get

$$\underbrace{\begin{pmatrix} 1 & -\beta'_2 \\ 0 & I_{k_2} \end{pmatrix}}_{=A} \underbrace{\begin{pmatrix} y_i \\ x_{2i} \end{pmatrix}}_{=w_i} = \underbrace{\begin{pmatrix} \beta'_1 & 0 \\ \Gamma'_{12} & \Gamma'_{22} \end{pmatrix}}_{=B} \underbrace{\begin{pmatrix} x_{1i} \\ z_{2i} \end{pmatrix}}_{=z_i} + \underbrace{\begin{pmatrix} e_i \\ u_{2i} \end{pmatrix}}_{=\eta_i}$$

This now puts all endogenous variables on the left hand side and the exogenous variables on the right. To use the maximum likelihood approach, we need some assumptions on η_i . Specifically, we assume that η_i is conditionally IID normal and

$$\eta_i | z_i \sim N(0, \Sigma_\eta)$$

If we apply these facts, then we can derive the log-likelihood function for η_i

Lemma 1.1 (Multivariate Normal Density). *Let $X \in \mathbb{R}^k$ be a multivariate normal distribution s.t. $X \sim N(\mu, \Sigma)$. Then the density function can be written as*

$$f_X(x) = (2\pi)^{-\frac{k}{2}} \det(\Sigma)^{-\frac{1}{2}} \times e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}$$

Lemma 1.2 (Jacobian Transformation). *If $X \in \mathbb{R}^k$ is a continuous random vector and $g : \mathbb{R}^k \rightarrow \mathbb{R}^k$ is one-to-one and onto function, the density of $Y = g(X)$ is given by*

$$f_Y(y) = f_X(g^{-1}(y)) |J|$$

where $J = \det \begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \dots & \frac{\partial x_1}{\partial y_k} \\ \dots & \dots & \dots \\ \frac{\partial x_k}{\partial y_1} & \dots & \frac{\partial x_k}{\partial y_k} \end{pmatrix}$. Alternatively, we can write

$$f_X(x) = f_Y(g(x))|H|, H = \det \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \dots & \frac{\partial y_1}{\partial x_k} \\ \dots & \dots & \dots \\ \frac{\partial y_k}{\partial x_1} & \dots & \frac{\partial y_k}{\partial x_k} \end{pmatrix}$$

So to work through the likelihood function, we take these steps

1. **Re-express the PDF:** The PDF pf η_i is

$$(2\pi)^{-\frac{k_2+1}{2}} \det(\Sigma)^{-\frac{1}{2}} \times e^{-\frac{1}{2}(\eta_i)' \Sigma_\eta^{-1}(\eta_i)}$$

Note that $\eta_i = Aw_i - Bz_i$, Then we can write the pdf of w_i in terms of A and B as ²

$$(2\pi)^{-\frac{k_2+1}{2}} \det(\Sigma)^{-\frac{1}{2}} \times e^{-\frac{1}{2}(Aw_i - Bz_i)' \Sigma_\eta^{-1}(Aw_i - Bz_i)} |A|$$

by Applying a Jacobian Transformation. Note that the determinant of the A matrix is 1. So this term would be dropped. By IID assumption, the log likelihood for $i = 1, \dots, n$ becomes

$$C - \frac{n}{2} \log(\det(\Sigma_\eta)) - \frac{1}{2} \sum_{i=1}^n (Aw_i - Bz_i)' \Sigma_\eta^{-1} (Aw_i - Bz_i)$$

2. **Concentrating out:** To find A, B that maximizes the log - likelihood, we concentrate out the Σ_η term by using its estimator,

$$\hat{\Sigma}_\eta = \frac{1}{n} \sum_{i=1}^n (Aw_i - Bz_i)(Aw_i - Bz_i)'$$

Then plug this estimator into the log-likelihood function to get

$$C - \frac{n}{2} \log(\det(\hat{\Sigma}_\eta)) - \frac{1}{2} \sum_{i=1}^n (Aw_i - Bz_i)' \hat{\Sigma}_\eta^{-1} (Aw_i - Bz_i)$$

²Note that $\chi_i = w_i - A^{-1}Bz_i$ is also another normal distribution. You can notice that $\eta_j = A\chi_i$. Try matching this with the second version of the Jacobian transformation.

