Lecture 4: Instrumental variables

P.J. Messe[1]

[1]Le Mans Université GAINS-TEPP, CEET, LEMNA

Master in Applied Econometrics

Outline

The endogeneity issue

- Back to the example in Lecture 2: estimating the causal effect of schooling (S_i: number of years of education) on earnings
  - Let family background, intelligence and motivation be denoted by a $1xK$ vector A_i
  - The regression of earnings on the variable of interest (S_i) controlling for this vector of covariates writes as:

$$Y_i = \alpha + \rho S_i + A_i \gamma + \epsilon_i$$

  - where γ is a $Kx1$ vector of population regression coefficients associated to each covariate.

The endogeneity issue

- If the CIA holds given the set of covariates A_i, ρ is the causal population ATE and $\hat{\rho}_{OLS}$ is a consistent estimator of this ATE.
  - BUT in practice A_i is hard to measure so the researcher can only estimate a "short" regression in which this strong predictors are omitted.

$$Y_i = \alpha + \rho S_i + \eta_i$$

The endogeneity issue

- If the CIA holds given the set of covariates A_i, ρ is the causal population ATE and $\hat{\rho}_{OLS}$ is a consistent estimator of this ATE.
    - BUT in practice A_i is hard to measure so the researcher can only estimate a "short" regression in which this strong predictors are omitted.

$$Y_i = \alpha + \rho S_i + \eta_i$$

- In that case S_i is **endogenous**: correlated with unobservables (including ability, motivation, ...)
- $cov(S_i, \eta_i) \neq 0$

The endogeneity issue

- Here, endogeneity issue arises because of omitted variables that strongly predict the dependent variable
- BUT this issue may result from
  - Simultaneity/reverse causality bias
  - Measurement errors

The role of instrumental variables

- The issue can be addressed with a variable (instrument variable: IV) z_i that is
  - **strongly correlated with the causal variable of interest** D_i
  - **Uncorrelated with unobservable variables** η_i (including ability, motivation, ...)
  - In that case, z_i can be **excluded** from the causal model of interest: it is an **exclusion restriction**

The role of instrumental variables

- Good instruments are often linked to institutional settings
  - Example: compulsory schooling laws (Angrist and Krueger, 1991): children have to enter school in the calendar year in which they turn 6 and have to remain in school until their 16th birthday
  - So we can expect that individuals of a given cohort born earlier (1st quarter of the year) tend to have lower average schooling levels than individuals born later (4th quarter of the year)

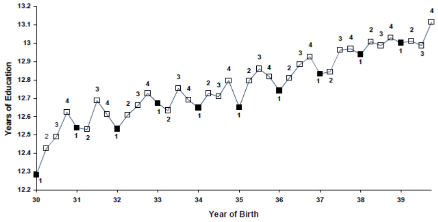
The role of instrumental variables

- Is quarter-of-birth of individuals of a given cohort a good instrument?
  - exogenous?: a priori no reason why quarter of birth influences ability: OK (**but not testable**)
  - Do quarter-of-birth strongly influence educational attainment (first-stage): testable with data

The relationship between quarter of birth and educational attainment using US census data (Angrist and Krueger, 1991)

Within each cohort (year-of-birth) average schooling levels strongly increase with the quarter-of-birth: strong first-stage

A. Average Education by Quarter of Birth (first stage)



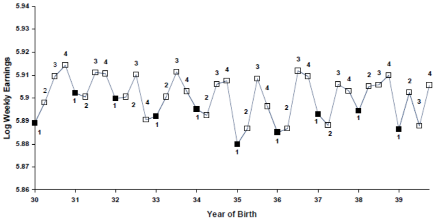
The role of instrumental variables

- IV models consist in estimating 2 simultaneous equations
  - The first-stage: regression of the variable of interest (here education) on the instrument z_i adjusting for a $1xK$ vector of covariates X_i
  - The reduced-form: regression of the dependent variable Y_i on the set of covariates X_i and the instrument z_i
  - X_i are possible strong predictors of Y_i (e.g. year of birth as earnings increase with age)

Graphical representation of the reduced-form equation of the Angrist and Krueger (1991) example

- Controlling for year of birth, earnings increase with quarter-of-birth z_i
- If z_i is truly exogenous, and since it strongly predicts schooling choices, this pattern can only be driven by the effect of education on earnings

B. Average Weekly Wage by Quarter of Birth (reduced form)



Outline

Two-Stage Least Squares (2SLS)

- Let $\hat{s}_i$ be the fitted value of the number of years of education S_i after regressing it on a set of included exogenous characteristics X_i and a set of excluded instruments z_i.

$$\hat{s}_i = X_i \hat{pi}_{10} + \hat{\pi}_{11} z_i$$

- where $\hat{pi}_{10}$ and $\hat{\pi}_{11}$ are OLS estimates from the first-stage.

