Introduction to Econometrics II: Recitation 2*

Seung-hun Lee†

January 29th, 2020

1 Classical Linear Models

1.1 Ordinary Least Squares

Throughout this lecture (and possibly beyond), we will assume a data generating process that looks like

$$y_i = x_i'\beta + e_i,\ x_i = \begin{pmatrix} x_{i1} \\ ... \\ x_{ik} \end{pmatrix},\ i = 1, ..., n$$

where x_i and β are both in $\mathbb{R}^k$ and y_i and e_i are scalars. In a matrix notation, this can be written as

$$y = X\beta + e,\ y = \begin{pmatrix} y_1 \\ ... \\ y_n \end{pmatrix} \in \mathbb{R}^n, X = \begin{pmatrix} x_1' \\ ... \\ x_n' \end{pmatrix} \in \mathbb{R}^{n\times k}, e = \begin{pmatrix} e_1 \\ ... \\ e_n \end{pmatrix} \in \mathbb{R}^n$$

To demonstrate the consistency and the limiting distribution of the OLS estimators, I will use some of these assumptions

Assumption 1.1 (Assumptions for Classical Linear Models)**.** *The following assumptions are used in showing consistency and the limiting distribution of OLS estimators*

*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.

A1 (y_i, x_i) *are IID across i's*

A2 $E(x_i e_i) = 0$

A2′ $E(e_i|x_i) = 0$ *(Problem set 1 includes a question that asks you to derive A2 from A2′)*

A3 $E(x_i x_i') = Q$ *is a positive definite matrix (hereafter PD matrix)*

A4 $E||x_i^4|| < \infty, E||y_i^4|| < \infty$

The OLS estimator can be found by minimizing the sum of squared errors. In other words

$$\hat{\beta} = \min_b \sum_{i=1}^{n} (y_i - x_i' b)^2 = \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1} \left(\sum_{i=1}^{n} x_i y_i \right)$$

or in matrix notation, $(X'X)^{-1}(X'y)$. The consistency and the limiting distribution of OLS estimators can be demonstrated as follows

Theorem 1.1 (Consistency of $\hat{\beta}$)**.** *Under assumptions* ***A1-A3****,* $\hat{\beta} \xrightarrow{p} \beta$

Proof. Rewrite $\hat{\beta}$ as $\beta + \left(\sum_{i=1}^n x_i x_i'\right)^{-1} \left(\sum_{i=1}^n x_i e_i\right)$. To carry out the asymptotic analysis on the summation terms, multiply $\frac{1}{n}$ to both. By **A1**, we can deduce that x_i and e_i are IID. Then, we can apply weak law of large numbers and continuous mapping theorem to show that

$$\frac{1}{n} \sum_{i=1}^{n} x_i x_i' \xrightarrow{p} E\left(x_i x_i'\right)$$

$$\left(\frac{1}{n} \sum_{i=1}^{n} x_i x_i'\right)^{-1} \xrightarrow{p} E\left(x_i x_i'\right)^{-1} (\because CMT)$$

$$\frac{1}{n} \sum_{i=1}^{n} x_i e_i' \xrightarrow{p} E\left(x_i e_i'\right)$$

By assumptions **A2,A3**, $\left(\frac{1}{n} \sum_{i=1}^n x_i x_i'\right)^{-1} \xrightarrow{p} Q^{-1}$ and $\frac{1}{n} \sum_{i=1}^n x_i e_i' \xrightarrow{p} 0$. By Slutsky's theorem, $\left(\sum_{i=1}^n x_i x_i'\right)^{-1} \left(\sum_{i=1}^n x_i e_i\right) \xrightarrow{p} 0$. Thus, $\hat{\beta} \xrightarrow{p} \beta$. □

Theorem 1.2 (Limiting distribution of $\hat{\beta}$)**.** *Under assumptions* ***A1-A4****, the limiting distribution of* $\hat{\beta}$ *is characterized by* $\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, Q^{-1}\Omega Q^{-1})$ *, where* $\Omega = E(x_i x_i' e_i^2)$