Since the last term is a scalar, we can take

$$(Aw_i - Bz_i)' \hat{\Sigma}_\eta^{-1} (Aw_i - Bz_i) = \text{tr} \left((Aw_i - Bz_i)' \hat{\Sigma}_\eta^{-1} (Aw_i - Bz_i) \right)$$

Also, using the fact that $\text{tr}(AB) = \text{tr}(BA)$,

$$\text{tr} \left((Aw_i - Bz_i)' \hat{\Sigma}_\eta^{-1} (Aw_i - Bz_i) \right) = \text{tr} \left(\hat{\Sigma}_\eta^{-1} (Aw_i - Bz_i) (Aw_i - Bz_i)' \right)$$

This reduced the likelihood function to

$$C - \frac{n}{2} \log(\det(\hat{\Sigma}_\eta)) - \frac{1}{2} \text{tr} \left(\hat{\Sigma}_\eta^{-1} \sum_{i=1}^n (Aw_i - Bz_i) (Aw_i - Bz_i)' \right) = C - \frac{n}{2} \log(\det(\hat{\Sigma}_\eta)) - \frac{n(k_2 + 1)}{2}$$

3. **Maximize:** The only thing left to maximize over now is $-\frac{n}{2} \log(\det(\hat{\Sigma}_\eta))$. This is equivalent to maximizing

$$-\det \left(\frac{1}{n} \sum_{i=1}^n (Aw_i - Bz_i)' (Aw_i - Bz_i) \right) = -\det \left(\frac{1}{n} \sum_{i=1}^n \begin{pmatrix} y_i - \beta_1' x_{1i} - \beta_2' x_{2i} \\ x_{2i} - \Gamma_{12}' x_{1i} - \Gamma_{22}' z_{2i} \end{pmatrix} \begin{pmatrix} y_i - \beta_1' x_{1i} - \beta_2' x_{2i} \\ x_{2i} - \Gamma_{12}' x_{1i} - \Gamma_{22}' z_{2i} \end{pmatrix}' \right)$$

The combination of β and Γ that maximizes this is the LIML estimator.³

1.1.3 k -class Estimators

Another way to compute the LIML estimator is to use a k -class estimator with a particular choice for k . In general, k -class estimator is defined as

$$\hat{\beta}_k = \arg \min_{\beta} (y - X\beta)' (I_n - kM_Z) (y - X\beta)$$

where $M_Z = I - Z(Z'Z)^{-1}Z' = I - P_Z$. The other way to express this, after some matrix differentiation, is

$$\begin{aligned} -2X'y + 2kX'M_Z y + 2(X'X)\beta - 2k(X'M_Z X)\beta &= 0 \\ \iff (X'(I_n - kM_Z)X)\beta &= X'(I_n - kM_Z)y \\ \implies \hat{\beta}_k &= (X'(I_n - kM_Z)X)^{-1} X'(I_n - kM_Z)y \end{aligned}$$

³Further steps involve heavy amounts of algebra involving multi-dimensional matrices. For details, I will have you refer to the recitation notes from Dong Woo last year, which I will upload.

In fact, we can show that OLS ($k = 0$) and 2SLS ($k = 1$) are k -class estimators. LIML is a k -class estimator with a parameter choice of some $k > 1$. What we need to do find the associated value of k . To do so, define

$$W = (y \ X_2) \in \mathbb{R}^{n \times (k_2+1)}, \ M_1 = I - X_1(X_1'X_1)^{-1}X_1'$$

Then we compute the minimum eigenvalue of ⁴

$$(W'M_1W)(W'M_ZW)^{-1}$$

This minimum eigenvalue will be our choice of k , which will be denoted as \hat{k} , and the LIML estimator would be

$$\hat{\beta}_{LIML} = (X'(I_n - \hat{k}M_Z)X)^{-1}(X'(I_n - \hat{k}M_Z)y)$$

From here, we can also find that LIML is also a type of an IV estimator. We can see that by rewriting above equation as

$$\hat{\beta}_{LIML} = (\tilde{X}'X)^{-1}(\tilde{X}'y)$$

where $\tilde{X} = (I_n - \hat{k}M_Z)X$. This also hints that the asymptotic properties of LIML and some other types of IV estimators are similar.

