LECTURE NOTE 10: INSTRUMENTAL VARIABLES

1 Instrumental Variables with Homogeneous Causal Effects

The method of instrumental variables (IV) can help us estimate the causal effect of X_i on Y_i. We first discuss IV when the causal effect is homogeneous (does not vary across individuals). Suppose we wish to estimate the effect of X_i on Y_i in the following equation:

$$Y_i = \beta_0 + \beta_1 X_i + U_i$$

However, X_i is endogenous, meaning that it is correlated with U_i. For instance, when X_i is years of schooling, and Y_i is earnings, researchers worry that both X_i and Y_i are correlated with academic ability, an unobserved variable in the error term U_i. We can represent this situation in the following causal graph:

$$\begin{array}{ccc} X_i & \rightarrow & Y_i \\ \uparrow & \nearrow & \\ U_i & & \end{array}$$

Because $cov(X_i, U_i) \neq 0$, we cannot estimate the model by OLS.

Our solution is to find an instrument Z_i that causes X_i but is unrelated to U_i:

$$\begin{array}{ccccc} Z_i & \rightarrow & X_i & \rightarrow & Y_i \\ & & \uparrow & \nearrow & \\ & & U_i & & \end{array}$$

As the diagram suggests, two conditions are necessary for Z_i to be a valid instrument:

1. *Instrument relevance*: $cov(Z_i, X_i) \neq 0$. Z_i must be correlated with X_i.
2. *Instrument exogeneity*: $cov(Z_i, U_i) = 0$. Z_i must be uncorrelated with other determinants of Y_i. This condition is also called an *exclusion restriction* because, conditional on X_i, Z_i can be excluded from the system. Most disagreements about instrument validity concern this assumption.

How exactly can we use Z_i to estimate β_1? The answer becomes apparent when we derive the covariance between Z_i and Y_i:

$$cov(Z_i, Y_i) = cov(Z_i, \beta_0 + \beta_1 X_i + U_i) = \beta_1 cov(Z_i, X_i) + cov(Z_i, U_i)$$

By assumption, $cov(Z_i, U_i) = 0$. So we can rearrange terms and obtain:

$$\beta_1 = \frac{cov(Z_i, Y_i)}{cov(Z_i, X_i)} = \frac{cov(Z_i, Y_i)/V[Z_i]}{cov(Z_i, X_i)/V[Z_i]}$$

The second equality tells us that β_1 is the ratio of two coefficients. The numerator is the coefficient on Z_i from a regression of Y_i on Z_i, while the denominator is the coefficient on Z_i from a regression of X_i on Z_i. This expression suggests the following procedure:

1. Run a regression of Y_i on Z_i. This regression is often called the *reduced form regression*.
2. Run a regression of X_i on Z_i. This regression is often called the *first stage regression*.
3. Divide the coefficient on Z_i in the reduced form by the coefficient on Z_i in the first stage.

The ratio of the reduced form coefficient to the first stage coefficient is called an instrumental variables estimator. It is a consistent estimator of β_1.

We can obtain the same estimator by another procedure, called *two-stage least squares* (*TSLS* or *2SLS*). Appropriately, two-stage least squares involves two steps:

1. *First stage*: Estimate the first stage regression, $X_i = \pi_0 + \pi_1 Z_i + V_i$, by OLS and generate a predicted value for X_i, $\hat{X}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i$.
2. *Second stage*: Estimate the causal model of interest by OLS, using the predicted value of X_i instead of the actual value of X_i. In other words, run: $Y_i = \beta_0^{TSLS} + \beta_1^{TSLS} \hat{X}_i + \varepsilon_i$.

Is case of a single endogenous variable and a single instrument, the TSLS estimator is the same as the ratio of the reduced form coefficient to the first stage coefficient.[1] The variance of the TSLS estimator must take into account uncertainty from both the first and second stage regressions. As a result, if one estimates the second stage as a standalone OLS regression, the standard errors will be too small. We will not dwell on these details in this course, but the TSLS estimator has the asymptotic (large-sample) distribution:

$$\hat{\beta}_1^{TSLS} \sim \mathcal{N}\left(\beta_1, \frac{1}{N}\frac{V\left[(Z_i - E[Z_i])U_i\right]}{\left(cov(Z_i, X_i)\right)^2}\right)$$

The standard error of $\hat{\beta}_1^{TSLS}$ is the square root of the variance. In large samples, we can compute p-values and confidence intervals as usual.

To implement TSLS in R, you can still use the `feols()` function in the `fixest` package. For endogenous covariate `x`, instrument `z`, and two control variables `w1` and `w2`, type: `feols(y ~ w1 + w2 | x ~ z, data = df)`. In Stata, use the `ivregress` command. For the same regression, type: `ivregress 2sls y (x = z) w1 w2`. In both cases, you can modify the standard errors (robust, clustered, etc.) as usual.

[1] If you are interested, you can derive this result yourself by plugging $\hat{X}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i$ into $\hat{\beta}_1^{TSLS} = cov(\hat{X}_i, Y_i)/V[\hat{X}_i]$.

When Z_i is a binary variable, the IV estimator simplifies considerably. Define $\bar{Y}_0$ and $\bar{Y}_1$ as the means of Y_i in the sub-samples with $Z_i = 0$ and $Z_i = 1$, respectively. Define $\bar{X}_0$ and $\bar{X}_1$ similarly. Then:

$$\hat{\beta}_1^{Wald} = \frac{\bar{Y}_1 - \bar{Y}_0}{\bar{X}_1 - \bar{X}_0}$$

is a consistent estimator for β_1. This ratio is known as the *Wald estimator*.

