Econometrics of Policy Evaluation: Instrumental Variables (IVs)

Cristian Huse

- Remember **exogeneity**, the core OLS assumption: $E(X'u) = 0$.
  - This means the treatment/regressor (X) must be uncorrelated with the error term (u).
- **Endogeneity** is the failure of this assumption: $E(X'u) \neq 0$.[1] When does it happen?
  - **Omitted Variable Bias:** An unobserved variable (W, like "ability" or "motivation") affects both X (e.g., "Education") and Y (e.g., "Wages"). This W is part of the error term u, causing correlation. This **unobserved variable** is the source of **selection bias** identified in the **FPCI.**
  - **Simultaneity/Reverse Causality:** X causes Y, but Y also causes X.
- **Conclusion:**
  - When we have endogeneity (or "selection on unobservables"), OLS is biased and inconsistent.
  - We need a new tool: Instrumental Variables.

[1] In what follows, the following will be used as equivalent: $Cov(X, u) \neq 0$, $E[u|X] \neq 0$ and $E(X'u) \neq 0$. They all denote the failure of the exogeneity assumption.

- Randomized assignment (implicitly) assumes that an entity (policy maker, program administrator etc) has the power to assign units to treatment and control groups, with those assigned fully complying to the decision imposed on them
  - i.e., units that are assigned to the treatment and control groups **fully comply** with their assignment
    $\Rightarrow$ **Full compliance assumption**
- Full compliance is easier to attain in lab settings/medical trials, where the researcher can carefully make sure that...
  1. all subjects in the treatment group take a given treatment; and
  2. none of the subjects in the control group take it

- **Full compliance is often unrealistic in real life**
  - Some programs allow potential participants to choose to enroll (self-selection into treatment) and thus are not able to exclude potential participants who want to enroll
  - Some programs have a budget that is big enough to supply the program to the entire eligible population, so randomly assigning people to treatment and control groups – and excluding potential participants for the sake of an evaluation – would not be ethical
- Given the strength of the full compliance assumption, one needs alternative ways to perform program evaluation: focus here is on the **IV method**

- In policy evaluation, one always compares TGs and CGs:
  - Full compliance ⇒ ATE (Average Treatment Effects)
- **Imperfect compliance** can take two forms:
  - Some individuals assigned to the TG choose not to enroll or are otherwise left untreated
    - **Example:** in teacher-training, some teachers assigned to the TG do not show up to do the course (perhaps because they are unmotivated)
  - Some individuals assigned to the CG manage to participate in the program
    - **Example:** in teacher-training, the most motivated teachers in the control group manage to attend the course somehow. Thus, the most motivated teachers in the TG would have no counterparts in the control group, and so it would not be possible to estimate the impact of the training for that segment of motivated teachers
- Without full compliance, can estimate the effect of offering a program or the effect of participating in the program

Imperfect Compliance is Endogeneity

- The model one wishes to estimate is $Y_i = \beta_0 + \beta_1 P_i + \varepsilon_i$
- **The Problem:** The regressor P_i is endogenous.
- **Why? Self-Selection.** The decision to participate is a choice, not a random assignment.
  - This choice is driven by unobserved variables (e.g., motivation, ambition, need) that are hidden in the error term (ε_i).
  - These same unobserved variables (motivation) are also correlated with the outcome (Y_i).
- **Conclusion:** $Cov(P_i, \varepsilon_i) \neq 0$. This is the classic endogeneity problem. A simple OLS regression of outcome on participation is biased and inconsistent.

The Solution: Random Assignment as an Instrument

- To solve the endogeneity of "Actual Participation" (P_i), we need an Instrumental Variable (IV).
- **The Instrument** (Z_i): The random assignment (or "offer") of the program.
- **Why is it a great instrument?**
  - **Relevance:** The offer (Z_i) is a strong predictor of participation (P_i). (This is the first stage).
  - **Exclusion Restriction:** The offer (Z_i) is random, so it is exogenous and uncorrelated with the unobserved "motivation" factors in ε_i.
- This leads to our two key estimates...

