# ECMA 31320 Pset 2

CHRIS LIAO
NOAH SOBEL-LEWIN

April 6, 2022

## 1

### a

$y$ has dimensions $n \times 1$.
$X_2$ has dimensions $n \times (p - p_1)$.
$\widehat{\beta}_1$ has dimensions $p_1 \times 1$.

### b

We can rewrite $X = \begin{bmatrix} X_1 & X_2 \end{bmatrix}$ and $\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$ so $X\beta = \begin{bmatrix} X_1 & X_2 \end{bmatrix} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = X_1\beta_1 + X_2\beta_2$. Then,

$$\begin{aligned}
&\min_{\beta}(y - X\beta)'(y - X\beta) \\
&= \min_{\beta_1,\beta_2}(y - X_1\beta_1 - X_2\beta_2)'(y - X_1\beta_1 + X_2\beta_2)
\end{aligned}$$

so we can find the normal equations by minimizing with respect to either $\beta_1$ and $\beta_2$.

$$\begin{aligned}
\frac{\partial}{\partial\beta_1}(y - X_1\beta_1 - X_2\beta_2)'(y - X_1\beta_1 + X_2\beta_2) \\
\text{so} - 2X_1'(y - X_1\beta_1 + X_2\beta_2) = 0 \\
\iff X_1'(y - X_1\beta_1 + X_2\beta_2) = 0 \\
\iff X_1'y = X_1'(X_1\beta_1 + X_2\beta_2) = 0 \\
\iff X_1'X_1\beta_1 + X_1'X_2\beta_2 = X_1'y
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial}{\partial \widehat{\beta}_2}(y - X_1\widehat{\beta}_1 - X_2\beta_2)'(y - X_1\widehat{\beta}_1 + X_2\beta_2) \\
\text{so} - 2X_2'(y - X_1\widehat{\beta}_1 + X_2\widehat{\beta}_2) = 0 \\
\iff X_2'(y - X_1\widehat{\beta}_1 + X_2\widehat{\beta}_2) = 0 \\
\iff X_2'y = X_2'(X_1\widehat{\beta}_1 + X_2\widehat{\beta}_2) = 0 \\
\iff X_2'X_1\widehat{\beta}_1 + X_2'X_2\widehat{\beta}_2 = X_2'y
\end{aligned}
$$

## c

$$
\begin{aligned}
X_1(X_1'X_1)^{-1}(X_1'X_1\beta_1 + X_1'X_2\beta_2) = X_1(X_1'X_1)^{-1}X_1'y \qquad \iff \\
X_1(X_1'X_1)^{-1}X_1'X_1\beta_1 + X_1(X_1'X_1)^{-1}X_1'X_2\beta_2 = X_1(X_1'X_1)^{-1}X_1'y \qquad \iff \\
X_1\beta_1 + P_1X_2\beta_2 = P_1y
\end{aligned}
$$

## d

$$
\begin{aligned}
y &= XB + \hat{\epsilon} && \iff \\
y &= X_1\beta_1 + X_2\beta_2 + \hat{\epsilon} && \iff \\
M_1y &= M_1X_1\beta_1 + M_1X_2\beta_2 + M_1\hat{\epsilon} \\
\widetilde{y} &= \widetilde{X_2}\widehat{\beta}_2 + (I - P_1)\hat{\epsilon} \\
\widetilde{y} &= \widetilde{X_2}\widehat{\beta}_2 + \hat{\epsilon} - P_1\hat{\epsilon} \\
\widetilde{y} &= \widetilde{X_2}\widehat{\beta}_2 + \hat{\epsilon} \text{ by orthogonality of } \widehat{\epsilon}
\end{aligned}
$$

## e

$$
\begin{aligned}
&\widetilde{y} = \widetilde{X}_2\widehat{\beta}_2 + \widehat{\epsilon} \\
&\iff (\widetilde{X}_2'\widetilde{X}_2)^{-1}\widetilde{X}_2'\widetilde{y} = (\widetilde{X}_2'\widetilde{X}_2)^{-1}\widetilde{X}_2'\widetilde{X}_2\widehat{\beta}_2 + (\widetilde{X}_2'\widetilde{X}_2)^{-1}\widetilde{X}_2'\widehat{\epsilon} \\
&\iff (\widetilde{X}_2'\widetilde{X}_2)^{-1}\widetilde{X}_2'\widetilde{y} = \widehat{\beta}_2 + (\widetilde{X}_2'\widetilde{X}_2)^{-1}\widetilde{X}_2'\widehat{\epsilon} \\
&\iff (\widetilde{X}_2'\widetilde{X}_2)^{-1}\widetilde{X}_2'\widetilde{y} = \widehat{\beta}_2 + (\widetilde{X}_2'\widetilde{X}_2)^{-1}\widetilde{X}_2'\widehat{\epsilon} = \widehat{\beta}_2
\end{aligned}
$$

