Slides for the course

Statistics and econometrics

Part 7: The multiple regression model

European University Institute

Andrea Ichino

October 8, 2014

Outline

Section 1

The Conditional Independence Assumption

Controlling for observables and causal PRF

Even if a non causal PRF is in any case a useful tool, our main goal is to estimate a PRF that can be interpreted causally.

We now consider cases in which it is reasonable to make the *Conditional Independence Assumption*.

This assumption says that controlling for a set of observable variables, the PRF may have a causal interpretation.

In this part of the course we want to understand:

- the meaning of this assumption;
- how it relates to multiple regression.

Later in the econometric sequence you will see other assumptions that allow you to estimate consistently causal parameters.

Training and earnings

Consider the causal model:

$$y = \mu + \tau_1 x_1 + \nu, \tag{1}$$

where y is earnings and x_1 is training, and the population regression function

$$y = \beta_0 + \beta_1 x_1 + u \tag{2}$$

By definition of the PRF, β_1 is

$$\begin{aligned} \beta_1 = \frac{Cov(y, x_1)}{V(x_1)} &= \frac{Cov(\mu + \tau_1 x_1 + \nu, x_1)}{V(x_1)} \\ &= \tau_1 + \frac{Cov(\nu, x_1)}{V(x_1)} \end{aligned} \tag{3}$$

which, is not equal to τ_1 if X_1 is not randomly assigned and thus

$$Cov(\nu, x_1) \neq 0.$$

The Conditional Independence Assumption

Suppose that there exist a variable X_2 such that:

- ν can be decomposed (in the PRF sense) into

$$\nu = \delta_2 X_2 + \omega \tag{4}$$

  where, by definition of the PRF, $E(X_2\omega)$ and $E(\omega) = 0$;

- and ω is such that, for each value of X_2, X_1 is randomly assigned,

$$E(X_1\omega | X_2) = 0 \tag{5}$$

  that is, X_2 is the only reason why ν and X_1 are correlated, which is an identifying assumption (not valid by construction).

This is for example the setting of Black, Smith, Berger and Noel, “Is the threat of reemployment services more effective than the services themselves” AER, September 2003.

The Conditional Independence Assumption (cont.)

Under the CIA, we will now show that the PRF:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \tag{6}$$

is causal for the effect of x_1 on y, i.e.:

$$\beta_1 = \tau_1$$

Note that $\beta_2 = \delta_2$ and does not have any causal interpretation.

It is crucial to understand that the CIA is a solution as long as $\delta_2 X_2$ captures all the information in ν such that controlling for it X_1 is randomly assigned

A detail on which we will come back in the spring course: meaning of "fully saturated PRF".

Section 2

The Population Multiple Regression Function

The Population Multiple Regression Function

Consider the population regression of y on both x_1 and x_2:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \tag{7}$$

where

$$(\beta_0, \beta_1, \beta_2) = arg \min_{b_0, b_1, b_2} E\left[(y - b_0 - b_1 x_1 - b_2 x_2)^2\right] \tag{8}$$

i.e. where the population parameters are defined to minimize the square of the difference between y and the PMRF itself.

We want to show that if the CIA holds β_1 is the causal effect of x_1 on y

The same is true symmetrically if we are interested in the effect of x_2.

If we can estimate consistently the PMRF, we get consistent estimates of the causal parameters of interest as well.

The coefficients of the PMRF

The First Order Conditions for problem 8 are:

$$E\left[x_1(y-\beta_0-\beta_1 x_1-\beta_2 x_2)\right] = 0 \quad (9)$$

$$E\left[x_2(y-\beta_0-\beta_1 x_1-\beta_2 x_2)\right] = 0 \quad (10)$$

$$E\left[(y-\beta_0-\beta_1 x_1-\beta_2 x_2)\right] = 0 \quad (11)$$

The first two conditions are symmetric: let's focus on 9.

The coefficients of the PMRF (cont.)

Consider the simple linear PRF of x_1 on x_2. We can always write:

$$x_1 = \hat{x}_1 + \hat{r}_1 \tag{12}$$

which we can substitute in 9 to get

$$E\left[(\hat{x}_1 + \hat{r}_1)(y - \beta_0 - \beta_1 x_1 - \beta_2 x_2)\right] = E\left[(\hat{x}_1 + \hat{r}_1)u\right] = 0 \tag{13}$$

By definition of the PRF, $E(\hat{x}_1 u) = 0$, since $\hat{x}_1$ is a linear function of x_2. Moreover $E(\hat{r}_1 x_2) = 0$ given 12 and $E(\hat{r}_1 \beta_0) = 0$, and 13 becomes:

$$E\left[\hat{r}_1(y - \beta_1 x_1)\right] = 0 \tag{14}$$

The coefficients of the PMRF (cont.)