Proof. We can write $\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{1}{n} \sum_{i=1}^n x_i x_i'\right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n x_i e_i\right)$. We know from the previous theorem that $\left(\frac{1}{n}\sum_{i=1}^n x_i x_i'\right)^{-1} \xrightarrow{p} Q^{-1}$. So we need to work on $\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n x_i e_i\right)$. From the central limit theorem, we can obtain the limiting distribution of $\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n x_i e_i\right)$

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n x_i e_i \xrightarrow{d} N(0,\Omega)$$

since $E(x_i e_i) = 0$ by **A2** and $var(x_i e_i) = E(x_i e_i e_i x_i') - (E(x_i e_i))^2 = E(x_i x_i' e_i^2) = \Omega$ (In using CLT, we need assumption **A4** so that the variance-covariance matrix obtained from here is finite.) Then, by Slutsky's theorem,

$$\sqrt{n}(\hat{\beta}-\beta) = \left(\frac{1}{n}\sum_{i=1}^n x_i x_i'\right)^{-1}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n x_i e_i\right) \xrightarrow{d} Q^{-1}N(0,\Omega) = N(0, \underbrace{Q^{-1}\Omega Q^{-1}}_{V})$$

$\square$

If we are interested in a particular element of β, namely β_j, we will need to work on the following limiting distribution

$$\sqrt{n}(\hat{\beta}_j - \beta_j) \xrightarrow{d} N(0, V_{jj}) \ (V_{jj} \text{ is the } (j,j)\text{th element of V})$$

1.2 Hypothesis Tests

- **Single element**: Consider the following setting

$$H_0 : \beta_j = \beta_j^0, \ \ H_1 : \beta_j \neq \beta_j^0$$

From the limiting distribution of β_j, we can show that the test statistic is distributed as a standard normal under H_0. It is characterized as

$$\frac{\hat{\beta}_j - \beta_j^0}{\sqrt{\widehat{V}_{jj}/n}} \xrightarrow{d} N(0,1)$$

where $\widehat{V}$ is estimated version of the variance (more on that later).

- **Multiple element**: Sometimes, we may be interested in the features of the linear combinations of the elements of β, or even multiple restrictions, Let $R \in \mathbb{R}^{k\times r}$ characterize such restrictions. Then we can write

$$H_0 : R'\beta = c, \;\; H_1 : \neg H_0$$

Then from the limiting distribution of $\hat{\beta}$, we can apply Slutsky's theorem to get the necessary limiting distribution

$$\sqrt{n}(R'\hat{\beta} - R'\beta) = \sqrt{n}R'(\hat{\beta} - \beta) \xrightarrow{d} N(0, R'VR)$$

Since $R'\beta = c$ under H_0, we can obtain the following Wald test statistic. (For convenience, we also assume that V is known)

$$n(R'\hat{\beta} - c)'(R'VR)^{-1}(R'\hat{\beta} - c) \xrightarrow{d} \chi^2_r$$

To see why we ended up with a χ^2_r distribution, we need to understand the following.

Theorem 1.3 (Chi-squared distribution). *If $\eta \sim N(0, A)$, where A is PD, then*

$$\eta' A^{-1} \eta \sim \chi^2_{rank(A)}$$

- **Notes on** $\widehat{V}$: The definition of $\widehat{V}$ is $\widehat{V} \equiv \widehat{Q}^{-1}\widehat{\Omega}\widehat{Q}^{-1}$, where
  - $\widehat{Q}$ is a sample analogue of Q, written as $\frac{1}{n}\sum_{i=1}^n x_i x_i'$
  - $\widehat{\Omega}$ is a sample analogue of $E(x_i x_i' e_i^2)$, also with the consideration that the true value of e_i is replaced with the residual $\hat{e}_i$. Therefore, we can write $\frac{1}{n}\sum_{i=1}^n x_i x_i' \hat{e}_i^2$

2 Instrumental Variables Method

We still work with the data generating process $y_i = x_i'\beta + e_i$, but now with the possibility that $E(x_i e_i) \neq 0$. In other words, the error term and the regressor can now be correlated (or the regressors are endogenous). Since we required **A2** assumption in showing that OLS estimators are consistent, the fact that $E(x_i e_i)$ is not necessarily zero implies that OLS may no longer be consistent. In this section, we study cases in which regressors can become endogenous and how instrumental variables allow us to address this problem.

