Recitation Notes 6

Konrad Menzel

October 22, 2006

1 Random Coefficient Models

1.1 Motivation

In the empirical literature on education and earnings, the main object of interest is the human capital earnings function (HCEF)

$$Y_i = \alpha + \beta S_i + \gamma_1 X_i + \gamma_2 X_i^2 + \varepsilon_i$$

Typically, IV estimates of the return to education are higher than OLS even though we'd thought that OLS already suffered from an upward ability bias (for a summary of results from prominent studies, see tables from Card's handbook chapter on next page). There are several explanations for this empirical regularity

1. OLS suffers from measurement error, IV does not
2. publication bias: a study by Ashenfelter and Harmon finds a positive correlation between the point estimate and its sampling error across studies[1]
3. Card's favorite story: instruments in the literature affect populations with relatively low levels of, and therefore high returns to education

In this recitation, I am planning to elaborate a little on the last point.

1.2 Heterogeneous Treatment Effects

Assume you have data about a training program for unemployed workers, and you want to tell a policy maker whether the program was successful so that the government should continue to finance it. Say, your main outcome of interest is earnings six months after the training program, y_i, and you know whether a particular person participated ($D_i = 1$) or not ($D_i = 0$).
Now, so far we have only looked at regressions of the type

$$Y_i = \alpha + D_i\beta + \varepsilon_i$$

This implies that across the whole population, everyone who participates earns exactly β dollars more than if he hadn't received training. This doesn't make much sense in most real-world applications. For

[1] The story behind the publication bias typically goes follows: it's hard to publish a study with a small t-statistic for the return to education. A study with a very imprecise estimator can only arrive at a significant t-statistic if the value of the coefficient is higher (in absolute value, but the return to education - at least in publishable studies - is usually positive). Therefore if studies are selected on having a large t-statistic, the imprecise estimators we see published must be on average higher than the more precise ones (and typically the IV standard error is several orders of magnitude higher than that of OLS).

Figure 1: Source: Card, D. (1999): The Causal Effect of Education on Earnings, *Handbook of Labor Economics*, ch **30**, pp.1835-36

Author	Sample and Instrument		Schooling Coefficients OLS	IV
1. Angrist and Krueger (1991)	1970 and 1980 Census Data, Men. Instruments are quarter of birth interacted with year of birth. Controls include quadratic in age and and indicators for race, marital status, urban residence	1920-1929 cohort in 1970	0.070 (0.000)	0.101 (0.033)
		1930-1939 cohort in 1980	0.063 (0.000)	0.060 (0.030)
		1940-1949 cohort in 1980	0.052 (0.000)	0.078 (0.030)
2. Staiger and Stock (1997)	1980 Census, Men. Instruments are quarter of birth interacted with state and year of birth. Controls are same as in Angrist and Krueger, plus indicators for state of birth LIML estimates	1930-1939 cohort in 1980	0.063 (0.000)	0.098 (0.015)
		1940-1949 cohort in 1980	0.052 (0.000)	0.088 (0.018)
3. Kane and Rouse (1993)	NLS Class of 1972, Women. Instruments are tuition at 2 and 4-year state colleges and distance to nearest college. Controls include race, part-time status, experience *Note: Schooling measured in units of college credit equivalents*	Models without test scores or parental education	0.080 (0.005)	0.091 (0.033)
		Models with test scores and parental education	0.063 (0.005)	0.094 (0.042)
4. Card (1995b)	NLS Young Men (1966 Cohort) instrument is an indicator for a nearby 4-year college in 1966, or the interaction of this with parental education. Controls include race, experience (treated as endo-genous), region and parental education	Models that use college proximity as instrument (1976 earnings)	0.073 (0.004)	0.132 (0.049)
		Models that use college proximity x family background as instrument	——	0.097 (0.048)
5. Conneely and Uusitalo (1997)	Finnish men who served in the army in 1982, and were working full time in civilian jobs in 1994. Administrative earnings and education data. Instrument is living in university town in 1980. Controls include quadratic in experience and parental education and earnings.	Models that exclude parental education and earnings	0.085 (0.001)	0.110 (0.024)
		Models that include parental education and earnings	0.083 (0.001)	0.098 (0.035)
6. Maluccio (1997)	Bicol Multipurpose Survey (rural Philippines): male and female wage earners age 20-44 in 1994, whose families were interviewed in 1978. Instruments are distance to nearest high school and indicator for local private high school. Controls include quadratic in age.	Models that do not control for selection of employment status or location	0.073 (0.011)	0.145 (0.041)
		Models with selection correction for location and employment	0.063 (0.006)	0.113 (0.033)
7. Harmon and Walker (1997)	British Family Expenditure Survey 1978-1986 (men). Instruments are indicators for changes in the minimum school leaving age in 1947 and 1973. Controls include quadratic in age, year, survey and region and region		0.061 (0.001)	0.153 (0.015)