Substituting 12 in 14 we get:

$$E\left[\hat{r}_1(y - \beta_1(\hat{x}_1 + \hat{r}_1))\right] = 0 \tag{15}$$

Again because $E(\hat{r}_1\hat{x}_1) = E(\hat{r}_1 x_2) = 0$ we are left with

$$E\left[\hat{r}_1(y - \beta_1\hat{r}_1)\right] = 0 \tag{16}$$

which finally gives

$$\beta_1 = \frac{E(\hat{r}_1 y)}{E(\hat{r}_1^2)} = \frac{Cov(\hat{r}_1, y)}{V(\hat{r}_1)} \tag{17}$$

The PRF coefficient β_1 is equal to the covariance between y and the residuals of the PRF of x_1 on x_2, divided by the variance of these residuals.

The coefficients of the PMRF (cont.)

We now want to show that if the CIA is satistified

$$\beta_1 = \tau_1 \tag{18}$$

and the PRF of y on x_1 and x_2 has a causal interpretation for the effect of x_1.

The coefficient of the PMRF under the CIA

Substitute the causal model 1 in the numerator of 17:

$$\begin{aligned} E(\hat{r}_1 y) &= E(\hat{r}_1(\mu + \tau_1 x_1 + \nu)) \\ &= \tau_1 E(\hat{r}_1^2) + E(\hat{r}_1 \nu) \end{aligned} \tag{19}$$

since $E(\hat{r}_1 X_1) = E(\hat{r}_1(\hat{X}_1 + \hat{r}_1)) = E(\hat{r}_1^2)$

Now note that, given (4):

$$\begin{aligned} E(\hat{r}_1 \nu) &= E(\hat{r}_1(\delta_2 x_2 + \omega)) \\ &= \delta_2 E(\hat{r}_1 x_2) + E(\hat{r}_1 \omega)) = E(\hat{r}_1 \omega)) \\ &= E(E(\hat{r}_1 \omega | x_2)) = E(E(x_1 \omega | x_2)) = E(0) = 0 \end{aligned} \tag{20}$$

where the next to last equality holds because of the CIA (see equation (5)).

The coefficients of the PMRF under the CIA (cont.)

Substituting the results of the previous slide in 17

$$\beta_1 = \frac{E(\hat{r}_1 y)}{E(\hat{r}_1^2)} = \frac{\tau_1 E(\hat{r}_1^2)}{E(\hat{r}_1^2)} = \tau_1 \tag{21}$$

If the CIA holds the PMRF can be interpreted causally for β_1.

However, no causal interpretation can be given of β_2.

Summary

We have shown that if we are interested in the causal effect of x_1 on y the CIA may represent a solution.

The CIA says that the other variables $\{x_2, ..., x_k\}$ that we observe are detailed and exaustive enough to guarantee that if two subjects are equal in terms of these variables the value of x_1 is effectively assigned randomly to them.

The randomness of the assignment of x_1 given $\{x_2...x_k\}$ is what permits a causal interpretation of β_1.

In what follows in this course we assume that the CIA holds symmetrically for all variables, and therefore all the parameters of the PMRF can be interpreted causally.

Future courses will discuss alternative solutions.

Section 3

Interpretation of the coefficients of the PMRF

The partial Multiple Regression coefficient

Extending the analysis to many covariates x, consider:

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + u \tag{22}$$

$$(\beta_0, ..., \beta_k) = arg \min_{b_0,...,b_k} E\left[(y - b_1 x_1 - ... - b_k x_k)^2\right] \tag{23}$$

the generic parameter β_j (for $j > 0$) is

$$\beta_j = \frac{E(\hat{r}_j y)}{E(\hat{r}_j^2)} = \frac{Cov(\hat{r}_j, y)}{V(\hat{r}_j)} \tag{24}$$

This parameter measures the effect on y of the component of x_j that is orthogonal to the other x variables. In fact , it can be obtained by:

- regressing x_j on all the others x variables;
- taking the residuals of this regression $\hat{r}_j$;
- considering the simple PRF of y on the single variable $\hat{r}_j$;
- $\hat{r}_j$ captures the part of x_j that is orthogonal to the other x.

Section 4

SMRF and PMRF in matrix notation

The model

As for the case of the simple linear regression, we now suppose to have a random sample of observations on y and $x_1, ...x_k$ and we ask:

- whether we can extend the OLS estimator;
- whether OLS continues to have good properties.

Given multiple covariates it is convenient to use matrix notation.

$$Y = X\beta + U \tag{25}$$

- Y is the $n \times 1$ column vector of observations on the outcome y_i.
- X is the $n \times (k+1)$ matrix of observations x_{ij} on the jth covariate.
- U is the $n \times 1$ column vector of observations u_i.
- β is the $(k+1) \times 1$ column vector of the parameters.