2.1 Sources of Endogeneity

- **Measurement Error in Regressors:** Suppose that the linear model we want to estimate is as follows

$$y_i = {x_i^*}' \beta + e_i \text{ (We assume } E(x_i^* e_i) = 0)$$

However, we cannot observe x_i^*. Instead, we can observe $x_i = x_i^* + v_i$, where v_i has mean zero and independent of both x_i^* and e_i. So we have a classical measurement error in which x_i is unbiased, but noisy measure of x_i^*. If we use x_i instead,

$$y_i = (x_i - v_i)'\beta + e_i \quad = x_i'\beta \underbrace{-v_i'\beta + e_i}_{=u_i}$$

Then $E(x_i u_i)$ is as follows

$$E(x_i u_i) = E[x_i(-v_i'\beta + e_i)] = E[(x_i^* + v_i)(-v_i'\beta + e_i)] = -E(v_i v_i')\beta$$

So unless $\beta = 0$, or $E(v_i v_i') = 0$, $E(x_i u_i) \neq 0$. When we use OLS on this context, the probability limit of the OLS estimator would be

$$\begin{aligned}
\hat{\beta}_{OLS} &= \beta + E(x_i x_i')^{-1} E(x_i u_i) \\
&= \beta - E(x_i x_i')^{-1} E(v_i v_i)\beta \\
&= \beta - E[(x_i^* + v_i)(x_i^* + v_i)']^{-1} E(v_i v_i)\beta \\
&= \beta - [E(x_i^* {x_i^*}') + E(v_i v_i')]^{-1} E(v_i v_i)\beta \\
&= \frac{E(x_i^* {x_i^*}')}{E(x_i^* {x_i^*}') + E(v_i v_i')}\beta \ (\leq \beta)
\end{aligned}$$

The only time that $\frac{E(x_i^* {x_i^*}')}{E(x_i^* {x_i^*}') + E(v_i v_i')}\beta$ would equal β is when β itself is zero or when $E(v_i v_i') = 0$. The latter, however, implies that $var(v_i) = 0$ and the noise v_i has mean 0 and has a point mass at 0 - so no measurement error exists. In usual cases, the OLS estimator has a probability limit of something less than β. This is what is also known as **attenuation bias**.

> **Comment 2.1** (Comment on Measurement Errors)**.** *So how do we address the endogeneity problem?*

- *If there exists another noisy, but unbiased measure of x_i^*, namely $w_i = x_i^* + \delta_i$, we can use w_i to instrument for x_i. The condition is that η_i has mean zero and uncorrelated with $(x_i^*, e_i.v_i)$. Try verifying that this satisfies all IV conditions.*
- *If there is a measurement error in y_i, the only this it does is to change the component of e_i. Assuming all the old assumptions hold, this does not pose as much problem as having a measurement error in the regressor.*

- **Simultaneity Bias:** A classic example of this would be a supply and demand system type of setting:

$$q_i = \beta_1 p_i + u_i \tag{Supply}$$
$$q_i = -\beta_2 p_i + v_i \tag{Demand}$$

I will assume $e_i = (u_i \; v_i)'$ is IID, $E(e_i) = 0, E(e_i e_i') = I_2$ When you do some algebra, the equilibrium of this system is

$$p_i = \frac{v_i - u_i}{\beta_1 + \beta_2}, q_i = \frac{\beta_1 v_i + \beta_2 u_i}{\beta_1 + \beta_2}$$

So for both supply and demand equations, we have $E(p_i u_i) \neq 0$ and $E(p_i v_i) \neq 0$. When naively applying OLS to this equation, the result is as follows.

$$q_i = \beta^* p_i + \eta_i, \; E(p_i \eta_i) = 0 \implies \hat{\beta}^* = \frac{E(p_i q_i)}{E(p_i^2)} = \frac{\beta_1 - \beta_2}{2}$$

Thus, OLS estimators does not converge to either one of β_1 or β_2, resulting in a **simultaneity bias**.