OLS and IV estimates of the return to education with instruments based on features of the school system

Figure by MIT OCW.

example we'd think that, say, a basic literacy program wouldn't have much of an effect on more educated individuals, or that the causal effect of the number of children in a household on the mother's labor supply differs a lot with the mother's age, education and other characteristics.
So we'd rather like to write the model as

$$Y_i = \alpha_i + D_i\beta_i$$

which allows individuals not only to have different initial earnings levels, but each person could also benefit from the program to a different degree. Note that there is no ε_i in this equation. The reason for this is that the causal model for the potential outcomes should be understood as *deterministic*, i.e. the α_i and β_i are fixed for the individual. From this point of view the disturbance ε in the regression equation only captures *our* imperfect knowledge about the parameters in the potential outcomes for a particular individual i.
So if people participated in the program regardless of their initial earnings level (this may stand for "ability"), $\alpha_i \perp D_i$, and we could run OLS, remembering that for dummy regressions, [2]

$$\hat{\beta}_{LS} = \mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0] = \mathbb{E}[\alpha_i + \beta_i|D_i = 1] - \mathbb{E}[\alpha_i|D_i = 0] \overset{(\alpha_i \perp D_i)}{=} \mathbb{E}[\beta_i|D_i = 1]$$

Therefore, OLS estimates the *treatment effect on the treated* individuals for our training program. If treatment was truly the outcome of a randomization, i.e. $(\alpha_i, \beta_i) \perp D_i$, we have in addition

$$\text{plim}_N \hat{\beta}_{LS} = \mathbb{E}[\beta_i|D_i] = \mathbb{E}[\beta_i] =: \beta_{ATE}$$

which is called the *average treatment effect*, i.e. the effect for all individuals regardless of their actual treatment status. We might be tempted to think that it is a disadvantage of OLS that it doesn't pick up the treatment effect for the whole population, but for practical purposes, this typically isn't so. Often individuals which aren't reached by our program aren't too relevant for our evaluation question either - e.g. we wouldn't observe unemployed economics PhDs participating in a basic literacy program, but we wouldn't be too interested either in the effect of the literacy program on them because we'd never choose to send them to that program anyway on apriori grounds. The treatment effect on the treated answers the question about how much better off we are by running the program (ignoring the cost of running it) compared to a world in which we shut it down entirely and for everyone.
In many situations, there is always some degree of self-selection into, or imperfect compliance with a particular treatment, but sometimes we have a good instrument Z for participation, e.g. a randomized encouragement to participate, some exogenous eligibility rule, or some factor that shifts exogenously individuals' cost of taking up the program.
This assignment Z makes only some people switch from control to treatment (*compliers*), and from treatment to control (*defiers*). But there are also individuals which participate no matter what (*always-takers*) or don't in any case (*never-takers*). In order to formalize that, we denote the treatment a person would receive if $Z_i = 1$ with D_{1i}, and for $Z_i = 0$, we would observe D_{0i}. Now let's also assume

1. Independence: $(D_i, \beta_i, \varepsilon_i) \perp\!\!\!\perp Z_i$
2. Monotonicity: $D_{1i} \geq D_{0i}$ for all individuals

[2] This is the far too complicated derivation I gave in recitation, just for completeness:

$$\hat{\beta}_{LS} \quad \longrightarrow \quad \frac{\mathbb{E}[D_iY_i] - \mathbb{E}[Y_i]\mathbb{E}[D_i]}{\mathbb{E}[D_i] - \mathbb{E}[D_i]^2} = \frac{\mathbb{E}[D_i\beta_i] - \mathbb{E}[D_i\beta_i]\mathbb{E}[D_i]}{\mathbb{E}[D_i] - \mathbb{E}[D_i]^2} = \frac{\left(\mathbb{E}[D_i] - \mathbb{E}[D_i]^2\right)\mathbb{E}[\beta_i|D_i = 1]}{\mathbb{E}[D_i] - \mathbb{E}[D_i]^2} = \mathbb{E}[\beta_i|D_i = 1]$$

by the law of iterated expectations.