Two-Stage Least Squares (2SLS)

- Let $\hat{s}_i$ be the fitted value of the number of years of education S_i after regressing it on a set of included exogenous characteristics X_i and a set of excluded instruments z_i.

$$\hat{s}_i = X_i \hat{pi}_{10} + \hat{\pi}_{11} z_i$$

- where $\hat{pi}_{10}$ and $\hat{\pi}_{11}$ are OLS estimates from the first-stage.
- The coefficient on $\hat{s}_i$ in the regression of Y_i on X_i and $\hat{s}_i$ is the called **the Two-Stage Least Squares (2SLS) estimator**
  - The first stage consists in estimating $\hat{s}_i$ by OLS
  - The second stage consists in estimating the coefficient on $\hat{s}_i$ in the regression of Y_i on X_i and $\hat{s}_i$ by OLS

Advices to avoid common mistakes when implementing 2SLS

- CAUTION: the estimator of the 2SLS variance is built from the residual $\hat{\eta}_i$
- NOT from the residual of the second-stage after regressing the dependent variable on $\hat{s}_i$ and on X_i

Advices to avoid common mistakes when implementing 2SLS

- CAUTION: the estimator of the 2SLS variance is built from the residual $\hat{\eta}_i$
- NOT from the residual of the second-stage after regressing the dependent variable on $\hat{s}_i$ and on X_i

- Computing the estimator of the variance of the 2SLS coefficient including the fitted values of the endogenous variable into the second-stage equation and using the second-stage residual is a COMMON MISTAKE
- Caution when implementing 2SLS by hand

Advices to avoid common mistakes when implementing 2SLS

- Put the SAME exogenous covariates in the first and second stage
- If you introduce exogenous covariates in the second-stage that are not included in the first-stage, the 2SLS estimator of ρ is **inconsistent**

Outline

The over-identified case

- We start from a structural equation:

$$Y = X\beta + u$$

- where X is a NxK matrix of exogenous and endogenous covariates
- Let us denote Z a NxL matrix of included exogenous covariates AND the excluded instruments.

The over-identified case

- We start from a structural equation:

$$Y = X\beta + u$$

- where X is a NxK matrix of exogenous and endogenous covariates
- Let us denote Z a NxL matrix of included exogenous covariates AND the excluded instruments.
- The exogeneity assumption regarding Z implies the following set of moments conditions:

$$g_i(\beta) = Z_i' u_i = Z_i'(y_i - X_i\beta) = 0$$

- This yields the following system of equations:

$$\begin{pmatrix} Z'_{1i}(y_i - X_i\beta) & = & 0 \\ Z'_{2i}(y_i - X_i\beta) & = & 0 \\ \vdots & & \\ Z'_{li}(y_i - X_i\beta) & = & 0 \end{pmatrix}$$

The over-identified case

Three cases

1. $L < K$: number of excluded instruments is lower than the number of endogenous covariates. No solution.
2. $L = K$: as many excluded instruments as included endogenous variables: A unique solution exists, this is the standard IV estimator:
$$\hat{\beta}_{IV} = (Z'X)^{-1}Z'Y$$
3. $L > K$: more instruments than needed to estimate the effect of included endogenous variables: the model is overidentified.

The over-identified case

Three cases

1. $L < K$: number of excluded instruments is lower than the number of endogenous covariates. No solution.
2. $L = K$: as many excluded instruments as included endogenous variables: A unique solution exists, this is the standard IV estimator:
$$\hat{\beta}_{IV} = (Z'X)^{-1}Z'Y$$
3. $L > K$: more instruments than needed to estimate the effect of included endogenous variables: the model is overidentified.

This latter case (**overidentification**) is interesting as it allows to test the validity of the instruments

The over-identified case

In the over-identification case

- We can estimate β by Two-Stage Least Squares and then using a **Sargan test**
  - Regressing u on all instruments in Z
  - Under the null hypothesis that all instruments are uncorrelated with u, the Sargan' statistic is $N * R^2$ and has a large sample $\chi_2(r)$ distribution, r being the number of overidentifying restrictions

Outline

Weak instruments and under-identification

- If the regressor of interest is endogenous, OLS estimator is biased
- BUT even though we find an exogenous instrument, 2SLS may also be biased
  - **If the instrument is not or insufficiently correlated with the endogenous variable**
  - This raises under- or weak-identification issues

