

# ECON 710A - Problem Set 1

Alex von Hafften\*

2/1/2020

1. Suppose  $(Y, X')'$  is a random vector with  $Y = X'\beta_0 \cdot U$  where  $E[U|X] = 1$ ,  $E[XX']$  is invertible, and  $E[Y^2 + ||X||^2] < \infty$ . Furthermore, suppose that  $\{(Y_i, X'_i)'\}_{i=1}^n$  is a random sample from the distribution of  $(Y, X')'$  where  $\frac{1}{n} \sum_{i=1}^n X_i X'_i$  is invertible and let  $\hat{\beta}$  be the OLS estimator, i.e.,  $\hat{\beta} = (\frac{1}{n} \sum_{i=1}^n X_i X'_i)^{-1} \frac{1}{n} \sum_{i=1}^n X_i Y_i$ .

- (i) Interpret the entries of  $\beta_0$  in this model?

The entries of  $\beta_0$  are the average marginal effects of the elements of  $X$  on  $Y$  conditional on  $X$ . It is the average effect because  $E[U|X] = 1$ . For example, the  $i$ th element of  $\beta_0$  is the average marginal effects of the  $i$ th element of  $\beta_0$ .

$$E[Y|X] = E[X'\beta_0 \cdot U|X] = X'\beta_0 E[U|X] = X'\beta_0 \implies \frac{\partial E[Y|X]}{\partial X} = \beta_0$$

- (ii) Show that  $Y = X'\beta_0 + \bar{U}$  where  $E[\bar{U}|X] = 0$ .

Define  $\bar{U} := X'\beta_0 \cdot (U - 1)$ .

$$X'\beta_0 + \bar{U} = X'\beta_0 + X'\beta_0 \cdot (U - 1) = X'\beta_0 \cdot (1 + U - 1) = X'\beta_0 \cdot U = Y$$

$$E[\bar{U}|X] = E[X'\beta_0 \cdot (U - 1)|X] = X'\beta_0 \cdot (E[U|X] - 1) = X'\beta_0 \cdot (1 - 1) = 0$$

- (iii) Show that  $E[X(Y - X'\beta)] = 0$  iff  $\beta = \beta_0$  and use this to derive OLS as a method of moments estimator. ( $\implies$ ):

$$\begin{aligned} E[X(Y - X'\beta)] &= 0 \\ \implies E[X(X'\beta_0 + \bar{U} - X'\beta)] &= 0 \\ \implies E[XX'\beta_0] + E[X \cdot \bar{U}] - E[XX'\beta] &= 0 \\ \implies \beta_0 E[XX'] + E[E[X \cdot \bar{U}|X]] - \beta E[XX'] &= 0 \\ \implies \beta_0 E[XX'] + E[X \cdot E[\bar{U}|X]] - \beta E[XX'] &= 0 \\ \implies \beta_0 E[XX'] + E[X \cdot 0] - \beta E[XX'] &= 0 \\ \implies \beta_0 E[XX'] - \beta E[XX'] &= 0 \\ \implies \beta_0 &= \beta \end{aligned}$$

---

\*I worked on this problem set with a study group of Michael Nattinger, Andrew Smith, and Ryan Mather. I also discussed problems with Emily Case, Sarah Bass, and Danny Edgel.

( $\Leftarrow$ ):

$$\begin{aligned}
\beta &= \beta_0 \\
\Rightarrow E[X(Y - X'\beta)] &= E[X(X'\beta_0 + \bar{U} - X'\beta)] \\
&= E[X(X'\beta + \bar{U} - X'\beta)] \\
&= E[X\bar{U}] \\
&= E[E[X \cdot \bar{U}|X]] \\
&= E[X \cdot E[\bar{U}|X]] \\
&= E[X \cdot 0] \\
&= 0
\end{aligned}$$

Therefore,  $\beta_0$  is identified by the moment function  $g((Y, X')', \beta) = X(Y - X'\beta)$ . Thus, for the random sample  $\hat{\beta}^{MM}$  is unique solution to  $\frac{1}{n} \sum_{i=1}^n g((Y_i, X_i')', \beta) = 0$ :