1.1.4 Asymptotics of LIML

Given the above definition of $\hat{\beta}_{LIML} = (X'(I_n - \hat{k}M_Z)X)^{-1}(X'(I_n - \hat{k}M_Z)y)$, we can use this to study the asymptotic properties of LIML estimators. Note that

$$\begin{aligned} \hat{\beta}_{LIML} - \beta &= (X'(I_n - \hat{k}M_Z)X)^{-1}(X'(I_n - \hat{k}M_Z)e) \\ &= (X'(P_Z - (\hat{k} - 1)M_Z)X)^{-1}(X'(P_Z - (\hat{k} - 1)M_Z)e) \\ &(\because I - M_Z = P_Z, \implies I - \hat{k}M_Z = P_Z - (\hat{k} - 1)M_Z) \\ \implies \sqrt{n}(\hat{\beta}_{LIML} - \beta) &= \left(\frac{X'P_ZX}{n} - (\hat{k} - 1)\frac{X'M_ZX}{n} \right)^{-1} \left(\frac{X'P_ZX}{\sqrt{n}} - (\hat{k} - 1)\frac{X'M_Ze}{\sqrt{n}} \right) \end{aligned}$$

Note that

- Anderson and Rubin (1949) shows that $\hat{k} - 1 \xrightarrow{p} 0$, which we accept as given.

⁴This is shown in Anderson and Rubin (1949) and Amemiya (1985)

- $\frac{X'M_Z X}{n} = \frac{X'X}{n} - \frac{X'P_Z X}{n} = \frac{X'X}{n} - \frac{X'P_Z' P_Z X}{n} = \frac{X'X}{n} - \frac{(P_Z X)'(P_Z X)}{n} \leq \frac{X'X}{n} \xrightarrow{p} E(x_i x_i')$. Therefore, $\frac{X'M_Z X}{n}$ is bounded (and thus $O_p(1)$)
- $\frac{X'M_Z e}{\sqrt{n}}$ converges in distribution to a normal distribution (CLT), so it is $O_p(1)$
- From the third and first points, we can infer that $(\hat{k} - 1) \frac{X'M_Z e}{\sqrt{n}} = o_p(1)$

Combining these leads to the result that

$$\sqrt{n}(\hat{\beta}_{LIML} - \beta) = \left(\frac{X'P_Z X}{n} \right)^{-1} \frac{X'P_Z X}{\sqrt{n}} + o_p(1)$$

which is equivalent to $\sqrt{n}(\hat{\beta}_{2SLS} - \beta) + o_p(1)$. So for a fixed n , the 2SLS and LIML converge to a similar limiting distribution. However, when either instrument is weakly related to X or large in numbers, 2SLS bias becomes large, giving advantage to the LIML⁵.

1.2 Control Function Approach

This is another way to derive a 2SLS estimator. As before, I will assume $E(x_{1i}e_i) = 0, E(x_{2i}e_i) \neq 0$ and write the structural and reduced form regression as

$$y_i = x'_{1i}\beta_1 + x'_{2i}\beta_2 + e_i \quad (\text{Structural})$$

$$x_{2i} = \Gamma'_{12}x_{1i} + \Gamma'_{22}z_{2i} + u_{2i} \quad (\text{Reduced Form})$$

Also, we have a $z_i \in (x_{1i} \ z_{2i})' \in \mathbb{R}^l$ that satisfies $E(z_i e_i) = 0$. The key driving idea for the control function method is that we can write $E(x_{2i}e_i)$ differently. Specifically,

$$\begin{aligned} E[x_{2i}e_i] &= E[(\Gamma'_{12}x_{1i} + \Gamma'_{22}z_{2i} + u_{2i})e_i] \\ &= \Gamma'_{12}E(x_{1i}e_i) + \Gamma'_{22}E(z_{2i}e_i) + E(u_{2i}e_i) \\ &= E(u_{2i}e_i) \quad (\because \text{The first by exogeneity, the second by IV conditions}) \\ \therefore E[x_{2i}e_i] \neq 0 &\iff E[u_{2i}e_i] \neq 0 \end{aligned}$$

From here, we consider a linear projection of e_i onto u_{2i} , which we write as

$$e_i = u'_{2i}\alpha + \epsilon_i \quad (\text{LP})$$

⁵This is the key takeaway from Dr. Pischke's slides.