The Solution: Random Assignment as an Instrument

- **Intention-to-Treat (ITT):**
  - This is the **Reduced Form** (Y on Z_i), which estimates the causal effect of the offer:

$$ITT = E(Y|Z = 1) - E(Y|Z = 0)$$

  - **In words:**
    - The ITT equation estimates the **causal effect of the assignment** (the offer) **on the outcome**.
    - ITT is the average outcome (Y) for the group **assigned to treatment** ($Z = 1$) minus the average outcome (Y) for the group **assigned to control** ($Z = 0$).
    - ITT simply compares the means of the two **randomized** groups, **ignoring whether they actually participated or not**.

The Solution: Random Assignment as an Instrument

- **Treatment-on-the-Treated (TOT) / LATE:**
  - This is the **IV Estimate** (Y_i on P_i, instrumented by Z_i).
  - It estimates the **causal effect of participation** for the subgroup of **"compliers"**.
  - It is calculated as ("Wald Estimator", $Cov(Y, Z)/Cov(P, Z)$):

$$TOT = \frac{ITT}{First - stage} = \frac{E(Y|Z = 1) - E(Y|Z = 0)}{E(P_i|Z_i = 1) - E(P_i|Z_i = 0)}$$

  - **Interpretation:** The TOT is the **effect of the program** on the people who **actually participated** because they were offered it.
    - TOT scales up the ITT to account for the fact that **not everyone complied**, giving us the estimated effect for the "complier" subgroup;
    - The first-stage measures the "net compliance rate" (*"How much did the offer actually increase participation?"*)
    - $E[P \mid Z = 1]$: The share of people who participated given they were offered the program.
    - $E[P \mid Z = 0]$: The share of people who participated even though they were not offered the program ("crossovers").

The Solution: Random Assignment as an Instrument

- **Example:**
  - If **offering a tutoring program** increases test scores by 5 points ($ITT = 5$)
  - But only 50% of the students offered the tutoring **actually attended** ($First - stage = 0.50$)
  - Then the **effect on the students who actually attended** must be 10 points ($5/0.50 = 10$)
- **Note:**
  - The TOT/LATE is almost always larger (in absolute value) than the ITT.

$$|LATE| = \frac{|ITT|}{\text{Compliance Rate}}$$

  - Since Compliance Rate ≤ 1, the LATE inflates the ITT.
    - If compliance is 100%, $LATE = ITT$.
    - If compliance is 50%, $LATE = 2 \times ITT$
    - If compliance is 1%, $LATE = 100 \times ITT$

- **Instrumental variables (IVs)** can help evaluating programs with imperfect compliance, voluntary enrollment, or universal coverage
- The IV method relies on some **external source of variation** to determine treatment status
  - An instrument influences the **likelihood of participating in a program**, but is outside of the participant's control and is unrelated to the participant's characteristics
- IVs can be generated **ex ante**:
  - Randomized promotion (or encouragement design)
  - "Randomized offering" of a program
- IVs can be used **ex post** to correct for non-compliance to a program (details below)

The Two Conditions for a Valid Instrument

- For a variable Z to be a valid instrument for an endogenous variable X (to explain outcome Y) it has to satisfy:

1. **The Relevance Condition:** The instrument must be a strong predictor of the endogenous variable.
   1. **Formally:** $Cov(Z, X) \neq 0$.
   2. **How we check it:** This is testable. We check the **first-stage regression**. We look for a high F-statistic (Rule of thumb: F $>$ 10).
2. **The Exclusion Restriction:** The instrument must only affect the outcome Y **through** its effect on X.
   1. **Formally:** $Cov(Z, u) = 0$.
   2. **This means two things:**
      1. Z has **no direct causal effect** on Y.
      2. Z is **not correlated** with the unobserved factors (u) that also affect Y.
   3. **How we check it: This is an untestable assumption.** It must be **defended with economic theory, institutional knowledge, and a compelling story.** This is where all IV arguments are won or lost.

- Consider the evaluation of a voluntary program (job training, environmental etc)
  - Any person is eligible to take part (**Universal eligibility**)
  - Some people choose to register (**Participants**)
  - Other people choose not to register (**Non-participants**)
- Intuitive (**but not correct**) ways to evaluate said program include
  - Compare **before and after** situation in the **participant group**
  - Compare situation of **participants and non-participants** after the intervention
  - Compare situation of **participants and non-participants before and after (DD)**

- Compare outcomes for treatment and control using

$$y = \beta_0 + \beta_1 P + \beta_2 x + u$$

where $P = 1$[person participates in program] and x are control variables (exogenous and observed).