Where the last equality follows from the fact that $\widehat{\epsilon}$ is orthogonal to $\widetilde{X}_2$. Note that this $\widehat{\beta}_2$ is identical to the coefficient on $\widetilde{X_2}$ in the best linear prediction problem of $\widetilde{y}$ given $\widetilde{X}_2$.

### f

We have determined that the coefficient, $\widehat{\beta_2}$, on $X_2$ in the best linear prediction problem of $y$ given $X_1$ and $X_2$ can be retrieved by first projecting $X_2$ and $y$ on the orthogonal complement of $S(X_1)$ to get $\widetilde{X_2}$ and $\widetilde{y}$ and then getting the coefficient on $\widetilde{X_2}$ in the best linear prediction problem of $\widetilde{y}$ given $\widetilde{X_2}$.

### g

The results above hold given we have data matrices $y$ and $X$. It is completely agnostic as to the data generating process and follows mechanically. Therefore, the answer to the following four questions is no.

1. no
2. no
3. no
4. no

## 2

### a

We want to show that $\mathbb{E}[\epsilon_i^2 x_i x_i'] = \mathbb{E}[x_i x_i']\mathbb{E}[\epsilon_i^2 \mid x_{1i} \cdots x_{Ki}]$.

$$\begin{aligned}
\mathbb{E}[\epsilon_i^2 x_i x_i'] &= \mathbb{E}[\mathbb{E}[\epsilon_i^2 x_i x_i' \mid x_{1i} \cdots x_{Ki}]] \\
&= \mathbb{E}[x_i x_i' \mathbb{E}[\epsilon_i^2 \mid x_{1i} \cdots x_{Ki}]]
\end{aligned}$$

First since $\mathbb{E}[y_i \mid x_{1i}, \cdots, x_{Ki}] = \beta_0 + \beta_1 x_{1i} + \cdots + \beta_K x_{Ki}$ then $\mathbb{E}[\epsilon \mid x_{1i}, \cdots, x_{Ki}] = 0$. Then,

$$\begin{aligned}
\mathbb{E}[\epsilon_i^2 \mid x_{1i} \cdots x_{Ki}] &= \mathbb{E}[\epsilon_i^2 \mid x_{1i} \cdots x_{Ki}] - \mathbb{E}[\epsilon_i \mid x_{1i} \cdots x_{Ki}]^2 \\
&= \mathrm{Var}[\epsilon_i \mid x_{1i} \cdots x_{Ki}] \\
&= \mathrm{Var}[\epsilon_i \mid x_{1i} \cdots x_{Ki}] \\
&= \mathrm{Var}[y - \beta_0 - \beta_1 x_{1i} - \cdots \beta_K x_{Ki} \mid x_{1i} \cdots x_{Ki}] \\
&= \mathrm{Var}[y \mid x_i] + \mathrm{Var}[-\beta_0 - \beta_1 x_{1i} - \cdots \beta_K x_{Ki} \mid x_{1i} \cdots x_{Ki}] \text{ by iid}
\end{aligned}$$

$$= \sigma^2 + 0 = \sigma^2$$

as conditioning on $x_{1i} \cdots x_{Ki}$ means $-\beta_0 - \beta_1 x_{1i} - \cdots \beta_K x_{Ki}$ is constant

Thus, $\mathbb{E}[\epsilon_i^2 x_i x_i'] = \mathbb{E}[x_i x_i' \sigma^2] = \mathbb{E}[x_i x_i']\sigma^2$. As such,

$$\mathbb{E}[x_i x_i']^{-1}\mathbb{E}[\epsilon_i^2 x_i x_i']\mathbb{E}[x_i x_i']^{-1} = \mathbb{E}[x_i x_i']^{-1}\mathbb{E}[x_i x_i']\sigma^2\mathbb{E}[x_i x_i']^{-1} = \sigma^2\mathbb{E}[x_i x_i']^{-1}$$