Note that X includes a column with all elements equal to 1 and the corresponding parameter is the constant β_0.

The basic set of necessary assumption

- MLR 1: The population regression function is linear in the parameters:

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

- MLR 2: The n observations on Y and X are a random sample of the population, so that

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

  where X_i is the *ith* row of X.

- MLR 3: There is no perfect collinearity, i.e no variable in X is constant (in addition to the constant term ...) and there is no exact linear dependency between any set of variables in X. Thus X has full rank equal to (k+1).

MLR-3 generalizes SLR-3 in the simple regression case.

The OLS estimator in matrix form

Under these assumptions, OLS solves the following problem

$$\hat{\beta} = \underset{b}{\text{argmin}}\, U'U = \underset{b}{\text{argmin}}[Y - Xb]'[Y - Xb] \tag{28}$$

where b is a $(k+1) \times 1$ column vector of possible parameter values.

There are $k+1$ FOC for this problem which we can write as

$$\frac{\partial U'U}{\partial b} = X'[Y - X\hat{\beta}] = 0 \tag{29}$$

$$X'X\hat{\beta} = X'Y \tag{30}$$

which gives the OLS estimator in matrix form

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{31}$$

where the full rank of X makes $X'X$ invertible.

Algebraic properties of OLS in matrix form

The fitted values are

$$\hat{Y} = X\hat{\beta} \tag{32}$$

and the estimated residuals are

$$\hat{U} = Y - \hat{Y} = Y - X\hat{\beta} \tag{33}$$

Therefore the first order condition 29 can also be written as

$$X'\hat{U} = 0 \tag{34}$$

and since the first row of X' is a row of ones (the constant), the sum of the OLS residuals is zero.

Section 5

Partialling out

Again on the interpretation of the PMRF

The matrix

$$H = Z(Z'Z)^{-1}Z' \tag{35}$$

is called a"projection matrix" because if you premultiply any vector Y by H, the result is the projection of the vector Y on the space spanned by Z.

Numerically it gives the least square prediction of Y given Z (see graphical interpretation of OLS).

$$Y_Z = HY = Z(Z'Z)^{-1}Z'Y = Z\hat{\psi} \tag{36}$$

for the PRF

$$Y = Z\psi + V \tag{37}$$

Note that H is symmetric and idempotent.

Projections

Consider the population regression:

$$Y = X\beta + U = W\delta + Z\gamma + U \tag{38}$$

where W is the main variable of interest and Z are control variables.

Consider the two projections

$$\begin{aligned} Y_Z &= HY = Z(Z'Z)^{-1}Z'Y = Z\tilde{\gamma} && (39) \\ W_Z &= HW = Z(Z'Z)^{-1}Z'W = Z\tilde{\rho} && (40) \end{aligned}$$

Consider the residuals from these two projections that we denote as

$$\begin{aligned} \tilde{Y} &= Y - Y_Z && (41) \\ \tilde{W} &= W - W_Z && (42) \end{aligned}$$

What happens if we regress $\tilde{Y}$ on $\tilde{W}$?

Partialling out matrices

Consider now the symmetric idempotent matrix M:

$$M = I - H = I - Z(Z'Z)^{-1}Z' \tag{43}$$

If you premultiply any vector by M you obtain the least square estimated residuals of the regression of the vector on Z.

$$\begin{aligned}
\tilde{Y} &= Y - Y_Z &(44)\\
&= MY = Y - Z(Z'Z)^{-1}Z'Y &(45)\\
\tilde{W} &= W - W_Z &(46)\\
&= MW = W - Z(Z'Z)^{-1}Z'W &(47)\\
\tilde{U} &= U - U_Z &(48)\\
&= MU = U - Z(Z'Z)^{-1}Z'U &(49)\\
\tilde{Z} &= Z - Z_Z &(50)\\
&= MZ = Z - Z(Z'Z)^{-1}Z'Z = 0 &(51)
\end{aligned}$$

Partialling out matrices (cont.)

Let's now premultiply the PMRF 38 by M:

$$\begin{aligned} MY &= MW\delta + MZ\gamma + MU \\ \tilde{Y} &= \tilde{W}\delta + \tilde{U} \end{aligned} \tag{52}$$

which explains why M is called a “partialling out" matrix.

Consider the OLS-MM estimator of 52

$$\begin{aligned} \hat{\delta} &= (\tilde{W}'\tilde{W})^{-1}\tilde{W}'\tilde{Y} \\ &= (W'M'MW)^{-1}W'M'MY = (W'MW)^{-1}W'MY \end{aligned} \tag{53}$$

It is obtained by regressing Y on the component of W which is orthogonal to Z and is numerically identical to the OLS-MM estimator of δ that we would obtain by estimating directly 38.

Also the standard error is numerically identical:
$Var(\hat{\delta}) = \sigma^2(W'MW)^{-1}$.