Slides for the course

Statistics and econometrics

Part 8: Properties of the multiple regression model

European University Institute

Andrea Ichino

October 12, 2020

Outline

Finite sample properties

Unbiasedness

Good and bad habits concerning control variables

Efficiency

The Gauss-Markov theorem for the PMRF

Asymptotic properties

How to write $\hat{\beta}$ so that we can apply the asymptotic results

Consistency

Asymptotic normality

Section 1

Finite sample properties

Subsection 1

Unbiasedness

Unbiasedness of the OLS estimator of the PMRF

The proof of unbiasedness is similar to the simple regression case;

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'Y \\ &= (X'X)^{-1}X'(X\beta + U) \\ &= (X'X)^{-1}X'X\beta + (X'X)^{-1}X'U \\ &= \beta + (X'X)^{-1}X'U\end{aligned} \tag{1}$$

Unbiasedness of the OLS estimator (cont.)

Taking the expectation

$$E(\hat{\beta}|X) \quad = \quad \beta + (X'X)^{-1}X'E(U|X) = \beta \tag{2}$$

which follows from the assumption:

- MLR 4: Conditioning on the entire X each u_i has zero mean

$$E(U|X) = 0 \tag{3}$$

The considerations we made about this assumption for the simple regression case hold here as well.

Omitted variable bias and irrelevant regressors

Suppose that we have omitted a variable Z which we think should be included for the CIA to hold. Thus:

$$U = Z\gamma + V \tag{4}$$

The expected value of the estimator for β is:

$$\begin{aligned} E(\hat{\beta}|X) &= \beta + (X'X)^{-1}X'E[U|X] \\ &= \beta + (X'X)^{-1}X'E[Z|X]\gamma + (X'X)^{-1}X'E[V|X] \\ &= \beta + (X'X)^{-1}X'E[Z|X]\gamma \end{aligned} \tag{5}$$

Omitted variable bias and irrelevant regressors (cont.)

The omission of Z generates a bias if

- $E(Z|X) \neq 0$;
- Z has a non-zero effect γ on the outcome.

The sign of the bias is easy to determine if X and Z include only one variable each. Not obvious otherwise.

Subsection 2

Good and bad habits concerning control variables

Are control variables always useful?

It may not be a good idea to add controls, specifically if they are themselves causally affected by the main variable of interest.

We should control for omitted variables, when the CIA holds:

$$Y = X\beta + Z\gamma + U \tag{6}$$

and if we run

$$Y = X\beta + V \tag{7}$$

we get a biased and inconsistent estimate

$$E(\hat{\beta}) = \beta + (X'X)^{-1}E[X'Z]\gamma \tag{8}$$

In this case, If we observe Z we should include it.

Controlling to increase precision

It may be a good idea to include Z even if $E[X'Z] = 0$, if the goal is not to avoid a bias but to increase efficiency.

Consider a random experiment in which a training program X is randomly assigned to estimate its effect on future earnings Y. The causal PRF is

$$Y = X\beta + U \tag{9}$$

Consider a set of predetermined demografic characteristics D, which by random assignment of X are orthogonal to X, but have a causal effect on Y.

Controlling to increase precision (cont.)

If we run the PMRF

$$Y = X\beta + D\gamma + V \tag{10}$$

the OLS estimator for β is:

$$\hat{\beta} = (X'MX)^{-1}X'MY \tag{11}$$

where $M = I - D(D'D)^{-1}D'$. Note that $MX = X$ because $D(D'D)^{-1}D'X = 0$: D and X are orthogonal. But

$$\sigma_U^2 = Var(U) = \gamma^2 Var(D) + Var(V) > Var(V) = \sigma_V^2$$

and therefore β is estimated more precisely using 10.

A first case of misleading control variable

Now suppose that D is instead the occupation chosen by the subject after training:

- $D = 1$ white collar
- $D = 0$ blue collar

The training program

- $X = 1$ trained
- $X = 0$ not trained

increases the chance of a white collar occupation.

Note that X is randomly assigned in the population, but not within the occupational group!

A first case of misleading control variable (cont.)

If we estimate

$$Y = X\beta + U \tag{12}$$

we get an unbiased and consistent estimate of β which is the overall causal effect of training, including the effect that runs through the occupational choice.