- What are the problems with this strategy?
  1. The existence of variables that we omit (for various reasons) but that are important.
  2. Decision to participate in program is **endogenous.**
     - ("self-selection into treatment")

- **Problem: Endogenous Decision to Participate**
  - Assume individuals have reasons to take part in the program
  - While we estimate $y = \beta_0 + \beta_1 P + \beta_2 x + u$, the **true model** is:

$$y = \gamma_0 + \gamma_1 P + \gamma_2 x + \eta$$
$$P = \pi_0 + \pi_1 x + \pi_2 M_2 + \xi$$

    i.e., the decision to participate is influenced by individual characteristics (denoted by M_2 above).

- **Is $\beta_{1,OLS}$ an unbiased and consistent estimator for γ_1?**
- To answer, compute the correlation between P and the error term

$$\begin{aligned} Corr(u, P) &= Corr(u, \pi_0 + \pi_1 x + \pi_2 M_2 + \xi) \\ &= \pi_1 Corr(u, x) + \pi_2 Corr(u, M_2) \\ &= \pi_2 Corr(u, M_2) \end{aligned}$$

- **In words:** If there is a correlation between the missing variables that determine participation (e.g., motivation) and outcomes not explained by observed characteristics, then the OLS estimator will be biased

- **Problem:** We cannot force participation. However, we can randomly assign a nudge/promotion to take part in a programme.
  - **Example:**
    - A Social Worker randomly visits 50% of households (Instrument Z).
    - If she is effective, many people she visits will enroll, i.e., there will be a correlation between receiving a visit and enrolling.
    - This is closely related to participation P but doesn't directly affect people's outcomes Y, **other than through its effect on participation**
    - Thus, randomized "encouragement" or "promotion" visits are an IV.

- This is an example of **Randomized promotion** – an IV method that allows us to estimate impact in an unbiased way.
  - It randomly assigns a promotion, or encouragement, to participate in the program.
  - It is a useful strategy to evaluate programs that are open to everyone who is eligible.
  - It leads to valid estimates of the counterfactual if the promotion campaign substantially increases take-up of the program without directly affecting the outcomes of interest.

- The instrument Z (promotion) is used to perform 2SLS estimation
  - **First-stage:** Regress (OLS) the endogenous variable on both the instrument and the exogenous explanatory variables (reduced-form equation)

$$x_i = \alpha_0 + \alpha_1 z_i + \alpha_2 w_i + \varepsilon_i$$

    and save the predicted values $\hat{x}_i = \hat{\alpha}_0 + \hat{\alpha}_1 z_i + \hat{\alpha}_2 w_i$
    - Note this relies on the **Relevance Condition** ($Cov(Z, P) \neq 0$)
  - **Second-stage:** The predicted value of x is substituted for its actual value in an OLS regression:

$$y_i = \beta_0 + \beta_1 \hat{x}_i + \beta_2 w_i + u_i$$

    and $\hat{\beta}_1$ is the IV estimator ($\hat{\beta}_{IV}$)
  - Since compliance was voluntary, the effect measured is only for those induced to participate by the Social Worker (the "Compliers").

- **Note:**
  - A "manual" 2SLS (running two OLS regressions and plugging $\hat{x}$ into the second) will produce incorrect standard errors. You must use a built-in command that calculates the correct asymptotic standard errors.
  - The IV estimator is consistent.
  - The IV estimator is asymptotically normally distributed (usual tests apply).
  - IV estimator has larger standard errors than does OLS.
    - This loss in efficiency results because 2SLS uses only a part of the variation in x.

What Does IV Estimate? The LATE

- IV does **not** estimate the ATE. It estimates the **Local Average Treatment Effect (LATE)** (Imbens & Angrist, 1994).
- To understand this, divide the population into three types (using your "Randomized Promotion" example):
  1. **Compliers:** People who only participate if they get the promotion (Z causes X for them).
  2. **Always-Takers:** People who participate anyway (promotion or not).
  3. **Never-Takers:** People who never participate (promotion or not).
- **Intuition:** It estimates the LATE because it focuses on the treatment effect for "compliers".