Under these assumptions

$$\begin{aligned}
\hat{\beta}_{Wald} &= \frac{\mathbb{E}[Y_i|Z_i=1]-\mathbb{E}[Y_i|Z_i=0]}{\mathbb{E}[D_i|Z_i=1]-\mathbb{E}[D_i|Z_i=0]} = \frac{\mathbb{E}[D_i\beta_i|Z_i=1]-\mathbb{E}[D_i\beta_i|Z_i=0]}{\mathbb{E}[D_i|Z_i=1]-\mathbb{E}[D_i|Z_i=0]} \\
&= \frac{\mathbb{E}[D_{1i}\beta_i|Z_i=1]-\mathbb{E}[D_{0i}\beta_i|Z_i=0]}{\mathbb{E}[D_{1i}|Z_i=1]-\mathbb{E}[D_{0i}|Z_i=0]} \overset{indep.}{=} \frac{\mathbb{E}[(D_{1i}-D_{0i})\beta_i]}{\mathbb{E}[D_{1i}-D_{0i}]} \\
&\overset{LIE}{=} \frac{\mathbb{E}[\beta_i|D_{1i}=1,D_{0i}=0]\mathbb{P}(D_{1i}=1,D_{0i}=0)-\mathbb{E}[\beta_i|D_{1i}=0,D_{0i}=1]\mathbb{P}(D_{1i}=0,D_{0i}=1)}{\mathbb{P}(D_{1i}=1,D_{0i}=0)-\mathbb{P}(D_{1i}=0,D_{0i}=1)} \\
&\overset{monot.}{=} \frac{\mathbb{E}[\beta_i|D_{1i}>D_{0i}]\mathbb{P}(D_{1i}>D_{0i})}{\mathbb{P}(D_{1i}>D_{0i})} = \mathbb{E}[\beta_i|D_{1i}>D_{0i}] =: \beta_{LATE}
\end{aligned}$$

That is, Wald (and thereby all instrumental variables estimators) estimate the average effect of the treatment on the subpopulation of compliers with a particular instrument. The important point to take away from this is that each instruments has a different set of compliers, so e.g. in the Angrist and Evans paper on the twin births and same-sex instruments we'd expect the LATE for the effect of the third child on mothers of twins to be different from the LATE on mothers whose first two children were of the same sex - the group of mothers that have a third child in order to balance the sex composition of their offspring is different from those mothers who have twins at their second birth (and therefore automatically have a third child). This is again not a weakness of the estimator, but we just have to be aware that each instrument defines its own Wald *estimand*. If Z is our policy intervention (e.g. offering a training program for which participation is voluntary), the instrumental variables estimator gives us directly the answer about the effect on those individuals which were affected by that policy, excluding people who would have received treatment in any case. And that's the actual question we'd typically want to ask if we evaluate the policy corresponding to Z: we want to know by how much e.g. offering that particular training program makes everyone better off given that on the one hand in a world without that intervention there could still be close substitutes available, and that on the other hand, many people wouldn't want to take part in the program either way.

1.3 Example: General Exam 2003, Question I

This part is based on Question I from the 2003 Labor Generals. We are given a human capital earnings function

$$y_i = \alpha_i + b_i S_i - \frac{1}{2}k_1 S_i^2$$

and individuals face a convex cost of schooling

$$c_i = \kappa_i + r_i S_i + \frac{1}{2}k_2 S_i^2$$

where r_i is often interpreted as the individual's discount rate, or the opportunity cost of capital (for empirical purposes, earnings y_i is actually understood to be in logs, but for the theoretical derivations, it should be just earnings). In order to determine the optimal level of schooling, the individual equalizes the marginal return to schooling to the marginal cost,

$$b_i - k_1 S_i = r_i + k_2 S_i \Longleftrightarrow S_i^* = \frac{b_i - r_i}{k_1 + k_2}$$