In this case, it would be a bad idea to run

$$Y = X\beta + D\gamma + V \tag{13}$$

unless the efficiency gain were huge.

A first case of misleading control variable (cont.)

If we did run 13, we would get

$$\hat{\beta} = (X'MX)^{-1}X'MY \neq (X'X)^{-1}X'Y \tag{14}$$

To understand the bias note that 13 is equivalent to comparing trained and not trained for given occupation.

In what follows

- D_0 is the occupation you choose given that you were not trained;
- D_1 is the occupation you choose given that you were trained;
- Y_0 earnings you get given that you were not trained;
- Y_1 earnings you get given that you were trained.

A first case of misleading control variable (cont.)

$$
\begin{aligned}
E(Y|X &= 1, D = 1) - E(Y|X = 0, D = 1) \qquad (15)\\
&= E(Y_1|X = 1, D_1 = 1) - E(Y_0|X = 0, D_0 = 1)\\
&= E(Y_1|D_1 = 1) - E(Y_0|D_0 = 1)\\
&= E(Y_1 - Y_0|D_1 = 1) + [E(Y_0|D_1 = 1) - E(Y_0|D_0 = 1)]
\end{aligned}
$$

where the second equality derives from the joint independence of Y_1, D_1, Y_0, D_0 from X.

The bias is represented by the selection effect $[E(Y_0|D_1 = 1) - E(Y_0|D_0 = 1)]$ which reflects the fact that composition of the pool of white collar workers has changed because of training even in the counterfactual case of no training.

A second case of misleading control variable

Let's now go back to the case in which the true causal PRF is

$$Y = \alpha + X\beta + Z\gamma + U \tag{16}$$

Z is predetermined ability, X is education, Y is earnings, but we observe only a measure $\tilde{Z}$ of Z taken after education has finshed (e.g. IQ):

$$\tilde{Z} = \pi_0 + X\pi_1 + Z\pi_2 + e \tag{17}$$

A second case of misleading control variable (cont.)

Substituting 17 in 16 we get

$$Y = \left(\alpha - \gamma\frac{\pi_0}{\pi_2}\right) + \left(\beta - \gamma\frac{\pi_1}{\pi_2}\right) X + \frac{\gamma}{\pi_2}\tilde{Z} + U \tag{18}$$

And OLS would be biased and inconsistent for the causal parameters.

Depending on assumptions, we can still say something on β.

But, clearly, timing is crucial in the choice of control variables.

Subsection 3

Efficiency

Variance of the OLS estimator of the PMRF

Consider first the simple case of homoschedasticity:

- MLR 5: The variance-covariance matrix of U is

$$Var(U|X) = E(UU'|X) = \sigma^2 I_n \tag{19}$$

where I_n is the $n \times n$ identity matrix.

Note that this assumption has two important components:

- The variance of u_i should not depend on any variable x_j.
- The covariance between u_t and u_s should be zero for any $t \neq s$. Note that this condition:
  - typically does not hold in time series because of serial correlation;
  - it is unlilkely to hold even in a cross section, and you will learn how to deal with it.

Variance-covariance matrix

$$\begin{aligned} Var(\hat{\beta}|X) &= E[(\hat{\beta}-\beta)(\hat{\beta}-\beta)'|X] \qquad (20) \\ &= E[(X'X)^{-1}X'UU'X(X'X)^{-1}|X] \\ &= (X'X)^{-1}X'E[UU'|X]X(X'X)^{-1} \\ &= (X'X)^{-1}X'\sigma^2 I_n X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1} \end{aligned}$$

which is a $(k+1)\times(k+1)$ matrix.

The OLS estimator is more precise:

- the smaller is the variance of the unobservable components.
- the larger is the total variation in the observable regressors X.
- the smaller is the collinearity among observables X.

An alternative useful way to write the variance

Following Wooldridge (Appendix to Chapter 3):

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)} \tag{21}$$

- $SST_j = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2$ is the total sample variation of x_j.
- R_j^2 is the R-squared of the regression of x_j on the other regressors.

This expression highlights three important components of the OLS:

- variance of the unobservable components;
- variance of the regressors;
- multicollinearity between the regressors.

Is it always a good idea to include more regressors?