Identification Conditions

- **Condition for Identification:**
  - We need at least as many excluded instruments (Z) as we have endogenous regressors (X).
  - **Rank Condition:** $Cov(Z, X) = 0$ (Relevance)
  - **Order Condition:** Let L be the number of instruments and K be the number of endogenous regressors.
- **The Three Cases:**
  - **Underidentified** ($L < K$): Impossible to solve. (e.g., 1 endogenous variable, 0 instruments).
  - **Just-identified** ($L = K$): Unique solution. (e.g., 1 endogenous variable, 1 instrument).
    - Estimator: Simple IV.
  - **Over-identified** ($L > K$): More equations than unknowns. No unique solution.
    - Estimator: 2SLS or GMM (chooses the "best" combination of instruments).
    - Benefit: Allows for testing exogeneity (Sargan/Hansen J-Test).
    - **Over-identifying restrictions** are the additional restrictions generated by the "extra" instruments.

- A potential instrument can be flawed in two important ways:
  1. An instrument can be **invalid** if it is itself correlated with the error term in the equation of interest
     - Invalid instruments yield biased and inconsistent estimates
     - *A priori*, any potential instrument is suspected to be invalid
  2. An instrument can be too weakly correlated with the explanatory variable of interest (**weak** instrument)
     - Will not overcome OLS bias (since not enough exogenous variation)
     - Weak instrument yield misleading statistical significance
- **Searching for an IV *ex post* is both hard and risky!**
  - Generating an IV with information campaign designed *ex ante.*
    - If everyone is eligible to participate in treatment.
    - But some have more information than others (Those with more information will be more likely to participate) .
    - Provision of "additional information" on a random basis.

- **Efficiency-bias Trade-off:**
  - An increased number of identifying restrictions (i.e. more instruments) confers the benefit of a higher R^2 in the first-stage estimation
    - i.e., a larger exogenous part of the endogenous variable is captured by the first stage
  - Yields standard errors closer to those of OLS $\rightarrow$ **efficiency**
  - **Problem:** Including a bad instrument will **bias** the estimates, thus potentially a high-risk strategy
- **Note:**
  - The IV method is a general strategy to address endogeneity
  - There are several IV estimators which differ, for instance, in how they combine (weigh) instruments
    - e.g., 2SLS, GMM, LIML, FIML

- Valid instruments are typically (or ideally) derived from a natural or random experiment
- Instrument needs to be **exogenous** (uncorrelated with the error term u)
  - i.e., $Corr(Z, u) \neq 0 \Rightarrow$ "bad instrument"
- This requirement needs a strong **theoretical** argument (and can in general not be tested)
- Theoretical argument has to convincingly
  - Describe how the instrument influences the endogenous variable
    - i.e., what is its influence after controlling for the effect through other included regressors?
  - Rule out any direct effect of the instrument on the dependent variable or any effect running through omitted variables. This is sometimes called exclusion restriction
  - Rule out any reverse effect of the dependent variable on the instrument

- The exogeneity of instruments can in general not be tested
  - **Natural (but wrong, see below) idea:** add the instrument to the model and use OLS to test whether it has explanatory power in the equation
  - **Don't!** The presence of the endogenous variable will bias the OLS estimate!
  - However, over-identified equations do allow for a variant of this test...
- In case we have more instruments than necessary, we can perform the so-called **J-test of overidentifying restrictions (Sargan test)**
  - This tests whether all instruments are exogenous assuming that at least one instrument is exogenous
  - i.e., the Sargan test will **not** detect a situation in which all the instruments are endogenous (problematic in practice)

- **Overview of the Sargan Test** (H_0 : model correctly specified)
  - Estimate your IV regression
  - Regress the residuals on Z, i.e., on all exogenous variables of your model (including a constant $\rightarrow$ will need R^2 below)
  - Under the null that the regressors are truly exogenous, the correlation between said regressors and error term should be zero
    - (in practice: should be "small")
  - Now calculate $TR^2 \sim \chi^2$ with df equal to the amount of over-identification
- **Note:**
  - If we reject the null hypothesis, then our logic for choosing the IVs must be re-examined
  - If we fail to reject the null hypothesis, then we can have some confidence in the set of instruments used – up to a point, since the test does not tell us which IVs fail the exogeneity requirement (e.g., it could be one of them or all of them)