Thus, we can say that

$$\sqrt{N}(\widehat{\beta} - \beta) \xrightarrow{\text{d}} \mathcal{N}(0, \sigma^2\mathbb{E}[x_i x_i']^{-1})$$

## b

### i

$$\begin{aligned}\widehat{\widehat{\beta}} &= (X'X)^{-1}X'\widetilde{y} \\ &= (X'X)^{-1}X'(y + u_y) \\ &= (X'X)^{-1}X'(X\beta + \epsilon + u_y) \\ &= \beta + (X'X)^{-1}X'\epsilon + (X'X)^{-1}X'u_y\end{aligned}$$

In order for $\widehat{\widehat{\beta}}$ to be a consistent estimator, we require that $u_y$ satisfy the same conditions as $\epsilon$ because in our consistency argument it is indistinguishable from $\epsilon$. Namely, we need that $\mathbb{E}[u_y] = 0$ and $\mathbb{E}[u_y x_k] = 0$ for each $k \in \{1, \ldots, K\}$.

### ii

Because of our assumptions on $u_y$, it is reasonable to write $\widetilde{\epsilon} \equiv \epsilon + u_y$ and utilize our formula for the asymptotic distribution under homoskedasticity, noting that,

$$\mathbb{E}[\widetilde{\epsilon}^2 \mid x_i] = \mathbb{E}[\epsilon^2 \mid x_i] + 2\mathbb{E}[\epsilon u_y \mid x_i] + \mathbb{E}[u_y^2 \mid x_i] = \mathbb{E}[\epsilon^2 \mid x_i] + \mathrm{Var}[u_y \mid x_i] = \sigma^2 + \sigma_u^2$$

Therefore, the asymptotic variance is larger as

$$\begin{aligned}\sqrt{N}(\widehat{\widehat{\beta}} - \beta) &\xrightarrow{\text{d}} \mathcal{N}(0, \mathbb{E}[\widetilde{\epsilon}^2]\mathbb{E}[x_i x_i']^{-1}) \\ &\equiv \mathcal{N}(0, (\sigma^2 + \sigma_u^2)\mathbb{E}[x_i x_i']^{-1})\end{aligned}$$

### iii

We should work with $y$ rather than $\widetilde{y}$ because the asymptotic variance of $\widehat{\beta} < \widehat{\widehat{\beta}}$ and they are both consistent estimators of $\beta$.

## c

### i

First we want that $\{u_{Ki}, x_{Ki}, y_i\}$ and $\mathbb{E}[u_k] = 0$. $\sum \widetilde{x}'\widetilde{x} = \sum xx' + 2\sum xu_k + \sum uu'$ but $\sum uu' \neq 0$

$$\begin{aligned}
\widehat{\widehat{\beta}} &= (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\widetilde{y} \\
&= (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'(y) \\
&= (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'(X\beta + \epsilon) \\
&= (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'(\widetilde{X}\beta - U_y\beta + \epsilon) \\
&= \beta - (\widetilde{X}'\widetilde{X})^{-1}(\widetilde{X}'U_y)\beta + (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\epsilon \\
&= \beta - (\widetilde{X}'\widetilde{X})^{-1}((X + U_y)'U_y)\beta + (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\epsilon \\
&= \beta - (\widetilde{X}'\widetilde{X})^{-1}(X'U_y + U_y'U_y)\beta + (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\epsilon \\
&= \beta - (\widetilde{X}'\widetilde{X})^{-1}(X'U_y) + (\widetilde{X}'\widetilde{X})^{-1}(U_y'U_y)\beta + (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}'\epsilon
\end{aligned}$$

Where $U_y = \begin{bmatrix} 0 & \cdots & 0 & u_{K_1} \\ 0 & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & u_{K_N} \end{bmatrix}$ Even with the assumption that $u_K$ is orthogonal to $\epsilon$ and $X$ and $x_k$ is orthogonal to $u_k$, we still get:

$$\widehat{\widehat{\beta}} = \beta + (\widetilde{X}'\widetilde{X})^{-1}(U_y'U_y)\beta$$

where $U_y'U_y$ is zero everywhere except the element in the $k-th$ row and $k-th$ column, which is equal to $\sum_{i=1}^{n} u_{K_i}^2$. Therefore our estimate for $\beta$ will not be consistent unless $\sum_{i=1}^{n} u_{K_i}^2$ converges to 0 which requires that $u_{K_i} \equiv 0$ for all $i$. Thus our measurement error term must be a degenerate random variable.