Weak instruments and under-identification

A simple rule of thumb based on the F-test of the first-stage

- Statistic use to test joint significance of the first-stage coefficients

$$F = \frac{(SSR_R - SSR_{UR})/(L)}{SSR_{UR}/[N - L - 1]}$$

- where SSR_R and SSR_{UR} are the sum of squared residuals respectively for
  - The restricted model (R): only regressing endogenous regressor on a constant in the first-stage)
  - The unrestricted model (UR): regressing endogenous regressor on the full set of L instruments

Weak instruments and under-identification

A simple rule of thumb based on the F-test of the first-stage

- Statistic use to test joint significance of the first-stage coefficients

$$F = \frac{(SSR_R - SSR_{UR})/(L)}{SSR_{UR}/[N - L - 1]}$$

- where SSR_R and SSR_{UR} are the sum of squared residuals respectively for
  - The restricted model (R): only regressing endogenous regressor on a constant in the first-stage)
  - The unrestricted model (UR): regressing endogenous regressor on the full set of L instruments

- If $F < 10$, it casts some doubts on the strongness of the first-stage (instrument is not valid)

Weak instruments and under-identification

Note: the F-stat can also be computed with the R^2, i.e. the fraction of observed variance of the dependent variable explained by the model, obtained for the restricted and unrestricted models

- Remember that the R^2 writes as:

$$R^2 = 1 - (SSR/TSS)$$

- where SSR is the sum of squared residuals and TSS is the Total Sum of Squares, i.e. the observed variance of the dependent variable.

Weak instruments and under-identification

Note: the F-stat can also be computed with the R^2, i.e. the fraction of observed variance of the dependent variable explained by the model, obtained for the restricted and unrestricted models

- Remember that the R^2 writes as:

$$R^2 = 1 - (SSR/TSS)$$

- where SSR is the sum of squared residuals and TSS is the Total Sum of Squares, i.e. the observed variance of the dependent variable.

- The F-stat can be written as:

$$F = \frac{(R^2_{UR} - R^2_R)/(L)}{(1 - R^2_{UR})/[N - L - 1]}$$

Weak instruments and under-identification

However even though $F > 10$, the first-stage is not necessarily strong

- It tests the joint significance of ALL first-stage coefficients
- BUT it could be that the coefficients associated to the included exogenous regressors are high so F can be high and the instruments are not correlated with the endogeneous regressor

Weak instruments and under-identification

However even though $F > 10$, the first-stage is not necessarily strong

- It tests the joint significance of ALL first-stage coefficients
- BUT it could be that the coefficients associated to the included exogenous regressors are high so F can be high and the instruments are not correlated with the endogeneous regressor

- In the case of a single endogenous regressor, the Cragg-Donald (1993) Statistic is a F-stat to test H_0: excluded instruments do not enter the first-stage regression
- **A good test for under-identification**

Outline

The Wald estimator

- In a case of one endogenous regressor S_i, no covariates and one BINARY instrument $Z_i \in \{0;1\}$

$$Y_i = \alpha + \rho S_i + \eta_i$$

  - Let N_t be the number of units for whom $Z_i = 1$ and N_c the number of units for whom $Z_i = 0$
  - Recall (cf lecture 2) that the OLS estimator of a regression of Y_i on an intercept and one dummy variable Z_i is:

$$\frac{cov(Y_i, Z_i)}{Var(Z_i)} = \frac{1}{N_t}\sum_{i:Z_i=1} Y_i - \frac{1}{N_c}\sum_{i:Z_i=0} Y_i$$

The Wald estimator

- In a case of one endogenous regressor S_i, no covariates and one BINARY instrument $Z_i \in \{0;1\}$

$$Y_i = \alpha + \rho S_i + \eta_i$$

  - Let N_t be the number of units for whom $Z_i = 1$ and N_c the number of units for whom $Z_i = 0$
  - Recall (cf lecture 2) that the OLS estimator of a regression of Y_i on an intercept and one dummy variable Z_i is:

$$\frac{cov(Y_i, Z_i)}{Var(Z_i)} = \frac{1}{N_t} \sum_{i:Z_i=1} Y_i - \frac{1}{N_c} \sum_{i:Z_i=0} Y_i$$

- Given the formula of the IV estimator, we could write it as:

$$\hat{\rho}^W = \frac{\frac{1}{N_t} \sum_{i:Z_i=1} Y_i - \frac{1}{N_c} \sum_{i:Z_i=0} Y_i}{\frac{1}{N_t} \sum_{i:Z_i=1} S_i - \frac{1}{N_c} \sum_{i:Z_i=0} S_i}$$

The Wald estimator

- $\hat{\rho}^W$ is **the Wald estimator**
  - The numerator is the reduced-form difference in means
  - The denominator is the first-stage difference in means