- Instruments with low correlation with the endogenous variable are weak instruments – i.e. $Corr(Z, P)$ low
- Empirical and theoretical evidence that IV estimation with weak instruments may perform worse than OLS (Stock, Wright and Yogo, 2002)
  - May generate biases, even for large samples
  - May lead to far too small standard errors in IV estimation (misleading)
- The weakness of instruments can be assessed in the first-stage regression
  - Rule of thumb: The F-statistic of a joint test whether all instruments are significant should be bigger than 10 in case of a single endogenous variable
  - **Example:** Single instrument/single endogenous variable, t-stat should be bigger than 3.2 (should always be reported!)
  - Also, check coefficient of included instrument in first-stage regression – want something significant here!

- **Aim:** Test whether there is an endogeneity problem in an application
- **(Durbin-Wu-)Hausman test** (if we have a set of valid instruments at hand)
- Compare the estimates obtained with OLS and IV:

$$H = (\hat{\beta}_{IV} - \hat{\beta}_{OLS})'[\hat{Var}(\hat{\beta}_{IV}) - \hat{Var}(\hat{\beta}_{OLS})](\hat{\beta}_{IV} - \hat{\beta}_{OLS}) \sim \chi^2_J$$

  where J is the number of endogenous regressors
- **Null hypothesis:** There are no endogenous variables or that endogeneity is not a problem for OLS
  - **Intuition:** If that is the case, distance between OLS and IV estimates is low, so test statistic is low and null is not rejected
- **Note:**
  - If the instruments are not valid, the Hausman test is not valid
  - We can perform this test if we already have the means to solve the endogeneity problem

Returns to Schooling

- **Paper:** Angrist & Krueger (1991).
- **Problem:** Effect of Education (X) on Wages (Y).
- **Endogeneity:** Unobserved "Ability" (u) affects both.
- **Instrument:** Quarter of Birth (QOB).
- **Relevance (The Story):** Compulsory schooling laws + fixed school start dates mean students born late in the year are forced to get slightly more schooling than students born early in the year.
- **Exclusion Restriction (The Debate):** Is QOB truly random? Does it affect wages only through education?
- **LATE:** The effect of education only for those people whose years of schooling were changed by their quarter of birth.

Institutions and Growth

- **Paper:** Acemoglu, Johnson, & Robinson (2001).
- **Problem:** Effect of "Good Institutions" (X) on modern GDP (Y).
- **Endogeneity:** Rich countries can afford good institutions (reverse causality).
- **Instrument:** European Settler Mortality rates (1600s-1800s).
- **Relevance (The Story):** High mortality $\rightarrow$ Europeans set up "extractive" institutions. Low mortality $\rightarrow$ Europeans settled and set up "good" institutions (e.g., property rights). These institutions persisted.
- **Exclusion Restriction:** Historical mortality rates from centuries ago only affect current GDP through their persistent effect on current institutions.

Natural Experiments

- **Paper:** Duflo & Pande (2007).
- **Problem:** Effect of Dams (X) on Poverty (Y).
- **Endogeneity:** Dams are built in non-random locations (e.g., in richer or poorer districts).
- **Instrument:** River Gradient (steepness).
- **Relevance:** Large dams can only be viably constructed on rivers with a sufficient gradient.
- **Exclusion Restriction:** River gradient is a geographic feature that only affects poverty through its suitability for a dam, not through any other channel (e.g., gradient itself doesn't make a district poor).