An unbiased estimator of σ^2

We want to show that

$$\hat{\sigma}^2 = \frac{1}{n-k-1} \hat{U}'\hat{U} \tag{22}$$

is unbiased for σ^2. Note that for $k = 1$ this is the same estimator that we have studied for the simple linear regression case.

$$\begin{aligned} \hat{U} &= Y - X\hat{\beta} \\ &= Y - X(X'X)^{-1}X'Y \\ &= MY = M(X\beta + U) = MU \end{aligned} \tag{23}$$

Where $M = I - X(X'X)^{-1}X'$ is a symmetric and idempotent matrix:

- $M' = M$
- $M'M = M$
- $MX = 0$
- $MY = MU$

An unbiased estimator of σ^2 (cont.)

$$\begin{aligned} E[\hat{U}'\hat{U}|X] &= E[U'M'MU|X] && (24)\\ &= E[tr(U'MU)|X] && \text{because a scalar is equal to its trace}\\ &= E[tr(MUU')|X] && \text{because of the property of the trace}\\ &= tr(ME[UU'|X]) = tr(M)\sigma^2 = (n-k-1)\sigma^2 \end{aligned}$$

which proves the result. The last equality follows because

$$\begin{aligned} tr(M) &= tr(I_n) - tr(X(X'X)^{-1}X') && (25)\\ &= tr(I_n) - tr((X'X)^{-1}X'X)\\ &= tr(I_n) - tr(I_{k+1}) = n-k-1 \end{aligned}$$

In a sample of size n that we use to estimate $k+1$ parameters β, we are left with only $n-k-1$ “degrees of freedom" to estimate σ^2.

Subsection 4

The Gauss-Markov theorem for the PMRF

The Gauss-Markov theorem

Under the assumptions

- MLR 1: The PRF is linear in the parameters:

$$Y = X\beta + U \tag{26}$$

- MLR 2: The n observations on Y and X are a random sample

$$y_i = X_i\beta + u_i \tag{27}$$

- MLR 3: There is no collinearity and X has full rank equal to (k+1).
- MLR 4: Conditioning on the entire X each u_i has zero mean

$$E(U|X) = 0 \tag{28}$$

- MLR 5: The variance-covariance matrix of unobservables is

$$Var(U|X) = E(UU'|X) = \sigma^2 I_n \tag{29}$$

Then the OLS estimator $\hat{\beta}$ is the best linear unbiased estimator.

Proof of the Gauss Markov theorem

Consider a generic alternative linear unbiased estimator

$$\tilde{\beta} = A'Y \tag{30}$$

where A is a $n \times (k+1)$ matrix. Linearity in Y implies that A is a function of X but cannot be a function of Y. Since $\tilde{\beta}$ is unbiased it must be the case that:

$$\begin{aligned} E(\tilde{\beta}|X) &= A'X\beta + A'E(U|X) && (31) \\ &= A'X\beta && \text{because } E(U|X) = 0 \\ &= \beta \end{aligned}$$

and therefore $A'X = I_{k+1}$ and $\tilde{\beta}$ characterizes the class of linear (in Y) unbiased estimators.

Proof of the Gauss Markov theorem (cont)

The variance of $\tilde{\beta}$ is:

$$\begin{aligned} Var(\tilde{\beta}|X) &= E[(\tilde{\beta}-\beta)(\tilde{\beta}-\beta)'|X] \\ &= E[A'UU'A|X] \\ &= \sigma^2(A'A) \end{aligned} \tag{32}$$

Proof of the Gauss Markov theorem (cont)

$$\begin{aligned} Var(\tilde{\beta}|X) &- Var(\hat{\beta}|X) = \sigma^2[A'A - (X'X)^{-1}] \qquad (33) \\ &= \sigma^2[A'A - A'X(X'X)^{-1}X'A] \quad \text{because } A'X = I_{k+1} \\ &= \sigma^2 A'[I_n - X(X'X)^{-1}X']A \\ &= \sigma^2 A'MA \end{aligned}$$

Since M is symmetric and idempotent, $A'MA$ is positive semidefinite for any conformable A, which proves the result.

The OLS-MM estimator $\hat{\beta}$ has the smallest variance in the class of linear unbiased estimators.