### ii

Yes, because then, $\widetilde{x_k} = x_k$ as the measurement error is a degenerate random variable so there is no difference between working with $\widetilde{x_k}$ and $x_k$.

### iii

We have shown prior that you need stronger assumptions to consistently estimate $\beta$, namely that $u_y \equiv 0$.

### iv

$$
\begin{aligned}
\operatorname*{plim}_{N\to\infty} \widetilde{\beta_1} &= \frac{\text{Cov}[\widetilde{x_{1i}}, y_i]}{\text{Var}[\widetilde{x_{1i}}]} \\
&= \frac{\text{Cov}[x_{1i}, y_i] + \text{Cov}[u_{1i}, y_i]}{\text{Var}[x_{1i}] + \text{Var}[u_{1i}] + 2\text{Cov}[x_{1i}, u_{1i}]} \\
&= \frac{\text{Cov}[x_{1i}, \beta_1 x_{1i}] + \text{Cov}[x_{1i}, \epsilon_{1i}] + \text{Cov}[u_{1i}, \beta_1 x_{1i}] + \text{Cov}[u_{1i}, \epsilon_i]}{\text{Var}[x_{1i}] + \text{Var}[u_{1i}] + 2\text{Cov}[x_{1i}, u_{1i}]} \\
&= \frac{\beta_1(\text{Var}[x_{1i}] + \text{Cov}[u_{1i}, x_{1i}]) + \text{Cov}[u_{1i}, \epsilon_i]}{\text{Var}[x_{1i}] + \text{Var}[u_{1i}] + 2\text{Cov}[x_{1i}, u_{1i}]}
\end{aligned}
$$

When measurement error is “random” then $\text{Cov}[u_{1i}, x_{1i}] = 0$ and $\text{Cov}[u_{1i}, \epsilon_{1i}] = 0$ so

$$
\operatorname*{plim}_{N\to\infty} \widetilde{\beta_1} = \frac{\beta_1 \text{Var}[x_{1i}]}{\text{Var}[x_{1i}] + \text{Var}[u_{1i}]}
$$

This result lines up with our result from part i earlier in this problem - unless the error term is a degenerate random variable where $\text{Var}[u_i] = 0$ then we still cannot consistently estimate $\beta_1$ with $\widetilde{\beta_1}$

# 3

## a

### i

Yes

### ii

Let $x \equiv (1, x_1, \ldots, x_K)$ and $\beta \equiv (\beta_0, \ldots, \beta_K)$. For convenience $Z$ is the scalar random variable $Z_K$ and $z_i$ is the realization of $Z_K$.

$$
\widehat{\text{Cov}}[z, y - x'\beta] = \frac{1}{N}\sum_i z_i(y_i - x_i'\beta) - \overline{z}(\overline{y} - \overline{x'}\beta)
$$

$$\iff \widehat{\text{Cov}}[z, y - x'\beta] - \text{Cov}[Z, y - X'\beta] =$$
$$\frac{1}{N}\sum_i z_i(y_i - x_i'\beta) - \mathbb{E}[Z(y - X'\beta)] + \mathbb{E}[Z]\mathbb{E}[y - X'\beta] - \overline{z}(\overline{y} - \overline{x}'\beta)$$

Moreover, the law of large numbers and the continuous mapping theorem tells us that the difference of the latter two terms converges in probability to 0, so for now we will just consider the first two. Once we have their distribution, we can apply Slutsky's theorem to conclude that we need not care about the latter two terms.

$$\begin{aligned}
\sqrt{N}(\widehat{\text{Cov}}[z, y - x'\beta] - \text{Cov}[Z, y - X'\beta]) &= \sqrt{N}\left(\frac{1}{N}\sum_i z_i(y_i - x_i'\beta) - \mathbb{E}[Z(y - X'\beta)]\right) \\
&= \sqrt{N}\left(\frac{1}{N}\sum_i z_i y_i - \mathbb{E}[ZY]\right) \\
&+ \sqrt{N}\left(\frac{1}{N}\sum_i z_i x_i'\beta - \mathbb{E}[ZX']\beta\right)
\end{aligned}$$