- **Example. ITT Estimates (intention-to-treat)**
  - Use data collected **after** the start of the program (`round=1`)
  - Regress the outcome of interest, `health expenditures`, on the binary variable `treatment_locality` (=1 if household is in treatment area, regardless of whether it took up the program)
  - The **ITT estimate** is -6.4, i.e., HHs in villages where HISP was offered on average spent $6.4 less on health expenditures than HHs in villages where HISP was not offered
  - Remember that 59 percent of HHs enrolled, but 41 percent of HHs did not participate. Therefore, when computing average outcomes in the TG, we obtain a weighted average of outcomes among the 59 percent of HHs that participated in the program and the 41 percent of HHs that chose not to participate in the program
  - The **ITT estimate** is obtained by taking the difference between the weighted average in the TG and the same weighted average of outcomes in the CG. As a result, it captures the "intent-to-treat" impacts of offering the program to an average HH in a treatment locality, independently of who participates within the locality

```
Stata Example 5. "Intent-to-Treat" (ITT) Estimates

* In this context, the program is randomized at the village level.
* While everyone is eligible for the program in treatment communities, not everyone
participates.

*Select the relevant data
use "evaluation.dta", clear
drop eligible

* You can estimate 'intent-to-treat estimates', i.e. program impact at the village-
level irrespective of who takes up the program or not.

reg health_expenditures treatment_locality if round ==1, cl(locality_identifier)

Linear regression                               Number of obs =    9914
                                                F(  1,   199) =  163.93
                                                Prob > F      =  0.0000
                                                R-squared     =  0.0726
                                                Root MSE      =  11.451

                        (Std. Err. adjusted for 200 clusters in locality_identifier)
------------------------------------------------------------------------------
                    |               Robust
health_expenditu~s |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]

--------------------+---------------------------------------------------------------
treatment_locality |  -6.406008   .5003313   -12.80   0.000     -7.39264   -5.419376
             _cons |   20.06416   .3763165    53.32   0.000     19.32208    20.80624
------------------------------------------------------------------------------
```

- **Example. ITT Estimates cont'd**
  - In contrast, the "**local average treatment effect**" seeks to estimate treatment effects only on the subgroup of units that actually comply with their treatment assignment
  - In general, $|\delta_{LATE}| \geq |\delta_{ITT}|$ since $\delta_{ITT} = Take - up \times \delta_{LATE}$ and $Take - up \in [0, 1]$, i.e., it is a probability
  - See Gertler et al (2016)'s **Technical Companion** for a detailed discussion

- **Example. LATE**
  - The IV is `treatment_locality`, (=1 if HISP was randomly offered to households in a given locality, 0 otherwise)
  - In the first-stage regression, a variable capturing whether a household was enrolled or not in the program is regressed on the randomization dummy (`treatment_locality`):

$$P_i = \gamma_0 + \gamma_1 X_i + \gamma_2 Z_i + \eta_i$$

("**effect of offering the program on actual participation**")

  - The estimate of 0.598 indicates that approximately 59.8% of households enrolled in HISP when the program was offered in their locality

Stata Example 6: "Local Average Treatment Effect" (LATE) 2SLS IV Estimates[13]

```
* You can back out 'local average treatment effect' estimates on complier units that
do take-up the program in treatment communities

ivreg health_expenditures (enrolled = treatment_locality) if round ==1, first

First-stage regressions
-----------------------
```

```
      Source |       SS       df       MS              Number of obs =    9914
-------------+------------------------------           F(  1,  9912) = 7361.23
       Model |  885.675858     1  885.675858           Prob > F      =  0.0000
    Residual |   1192.5756  9912  .120316344           R-squared     =  0.4262
-------------+------------------------------           Adj R-squared =  0.4261
       Total |  2078.25146  9913  .209649093           Root MSE      =  .34687
```

```
------------------------------------------------------------------------------
      enrolled_ro |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
treatment_locality |   .5977823   .0069674    85.80   0.000     .5841248    .6114397
             _cons |   2.12e-14   .0049282     0.00   1.000    -.0096602    .0096602
------------------------------------------------------------------------------
```