Section 2

Asymptotic properties

Subsection 1

How to write $\hat{\beta}$ so that we can apply the asymptotic results

Rewriting $\hat{\beta}$

Given the PRF:

$$Y = X\beta + U$$

the OLS estimator can be written as:

$$\begin{aligned}
\hat{\beta} &= (X'X)^{-1}X'Y \\
&= \beta + (X'X)^{-1}X'U \\
&= \beta + \left(\frac{1}{n}\sum_{i=1}^{n} X_i'X_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} X_i'u_i\right)
\end{aligned} \tag{34}$$

where X_i is the $1 \times k+1$ vector of the regressors observed for subject i. s

Rewriting $\hat{\beta}$ (cont.)

Rearranging (34) we get:

$$\sqrt{n}(\hat{\beta} - \beta) = \left(\frac{1}{n} \sum_{i=1}^{n} X_i' X_i \right)^{-1} \frac{1}{\sqrt{n}} \left(\sum_{i=1}^{n} X_i' u_i \right) \tag{35}$$

We can now apply asymptotic results to (35) to derive the asymptotic properties of the estimator.

Subsection 2

Consistency

Proof of consistency

Consistency can be demonstrated using in (34)

- the Continuous Mapping Theorem;
- the Law of Large Numbers.

Exploiting the fact that probability limits go through continuous functions and substituting population moment to sample moment, $\hat{\beta}$ converges in probability to:

$$\begin{aligned}\hat{\beta} &\xrightarrow{p} \beta + (E(X_i'X_i))^{-1}(E(X_i'u_i)) \\ &= \beta\end{aligned} \tag{36}$$

where the last equality holds because $E(X_i'u_i) = 0$ by definition of the PRF.

Subsection 3

Asymptotic normality

The asymptotic distribution of $\hat{\beta}$

Applying Slutsky to 35, $\sqrt{n}(\hat{\beta} - \beta)$ has the same distribution of

$$(E(X_i'X_i))^{-1}\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i'u_i\right) \tag{37}$$

Since

$$\frac{1}{n}\sum_{i=1}^{n} X_i'u_i \qquad \xrightarrow{p} \qquad E(X_i'u_i) = 0 \tag{38}$$

then

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i'u_i\right) \qquad \xrightarrow{d} \qquad \text{Normal}(0, E(X_i'X_iu_i^2)) \tag{39}$$

because it is a root-n blown up and centered sample moment, for which we can use the Central Limit Theorem.

Note that $E(X_i'X_iu_i^2)$ is a $(k+1) \times (k+1)$ matrix.

The asymptotic distribution of $\hat{\beta}$ (cont.)

It then follows that:

$$\sqrt{n}(\hat{\beta}-\beta) \xrightarrow{d} \text{Normal}(0, [E(X_i'X_i)^{-1}][E(X_i'X_i)u_i^2][E(X_i'X_i)^{-1}]) \tag{40}$$

where $[E(X_i'X_i)^{-1}][E(X_i'X_i)u_i^2][E(X_i'X_i)^{-1}]$ is again a $(k+1)\times(k+1)$ matrix.

Note that we have not assumed homoscedasticity.

We have only assumed identically and independently distributed random observations, which is what CLT and LLN require to hold.

These asymptotic standard errors are called “Robust", or “Huber - Eicker - White" standard errors (White (1980)) and provide accurate hypothesis tests in large sample with minimal assumptions.

The asy-distribution of $\hat{\beta}$ with homoschedasticity

If we are willing to assume homoschedasticity then

$$E(u_i^2|X) = \sigma^2 \tag{41}$$

and the “Robust" variance covariance matrix in 40 simplifies to

$$\begin{aligned} [E(X_i'X_i)^{-1}][E(X_i'X_i)u_i^2][E(X_i'X_i)^{-1}] &= \\ [E(X_i'X_i)^{-1}][E(X_i'X_iE(u_i^2|X)][E(X_i'X_i)^{-1}] &= \\ \sigma^2[E(X_i'X_i)^{-1}][E(X_i'X_i)][E(X_i'X_i)^{-1}] &= \\ \sigma^2[E(X_i'X_i)^{-1}] & \end{aligned} \tag{42}$$

and

$$\sqrt{n}(\hat{\beta} - \beta) \quad \xrightarrow{d} \quad \text{Normal}(0, \sigma^2[E(X_i'X_i)^{-1}]) \tag{43}$$