14.32 Econometric Data Science
Professor Anna Mikusheva
September, 2019

Lectures 22-24

Treatment effect and experiments

An experiment is designed and implemented consciously by a human researcher. An experiment entails the conscious use of treatment and control groups with random assignment. We also will talk about quasi-experiments (or natural experiments), when the source of randomization comes from nature, that is, 'as if' randomly assigned.

Program evaluation, the field of Statistics which aims to evaluate the effects of policy interventions, had the very useful language/notion of potential outcomes. That language will be very useful for us to understand the background of our methods and regression analysis on a deeper level (though in more simple settings).

Potential outcomes

Let us consider a binary variable D_i which will represent the status of treatment (equal 1 if treated, and 0 otherwise). For example, D_i can be a hospital stay, being in a small class or earning a college diploma. We would assume that for any individual in the sample there are two potential outcomes Y_{0i} - an outcome if the individual is not treated, and Y_{1i} the outcome of the same individual if treated. We observe only one of outcomes:

$$Y_i = Y_{i0} + (Y_{1i} - Y_{0i})D_i.$$

We are interested in learning the effect of treatment $Y_{1i} - Y_{0i}$, which is typically a random variable, so the proper object of interest can be 'average treatment effect' $E\left[Y_{1i} - Y_{0i}\right]$.

Let us see what happens when we do a comparison of averages, that is the average outcome over treated minus the same for non-treated. The challenge here is that they are estimated on different populations:

$$\begin{aligned} E[Y_i|D_i = 1] - E[Y_i|D_i = 0] &= E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 0] = \\ &= \left\{E[Y_{1i}|D_i = 1] - E[Y_{0i}|D_i = 1]\right\} + \left\{E[Y_{0i}|D_i = 1] - E[Y_{0i}|D_i = 0]\right\}. \end{aligned}$$

The first term here, $E[Y_{1i} - Y_{0i}|D_i = 1]$, is called the 'average treatment effect on the treated' and is a valid object of interest as well. However, the second term is known as selection bias: it compares how different the two groups are; if they are different enough it will completely bias the results.

The experiment uses a random assignment of the treatment, that is, D_i is chosen independently from two potential outcomes (Y_{1i}, Y_{0i}). In such a case

$$E[Y_{0i}|D_i = 1] = E[Y_{0i}|D_i = 0] = E[Y_{0i}] \quad \text{and} \quad E[Y_{1i}|D_i = 1] = E[Y_{1i}|D_i = 0] = E[Y_{1i}].$$

Thus, the difference of averages obtained in a correctly run experiment estimates the average treatment effect.

Estimation in randomized control trials under homogeneous effect assumption

Examples of economic experiments:
(1) Oregon Experiment of randomly broadening Medicaid;
(2) Tenessy STAR experiment about class size;
(3) A number of experiments run by MIT J-PAL in developing countries.

Assumption about homogeneous treatment effect. One assumption underlying much of regression analysis done prior to this lecture is the assumption of homogeneous treatment, that is, $Y_{1i} - Y_{0i} = \beta$ is the same for all individuals. This is a very strong assumption, and in this lecture we will explicitly say when we make it or when we do not. If homogeneous treatment is assumed then

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})D_i = EY_{i0} + \beta D_i + (Y_{0i} - EY_{0i}) = \alpha + \beta D_i + e_i.$$

OLS on this regression estimates

$$E(Y_i|D_i = 1) - E(Y_i|D_i = 0) = \beta + (E(e_i|D_i = 1) - E(e_i|D_i = 0)),$$

where the last term is 'selection bias'. Thus, even in the case of a homogeneous treatment effect the OLS (difference between two averages for treated and non-treated) does not estimate the correct quantity without making additional assumptions. The additional assumption here is

$$E(e_i|D_i) = 0,$$

which is the exogeneity assumption we have maintained through the whole course.

Adding controls. Assume we have data on a randomized control trial, and let us discuss why, nevertheless, we may not just compare the averages of treated and non-treated, but rather add some controls. There are at least three reasons why one may wish to run OLS regression with additional controls:

(1) Check randomization: run two OLS regressions:

$$Y_i = \alpha + \beta D_i + e_i \text{ and } Y_i = \alpha + \beta D_i + \gamma W_i + e_i^*,$$

and then compare two estimates of $\widehat{\beta}$, if different, then there was some challenges with the randomization.

(2) Higher efficiency (?) Assume homogeneous treatment. There is no exact prediction of whether or not standard errors in 'long' regression $Y_i = \alpha + \beta D_i + \gamma W_i + e_i^*$ will be larger or smaller than in the short version. On one side, the number of estimated coefficients increases (may decrease accuracy), but the variance of the error term (unexplained term) is smaller as well. If W is a strong predictor of the outcome Y, it is likely that the long regression will result in a more accurate estimate of β.