The Wald estimator

- $\hat{\rho}^W$ is **the Wald estimator**
  - The numerator is the reduced-form difference in means
  - The denominator is the first-stage difference in means
- The Wald estimator in the Angrist and Krueger study:

Table 4.1.2: Wald estimates of the returns to schooling using quarter of birth instruments

	(1) Born in the 1st or 2nd quarter of year	(2) Born in the 3rd or 4th quarter of year	(3) Difference (std. error) (1)-(2)
ln (weekly wage)	5.8916	5.9051	-0.01349 (0.00337)
Years of education	12.6881	12.8394	-0.1514 (0.0162)
Wald estimate of return to education			0.0891 (0.0210)
OLS estimate of return to education			0.0703 (0.0005)

Notes: Adapted from a re-analysis of Angrist and Krueger (1991) by Angrist and Imbens (1995). The sample includes native-born men with positive earnings from the 1930-39 birth cohorts in the 1980 Census 5 percent file. The sample size is 329,509.

Outline

A generalized potential outcomes framework

- $Y_i(d, z)$: the potential outcome of an individual i were this person to have:
  - Treatment status $D_i = d$
  - Instrument value $Z_i = z$

A generalized potential outcomes framework

- $Y_i(d, z)$: the potential outcome of an individual i were this person to have:
  - Treatment status $D_i = d$
  - Instrument value $Z_i = z$

- The observed treatment status is:

$$D_i = D_{0i} + (D_{1i} - D_{0i})Z_i = \pi_0 + pi_1 Z_i + \zeta_i$$

- D_{0i} tells us whether the unit i is treated if $Z_i = 0$
- D_{1i} tells us whether the unit i is treated if $Z_i = 1$

A generalized potential outcomes framework

- In that case, Z_i stands for the treatment assignment
- And D_i indicates whether the treatment is effectively received

A generalized potential outcomes framework

- In that case, Z_i stands for the treatment assignment
- And D_i indicates whether the treatment is effectively received
- Some examples
  - Active Labor Market policies: most often take-up rate < 1)
  - Randomized encouragement (Z_i) through supplementary information on a program (D_i)
  - Military service and earnings (Angrist, 1990): In the Vietnam-war era (1960's-1970's), conscription has been based on random numbers assigned based on birthdays.
    - Those with lottery numbers below an eligibility ceiling were eligible for the draft.
    - Those with numbers above the ceiling were not eligible.

A generalized potential outcomes framework

In that case, we can group units into 4 categories $G_i \in \{C; NT; AT; D\}$ according to their **compliance types**

	$D_i = 1, Z_i = 1$	$D_i = 0, Z_i = 1$
$D_i = 1, Z_i = 0$	Always-Takers (AT)	Defiers (D)
$D_i = 0, Z_i = 0$	Compliers (C)	Never-Takers (NT)

A generalized potential outcomes framework

In the Angrist paper using Vietnam-era draft lottery as instrument for conscription

- Some men with numbers below the draft ceiling have been exempted from service (e.g. for health reasons): never-takers or defiers
- Some men with numbers below the draft ceiling serve in the military: compliers or always-takers
- Some men with numbers above the draft ceiling (a priori exempted) volunteered for service: always-takers or defiers

A generalized potential outcomes framework

We need additional assumptions with respect to the case of constant treatment effect

- **Random assignment**: $Z_i \perp (\{Y_i(d,z) \forall d,z\}, D_{1i}, D_{0i})$
  - This allows to identify the average effect of assignment: **the Intention-to-Treat (ITT) effect**

$$E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0) = E(Y_i(D_{1i}, 1) - Y_i(D_{0i}, 0))$$

A generalized potential outcomes framework

We need additional assumptions with respect to the case of constant treatment effect

- **Random assignment**: $Z_i \perp (\{Y_i(d,z) \forall d,z\}, D_{1i}, D_{0i})$
  - This allows to identify the average effect of assignment: **the Intention-to-Treat (ITT) effect**

$$E(Y_i|Z_i = 1) - E(Y_i|Z_i = 0) = E(Y_i(D_{1i}, 1) - Y_i(D_{0i}, 0))$$

- In the Angrist (1990) study, draft lottery number instruments are random.

