Econometrics, PS6

Aleksandra Malova
discussion with Giacomo Mangiante

06 May 2019

1 Application: Matching vs Regression

1.1 (a) Can gender be considered as treatment? Explain.

Theoretically gender can be considered as treatment since gender can be described as binary variable and thus individual of a certain gender can be considered as an individual with a certain treatment status. However, for all practical purposes gender is unlikely to be a good treatment because usually individual is associated with the certain gender since birth (the cases of gender change of course plausible but not numerous I guess). What is inherited from birth is hardly to disentangle from other characteristics of an individual, thus it's hard to imagine a setup in which gender will be independent of all other features of individual. Thus, gender seems to be useless as treatment because main interest in analyzing treatment is when it is independent of potential outcome but it's hardly to be the case.

1.2 (b) Assuming full independence between treatment and potential outcomes, what is the average treatment effect of gender on admission?

Population formula for the average treatment effect is

$$\Delta^{ATE} = E(Y_{i1}|R_i = 1) - E(Y_{i0}|R_i = 0)$$

where $Y_i = \begin{cases} \text{Y}_{i0} \text{ if } R = 0 \\ \text{Y}_{i1} \text{ if } R = 1 \end{cases}$ the outcome variable representing admission status. Outcome variable is a binary variable that takes two values, $Y_i = 0$ if not admitted and $Y_i = 1$ if admitted; R_i is binary treatment variable, $R_i = 1$ if individual is male and $R_i = 0$ if the individual is female. Sample analogue of the population formula is

$$\hat{\Delta}^{ATE} = \frac{\sum\limits_{i=1}^{N} Y_i \cdot R_i}{\sum\limits_{i=1}^{N} R_i} - \frac{\sum\limits_{i=1}^{N} Y_i \cdot (1 - R_i)}{\sum\limits_{i=1}^{N} (1 - R_i)}$$

Given the data in the table we can compute $\hat{\Delta}^{ATE} = 0.445188 - 0.303542 = 0.141645$.

1.3 (c) Do you find the assumption of full independence plausible? Why, or why not?

The assumption of full independence of gender and the admission outcome may be plausible if we, first, know that admission board had no access to the gender status of the applicant before taking the admission decision. This can rule out personal preferences and attitude bias of the admission board. This is actually verifiable thing. Second, we need to assume that gender has no impact on the educational outcomes of the applicants which were taken into account for the admission process. This second assumption is neither verifiable (that's why it is an assumption) nor innocuous. Although we may believe that gender per se doesn't influence on mental abilities or abilities to study, however, social treatment of males and females is different, thus females may be lacking education, were discouraged for gaining the education etc. Given all these issues (mainly the second aspect) it's hard to find the full independence of gender status and admission decision plausible.

1.4 (d) A weaker assumption is conditional independence. Assume that treatment and potential outcomes are independent conditional on the field of study and compute the unconditional average treatment effect (ATE), as well as the treatment effect on the treated (ATT). These are a matching estimators.

Effect of treatment on treated given observable characteristics X_i:

$$\delta^{ATT} = E(Y_{i1} - Y_{i0}|D_i = 1) = E(E(Y_{i1}|X_i, D_i = 1) - E(Y_{i0}|X_i, D_i = 1)|D_i = 1) =$$

$$= E(E(Y_{i1}|X_i, D_i = 1) - E(Y_{i0}|X_i, D_i = 0)|D_i = 1) = E(\delta_x|D_i = 1)$$

where δ is random X-specific difference in mean outcome by treatment status for each value of X_i. For discrete case:

$$E(Y_{i1} - Y_{i0}|D_i = 1) = \sum_x \delta_x \cdot P(X_i = x|D_i = 1)$$

where $P(X_i = x|D_i = 1)$ is a probability mass function for X_i given $D_i = 1$. For computation purposes let's arrange a table (Table 1.).

Given two last columns of the table we can compute ATT as a scalar product of these columns and it yields $\delta^{ATT} = -.071$

To compute ATE we can use the following formula:

$$\delta_{ATE} = E(Y_{i1}|X_i, D_i = 1) - E(Y_{i0}|X_i, D_i = 0)) = \sum_x \delta_x \cdot P(X_i = x)$$

Table 1: Data needed for ATT and ATE computation

$\mathrm{X}_i = x$	δ_x	$P(X_i = x \mid D_i = 1)$	$P(X_i = x)$
A	-.203	.307	-.206
B	-.050	.208	.129
C	.029	.121	.203
D	-.018	.155	.175
E	.038	.071	.129
F	-.011	.139	.158

where $P(X_i = x)$ is the marginal distribution of X_i across whole sample (not only treated as it was for ATT). Thus we get $= \delta_{ATE} = -0.043$

1.5 (e) Why do your answers in b. and d. differ?