(3) Conditional randomization. Often full randomization is impossible, but rather only blocks randomization. Imagine that the experiment is possible only if it delivers some public good. For example, imagine that assignment to a class of a small size is forced to be skewed towards low-income kids but otherwise is random: before the experiment, kids are divided into two groups according to income, then two random assignments are done within each group, where the probability of being selected to a treatment is different for the two groups. Assume homogeneous treatment. This setting is described well by the following assumption:

$$E(e_i|D_i, W_i) = E(e_i|W_i).$$

The last quantity does not have to be zero, and often will not be zero if W is a determinant of the outcome Y. In such a case we need to run a long regression:

$$Y = \alpha + \beta D + \gamma W + e.$$

Here is a good time to explain the difference between a regressor of interest and one of control and why an estimator on β is consistent, while γ does not have to be consistent and does not have a causal interpretation. Assume that $E(e|W) = \delta_0 + \delta_1 W$(this is not restrictive if W is binary). Assume that γ is the causal coefficient (effect from W on Y in the idealized experiment). Consider

$$E(Y|D, W) = \alpha + \beta D + \gamma W + E(e|D, W) = \alpha + \beta D + \gamma W + E(e|W) =$$

$$= \alpha + \beta D + \gamma W + \delta_0 + \delta_1 W = (\alpha + \delta_0) + \beta D + (\gamma + \delta_1)W,$$

If we define $\widetilde{\alpha} = \alpha + \delta_0$ and $\widetilde{\gamma} = \gamma + \delta_1$. Then the regression

$$Y = \widetilde{\alpha} + \beta D + \widetilde{\gamma} W + e^*$$

satisfies the assumption of OLS, $E(e^*|D, W) = 0$, and thus the OLS regression of Y on D and W estimates β and $\widetilde{\gamma}$ (but not γ) consistently.

IV regression Assume homogeneous treatment. Often in experiments, the researcher faces a problem of compliance: people assigned to the treatment do not complete it. Let D_i be 'assigned treatment status', while X_i is the 'actual treatment received'. The question is what the researcher should do. Running the OLS regression

$$Y = \alpha + \beta X + e$$

does not seem to be appealing since it is plausible that the error is correlated with X. However, D is randomly assigned (not correlated with e) and usually very highly correlated with X (since most participants comply with assignment). Thus, D can serve as an instrument in the regression of Y on X. This setting is a very classical instrumental variable case.

Heterogeneous treatment effect

Heterogeneous treatment effect refers to the situation when the treatment effect varies for different individuals, that is, $Y_{1i} - Y_{0i} = \beta_i$ which varies for different i. Before (even when discussing randomized control trials) we made the assumption of a homogeneous treatment effect. This section discusses what in fact we actually estimate when an effect is heterogeneous, by using OLS (with or without controls) or when using IV.

Heterogeneous treatment effect in OLS

Heterogeneous treatment with a binary regressor. Assume there is a binary treatment D_i, then

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})D_i = E[Y_{0i}] + \beta_i D_i + (Y_{0i} - E[Y_{0i}]) = \alpha + \beta_i D_i + e_i$$

follows a regression line with individual slopes. We have made no assumptions so far. Now assume that treatment is not randomly assigned, but does satisfy exogeneity $E[e_i|D_i] = 0$ (it is a less than full random assignment). The question is what is the quantity that the OLS estimates. The OLS in this case is $\bar{Y}_{treated} - \bar{Y}_{control}$. Thus in the limit it estimates

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = \alpha + E[\beta_i|D_i = 1] + E[e_i|D_i = 1] - (\alpha + E[e_i|D_i = 0]) = E[\beta_i|D_i = 1].$$

Thus, it estimates the treatment effect on the treated!

The general case of OLS with heterogeneous treatment effect. Assume we have a regression with heterogeneous slopes:

$$Y_i = \alpha + \beta_i X_i + e_i$$

and assume exogeneity $E(e_i|X_i) = 0$. Then the limit of OLS (quantity it estimates) is

$$\lim \widehat{\beta} = \frac{cov(Y_i, X_i)}{Var(X_i)} = \frac{cov(\beta_0 + \beta_i X_i + e_i, X_i)}{Var(X_i)} = \frac{cov(\beta_i X_i, X_i)}{Var(X_i)} = E[\beta_i \omega_i],$$

where $\omega_i = \frac{1}{Var(X)}(X_i - EX)X_i$. That is, the estimand is 'weighted average treatment effect' where the weights depend on the deviation of the regressor from the average: more extreme regressor values have more weight. Such a quantity is hard to interpret.

If, however, we assume more than just exogeneity, but also random assignment, that is, the level of the regressor is randomly assigned independently of outcome (β_i and e_i), then

$$cov(\beta_i X_i, X_i) = E[\beta_i X_i(X_i - EX)] = E[\beta_i]Var(X_i).$$

Thus

$$\widehat{\beta}_{OLS} \to^p E[\beta_i].$$

The estimand in this case is the average treatment effect.