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n g((Y_i, X_i')', \hat{\beta}^{MM}) &= 0 \\
\frac{1}{n} \sum_{i=1}^n X_i(Y_i - X_i'\hat{\beta}^{MM}) &= 0 \\
\frac{1}{n} \sum_{i=1}^n X_i Y_i - \frac{1}{n} \sum_{i=1}^n X_i X_i' \hat{\beta}^{MM} &= 0 \\
\frac{1}{n} \sum_{i=1}^n X_i X_i' \hat{\beta}^{MM} &= \frac{1}{n} \sum_{i=1}^n X_i Y_i \\
\hat{\beta}^{MM} &= \left( \frac{1}{n} \sum_{i=1}^n X_i X_i' \right)^{-1} \left( \frac{1}{n} \sum_{i=1}^n X_i Y_i \right)
\end{aligned}$$

Notice that  $\hat{\beta}^{MM} = \hat{\beta}^{OLS}$ .

(iv) Show that the OLS estimator is conditionally unbiased, i.e., that  $E[\hat{\beta}|X_1, \dots, X_n] = \beta_0$ .

$$\begin{aligned}
E[\hat{\beta}|X_1, \dots, X_n] &= E\left[\left(\frac{1}{n} \sum_{i=1}^n X_i X_i'\right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n X_i Y_i\right) \middle| X_1, \dots, X_n\right] \\
&= \left(\frac{1}{n} \sum_{i=1}^n X_i X_i'\right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n X_i E[Y_i|X_i]\right) \\
&= \left(\frac{1}{n} \sum_{i=1}^n X_i X_i'\right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n X_i E[X_i' \beta_0 + \bar{U}_i|X_i]\right) \\
&= \left(\frac{1}{n} \sum_{i=1}^n X_i X_i'\right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n X_i (X_i' \beta_0 + E[\bar{U}_i|X_i])\right) \\
&= \left(\frac{1}{n} \sum_{i=1}^n X_i X_i'\right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n X_i X_i' \beta_0\right) \\
&= \beta_0 \left(\frac{1}{n} \sum_{i=1}^n X_i X_i'\right)^{-1} \left(\frac{1}{n} \sum_{i=1}^n X_i X_i'\right) \\
&= \beta_0
\end{aligned}$$

(v) Show that the OLS estimator is consistent, i.e., that  $\hat{\beta} \rightarrow_p \beta_0$  as  $n \rightarrow \infty$ .

By the law of large numbers in  $\mathcal{L}^1$ ,

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n X_i X_i' &\rightarrow_p E[XX'] \\
\frac{1}{n} \sum_{i=1}^n X_i Y_i &\rightarrow_p E[XY]
\end{aligned}$$

Since  $E[XX']$  is invertible, by the continuous mapping theorem,

$$\begin{aligned}
\hat{\beta} &\rightarrow_p (E[XX'])^{-1} E[XY] \\
&= (E[XX'])^{-1} E[X(X' \beta_0 + \bar{U})] \\
&= \beta_0 (E[XX'])^{-1} E[XX'] + E[X \cdot \bar{U}] \\
&= \beta_0
\end{aligned}$$

2. Let  $X$  be a random variable with  $E[X^4] < \infty$  and  $E[X^2] > 0$ . Furthermore, let  $\{X_i\}_{i=1}^n$  be a random sample from the distribution of  $X$ .

(i) For which of the following four statistics can you use the law of large numbers and continuous mapping theorem to show convergence in probability as  $n \rightarrow \infty$ ,

(a)  $\frac{1}{n} \sum_{i=1}^n X_i^3$

Note that  $E[X^4] < \infty \implies E[X^3] < \infty$  and  $E[-X^3] < \infty$ , so  $E[|X^3|] < \infty$ . By the law of large numbers,

$$\frac{1}{n} \sum_{i=1}^n X_i^3 \rightarrow_p E[X^3]$$

(b)  $\max_{1 \leq i \leq n} X_i$

We cannot use the law of large numbers and the continuous mapping theorem to show convergence in probability for this statistic.