where $E(u_{2i}\epsilon_i) = 0$ and the population analogue of $\alpha = E(u_{2i}u'_{2i})^{-1}E(u_{2i}e_i)$. What we now do is to substitute the e_i term in the (Structural) equation with the (LP) equation to get

$$y_i = x'_{1i}\beta_1 + x'_{2i}\beta_2 + u'_{2i}\alpha + \epsilon_i \quad (\text{CFA})$$

where the following are satisfied

$$\begin{aligned} E[x_{1i}\epsilon_i] &= E[x_{1i}(e_i - u'_{2i}\alpha)] = 0 - 0 = 0 \\ E[u_{2i}\epsilon_i] &= E[u_{2i}(e_i - u'_{2i}\alpha)] \\ &= E[u_{2i}e_i] - E[u_{2i}u'_{2i}]E[u_{2i}u'_{2i}]^{-1}E[u_{2i}e_i] = 0 \\ E[x_{2i}\epsilon_i] &= E[x_{2i}(e_i - u'_{2i}\alpha)] = E[x_{2i}e_i] - E[x_{2i}u'_{2i}]\alpha \\ &= E[u_{2i}e_i] - \Gamma'_{12}E[x_{1i}u'_{2i}]\alpha - \Gamma'_{22}E[z_{2i}u'_{2i}]\alpha - E[u_{2i}u'_{2i}]\alpha \\ &= E[u_{2i}e_i] - 0 - 0 - E[u_{2i}e_i] = 0 \\ &(\because \text{LS assumption from Reduced Form, Recitation 2}) \end{aligned}$$

So the key takeaway is that now x_{2i} is exogenous with ϵ_i . In words, x_{2i} is correlated with e_i through u_{2i} and ϵ_i is the error after e_i has been projected onto u_{2i} .

If we know the true u_{2i} , the (CFA) equation can be regressed with an OLS straight away. This is not usually the case, so we need to use the residual \hat{u}_{2i} which we can get from the (Reduced Form) equation. We work in these steps

1. **Obtain \hat{u}_{2i} :** This is done by regressing (Reduced Form) equation.
2. **Work with (CFA):** However, instead of u_{2i} , use \hat{u}_{2i} . Then we can run an OLS on the REWRITTEN (CFA) equation.

Once this is done, we can show that the estimates from control function approach is numerically identical to 2SLS.⁶ One thing to note is that with this setup, we can conduct a test of endogeneity. Formally we want to test

$$H_0 : E(x_{2i}e_i) = 0, H_1 : E(x_{2i}e_i) \neq 0$$

If it is the case that $E(x_{2i}e_i) = 0$, then $E(u_{2i}e_i) = 0$. Then, by how we constructed α , this implies that $\alpha = 0$. We can then show that the Wald statistics, under H_0 , is distributed as

$$\hat{\alpha}'(\text{var}(\hat{\alpha}))^{-1}\hat{\alpha} \xrightarrow{d} \chi^2_{k_2}$$

⁶A more detailed explanation can be found in Hansen (2019) pages 426-427.

To test the null against the alternative hypothesis, define $C_{1-\alpha}$ as $\Pr(\chi_{k_2}^2 \leq C_{1-\alpha}) = 1 - \alpha$. Then, we can reject the H_0 hypothesis if $\hat{\alpha}'(\text{var}(\hat{\alpha}))^{-1}\hat{\alpha} > C_{1-\alpha}$.

2 Topics on Instrumental Variables

2.1 Hausmann Tests

Hausmann Test can be utilized as a general test of specification. In this context, we are interested in using Hausmann Tests for exogeneity / endogeneity of the regressors. Assume the following data generating process

$$y_i = x'_{1i}\beta_1 + x'_{2i}\beta_2 + e_i$$

and we are interested in checking the exogeneity / endogeneity of x_{2i} . So we test $H_0 : E(x_{2i}e_i) = 0$ against $H_1 : E(x_{2i}e_i) \neq 0$. Consider these properties of 2SLS and OLS estimators

- $\hat{\beta}_{OLS}$: Consistent and minimal variance under H_0 , inconsistent under H_1
- $\hat{\beta}_{2SLS}$: Consistent in either H_0 or H_1 . Inefficient under H_0 .