Heterogeneous treatment effect in TSLS regression

Let us consider an IV model with heterogeneity both stages:

$$Y_i = \alpha + \beta_i X_i + e_i; \quad X_i = \gamma + \pi_i Z_i + u_i.$$

Let us assume the following: (i) β_i, π_i are distributed independently from Z_i, e_i, U_i; (ii) $E(u|Z) = E(e|Z) = 0$; (iii)$E\pi_i \neq 0$. Then

$$\widehat{\beta}_{TSLS} \to^p \frac{cov(Y_i, Z_i)}{cov(X_i, Z_i)} = \frac{cov(\alpha + \beta_i X_i + e_i, Z_i)}{cov(\gamma + \pi_i Z_i + u_i, Z_i)} =$$
$$= \frac{cov(\alpha + \beta_i(\gamma + \pi_i Z_i + u_i) + e_i, Z_i)}{cov(\gamma + \pi_i Z_i + u_i, Z_i)} = \frac{E[\pi_i \beta_i]Var(Z_i)}{E[\pi_i]Var(Z_i)} = \frac{E[\pi_i \beta_i]}{E[\pi_i]} = E\left[\beta_i \omega_i\right],$$

where $\omega_i = \frac{1}{E[\pi]}\pi_i$. Thus, the estimand is the weighted average of the treatment effect, where the weights are such that the individuals most influenced by Z have the greatest weight. This quantity is referred to as the local average treatment effect.

What is interesting here: it is suggested that if both stages are heterogeneous, then the value of the estimand depends on the instruments. If one chooses a different instrument, then s/he gets a different quantity.

Heterogeneity of both stages is important. If the main equation is homogeneous ($\beta_i = \beta$), then

$$\frac{E[\beta_i \pi_i]}{E[\pi_i]} = \frac{\beta E[\pi_i]}{E[\pi_i]} = \beta.$$

If the first stage is heterogeneous, that is, if $\pi_i = \pi$, then

$$\frac{E[\beta_i \pi_i]}{E[\pi_i]} = \frac{\pi E[\beta_i]}{\pi} = E[\beta_i].$$

To come up with this interesting effect, when the estimand depends on instruments, we also need the heterogeneity to align. Assume that π_i is independent of β_i, then

$$\frac{E[\beta_i \pi_i]}{E[\pi_i]} = \frac{E[\beta_i]E[\pi_i]}{E[\pi_i]} = E[\beta_i].$$

Heterogeneous treatment effect in TSLS in bivariate case.

As a special case of what we have previously discussed, let us consider a situation in which both treatment X and instrument Z are binary.

As an example we now consider the question of the effect on test scores of attending a public vs a charter school, using data on the KIPP Charter school (ref. here). The research question is hard to address as the choice of public vs charter school is done in a non-random manner, but rather correlated with the students' characteristics. We will use an exogenous variation in charter school attendance due to a lottery. Apparently, since the number of applications to the charter school exceeds the number of available seats, admission is done through a publicly run lottery. Let Y- be the test scores we see for a student, X is the dummy for charter school attendance, and Z is whether a student won the lottery for admission to the charter school. This is not a perfectly run randomized control trial though, as not all students who won the lottery chose to attend the school; there is also a small number of students who did not win but ended up attending.

Let's introduce TSLS notation. The structural equation is

$$Y_i = \alpha + \beta_i X_i + e_i; \qquad X_i = \gamma + \pi_i Z_i + u_i.$$

If we plug one into the other, we get the reduced form equation:

$$Y_i = \mu + \delta_i Z_i + v_i.$$

The TSLS in this case is

$$\widehat{\beta}_{TSLS} = \frac{s_{YZ}}{s_{XZ}} = \frac{s_{YZ}/s_Z^2}{s_{XZ}/s_Z^2} = \frac{\widehat{\delta}_{OLS}}{\widehat{\pi}_{OLS}},$$

where $\widehat{\delta}_{OLS}$ and $\widehat{\pi}_{OLS}$ are OLS estimates in the reduced form and the first stage correspondingly. Given that both regressions are on the dummies, we get:

$$\widehat{\beta}_{TSLS} \to^p \frac{E(Y_i|Z_i=1) - E(Y_i|Z_i=0)}{E(X_i|Z_i=1) - E(X_i|Z_i=0)}. \tag{1}$$

Now let us look at the first stage from a different perspective as described by the following table:

	$Z_i = 1, X_i = 0$ won lottery did not attend	$Z_i = 1, X_i = 1$ won lottery attended
$Z_i = 0, X_i = 0$ did not win, did not attend	never takers	compliers
$Z_i = 0, X_i = 1$ did not win, attended	defiers	always takers

Notice that in the data we cannot determined which individual is in which group. To do that, we need to see both states (when he wins and when he does not). Let us assume that there are no defiers in population. Then the only people who contribute to either the numerator or the denominator of (1) are those whose action is affected by the instrument (those moved by instrument), that is, the compliers. For them, the move is from potential outcome Y_{0i} to Y_{1i}. Thus

$$\widehat{\beta}_{TSLS} \to^p E(\beta_i|compliers).$$