The Wald estimator highlights a link between IV estimation and the analysis of randomized controlled trials. To see this link, consider a program evaluation with eligibility randomization. Let Z_i be eligibility, and let X_i be actual participation in the program. Under the assumptions discussed in Lecture Note 9, ineligibles cannot participate in the program, so $\bar{X}_0 = 0$. Meanwhile, $\bar{X}_1$ is the participation rate among eligibles, and $\bar{Y}_1 - \bar{Y}_0$ is the ITT, so $\hat{\beta}_1^{Wald}$ is a consistent estimator of the TOT.

You may remember from the 120 series that you can have multiple endogenous covariates, multiple instruments, and exogenous control variables in an IV regression. In this case, TSLS runs multiple first-stage regressions, one for each endogenous covariate, and then runs a single second-stage regression using the predicted values from the first-stage. Every first-stage regression includes ALL of the instruments and ALL of the exogenous control variables. The second-stage regression includes the predicted values and ALL of the exogenous covariates. The number of instruments must be at least as large as the number of covariates. In this course, we will generally stick to working with one instrument and one endogenous regressor.

2 Instrumental Variables with Heterogeneous Causal Effects

We now consider the meaning of the IV (or TSLS) estimator when the causal effect of X_i on Y_i is heterogeneous. To do so, we return to the potential outcomes framework of Lecture Note 9. For simplicity, we start with the case of a single binary instrument, Z_i, and a single binary endogenous regressor, X_i. Because we are assuming that X_i is binary, we will refer to X_i as individual i's treatment status. Denote the *instrument level* as z and the *treatment level* as x. Let $Y_i(x, z)$ be the potential outcome for individual i at treatment level x and instrument level z, and let $X_i(z)$ be the potential treatment status for individual i at instrument level z. As in Lecture Note 9, every individual has her own set of potential outcomes and potential treatment statuses. In our data, we observe only $Y_i = Y_i(X_i(Z_i), Z_i)$, $X_i = X_i(Z_i)$, and Z_i.

Suppose we run two-stage least squares, using Z_i as an instrument for X_i. We know how to interpret $\hat{\beta}_1^{TSLS}$ when the effect of X_i is the same for all individuals, but the interpretation with heterogeneous effects is slightly more complex. To make such an interpretation possible, we make three assumptions:

1. Independence: $\{Y_i(x, z), X_i(z)\} \perp Z_i$. Potential outcomes and potential treatment levels are independent of the instrument.

2. Exclusion restriction: $Y_i(x,0) = Y_i(x,1)$. Conditional on x, the value of z does not affect the outcome. We can thus exclude z from the potential outcomes function: $Y_i(x,z) = Y_i(x)$.
3. Monotonicity: Either $X_i(0) \geq X_i(1)$ for all i or $X_i(0) \leq X_i(1)$ for all i. The instrument level affects the treatment level in weakly the same direction for all individuals. Individuals who are induced be be treated (i.e., those for whom $X_i(0) \neq X_i(1)$) are known as *compliers*.

Assumption (1) implies that the reduced form regression and the first stage regression are *identified*, meaning that we can interpret the coefficients on Z_i as causal. Assumption (2) implies that the effect of Z_i on Y_i is entirely mediated by X_i. Note that although assumption (2) is called an exclusion restriction, it is different from the exclusion restriction in Section 1 (which we also called the instrument exogeneity assumption). The exclusion restriction in Section 1 required that $cov(Z_i, U_i) = 0$. Here, assumption (2) is insufficient for that requirement. Assumptions (1) and (2) together guarantee that $cov(Z_i, U_i) = 0$.

Assumption (3), the monotonicity assumption, is key. It states that if Z_i increases X_i for any one individual, then for no individual does Z_i decrease X_i. In a policy experiment with eligibility randomization, monotonicity implies that eligibility does not prevent any individual from participating in the program. Lecture Note 9 assumed that ineligibles could not participate in the program, in which case monotonicity is guaranteed.

Under these assumptions, the TSLS estimator converges in probability to an estimand called the *local average treatment effect* ($LATE$). The $LATE$ is the average treatment effect among compliers:

$$\hat{\beta}_1^{TSLS} \xrightarrow{p} LATE = E[Y_i(1) - Y_i(0) | X_i(0) \neq X_i(1)]$$

The $LATE$ is closely related to the TOT in a randomized experiment. In particular, both the $LATE$ and the TOT measure the average treatment effect among individuals who were induced to be treated. When ineligibles cannot participate in the program, individuals who were induced to be treated *are* the treated population. For the $LATE$, we allow for the existence of treated individuals who would have been treated even in the absence of the experiment. (We call such individuals *always-takers*.) In this case, individuals who were induced to be treated are a subset of the treated population. But in the absence of always-takers, the $LATE$ and the TOT are the same.

We can generalize the local average treatment effect to non-binary instruments and non-binary endogenous variables, as well as multiple instruments and multiple endogenous variables. When X_i and Z_i are non-binary, the TSLS estimator again converges to a particular weighted average treatment effect. The result is easiest

to see through the lens of the first and second stage regressions:

$$\begin{aligned} X_i &= \pi_{0i} + \pi_{1i} Z_i + V_i \\ Y_i &= \beta_{0i} + \beta_{1i} X_i + U_i \end{aligned}$$

Under the assumptions above, two-stage least squares leads to:

$$\hat{\beta}_1^{TSLS} \xrightarrow{p} E\left[\frac{\pi_{1i}}{E\left[\pi_{1i}\right]}\beta_{1i}\right]$$

which is a weighted average of β_{1i}, giving more weight to individuals whose X_i was more affected by the instrument. With heterogeneous causal effects, instrumental variables methods uncover a weighted average of the causal effects that gives more weight to individuals who are more sensitive to the instrument.