Under H_0 , $\sqrt{n}(\hat{\beta}_{2SLS} - \hat{\beta}_{OLS})$ converges in distribution to $N(0, \text{var}(\hat{\beta}_{2SLS} - \hat{\beta}_{OLS}))$. Hausmann also shows that in H_0 , $\text{var}(\hat{\beta}_{2SLS} - \hat{\beta}_{OLS}) = \text{var}(\hat{\beta}_{2SLS}) - \text{var}(\hat{\beta}_{OLS})$ holds in general (you are asked to show this for a scalar case). Given this, we can write the Hausmann test statistic as

$$(\hat{\beta}_{2SLS} - \hat{\beta}_{OLS})'(\text{var}(\hat{\beta}_{2SLS}) - \text{var}(\hat{\beta}_{OLS}))^{-}(\hat{\beta}_{2SLS} - \hat{\beta}_{OLS}) \xrightarrow{d} \chi_{k_2}^2$$

where $(\text{var}(\hat{\beta}_{2SLS}) - \text{var}(\hat{\beta}_{OLS}))^{-}$ indicates a generalized inverse. We can use the test statistic, construct a critical value and reject/not reject H_0 based on comparison.

Note that generalized inverse is required since $(\text{var}(\hat{\beta}_{2SLS}) - \text{var}(\hat{\beta}_{OLS}))$ is usually not a full (column) rank matrix. The following trick is useful in showing this.

Lemma 2.1 (Breakdown of OLS coefficients, from Hansen (2019): pp. 80-82). *The OLS estimator for β_1, β_2 in $y = X_1\beta_1 + X_2\beta_2 + e$ can be written as*

$$\hat{\beta}_1 = (X'_1 M_2 X_1)^{-1} (X'_1 M_2 y), \quad \hat{\beta}_2 = (X'_2 M_1 X_2)^{-1} (X'_2 M_1 y)$$

where $M_1 = I - X_1(X_1'X_1)^{-1}X_1'$ and M_2 is defined similarly.

Proof. I will provide some sketches of this proof. Suppose that we want to find β_1 's OLS estimate. Then the problem is identical to

$$\hat{\beta}_1 = \arg \min_{\beta_1} (\min_{\beta_2} SSE(\beta_1, \beta_2))$$

In words, we concentrate out β_2 and express it as a function of β_1 . We put this expression back into the original equation and find β_1 from there. As a result, we can get that

$$\beta_2 = (X_2'X_2)^{-1}(X_2'(y - X_1\beta_1))$$

Put this back in to obtain the OLS estimation for β_1 , which is

$$\hat{\beta}_1 = (X_1'M_2X_1)^{-1}(X_1'M_2y)$$

It is also useful to note that M_1 is symmetric and idempotent. □

In our context, $\hat{\beta}_{1,OLS}$ would be $(X_1'X_1)^{-1}(X_1'(y - X_2\hat{\beta}_{2,OLS}))$ and its 2SLS estimator is $(\hat{X}_1'\hat{X}_1)^{-1}(\hat{X}_1'(y - \hat{X}_2\hat{\beta}_{2,2SLS}))$. Since $Z = [X_1 \ X_2]' \in \mathbb{R}^{n \times l}$, we can write that $X_1 = Z\Gamma$ with $\Gamma = \begin{pmatrix} I_{k_1} \\ 0 \end{pmatrix} \in \mathbb{R}^{l \times k_1}$. So we can write $\hat{X}_1 = Z(Z'Z)^{-1}Z'X_1 = Z\Gamma = X_1$. Therefore,

$$\begin{aligned} \hat{\beta}_{1,OLS} - \hat{\beta}_{1,2SLS} &= (X_1'X_1)^{-1}X_1'(\hat{X}_2\hat{\beta}_{2,2SLS} - X_2\hat{\beta}_{2,OLS}) \\ &= (X_1'X_1)^{-1}X_1'(P_ZX_2\hat{\beta}_{2,2SLS} - X_2\hat{\beta}_{2,OLS}) \\ &= (X_1'X_1)^{-1}X_1'P_ZX_2\hat{\beta}_{2,2SLS} - (X_1'X_1)^{-1}X_1'X_2\hat{\beta}_{2,OLS} \\ &= (X_1'X_1)^{-1}X_1'X_2\hat{\beta}_{2,2SLS} - (X_1'X_1)^{-1}X_1'X_2\hat{\beta}_{2,OLS} (\because X_1'P_Z = (P_ZX_1)' = X_1') \\ &= (X_1'X_1)^{-1}X_1'X_2(\hat{\beta}_{2,2SLS} - \hat{\beta}_{2,OLS}) \end{aligned}$$

So $\hat{\beta}_{1,OLS} - \hat{\beta}_{1,2SLS}$ is merely obtained by linearly transforming $\hat{\beta}_{2,OLS} - \hat{\beta}_{2,2SLS}$. Therefore, $var(\hat{\beta}_{2,2SLS} - \hat{\beta}_{2,OLS})$ is not full rank.