- **Example. LATE cont'd**
  - The second-stage regression uses the predicted enrollment from the first stage as a regressor to explain variation in the outcomes of interest

$$Y_i = \alpha + \delta_{LATE}\hat{P}_i + \gamma X_i + \epsilon_i$$

    - The estimate for δ_{LATE} suggests that participation in the HISP program lowers health expenditures by \$10.7. This is the same number that we obtained by multiplying the ITT estimate by the share of units enrolled in the treatment group using $\delta_{ITT} = Take - up \times \delta_{LATE}$

```
Instrumental variables (2SLS) regression

      Source |       SS       df       MS              Number of obs =    9914
-------------+------------------------------           F(  1,  9912) =  944.81
       Model |  334474.116     1  334474.116           Prob > F      =  0.0000
    Residual |  1067039.56  9912  107.651288           R-squared     =  0.2387
-------------+------------------------------           Adj R-squared =  0.2386
       Total |  1401513.68  9913  141.381386           Root MSE      =  10.376

------------------------------------------------------------------------------
health_exp~s |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   Enrolled  |  -10.71629    .348636   -30.74   0.000    -11.39969   -10.03289
       _cons |   20.06416   .1474116   136.11   0.000      19.7752    20.35312
------------------------------------------------------------------------------
Instrumented:  enrolled
Instruments:   treatment_locality
------------------------------------------------------------------------------
```

[13] The Stata command used here is ivreg. A more recent version of the command is called ivregress 2sls. The do-file for the technical companion provides the syntax for both commands. Note that the option "first" added at the end of the command ensures that the results of the first-stage of the two-stage estimation process are also displayed.

- **Example. Randomized Promotion**
  - Randomized promotion is another example of an IV
    - The program itself is not is randomized, but there is, e.g., an information campaign that can increase program take-up
  - When randomized promotion is used, the approach essentially seeks to generate a valid IV that is correlated with program participation, but uncorrelated with the outcome of interest – aside from its effect through participation
  - First-stage: as above – want to isolate the effect of randomized promotion on program participation
  - Second stage: regress the outcome of interest on the predicted value of program take-up from the first stage
    - i.e., the variation in program take-up that is driven by the randomized promotion campaign
- **Note**
  - Assume HISP is offered universally throughout the country, i.e., it is not randomized. Next, assume a promotion campaign takes place in a random sub-sample of villages, aimed at increasing awareness of HISP. This promotion campaign is an IV because (1) it seeks to increase enrollment in HISP; (2) it does not directly affect the outcome indicator of interest (health expenditures)

Stata Example 7. 2SLS IV Estimates for Randomized Promotion[14]

```
* In this context, everyone is eligible for the program. You compare what happens in
promoted and non-promoted villages.

*Select the relevant data
use "evaluation.dta", clear
drop eligible
drop treatment_locality
drop enrolled

ivreg health_expenditures (enrolled_rp = promotion_locality) if round ==1, first

First-stage regressions
-----------------------

      Source |       SS       df       MS              Number of obs =    9914
-------------+------------------------------           F(  1,  9912) = 2484.60
       Model |  411.879408     1  411.879408           Prob > F      =  0.0000
    Residual |  1643.13855  9912  .165772654           R-squared     =  0.2004
-------------+------------------------------           Adj R-squared =  0.2003
       Total |  2055.01795  9913  .207305352           Root MSE      =  .40715

------------------------------------------------------------------------------
      enrolled_rp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
------------------+-----------------------------------------------------------
promotion_locality |   .4077847   .0081809    49.85   0.000     .3917484     .423821
            _cons |   .0842476   .0058578    14.38   0.000      .072765    .0957301
------------------------------------------------------------------------------

Instrumental variables (2SLS) regression

      Source |       SS       df       MS              Number of obs =    9914
-------------+------------------------------           F(  1,  9912) =  337.77
       Model |  310737.314     1  310737.314           Prob > F      =  0.0000
    Residual |  1090776.36  9912  110.046042           R-squared     =  0.2217
-------------+------------------------------           Adj R-squared =  0.2216
       Total |  1401513.68  9913  141.381386           Root MSE      =   10.49

------------------------------------------------------------------------------
health_exp~s |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 enrolled_rp |  -9.499769   .5168948   -18.38   0.000    -10.51299    -8.48655
       _cons |   19.64571   .1846287   106.41   0.000      19.2838    20.00762
------------------------------------------------------------------------------
Instrumented:  enrolled_rp
Instruments:   promotion_locality
```

[14] The Stata command used here is ivreg. A more recent version of the command is called ivregress 2sls. The do-file for the technical companion provides the syntax for both commands. Note that the option "first" added at the end of the command ensures that the results of the first-stage of the two-stage estimation process are also displayed.