Answers differ because in (b) we assumed full independence of the outcome from the treatment and in the second case we assumed independence only conditional on some covariates, in our case conditional on field of study. If the second assumption is the true one then ATE calculated in (b) suffers from bias because acceptance to the university can systematically differ from field to field for males and females. However, given the setup of the problem it's hard to say whether the bias will be positive or negative in this case because of the lack of information.

1.6 (f) Run the regression $accept = \beta_0 + \beta_1 \cdot male + \beta_2 \cdot fieldB + ... + \beta_6 \cdot fieldF + \epsilon$

See Table 2 with the regressions output

1.7 (g) Read chapter 2.3.2 of Angrist and Krueger (1999). Are your results from OLS and matching the same? If not, why?

Regression estimates are larger than matching estimates. Matching and regression estimates control for the same variables, moreover, regression estimands can be seen as a sort of matching estimands. However "the regression estimands differ from the matching estimands in the weights used to sum the covariate-specific effects, δ_x into a single effect. In particular, matching uses the distribution of covariates among the treated to weight covariate-specific estimates into an estimate of the effect of treatment on the treated, while regression produces a variance-weighted average of these effects. Namely regression estimand weights each covariate-specific treatment effect by $[P(X_i = x|D_i = 1) \cdot (1 - P(X_i =$

Table 2: Regression table

	(1)
	Accept
Male	-0.0185
	(-1.20)
fieldB	-0.00964
	(-0.41)
fieldC	-0.303***
	(-13.64)
fieldD	-0.310***
	(-14.03)
fieldE	-0.402***
	(-16.13)
fieldF	-0.586***
	(-25.75)
Constant	0.660***
	(33.18)
Observations	4527

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

$x|D_i = 1)] \cdot P(X_i = x)$. The matching estimand for the effect of treatment on the treated weights treatment effect proportional to the probability of treatment at each value of the covariates. Thus, treatment-on-the-treated estimand puts the most weight on covariate cells containing those who are most likely to be treated. In contrast, regression puts the most weight on covariate cells where the conditional variance of treatment status is largest." (Angrist and Pischke, 2006).

2 Application: Instrumental variables

2.1 (a) What are the assumptions (expressed in the standard structural notation, not the potential outcomes one) needed for β_2 to be consistently estimated? Describe why these assumptions are not likely to be satisfied in this regression using our sample from the census.

For β_2 to be consistently estimated by OLS in a regression $Health_i = \beta_1 + \beta_2 \cdot educ_i + \beta_3 \cdot X_i + \mu_i$ there are three assumptions needed:

- **Ergodic stationarity**, which means that vector $\{Health_i, educ_i, X_i\}$ is jointly stationary and ergodic
- **Predetermined regressors**, which means that the regressors are orthogonal to the contemporaneous error term $E(\{educ_i, X_{ik}\} \cdot \mu_i) = 0$ for all i and $k = 1...K$, where K denotes the number of covariates
- **Rank condition**, which means that the matrix of regressors is non-singular

The assumption about predetermined regressors is not likely to be satisfied. It can be the case of the reverse causality that breaks this assumption, namely that initially more healthy people tend to get more education and at the same time more educated people are more alerted about their health and care about it more. There also can be an issue of unobserved regressors that affect both health and education, for example, family income when the individuals were children. Richer families tend to care more both about child's health (which affects future health status) and also about education. Thus people from richer families tend to have initially better health and better education.

2.2 (b) Theoretically, if a valid instrument is found, can this model solve the problems mentioned in point a?

Yes, the valid instrument can create exogenous variation in education level of people, that is unrelated to other factors that influence health outcomes and captured by an error term, which can help to trace out the influence of education level on a health status independent on potential confounding factors.

2.3 (c) What are the necessary assumptions for your instrumental variables strategy to yield consistent estimates for the health returns to education? Present the assumptions in our particular case both in the structural and in the potential outcomes framework.