2.2 Subset Endogeneity Test

This section can be seen as an extension to the endogeneity test we showed earlier. Instead, we break down x_{2i} into two parts - one that is 'potentially' endogenous (x_{2i}) and one that is endogenous for sure (x_{3i}). Then, we again test for $H_0 : E(x_{2i}e_i) = 0$ against

$H_1 : E(x_{2i}e_i) \neq 0$. We now go over how control function approach can be utilized in this context. We have the set of required equations as follows

$$y_i = x'_{1i}\beta_1 + x'_{2i}\beta_2 + x'_{3i}\beta_3 + e_i \quad (\text{Structural})$$

$$x_{2i} = \Gamma'_2 z_i + u_{2i} \quad (\text{Reduced Form 2})$$

$$x_{3i} = \Gamma'_3 z_i + u_{3i} \quad (\text{Reduced Form 3})$$

We then project e_i onto u_{2i}, u_{3i} to obtain

$$e_i = u'_{2i}\alpha_2 + u'_{3i}\alpha_3 + \epsilon_i \quad (\text{LP2})$$

where the following is satisfied

$$\begin{pmatrix} E(u_{2i}u'_{2i}) & E(u_{2i}u'_{3i}) \\ E(u_{3i}u'_{2i}) & E(u_{3i}u'_{3i}) \end{pmatrix} \begin{pmatrix} \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} E(u_{2i}e_i) \\ E(u_{3i}e_i) \end{pmatrix}$$

then the control function approach would require us to apply an OLS to the following

$$y_i = x'_{1i}\beta_1 + x'_{2i}\beta_2 + x'_{3i}\beta_3 + u'_{2i}\alpha_2 + u'_{3i}\alpha_3 + \epsilon_i \quad (\text{CFA2})$$

As before, $E(x_{2i}e_i) = 0$ implies $E(u_{2i}e_i) = 0$. This allows us to replace the H_0 with

$$H_0 : E[u_{2i}u'_{2i}]\alpha_2 + E[u_{2i}u'_{3i}]\alpha_3 = 0$$

This is a special case of $R\beta = c$ type of hypothesis test studied in Recitation 1. However, since we cannot usually know what the true value of u_{2i}, u_{3i} are, we use the residuals, and replace R with \hat{R} constructed from the sample analogue using residuals.

2.3 Overidentification Test

Consider the following setup

$$y_i = x'_i\beta + e_i, \quad (\dim(x_i) = k < \dim(z_i) = l)$$

where we have an overidentified model, with number of instruments being larger than endogenous variables. We can test for the validity of the instruments by testing $H_0 : E(z_i e_i) = 0$

vs. $H_1 : E(z_i e_i) \neq 0$. To do this, we construct a linear projection equation by projecting e_i onto z_i , obtaining

$$e_i = z_i' \alpha + \epsilon_i$$

Thus, $\alpha = E(z_i z_i')^{-1} E(z_i e_i)$. Because of this, the null and alternative hypotheses can be expressed by $H_0 : \alpha = 0$ vs. $H_1 : \alpha \neq 0$. This leaves us with the challenge of estimating α . Since we do not know what the true value of the error terms are, we take the following steps.

1. **Obtain \hat{e}_i :** This can be done using 2SLS estimates of β . As a result,

$$\hat{e}_i = y_i - x_i' \hat{\beta}_{2SLS} \implies \hat{e} = y - X \hat{\beta}_{2SLS}$$

2. **Obtain $\hat{\alpha}$:** Replace e with \hat{e} to get

$$\hat{\alpha} = (Z'Z)^{-1} Z' \hat{e}$$

3. **Sargan Test:** We will assume homoskedasticity (otherwise, the test statistic does not converge to χ^2 distribution). Then we make use of the following test statistic

$$S = \hat{\alpha}' (\text{var}(\hat{\alpha}))^{-1} \hat{\alpha} = \frac{\hat{e}' Z (Z' Z)^{-1} Z' \hat{e}}{\hat{\sigma}^2}$$

where $\hat{\sigma}^2$ can be obtained from two paths - one using $\frac{1}{n} \hat{e}' \hat{e}$ and the other using $\frac{1}{n} \hat{e}' \hat{e} = \frac{1}{n} (\hat{e} - Z \hat{\alpha})' (\hat{e} - Z \hat{\alpha})$. While they are slightly different in that the first estimate has larger variances⁷, they are asymptotically equal. Under the null, $S \xrightarrow{d} \chi^2_{l-k}$