A generalized potential outcomes framework

We need additional assumptions with respect to the case of constant treatment effect

- **Exclusion restriction**: the assignment does not affect outcome other than through the treatment received
  - So potential outcomes may be indexed solely against treatment status

$$Y_i(1,1) = Y_i(1,0) = Y_{1i} \quad \text{and} \quad Y_i(0,1) = Y_i(0,0) = Y_{0i}$$

A generalized potential outcomes framework

We need additional assumptions with respect to the case of constant treatment effect

- **Exclusion restriction**: the assignment does not affect outcome other than through the treatment received
  - So potential outcomes may be indexed solely against treatment status

$$Y_i(1,1) = Y_i(1,0) = Y_{1i} \quad \text{and} \quad Y_i(0,1) = Y_i(0,0) = Y_{0i}$$

- In the Angrist (1990) study, this does not necessarily hold if for instance low lottery number would have encouraged men to stay in college longer
- In that case, instrument is correlated with earnings through two channels: increased likelihood of military service AND increased likelihood of college attendance

A generalized potential outcomes framework

We need additional assumptions with respect to the case of constant treatment effect

- **Monotonicity (Imbens and Angrist, 1994)**: there are no defiers: $D_{1i} \geq D_{0i}$
  - There could be never-takers ($D_i = 0$ while $Z_i = 1$) BUT $D_i = 0$ is not caused by $Z_i = 1$

A generalized potential outcomes framework

Without monotonicity the true compliance status is not observed on all units

Z	D	Compliance type
0	0	Compliers or never-takers?
0	1	Always-takers of defiers?
1	0	Never-takers or defiers?
1	1	Compliers or always-takers?

A generalized potential outcomes framework

Assuming monotonicity (no defiers) it is possible to estimate the distribution of compliance types:

Z	D	Compliance type
0	0	Compliers or never-takers?
0	1	**Always-takers**
1	0	**Never-takers**
1	1	Compliers or always-takers?

A generalized potential outcomes framework

Assuming monotonicity (no defiers) it is possible to estimate the distribution of compliance types:

Z	D	Compliance type
0	0	Compliers or never-takers?
0	1	**Always-takers**
1	0	**Never-takers**
1	1	Compliers or always-takers?

Let N_0 denote the number of individuals not assigned into treatment ($Z_i = 0$) and N_1 the number of individuals assigned to treatment ($Z_i = 1$)

A generalized potential outcomes framework

- The probability of being an always-taker $\pi_{AT} = P(D_{0i} = D_{1i} = 1)$ is estimated by:

$$\hat{\pi}_{AT} = \frac{1}{N_0} \sum_{i}^{N} (1 - Z_i) D_i$$

- The probability of being a never-taker $\pi_{NT} = P(D_{0i} = D_{1i} = 0)$ is estimated by:

$$\hat{\pi}_{NT} = \frac{1}{N_1} \sum_{i}^{N} Z_i (1 - D_i)$$

A generalized potential outcomes framework

- This allows to estimate the probability of being a **complier** $\pi_C = P(D_{0i} = 0, D_{1i} = 1) = E(D_i|Z_i = 1) - E(D_i|Z_i = 0)$

$$\hat{\pi}_C = 1 - \hat{\pi}_{AT} - \hat{\pi}_{NT}$$

- Note that the ratio of the ITT over the probability of being a complier is:

$$\frac{ITT}{\pi_C} = \frac{E(Y_i|Z_i = 1) - E(Y_i|D_i = 0)}{E(D_i|Z_i = 1) - E(D_i|Z_i = 0)}$$

A generalized potential outcomes framework

- This allows to estimate the probability of being a **complier** $\pi_C = P(D_{0i} = 0, D_{1i} = 1) = E(D_i|Z_i = 1) - E(D_i|Z_i = 0)$

$$\hat{\pi}_C = 1 - \hat{\pi}_{AT} - \hat{\pi}_{NT}$$

- Note that the ratio of the ITT over the probability of being a complier is:

$$\frac{ITT}{\pi_C} = \frac{E(Y_i|Z_i = 1) - E(Y_i|D_i = 0)}{E(D_i|Z_i = 1) - E(D_i|Z_i = 0)}$$

- This is the Wald estimand and this corresponds to the **Local Average Treatment Effect: LATE** (Imbens and Angrist, 1994)

A generalized potential outcomes framework

The LATE theorem (Angrist and Imbens, 1994) says that:

- Under the assumptions of randomness of instrument, exclusion restrictions and monotonicity and provided that the first-stage is strong

$$\frac{E(Y_i|Z_i=1)-E(Y_i|D_i=0)}{E(D_i|Z_i=1)-E(D_i|Z_i=0)} = E(Y_{\mathbf{1}i}-Y_{\mathbf{0}i}|D_{\mathbf{1}i}>D_{\mathbf{0}i}) = E(Y_{\mathbf{1}i}-Y_{\mathbf{0}i}|G_i=C)$$

- IV estimates the average causal effect of the treatment on the COMPLIERS (**only a local effect**)