If $k_2 = 0$, the marginal return to schooling at the level of education chosen by the individual is

$$\frac{\partial y_i}{\partial S_i} = b_i - k_1 S_i^* = r_i$$

which is also the equilibrium condition for the classic model we saw in class. But if the marginal cost of education is strictly convex ($k_2 > 0$), we get instead

$$\frac{\partial y_i}{\partial S_i} = b_i - k_1 \frac{b_i - r_i}{k_1 + k_2} = \frac{k_2}{k_1 + k_2} b_i + \frac{k_1}{k_1 + k_2} r_i$$

which is a convex combination of the discount rate and b_i.
Now we want to estimate the return to schooling using a binary instrumental variable Z to address omitted variable bias and potential measurement error issues. More specifically, we assume that the instrumental variable shifts the capital cost (or discount rate for each individual, i.e. the individual faces $r_i = r_{0i}$ if the instrument $Z_i = 0$, and $r_i = r_{1i}$ if it takes the value 1. If the instrument Z is independent of (b_i, r_{0i}, r_{1i}), we can rewrite

$$\begin{aligned}
\text{plim}_N \hat{\beta}_{IV} &= \frac{\mathbb{E}[y_i|Z_i = 1] - \mathbb{E}[y_i|Z_i = 0]}{\mathbb{E}[S_i|Z_i = 1] - \mathbb{E}[S_i|Z_i = 0]} \\
&= \frac{\mathbb{E}[b_i S_i - k_1 S_i^2|Z_i = 1] - \mathbb{E}[b_i S_i - k_1 S_i^2|Z_i = 0]}{\mathbb{E}[S_i|Z_i = 1] - \mathbb{E}[S_i|Z_i = 0]} \\
&\overset{indep}{=} \frac{\mathbb{E}[\Delta S_i b_i] - \frac{1}{2} k_1 \mathbb{E}[\Delta(S_i^2)]}{\mathbb{E}[\Delta S_i]}
\end{aligned}$$

because of independence of (r_{0i}, r_{1i}, b_i) and Z_i. From the last part, we know that

$$\Delta S_i = \frac{b_i - r_{1i}}{k_1 + k_2} - \frac{b_i - r_{0i}}{k_1 + k_2} = \frac{r_{0i} - r_{1i}}{k_1 + k_2}$$

and

$$\begin{aligned}
\Delta(S_i^2) &= S_{1i}^2 - S_{0i}^2 \overset{(*)}{=} (S_{1i} + S_{0i})(S_{1i} - S_{0i}) \\
&= 2\bar{S}_i \Delta S_i = 2\bar{S}_i \frac{r_{0i} - r_{1i}}{k_1 + k_2}
\end{aligned}$$

where the "trick" in (*) is to use the third binomial formula $(a + b)(a - b) = a^2 - b^2$ (though brute force would have given the same result). Plugging this back into the plim for the Wald estimator, and noticing that the $(k_1 + k_2)$ terms cancel out,

$$\text{plim}_N \hat{\beta}_{IV} = \frac{\mathbb{E}[(r_{0i} - r_{1i})b_i] - k_1 \mathbb{E}[(r_{0i} - r_{1i})\bar{S}_i]}{\mathbb{E}[r_{0i} - r_{1i}]}$$

From this we can see that the Wald estimator weights individual returns by $r_{0i} - r_{1i}$, so that individuals who experience a high shift in the discount rate as a result of the intervention are "over-represented" relative to the other persons.
Card's argument about "discount rate bias" (he doesn't call it like that) is that the instrumental variables which are typically used in the literature induce individuals to increase their schooling levels who would have received relatively low *levels* schooling otherwise, but therefore also have relatively high *marginal returns*. This means that if we are interested in a "population average" of the parameters in the HCEF, our IV estimates of the return to schooling are biased upwards. On the other hand one could argue that the effect on the subpopulation which is affected by the instrument is a more interesting policy parameters as long as that subpopulation is similar to the population which would be affected by a particular policy measure. E.g. as its title already states, the Angrist and Krueger paper which used quarter of birth dummies as instruments for high-school dropouts adequately estimates the effect of mandatory schooling on kids for which the legal constraints are actually "binding"/who are at the margin of dropping out in the absence of mandatory schooling.