- **Omitted Variable Bias (OVB):** Suppose that we are interested in the determinant of wages (y_i). Also assume that education, x_i, and innate ability, a_i, determine wages in the following manner

$$y_i = x_i \beta_1 + a_i \beta_2 + e_i, \;\; E(x_i e_i) = 0, E(a_i e_i) = 0$$

However, instead of observing (y_i, x_i, a_i), we can only observe (y_i, x_i). the best we can do at the moment is to estimate the following equation

$$y_i = x_i \beta_1 + u_i, \text{ where } u_i = a_i \beta_2 + e_i$$

Then $E(x_i u_i)$ becomes

$$E(x_i u_i) = E(x_i(a_i\beta_2 + e_i)) = E(x_i a_i)\beta_2 + 0 = E(x_i a_i)\beta_2$$

Therefore, when 1)x_i and a_i are correlated and 2)$\beta_2 \neq 0$, x_i is endogenous with respect to u_i. On the flip side, if either one of the condition is not met, $E(x_i u_i) = 0$ again. Moreover, the OLS estimator acquired here has a probability limit of

$$\hat{\beta}_{OLS} = \beta_1 + E(x_i^2)^{-1}E(x_i u_i) = \beta_1 + E(x_i^2)^{-1}E(x_i a_i)\beta_2$$

So if 1) and 2) occurs, the above does not converge in probability to β_1. Also note that we can determine the direction of the bias by the sign of $E(x_i a_i)$ and β_2.

2.2 IV Estimators

Assume that the data generating process is as follows

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

where $E(x_{1i}e_i) = 0, E(x_{2i}e_i) \neq 0$, and $\dim(x_{1i}) = k_1, \dim(x_{2i}) = k_2,\ k_1 + k_2 = k$. In our case, x_{2i} is the collection of endogenous variables. It can be shown that consistency of the OLS estimators of β_2 and β_1 will not be guaranteed under this situation.

Example 2.1 (When $k_1 = k_2 = 1$). *In this case, we can write $\hat{\beta}_1$ as*

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_{2i}^2 \sum_{i=1}^n x_{1i}y_i - \sum_{i=1}^n x_{1i}x_{2i} \sum_{i=1}^n x_{2i}y_i}{\sum_{i=1}^n x_{1i}^2 \sum_{i=1}^n x_{2i}^2 - (\sum_{i=1}^n x_{1i}x_{2i})^2}$$

When we replace y_i with $x_{1i}\beta_1 + x_{2i}\beta_2 + e_i$, we end up with

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^n x_{2i}^2 \sum_{i=1}^n x_{1i}e_i - \sum_{i=1}^n x_{1i}x_{2i} \sum_{i=1}^n x_{2i}e_i}{\sum_{i=1}^n x_{1i}^2 \sum_{i=1}^n x_{2i}^2 - (\sum_{i=1}^n x_{1i}x_{2i})^2}$$

Since $E(x_{2i}e_i) \neq 0$, then $\frac{1}{n}\sum_{i=1}^n x_{2i}e_i$ converges to something that is not zero (whereas $\frac{1}{n}\sum_{i=1}^n x_{1i}e_i$ does converge in probability to 0). So the whole fraction term does not converge in probability to 0 and even $\hat{\beta}_1$ is not consistent.

Let $z_i \in \mathbb{R}^l = \begin{pmatrix} z_{1i} \\ z_{2i} \end{pmatrix} = \begin{pmatrix} x_{1i} \\ z_{2i} \end{pmatrix}$, where $\dim(z_{2i}) = l - k_1$ and $l - k_1 \geq k_2$. For z_i to be a valid IV, the following conditions must be satisfied

Definition 2.1 (IV conditions)**.** z_i is a valid IV if

1. **Exogeneity**: $E(z_i e_i) = 0$
   (a) **Exclusion**: $E(z_i y_i) = \beta_1 E(z_i x_{1i}) + \beta_2 E(z_i x_{2i})$, in other words, z_i should impact y_i through x_{1i} and x_{2i}
2. **Relevancy**: $rank[E(z_i x_i')] = \dim(x_i) = k$
3. **PD**: $E(z_i z_i') > 0$

We will derive IV estimators in two ways: Reduced form approach and 2SLS approach