Structural framework: Specify the system of equations for IV identification:

$$\begin{cases} h = m\beta + \mu \\ m = Z\pi + \nu \end{cases}$$

We assume β to be constant across all observations. The assumptions needed for β to be estimated consistently are the following:

- **Exogeneity**, which means that $cov(Z, \nu) = 0$
- **Exclusion restriction**, which means $cov(h, Z|m) = cov(\mu, Z) = 0$
- **Relevance**, which means $cov(m, Z) \neq 0$

Potential outcomes framework: We think of an outcome h as a continuous function of the treatment m, namely $h_{mi} \equiv f_i(m)$. Then we specify constant-effect causal model: $f_i(m) = \pi_0 + \pi_1 \cdot m + \mu_i$. In this model m doesn't have a subscript i because the model spits out what would be the outcome for a person i for any level of m, not only the realized value of m_i. All the i specific characteristics are captured with the error term μ_i. **First assumption that we make, writing down this model, is SUTVA, which means that the outcome of one unit is unaffected by assignment of the treatment to another unit.** When we substitute observed value of m_i into a causal model we get following regression: $h_i = \beta_0 + \beta_1 \cdot m_i + \mu_i$. However, m_i turned out to be correlated with μ_i. We have an instrument z which is correlated with the treatment variable m: $cov(m, z) \neq 0$. **This is the second assumption, which is called relevance or first-stage ($m_i = \pi_0 + \pi_1 \cdot z_i + \nu_i$, where $\pi_1 \neq 0$).** This instrument should be exogenous to the treatment variable, which means $cov(z_i, nu_i) = 0$ **(third, exogeneity assumption)**. The instrument should be also uncorrelated with any other determinants of the dependent variable, which is the third assumption, **exclusion restriction:** $cov(z_i, \mu_i) = 0$. And the last thing we assume in potential outcomes framework is the **homogeneity of the treatment effect**, which means $h_{m,i} - h_{m-1,i} = \pi_{1,i} = \pi_1$.

2.4 (d) Describe who the compliers, defiers, always takers and never takers are in AK91. What did you assume about them in point c?

In AK91 **compliers** are those who had extra year of studies when they were forced by law and wouldn't have studied an extra year otherwise; **defiers** could be those who didn't study an extra year while being forced by the law but who wanted to study otherwise (however, AK91 assumed the absence of defiers in their paper); **always-takers** are those who studies an extra year or more anyways, independent of being forced or not; **never-takers** are those who would have not studies an extra year in any case. In the context of using instruments with potential outcome framework with heterogeneous treatment effect we assume an absence of defiers through replacement the last assumption of heterogeneity of the treatment effect for the assumption about monotonicity of the treatment effect.

2.5 (e) AK91 uses Quarter of Birth as an instrument for education. They argue that birth quarter allows for an exogenous variation in the within-cohort year educational levels induced by age-based compulsory schooling laws. Discuss the arguments made by BJB95 and BJ96 concerning the validity of using quarter of birth as an instrument.

Both papers doubt the validity mainly based on the argument that compulsory school attendance laws are not the only factors behind the association between quarter of birth and educational attainment and earnings. Thus, if association between quarter of birth is related with both explanatory variable and the outcome by more subtle unobservable relation then exclusion restriction breaks and the identification strategy fails. I guess that even there is a relation between quarter of birth and educational attainment and earnings, it's more noise than signal in observing this relationship, thus empirically and for all practical purposes I guess that arguments presented by AK91 are convincing enough to trust their analysis.

2.6 (f) Theoretically, do you think that compulsory schooling laws (instead of quarter of birth) can be used directly as instruments? What assumption would the laws have to satisfy?

Yes, theoretically compulsory schooling laws if assigned randomly, say, across states, could be a good instrument by its own. But the main problem is that usually they are not assigned randomly across states. The state can accept higher mandatory level of schooling because of poor economic performance in the state, for example if people in this state tend to study less than in other states and thus work worse. Introduction of higher compulsory schooling may be an endogenous answer to the poor performance, which makes it a bad instrument in this case.

2.7 (g) Describe some potential mechanisms through which education might affect health.

- Better educated people may know more about health, especially they maybe more concerned about preventive measures, such as exercising, eating healthy, no smoking etc. Thus they may have better health than less educated people;
- More education may lead to increase in income, which allows better educated people to afford better health treatment (more expensive medicine in case of sickness or better nutrition in case of a good health) which lead to a better health;
- Better educated people may have more challenging and stressful job, which lead to poorer health outcomes;
- Better educated people may have more philosophical view on life and understand that we all die in the end, thus there is no point in maintaining healthy and boring life. They may think that it's more important to have fun, eat tasty (and unhealthy) food, drink alcohol and smoke, thus they can have poorer health.