A generalized potential outcomes framework

IV does not capture the average causal effect on all the treated or on all the non-treated individuals

- There are exceptions to this rule ONLY if the IV allows no always-takers or no never-takers

A generalized potential outcomes framework

IV does not capture the average causal effect on all the treated or on all the non-treated individuals

- There are exceptions to this rule ONLY if the IV allows no always-takers or no never-takers
- Angrist and Evans (1998) use multiple second births (twins) indicator to instrument the fact of having three children
- The goal is to estimate the effect of having three children on earnings among women with at least two children

A generalized potential outcomes framework

- In that case, there are no never-takers: if $Z_i = 1$, the woman has necessarily three children.
- BUT there are always-takers: even if $Z_i = 0$ a woman can decide to have a third children afterwards

A generalized potential outcomes framework

- In that case, there are no never-takers: if $Z_i = 1$, the woman has necessarily three children.
- BUT there are always-takers: even if $Z_i = 0$ a woman can decide to have a third children afterwards

- The LATE corresponds here to the average causal effect of the treatment on the non-treated individuals:

$$LATE = E(Y_{1i} - Y_{0i} | D_i = 0)$$

Outline

IV and non-compliance in RCTs

- In a completely randomized experiment, the OLS estimator captures the ATT (cf lecture 1 and 2)
  - BUT in many RCTs, participation is **voluntary** among those randomly assigned to receive treatment
  - The group that receives the treatment (compliers) is a self-selected subset of those offered treatment

IV and non-compliance in RCTs

- In a completely randomized experiment, the OLS estimator captures the ATT (cf lecture 1 and 2)
  - BUT in many RCTs, participation is **voluntary** among those randomly assigned to receive treatment
  - The group that receives the treatment (compliers) is a self-selected subset of those offered treatment

- In RCTs with non-compliance, there remains a selection bias (almost always positive): **OLS estimator is misleading**

IV and non-compliance in RCTs

- ALMP are generally characterized by non-compliance
  - Ex: the randomized evaluation of the Job Training Partnership Act shows that only 60% of those assigned to treatment received training.

IV and non-compliance in RCTs

- ALMP are generally characterized by non-compliance
  - Ex: the randomized evaluation of the Job Training Partnership Act shows that only 60% of those assigned to treatment received training.
- In that case, LATE using randomized assignment to treatment as an instrument of receiving treatment is a **consistent estimator of the Average Treatment Effect on the Treated (ATT)**
  - Since generally non-compliance concerns the treatment group.

IV and non-compliance in RCTs: the example of the evaluation of the JTPA

If we compare treated and non-treated individuals, the causal effect of JTPA on earnings is large ($\approx$ \$ 4000 for men)

Table 4.4.1: Results from the JTPA experiment: OLS and IV estimates of training impacts

	Comparisons by Training Status		Comparisons by Assignment Status		Instrumental Variable Estimates	
	Without Covariates (1)	With Covariates (2)	Without Covariates (3)	With Covariates (4)	Without Covariates (5)	With Covariates (6)
A. Men	3,970	3,754	1,117	970	1,825	1,593
	(555)	(536)	(569)	(546)	(928)	(895)
B. Women	2,133	2,215	1,243	1,139	1,942	1,780
	(345)	(334)	(359)	(341)	(560)	(532)

IV and non-compliance in RCTs: the example of the evaluation of the JTPA

BUT the Intention-to-Treat effect (ITT), i.e. the comparison of those assigned to treatment and those not assigned is strongly lower ($\approx$ \$1100 for men)

Table 4.4.1: Results from the JTPA experiment: OLS and IV estimates of training impacts

	Comparisons by Training Status		Comparisons by Assignment Status		Instrumental Variable Estimates	
	Without Covariates	With Covariates	Without Covariates	With Covariates	Without Covariates	With Covariates
	(1)	(2)	(3)	(4)	(5)	(6)
A. Men	3,970	3,754	1,117	970	1,825	1,593
	(555)	(536)	(569)	(546)	(928)	(895)
B. Women	2,133	2,215	1,243	1,139	1,942	1,780
	(345)	(334)	(359)	(341)	(560)	(532)

IV and non-compliance in RCTs: the example of the evaluation of the JTPA

The LATE (the ITT divided by the difference in compliance rates between treatment and control groups) estimates the ATT: roughly \$1800 for men

Table 4.4.1: Results from the JTPA experiment: OLS and IV estimates of training impacts