- **Reduced form**: In this approach, we assume that z_i is a least squares projection. So we can write

$$\begin{aligned} x_i &= \Gamma' z_i + u_i \ (\Gamma \in \mathbb{R}^{l \times k}) \\ \implies z_i x_i' &= z_i z_i' \Gamma + z_i u_i' \\ \implies E(z_i x_i') &= E(z_i z_i')\Gamma + E(z_i u_i') \end{aligned}$$

Since z_i is a least squares projection, $E(z_i u_i) = 0$. So

$$\Gamma = E(z_i z_i')^{-1} E(z_i x_i')$$

and the estimator for Γ would be its sample analogue, $\widehat{\Gamma} = \left(\frac{1}{n}\sum_{i=1}^n z_i z_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n z_i x_i'\right) = (Z'Z)^{-1}(Z'X)$.

The we get to the structural equation

$$y_i = x_i'\beta + e_i \iff y_i = (z_i'\Gamma + u_i')\beta + e_i \iff y_i = z_i' \underbrace{\Gamma\beta}_{=\lambda} + \underbrace{u_i'\beta + e_i}_{=v_i}$$

From the assumptions, we can show that $E(z_i v_i) = 0$. As such,

$$\lambda = E(z_i z_i')^{-1} E(z_i y_i)$$

with its sample analogue being $\hat{\lambda} = \left(\frac{1}{n}\sum_{i=1}^n z_i z_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^n z_i y_i\right) = (Z'Z)^{-1}Z'y$

In case where $k = l$, Z itself becomes invertible. Then we can show that $\beta = \Gamma^{-1}\lambda$ and thus

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

If otherwise, We note that $\widehat{\Gamma}\beta + \text{error} = \hat{\lambda}$ and show that

$$\hat{\beta}_{IV} = (\widehat{\Gamma}'\widehat{\Gamma})^{-1}\widehat{\Gamma}'\hat{\lambda}$$

- **2SLS**: Suppose the structural equation and the first-stage regression is as follows.

$$y = X\beta + e \qquad \text{(Structural)}$$
$$X = Z\Gamma + u \qquad \text{(First Stage)}$$

where $Z \in \mathbb{R}^{n\times l}, \Gamma \in \mathbb{R}^{l\times k}$. Z is our IV and we still maintain the least square projection assumption. We proceed as follows

1. Regress the first stage and obtain $\widehat{\Gamma} = (Z'Z)^{-1}Z'X$. Then the predicted value of X, denoted as $\widehat{X} = Z(Z'Z)^{-1}Z'X = P_Z X$.

> **Property 2.1** (Properties of Projection Matrix P_Z)**.** *Note the following*
>
> – *Symmetric:* $P_Z' = (Z(Z'Z)^{-1}Z')' = Z(Z'Z)^{-1}Z' = P_Z$
>
> – *Idempotent:* $P_Z^2 = Z(Z'Z)^{-1}Z'Z(Z'Z)^{-1}Z' = Z(Z'Z)^{-1}Z' = P_Z$

2. In the structural equation, replace X with $\widehat{X}$ and obtain

$$\begin{aligned}\hat{\beta}_{2SLS} &= (\widehat{X}'\widehat{X})^{-1}\widehat{X}'y = (X'P_Z'P_ZX)^{-1}(X'P_Z'y)\\ &= (X'P_ZX)^{-1}X'P_Zy = (\widehat{X}'X)^{-1}\widehat{X}'y\end{aligned}$$

which is effectively replacing Z in the previous approach with $\widehat{X}$.

2.2.1 Consistency and Limiting Distribution of 2SLS

To show the asymptotic properties of the 2SLS estimators, we need the following set of assumptions

Assumption 2.1 (2SLS Assumptions)**.** *This is a list of required assumptions.*

T1 (y_i, x_i, z_i) *are IID*

T2 *Finite second moments:* $E||y_i^2|| < \infty, E||x_i^2|| < \infty, E||z_i^2|| < \infty$

T3 $E(z_i z_i') > 0$

T4 $rank[E(z_i x_i')] = k$

T5 $E(z_i e_i) = 0$

T6 *Finite fourth moments:* $E||y_i^4|| < \infty, E||x_i^4|| < \infty, E||z_i^4|| < \infty$

T7 $E(z_i z_i' e_i^2) = \Omega > 0$

Then, we can show the consistency and asymptotic normality of 2SLS estimators.