We can determine the distributional limit of the expression above by defining:

$$g(a, b) : \mathbb{R}^{1+K} \to \mathbb{R} \equiv a + b'\beta$$

where $a \in \mathbb{R}$ and $b \in \mathbb{R}^K$. This is a linear function and hence continuous with gradient:

$$Dg(a, b) = \begin{bmatrix} 1 & \beta' \end{bmatrix} \equiv Dg$$

Using the central limit theorem, leveraging the fact that functions of independent random variables are independent and hence $\{z_i x_i'\}_i$ and $\{z_i y_i\}_i$ are i.i.d., we can observe that:

$$\sqrt{N}\left(\sum_i z_i y_i - \mathbb{E}[ZY]\right) \xrightarrow{D} \mathcal{N}(0, \text{Var}[ZY])$$
$$\sqrt{N}\left(\sum_i z_i x_i - \mathbb{E}[ZX]\right) \xrightarrow{D} \mathcal{N}(0, \text{Var}[ZX])$$

Therefore by the delta method, and utilizing our LLN result from the beginning of the question, we find

$$\sqrt{N}(\widehat{\text{Cov}}[z, y - x'\beta] - \text{Cov}[Z, y - X'\beta]) \xrightarrow{d} \mathcal{N}(0, Dg\Sigma Dg')$$

With $\Sigma$ is the variance covariance matrix of $ZY$ and $ZX$ We can test our hypothesis that $\text{Cov}[Z, y - X'\beta] = 0$ by our standard methods of hypothesis testing on the normal letting our null be the statement above.

## b

### i

We want to identify $\beta$ where

$$Y = X'\beta + \epsilon$$

and $\mathbb{E}[\epsilon] = 0$ where $\beta = [\beta_0, \beta_1, \cdots, \beta_k]'$ and $X = [1, x_1, \cdots, x_k]$. Then, in 2SLS IV, where $Z = [1, x_1, \cdots, x_{k-1}, z_k]$ we regress

$$X = \delta Z + \epsilon_{fs}$$

Where $\delta = \begin{bmatrix} & \text{Identity}_{K\times K} & \\ \eta_0 & \cdots & \eta_K \end{bmatrix}$
$\delta Z$ are known as the fitted values of $X$ and we regress $Y$ on them. so we get

$$Y = (\delta Z)'\beta + \epsilon = Z'\delta'\beta + \epsilon$$

We know that $\mathbb{E}[\epsilon] = 0$ by OLS first order conditions and $\mathbb{E}[Z\epsilon] = 0$ by OLS first order and exogeneity conditions. Also,

$$\delta = \mathbb{E}[ZZ']^{-1}\mathbb{E}[ZX']$$

and

$$\delta'\beta = \mathbb{E}[ZZ']^{-1}\mathbb{E}[ZY]$$

Then,

$$\begin{aligned} \beta = (\delta')^{-1}\delta'\beta &= (\mathbb{E}[ZZ']^{-1}\mathbb{E}[ZX'])^{-1} \cdot \mathbb{E}[ZZ']^{-1}\mathbb{E}[ZY] \\ &= \mathbb{E}[ZX']^{-1}\mathbb{E}[ZZ'] \cdot \mathbb{E}[ZZ']^{-1} \cdot \mathbb{E}[ZY] \\ &= \mathbb{E}[ZX']^{-1}\mathbb{E}[ZY] \end{aligned}$$

Note that this expression was derived using the assumptions that make $Z$ a valid instrument - relevance, exclusion, exogeneity.

### ii

$$\begin{aligned}
\mathbb{E}[z\epsilon^{IV}] &= \mathbb{E}[zy - zx'\mathbb{E}[zx']^{-1}\mathbb{E}[zy]] \\
&= \mathbb{E}[zy] - \mathbb{E}[zx'\mathbb{E}[zx']^{-1}\mathbb{E}[zy]] \\
&= \mathbb{E}[zy] - \mathbb{E}[zx']\mathbb{E}[zx']^{-1}\mathbb{E}[zy] \\
&= \mathbb{E}[zy] - \mathbb{E}[zy] \\
&= 0
\end{aligned}$$

### iii

The main problem with this strategy is our definition of $\epsilon_{IV}$ assumes that $\mathbb{E}[z\epsilon^{IV}] = 0$. Thus, showing $\mathbb{E}[z\epsilon^{IV}] = 0$ is circular and assumed so we have not actually proved anything. We required this condition to even use $Z$ as an instrument.