(c)  $\frac{\sum_{i=1}^n X_i^3}{\sum_{i=1}^n X_i^2}$

By the law of large numbers,

$$\frac{1}{n} \sum_{i=1}^n X_i^3 \rightarrow_p E[X^3] < \infty$$

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \rightarrow_p E[X^2] \in (0, \infty)$$

Because  $E[X^4] < \infty$ . By the continuous mapping theorem,

$$\frac{\sum_{i=1}^n X_i^3}{\sum_{i=1}^n X_i^2} = \frac{\frac{1}{n} \sum_{i=1}^n X_i^3}{\frac{1}{n} \sum_{i=1}^n X_i^2} \rightarrow_p \frac{E[X^3]}{E[X^2]}$$

(d)  $1\{\frac{1}{n} \sum_{i=1}^n X_i > 0\}$

Note that  $E[X^4] < \infty \implies E[|X|] < \infty$ . By the law of large numbers,

$$\frac{1}{n} \sum_{i=1}^n X_i \rightarrow_p E[X]$$

This indicator function is locally continuous for all neighbors that do not include zero:

- If  $E[X] > 0$ ,  $1\{\frac{1}{n} \sum_{i=1}^n X_i > 0\} \rightarrow_p 1$  by the CMT.
- If  $E[X] < 0$ ,  $1\{\frac{1}{n} \sum_{i=1}^n X_i > 0\} \rightarrow_p 0$  by the CMT.
- If  $E[X] = 0$ , the indicator function is not locally continuous, so we cannot apply the CMT.

(ii) For which of the following three statistics can you use the central limit theorem and continuous mapping to show convergence in distribution as  $n \rightarrow \infty$ ,

(a)  $W_n := \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i^2 - E[X_1^2])$

Note that  $E[X^4] < \infty \implies E[X^2] < \infty$ . Furthermore,  $Var(X_1^2) = E[X_1^4] - (E[X_1^2])^2 < \infty$  By the Lindeberg-Levy Lemma (CLT),

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i^2 - E[X_1^2]) \rightarrow_d N(0, E[X_1^4] - (E[X_1^2])^2)$$

(b)  $W_n^2$

By the continuous mapping theorem,  $W_n^2$  converges in distribution to a scaled chi-squared distribution.

(C)  $\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i^2 - \overline{X_n^2})$  where  $\overline{X_n^2} = \frac{1}{n} \sum_{i=1}^n X_i^2$

We cannot use the Lindeberg-Levy CLT or CMT to find the asymptotic distribution because they require that the statistic is shifted by its population expectation not its sample average.

(iii) Show that  $\max_{1 \leq i \leq n} X_i \rightarrow_p 1$  if  $X \sim \text{uniform}(0, 1)$ .

Define  $M_n := \max_{1 \leq i \leq n} X_i$ . Choose  $\varepsilon > 0$ . If  $\varepsilon \geq 1$ ,  $\Pr(|M_n - 1| \leq \varepsilon) = 1 \geq 1 - \varepsilon$ , so  $M_n$  converges in probability to 1. If  $\varepsilon < 1$ , choose  $N \geq \frac{\ln(\varepsilon)}{\ln(1-\varepsilon)}$ . For  $n > N$ ,

$$\begin{aligned} \Pr(|M_n - 1| \leq \varepsilon) &= \Pr(1 - M_n \leq \varepsilon) \\ &= \Pr(M_n > 1 - \varepsilon) \\ &= 1 - \Pr(M_n < 1 - \varepsilon) \\ &\geq 1 - \prod_{i=1}^n \Pr(X_i < 1 - \varepsilon) \\ &= 1 - (1 - \varepsilon)^n \\ &\geq 1 - \varepsilon \end{aligned}$$

Therefore,  $M_n$  converges in probability to 1.<sup>1</sup>

---

<sup>1</sup>Derivations for the last inequality:

$$\begin{aligned} 1 - (1 - \varepsilon) &\geq 1 - \varepsilon \\ \iff \varepsilon &\geq (1 - \varepsilon)^n \\ \iff \ln(\varepsilon) &\geq n \ln(1 - \varepsilon) \\ \iff n &\geq \frac{\ln(\varepsilon)}{\ln(1 - \varepsilon)} \\ \iff n &\geq N \end{aligned}$$

(iv) Show that  $\Pr(\max_{1 \leq i \leq n} X_i > M) \rightarrow 1$  for any  $M \geq 0$  if  $X \sim \text{exponential}(1)$ .