	Comparisons by Training Status		Comparisons by Assignment Status		Instrumental Variable Estimates	
	Without Covariates (1)	With Covariates (2)	Without Covariates (3)	With Covariates (4)	Without Covariates (5)	With Covariates (6)
A. Men	3,970	3,754	1,117	970	1,825	1,593
	(555)	(536)	(569)	(546)	(928)	(895)
B. Women	2,133	2,215	1,243	1,139	1,942	1,780
	(345)	(334)	(359)	(341)	(560)	(532)

Characterizing compliers

- IV estimates can use multiple instruments
- And the compliant subpopulations associated with each instrument may be very different

Characterizing compliers

- IV estimates can use multiple instruments
- And the compliant subpopulations associated with each instrument may be very different

- In that case, if we obtain similar IV estimates using different instruments, **we could conclude to homogeneous treatment effects**

- BUT how can we characterize compliant subpopulations? D_{1i} and D_{0i} are not observable at the same time.

Characterizing compliers

- The distribution of compliers' characteristics can be learned **from variation in the first-stage across covariate groups**
- The relative likelihood a complier is characterized by $X_{ik} = x$, where X_{ik} is a discrete covariate and x is its value, is given by

$$\frac{P(X_{ik} = x|D_{1i} > D_{0i})}{P(X_{ik} = x)} = \frac{E(D_i|Z_i = 1, X_{ik} = x) - E(D_i|Z_i = 0, X_{ik} = x)}{E(D_i|Z_i = 1) - E(D_i|Z_i = 0)}$$

Characterizing compliers

- The distribution of compliers' characteristics can be learned **from variation in the first-stage across covariate groups**
- The relative likelihood a complier is characterized by $X_{ik} = x$, where X_{ik} is a discrete covariate and x is its value, is given by

$$\frac{P(X_{ik} = x | D_{1i} > D_{0i})}{P(X_{ik} = x)} = \frac{E(D_i | Z_i = 1, X_{ik} = x) - E(D_i | Z_i = 0, X_{ik} = x)}{E(D_i | Z_i = 1) - E(D_i | Z_i = 0)}$$

- This is the ratio of the first stage among individuals with $X_{ik} = x$ to the overall first-stage.

Characterizing compliers

- Ex: Angrist and Evans (1998) use two instruments:
  - Multiple second births (twins)
  - Sex-composition of the first two children

Characterizing compliers

- Ex: Angrist and Evans (1998) use two instruments:
  - Multiple second births (twins)
  - Sex-composition of the first two children
- They look at the characteristics of the two associated compliant subpopulations

Characterizing compliers (Angrist and Evans, 1998)

Twins compliers are more educated than the average mother but this is the reverse story for sex-composition compliers

Table 4.4.3: Complier-characteristics ratios for twins and sex-composition instruments

Variable	$E[x]$ (1)	Twins at second birth: $E[x \mid D_1 > D_0]$ (2)	Twins at second birth: $P[x \mid D_1 > D_0]/P[X]$ (3)	First two children are same sex: $E[x \mid D_1 > D_0]$ (6)	First two children are same sex: $P[x \mid D_1 > D_0]/P[X]$ (5)
Age 30 or older at first birth	0.00291	0.00404	1.39 (0.0201)	0.00233	0.995 (0.374)
Black or hispanic	0.125	0.103	0.822 (0.00421)	0.102	0.814 (0.0775)
High school graduate	0.822	0.861	1.048 (0.000772)	0.815	0.998 (0.0140)
College graduate	0.132	0.151	1.14 (0.00376)	0.0904	0.704 (0.0692)

Notes: The table reports an analysis of complier characteristics for twins and sex-composition instruments. The ratios in columns 3 and 5 give the relative likelihood compliers have the characteristic indicated in each row. Data are from the 1980 Census 5% sample, including married mothers age 21-35 with at least two children, as in Angrist and Evans (1998). The sample size is 254,654 for all columns.

Outline

Limited endogenous variables and forbidden regression

- Back to the Angrist and Evans' paper (1998): estimating the motherhood's penalty.
  - Let D_i be the dummy indicating whether a woman has three children
  - Let X_i be a vector of exogenous covariates
  - Let Z_i be the excluded instrument: having already two boys or two girls

Limited endogenous variables and forbidden regression

- Back to the Angrist and Evans' paper (1998): estimating the motherhood's penalty.
  - Let D_i be the dummy indicating whether a woman has three children
  - Let X_i be a vector of exogenous covariates
  - Let Z_i be the excluded instrument: having already two boys or two girls