Theorem 2.1 (Consistency of 2SLS)**.** *Under assumptions **T1-T5**,* $\hat{\beta}_{2SLS} \xrightarrow{p} \beta$

Proof. We can rewrite 2SLS estimators as $\hat{\beta}_{2SLS} = \beta + (X'P_Z X)^{-1} X' P_Z e$, which becomes

$$\beta + \left[\frac{X'Z}{n} \left(\frac{Z'Z}{n} \right)^{-1} \frac{Z'X}{n} \right]^{-1} \left[\frac{X'Z}{n} \left(\frac{Z'Z}{n} \right)^{-1} \frac{Z'e}{n} \right]$$

Define $Q_{ZX} = E(z_i x_i'), Q_{XX} = E(x_i x_i')$. Then by weak law of large numbers,

$$\frac{Z'X}{n} \xrightarrow{p} Q_{ZX}, \frac{X'X}{n} \xrightarrow{p} Q_{XX}, \frac{Z'e}{n} \xrightarrow{p} 0$$

By applying Slutsky's theorem and continuous mapping theorem, all the terms right of β converge in probability to 0. Thus, $\hat{\beta}_{2SLS} \xrightarrow{p} \beta$ □

Theorem 2.2 (Limiting Distribution of 2SLS)**.** *Under assumptions **T1-T7**, the limiting distribution of the 2SLS estimator is characterized by* $\sqrt{n}(\hat{\beta}_{2SLS} - \beta) \xrightarrow{d} N(0, V_\beta)$

Proof. Note that

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) = \left[\frac{X'Z}{n} \left(\frac{Z'Z}{n} \right)^{-1} \frac{Z'X}{n} \right]^{-1} \left[\frac{X'Z}{n} \left(\frac{Z'Z}{n} \right)^{-1} \frac{Z'e}{\sqrt{n}} \right]$$

By CLT, $\frac{Z'e}{\sqrt{n}} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i e_i \xrightarrow{d} N(0, \Omega)$. Then apply Slutsky theorem and continuous mapping theorem to obtain

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) \xrightarrow{d} N(0, \underbrace{A_n^{-1} Q'_{ZX} Q_{ZZ}^{-1} \Omega Q_{ZZ}^{-1} Q_{ZX} A_n^{-1}}_{=V_\beta})$$

where $A_n = \left[\frac{X'Z}{n}\left(\frac{Z'Z}{n}\right)^{-1}\frac{Z'X}{n}\right]$ □

2.2.2 Remarks

- **Correct Standard Errors** When estimating $\Omega = E(z_i z_i' e_i^2)$, we are not aware of the true error. So we need to estimate this as well. Note that we need to use a proper residual. Namely, we must use

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

  This is correct, as $\hat{\beta}_{2SLS} \xrightarrow{p} \beta$. However, we frequently make a mistake of using

$$\tilde{e}_i = y_i - \hat{x}_i'\hat{\beta}_{2SLS}$$

  Since $\tilde{x}_i$ is not exactly x_i, this converges in probability to something else.

  Comment 2.2 (on STATA). *To see this, compare the standard errors from doing* `ivregress 2sls y (x=z)` *and 'hard-coding' 2SLS by using* `reg x z`, *then* `predict xhat`, *and then coding* `reg y xhat`.

- **Nonlinear Extension**: If we are instead interested in the properties of $g(\hat{\beta}_{2SLS})$, where $g(\cdot)$ is not necessarily nonlinear, we can apply delta method here.

- **Too Many IV?**: In practice, when we have too many IV's (large dimension for z_i), it may be possible for the instruments to 'perfectly' predict x_i. In other words, $\hat{x}_i$ becomes nearly identical to x_i. This can become a problem because the endogeneity bias that plagued original structural equation may not be mitigated.
  One way of getting around this is to use a regression method that involves penalty mechanism - like LASSO, for instance. We will also formally treat this issue when we learn over-identification tests in the near future.