2.8 (h) OLS regressions

See Table 3.

2.9 (i) Write down your 2SLS model (one equation for the health variable and one for education), justifying your choice of variables. Would you also add income as a covariate, why or why not?

2SLS system of equations is as following:

$$Health_i = \beta_0 + \beta_1 \cdot educrec_i + \beta_2 \cdot SMSA_i + \beta_3 \cdot married_i + \{dummies\} + \mu_i$$

$$educrec_i = \pi_0 + \pi_1 \cdot birthqtr_i + \pi_2 \cdot SMSA_i + \pi_3 \cdot married_i + \{dummies\} + \nu_i$$

It's important to include all covariates that are in the structural equation to the first stage, this justifies the specification of the first-stage equation.

No, I wouldn't add income as a covariate, since income is endogenously related to education.

2.10 (j) 2SLS regression

See table 4.

Table 3: OLS regression table

	(1) health	(2) health	(3) health
Educational attainment recode	-0.0270***	-0.0256***	-0.0252***
	(-118.57)	(-112.02)	(-109.90)
SMSA		0.00126	0.000690
		(0.96)	(0.52)
married		-0.0923***	-0.0930***
		(-77.93)	(-78.61)
New England Division		-0.00900***	-0.00937***
		(-4.14)	(-4.31)
Middle Atlantic Division		-0.0189***	-0.0195***
		(-11.57)	(-11.95)
East North Central Div.		-0.0144***	-0.0145***
		(-9.24)	(-9.25)
West North Central Div.		-0.0101***	-0.0101***
		(-5.23)	(-5.26)
South Atlantic Division		0.00445**	0.00454**
		(2.72)	(2.78)
East South Central Div.		0.0201***	0.0207***
		(9.69)	(9.97)
West South Central Div.		0.00135	0.00138
		(0.75)	(0.77)
Mountain Division		0.00251	0.00262
		(1.14)	(1.19)
Quarter of birth		-0.00167***	
		(-4.29)	
Birth year dummies	No	No	Yes
Constant	0.282***	0.359***	0.374***
	(173.87)	(147.81)	(146.89)
Observations	446241	446241	446241

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4: 2SLS regression table

	(1) IV birthqtr	(2) IV birthqtr	(3) IV birthqtr*birthyear	(4) IV birthqtr*birthstate
Educational attainment recode	-0.0725***	-0.0703***	-0.0675***	-0.0493***
	(-6.49)	(-6.23)	(-6.71)	(-7.23)
SMSA		0.0116***	0.0110***	0.00595**
		(3.80)	(3.92)	(3.09)
married		-0.0769***	-0.0779***	-0.0843***
		(-18.20)	(-20.44)	(-30.51)
New England Division		-0.0225***	-0.0217***	-0.0197***
		(-5.64)	(-5.87)	(-4.86)
Middle Atlantic Division		-0.0344***	-0.0335***	-0.0363***
		(-8.39)	(-8.96)	(-7.79)
East North Central Div.		-0.0403***	-0.0387***	-0.0333***
		(-6.05)	(-6.46)	(-6.82)
West North Central Div.		-0.0336***	-0.0322***	-0.0224***
		(-5.41)	(-5.73)	(-4.95)
South Atlantic Division		-0.0256***	-0.0237***	-0.00238
		(-3.31)	(-3.42)	(-0.97)
East South Central Div.		-0.0331*	-0.0298*	0.00378
		(-2.43)	(-2.44)	(0.89)
West South Central Div.		-0.0260***	-0.0244***	-0.00283
		(-3.66)	(-3.81)	(-1.07)
Mountain Division		-0.00689*	-0.00630*	-0.000778
		(-2.08)	(-2.02)	(-0.29)
Birth year dummies	No	Yes	Yes	Yes
State of birth dummies	No	No	No	Yes
Constant	0.596***	0.680***	0.662***	0.546***
	(7.75)	(8.86)	(9.65)	(11.34)
Observations	446241	446241	446241	446241

t statistics in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

2.11 (k) How strong is the first stage; is finite sample bias an issue; is the exclusion restriction likely to be satisfied?

See first-stage regressions in Table 5.