- The usual 2SLS first-stage is:

$$D_i = X_i \pi_{10} + \pi_{11} Z_i + \zeta_i$$

Limited endogenous variables and forbidden regression

- Since D_i is a dummy, the OLS first-stage is an approximation to the underlying nonlinear Conditional Expectation Function $E(D_i|X_i, Z_i)$.
  - We could use a non-linear first-stage to come closer to the CEF, like a Probit or a logit model.
  - And then plugging in the fitted value of D_i, $\hat{D}_i$ in the second-stage OLS equation

Limited endogenous variables and forbidden regression

- Since D_i is a dummy, the OLS first-stage is an approximation to the underlying nonlinear Conditional Expectation Function $E(D_i|X_i, Z_i)$.
  - We could use a non-linear first-stage to come closer to the CEF, like a Probit or a logit model.
  - And then plugging in the fitted value of D_i, $\hat{D}_i$ in the second-stage OLS equation

- BUT this is FORBIDDEN (Angrist and Pischke, 2008):
  - Only OLS first-stage is guaranteed to produce first-stage residuals that are uncorrelated with fitted values and covariates
  - The main issue with non-linear model: assuming the distribution of the first-stage CEF
  - With 2SLS, no need to worry about whether the first-stage CEF is really linear.

Limited dependent and endogenous variables and the use of a bivariate Probit

- An alternative approach: building up a causal story describing the process generating the dependent variable in detail
  - With a binary dependent variable ($Y_i = 1$ if the woman is employed), we can use a bivariate Probit model.

Limited dependent and endogenous variables and the use of a bivariate Probit

- An alternative approach: building up a causal story describing the process generating the dependent variable in detail
  - With a binary dependent variable ($Y_i = 1$ if the woman is employed), we can use a bivariate Probit model.

- Suppose that a woman decides to have a third child by comparing costs and benefits using a net benefit function. This latent variable would be linear in covariates, excluded instruments with an error term ζ_i.
- The bivariate Probit first-stage can be written as:

$$D_i = 1[X_i\pi_{10} + \pi_{11}Z_i > \zeta_i]$$

Limited dependent and endogenous variables and the use of a bivariate Probit

- Suppose that the employment status Y_i is defined by the following latent index arising from the comparison of the costs an benefits of working:

$$Y_i = 1[X_i\beta_0 + \beta_1 D_i > \epsilon_i]$$

- where ϵ_i is the second-stage error term.

Limited dependent and endogenous variables and the use of a bivariate Probit

- Suppose that the employment status Y_i is defined by the following latent index arising from the comparison of the costs an benefits of working:

$$Y_i = 1[X_i\beta_0 + \beta_1 D_i > \epsilon_i]$$

- where ϵ_i is the second-stage error term.
- The source of omitted variable bias in this setup is correlation between ζ_i and ϵ_i
  - unobserved determinants of childbearing could be correlated with unobserved determinants of employment

Limited dependent and endogenous variables and the use of a bivariate Probit

- Assuming that Z_i is independent of ζ_i and ϵ_i and that these residuals are normally distributed, the parameters can be identified maximizing the following log likelihood:

$$\sum Y_i ln\Phi_b\left(\frac{X_i\beta_0 + \beta_1 D_i}{\sigma_\epsilon}, \frac{X_i\pi_{10} + \pi_{11} Z_i}{\sigma_\zeta}; \rho_{\epsilon\zeta}\right)$$

$$+(1 - Y_i) ln[1 - \Phi_b\left(\frac{X_i\beta_0 + \beta_1 D_i}{\sigma_\epsilon}, \frac{X_i\pi_{10} + \pi_{11} Z_i}{\sigma_\zeta}; \rho_{\epsilon\zeta}\right)]$$

- where $\Phi_b(.,.,\rho_{\epsilon\zeta})$ is the bivariate normal cumulative distribution function with correlaton coefficient $\rho_{\epsilon\zeta}$.

Limited dependent and endogenous variables and the use of a bivariate Probit

- One big advantage of bivariate Probit model: estimating the ATE and not only the LATE

$$ATE = E\{\Phi[\frac{X_i\beta_0 + \beta_1}{\sigma_\epsilon}] - \Phi[\frac{X_i\beta_0}{\sigma_\epsilon}]\}$$

- where $\Phi[.]$ is the normal c.d.f.

Limited dependent and endogenous variables and the use of a bivariate Probit

- One big advantage of bivariate Probit model: estimating the ATE and not only the LATE

$$ATE = E\{\Phi[\frac{X_i\beta_0 + \beta_1}{\sigma_\epsilon}] - \Phi[\frac{X_i\beta_0}{\sigma_\epsilon}]\}$$

- where $\Phi[.]$ is the normal c.d.f.

- BUT this comes at a cost: assuming normality of the latent index error terms
- Without distributional assumption, the best we can do is estimating the LATE: **so we can also use a 2SLS model**