- **Example. Randomized Promotion cont'd**
  - The first stage identifies the effects of the promotion activities on program take-up: promotion activities increase program take-up by 40.8 percent
  - In the second stage, we regress the outcome variable on the predicted program participation from the first stage to obtain the LATE estimates. In this case, the results suggest that participation in the HISP program lowers health expenditures by $9.5

- Are the baseline characteristics balanced between the units that received the promotion campaign and those that did not?
  - Compare the baseline characteristics of the two groups
- Does the promotion campaign substantially affect the take-up of the program? It should
  - Compare the program take-up rates in the promoted and the non-promoted sub-samples
- Does the promotion campaign directly affect outcomes? It should not
  - This cannot usually be directly tested, so you need to rely on theory, common sense, and good knowledge of the setting of the impact evaluation for guidance.

- ATE is the straight difference in average outcomes between the group to whom you offered treatment, and the group to whom you did not offer treatment
- To compute the TOT (Effect of Treatment on the Treated), use the randomized offering as an IV (Z) for whether people accepted the treatment (P)
- **IV is a "local effect"**
  - IV methods identify the average gains to persons induced to change their choice by a change of the instrument (referred to as compliers)
  - However we cannot identify who these people are ("local average treatment effect" or LATE)
  - Different instruments will identify different parameters and answer different questions
  - Caution in extrapolating to the whole population

IV vs. Other Methods

- **DD** relies on the common trends assumption and requires panel data whereas IV relies on an **exclusion restriction**, a variable that affects the outcome only through the treatment.
- IV does not require panel data nor the common trends assumption. However, finding a valid instrument is often harder than finding a control group.
- **Fuzzy RD** arises when compliance is imperfect. It is estimated using an IV estimator where the instrument Z is the indicator of being above the cutoff.
  - LATE is the treatment effect for "compliers" (those whose status changes because of the cutoff/instrument)

IV vs. Other Methods

- **Matching** relies on the "selection on observables" (or Unconfoundedness) assumption, which posits that conditional on observed covariates X, treatment is random $(Y(0), Y(1) \perp D|X)$. In contrast, IV is explicitly designed for "selection on unobservables".
  - If we believe selection is driven only by observed variables like age or education, we should use Matching. We use IV only when we suspect unobserved factors like motivation or ability are driving selection.

- **Full compliance** is often unrealistic, making the claim that randomization yields you an ATE untenable
- **Challenge:** Finding a good instrument
  - An IV influences the likelihood of participating in a program, but is outside of the participant's control and is unrelated to the participant's characteristics
  - Searching for an IV *ex post* is both hard and risky
- An instrument can be flawed in two ways:
  - **Invalid** – it is correlated with the error term $\Rightarrow$ IV yields biased and inconsistent estimates
  - **Weak** – is weakly correlated with the instrumented variable $\Rightarrow$ OLS biased, misleading statistical inference
- Solid/convincing estimates as a combination of institutional knowledge and estimation method

- Gertler et al (2016). Impact Evaluation in Practice, 2nd. Edition. Washington, DC: Inter-American Development Bank and World Bank
  - Chapter 5
- Gertler et al (2016). Impact Evaluation in Practice, 2nd. Edition, Technical Companion (Version 1.0). Washington, DC: Inter-American Development Bank and World Bank.
  - p. 11-19

The IV Causal Diagram (DAG)

- A visual representation may help to understand the IV issue (let U be the confounder/unobservable variable).
  - Endogeneity problem: $X \leftarrow U \rightarrow Y$.
  - Adding the instrument: $Z \rightarrow X \leftarrow U \rightarrow Y$.
- **Explanation:**
  - $Z \rightarrow X$: This is the Relevance condition (an arrow exists).
  - Z has no arrow to Y: This is the Exclusion Restriction (no direct effect).
  - Z has no arrow to U: This is also the Exclusion Restriction (instrument is exogenous and not correlated with the confounder).
- **The Story:**
  - The instrument Z provides a "clean" or "exogenous" source of variation in X.
  - 2SLS works by isolating only the part of X that is "caused" by Z and using it to explain Y.