Some note of caution include:

- Even if the null hypothesis is rejected, the test alone is not enough to pick up which variable is violating the exclusion restriction.
- Since $E(z_i e_i) = 0 \iff E(z_i y_i) - E(z_i x_i') \beta = 0$, we are effectively solving a system of $l \times 1$ restrictions, despite β being a k -vector. Depending on the subset of z_i vectors used, the β may differ in finite samples. If the overidentification hypothesis is correct, the estimates that are different in finite samples are consistent for true β . You can think of the overidentification test as an attempt to see whether different β estimates are converging to a same parameter as $n \rightarrow \infty$

⁷It can be shown that $\hat{e}' \hat{e} = \hat{e}' \hat{e} - \hat{e}' Z (Z' Z)^{-1} Z' \hat{e}$ Since $\hat{e} \xrightarrow{p} e$ and $E(z_i e_i) = 0$, it can be shown that $\hat{e}' Z (Z' Z)^{-1} Z' \hat{e}$ goes to 0.

2.4 Weak Instrument Test

Assume that the structural and reduced form equations are (we are working with a scalar regressors)

$$y_i = x_i\beta + e_i$$

$$x_i = z_i\gamma + u_i$$

We say that there is a problem of weak instrument if $\gamma \simeq 0$. This can cause the 2SLS estimates to be biased and the distribution to be affected. Specifically, let $\gamma = \frac{\mu}{\sqrt{n}}$. For simplicity, I will assume

- $\text{var} \left(\begin{pmatrix} e_i \\ u_i \end{pmatrix} | z_i \right) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} = \Sigma$
- $E(z_i^2) = 1$
- $\frac{1}{\sqrt{n}} \sum_{i=1}^n \begin{pmatrix} z_i e_i \\ z_i u_i \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = N(0, \Sigma)$

Now I will show that neither OLS nor IV estimator is not consistent.

- OLS: Note that $\hat{\beta}_{OLS} - \beta$ can be written as $\frac{n^{-1} \sum_{i=1}^n x_i e_i}{n^{-1} \sum_{i=1}^n x_i^2}$, equivalent to

$$\begin{aligned} \frac{n^{-1} \sum_{i=1}^n x_i e_i}{n^{-1} \sum_{i=1}^n x_i^2} &= \frac{n^{-1} \sum_{i=1}^n (z_i \gamma + u_i) e_i}{n^{-1} \sum_{i=1}^n (z_i \gamma + u_i)^2} \\ &= \frac{n^{-1} \sum_{i=1}^n (z_i e_i \gamma + u_i e_i)}{n^{-1} \sum_{i=1}^n (z_i^2 \gamma^2 + u_i^2 + 2z_i u_i \gamma)} \\ &= \frac{n^{-1} \sum_{i=1}^n u_i e_i}{n^{-1} \sum_{i=1}^n u_i^2} + o_p(1) \\ &\xrightarrow{p} E(u_i e_i) / E(u_i^2) = \rho \end{aligned}$$

- IV: For $\hat{\beta}_{IV} - \beta$, we can write $\frac{n^{-1/2} \sum_{i=1}^n z_i e_i}{n^{-1/2} \sum_{i=1}^n z_i x_i}$ or equivalently

$$\begin{aligned} \hat{\beta}_{IV} - \beta &= \frac{n^{-1/2} \sum_{i=1}^n z_i e_i}{n^{-1/2} \sum_{i=1}^n z_i (z_i \gamma + u_i)} \\ &= \frac{n^{-1/2} \sum_{i=1}^n z_i e_i}{n^{-1/2} \sum_{i=1}^n z_i u_i + n^{-1} \sum_{i=1}^n z_i^2 \mu} \xrightarrow{d} \frac{\xi_1}{\xi_2 + \mu} \end{aligned}$$

which is not centered at 0, making it inconsistent.

In STATA, such test can be implemented by typing in `estat firststage` after 2SLS or LIML regression.