Table 5: First-stage F-stats

	(1) IV birthqtr	(2) IV birthqtr	(3) IV birthqtr*birthyear	(4) IV birthqtr*birthstate
F-stat	67.04	66.82	8.31	3.5

- First-stage seems to be weak in regressions where interactions were used as instruments.
- Based on this weak first-stage we can conclude that IV estimates may suffer from a finite-sample bias, which arises when correlation between endogenous variable and an instrument is weak. This is the case at least for first-stage for regressions (3) and (4) in table 5.
- Given the evidence presented in BJB95 and BJ96 quarter of birth seems to be correlated with educational attainment not only through compulsory schooling laws but also through other channels. The same channels may lead to the correlation between quarter of birth and earnings. So basically this question echoes the question 2(e) about validity of the instrument. Hence it's hard to say for sure whether exclusion restriction satisfied or not.

2.12 (l) If the instrument is valid, would one expect to find 2SLS estimates to be higher or lower than estimates form OLS? Argument may be build on measurement error, omitted variables or both

For omitted variables point of view the estimate from OLS tends to be higher than the estimate from 2SLS, since the coefficient will be estimated with the additional value that related to the relation between an error term and endogenous regressor. From measurement error point of view, OLS coefficient will be downward biased (attenuation bias). When the combination of these to effects is at play it's hard to say the direction of bias of the OLS coefficient relative to 2SLS.

2.13 (m) Interpret your 2SLS estimate in the face of heterogeneous treatment effects. In particular, whose return does the instrument capture? Considering this, would you expect the same relationship between the 2SLS and OLS estimates for health return to education?

In terms of heterogeneous treatment effect 2SLS estimate captures the effect for compliers. I expect the difference in the relationship between OLS and 2SLS to be positive, which means that those who has more education tend to be more healthy as predicted by 2SLS comapred to OLS.

2.14 (n) Is there a relationship between educationa dn health? Are you convinced with your analysis? Why?

I guess there is a relationship between education and health, however, I'm not convinced with my analysis because dependent variable is binary and we exploited linear probability model and linear 2SLS instead of probit for instance. I gues this makes the whole analysis basically useless. Although, health measurement as an absence of disability seems to be too rough measurement and actually doesn't reflect the mechanisms that maybe underlying the heath-education relationship.

3 Theory: Heterogeneous effects in IV models

3.1 (a) Under random assignment derive the $ATT = E(\tau_i|T_i = 1)$ and $ATE = E(\tau_i)$ in estimable quantities

ATT: $E(Y_i|T_i = 1) - E(Y_i|T_i = 0) = E(Y_{i1}|T_i = 1) - E(Yi0|T_i = 0) = E(Y_{i0} + \tau_i|T_i = 1) - E(Y_{i0}|T_i = 0) = E(Y_{i0}|T_i = 1) + E(\tau_i|T_i = 1) - E(Y_{i0}|T_i = 0) = E(\tau_i|T_i = 1)$

ATE: $E(Y_{i1} - Y_{i0}) = E(Y_{i1}) - E(Y_{i0}) = E(Y_{i0} + \tau_i) - E(Y_{i0}) = E(Y_{i0} + E(\tau_i) - E(Y_{i0} = E(\tau_i)$

3.2 (b) State the assumptions necessary to estimate the counterfactuals, $E(Yi0|T_i = 1)$ and $E(Yi1|T_i = 0)$ by selection-on-observables

The assumption necessary to estimate the counterfactuals is CIA - Conditional Independence Assumption. It says that treatment is "as good as randomlu assigned" conditional on observables X_i.

3.3 (c) What does the IV estimator converge to?

$$\hat{\tau}_i^{IV} = \frac{c\hat{o}v(Z,Y)}{c\hat{o}v(Z,T)} = \frac{\sum_{i=1}^{n}(Z_i - \bar{Z}) \cdot (Y_i - \bar{Y})}{\sum_{i=1}^{n}(Z_i - \bar{Z}) \cdot (T_i - \bar{T})}$$

$$plim\hat{\tau}\_i^{IV} = \frac{plim\frac{1}{n} \cdot \sum_{i=1}^{n}(Z_i - \bar{Z}) \cdot (Y_i - \bar{Y})}{plim\frac{1}{n} \cdot \sum_{i=1}^{n}(Z_i - \bar{Z}) \cdot (T_i - \bar{T})} = \frac{cov(Z,Y)}{cov(Z,T)} = \tau_i$$