Choose  $\varepsilon > 0$ . Let  $N \geq \frac{\ln(\varepsilon)}{\ln(1-e^{-M})}$ . For  $n > N$ ,

$$\begin{aligned} \left| \Pr \left( \max_{1 \leq i \leq n} X_i > M \right) - 1 \right| &= 1 - \Pr \left( \max_{1 \leq i \leq n} X_i > M \right) \\ &= \Pr \left( \max_{1 \leq i \leq n} X_i \leq M \right) \\ &= [\Pr(X_i \leq M)]^n \\ &= [1 - e^{-M}]^n \\ &\leq \varepsilon \end{aligned}$$

Thus,  $\Pr(\max_{1 \leq i \leq n} X_i > M) \rightarrow 1$ .<sup>2</sup>

---

<sup>2</sup>Derivations of the last inequality:

$$\begin{aligned} (1 - e^{-M})^n &\leq \varepsilon \\ \iff n \ln(1 - e^{-M}) &\leq \ln(\varepsilon) \\ \iff n > N \geq \frac{\ln(\varepsilon)}{\ln(1 - e^{-M})} \end{aligned}$$

3. Suppose that  $\{X_i\}_{i=1}^n$  is an iid sequence of  $N(0, 1)$  random variables. Let  $W$  be independent of  $\{X_i\}_{i=1}^n$  with  $\Pr(W = 1) = \Pr(W = -1) = 1/2$ . Let  $Y_i = X_i W$ .

(i) Show that  $\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \rightarrow_d N(0, 1)$  as  $n \rightarrow \infty$ .

Since  $\{X_i\}_{i=1}^n$  is an iid sequence of  $N(0, 1)$  random variables,  $E[X_i^2] = \text{Var}(X_i) = 1 < \infty$  and  $E[X_i] = 0$ . By Lindeberg-Levy,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \rightarrow_d N(0, 1)$$

(ii) Show that  $\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i \rightarrow_d N(0, 1)$  as  $n \rightarrow \infty$ .

Notice that a single draw of  $W$  determines whether the sign of  $X_i$  are flipped.

- If  $W = 1$ ,  $Y_i = X_i$  for all  $i = 1, \dots, n$ . Thus, by (i),  $\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i \rightarrow_d N(0, 1)$ .
- If  $W = -1$ ,  $Y_i = -X_i$  for all  $i = 1, \dots, n$ . Since  $N(0, 1)$  is symmetric,  $-X_i$  are also distributed iid  $N(0, 1)$ . Thus, by (i),  $\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i \rightarrow_d N(0, 1)$ .

(iii) Show that  $\text{Cov}(X_i, Y_i) = 0$ .

$$E[W] = (1) \Pr(W = 1) + (-1) \Pr(W = -1) = (1)(1/2) + (-1)(1/2) = 0$$

$$E[Y_i] = E[X_i W] = E[X_i] E[W] = 0$$

$$\text{Cov}(X_i, Y_i) = E[(X_i - E[X_i])(Y_i - E[Y_i])] = E[X_i Y_i] = E[X_i^2 W] = E[X_i^2] E[W] = (1)(0) = 0$$

(iv) Does  $V := \frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i, Y_i)' \rightarrow_d N(0, I_2)$  as  $n \rightarrow \infty$ ?

No. Conditional on  $W = 1$ ,  $X_i$  and  $Y_i$  are perfectly correlated, so  $\text{Cov}(X_i, Y_i) = 1$ . Conditional on  $W = -1$ ,  $X_i$  and  $Y_i$  are perfectly inversely correlated, so  $\text{Cov}(X_i, Y_i) = -1$ . Since  $\Pr(W = 1) = \Pr(W = -1) = 1/2$ , the unconditional covariance is zero, but  $V$  cannot converge to a joint normal distribution with zero covariance because the covariance is either always going to one or minus one depending on the value of  $W$ .

(v) How does this exercise related to the Cramer-Wold device introduced in lecture 2?

The Cramer-Wold device requires a sequence of random vectors  $\{W_n\}_{n=1}^\infty$  converge in distribution to the random vector  $W$  iff  $t'W_n$  converge in distribution to  $t'W$  for any nonrandom vector  $t$  with  $\|t\| = 1$ . In this exercise, there is no asymptotic distribution that works for all such  $t$  vectors.