Module 6: Noncompliance and Instrumental Variables

Fall 2021

Matthew Blackwell

Gov 2003 (Harvard)

Where are we? Where are we going?

- We’ve covered randomized experiments (no confounding).
- We’ve covered selection on observables (no unmeasured confounding).
- What if there is unmeasured confounding? What can we do?
- First approach we’ll explore: instrumental variables.
    - First: motivate IV through experiments and noncompliance.
    - Then: how does this relate to classical econometric methods like TSLS?

1/ Randomized experiments with noncompliance

Noncompliance

- GOTV experiment with door-to-door canvassing.
- Households are randomized so treatment assignment is unconfounded.
    - $Z_i = 1$ for assigned to treatment (canvassing attempted),
    - $Z_i = 0$ for assigned to control (no canvassing attempted).
- **Noncompliance**: units don't follow treatment assignment.
    - Units assigned to treatment take control or vice versa.
    - $D_i = 1$ for actually took treatment (heard canvasser message).
    - $D_i = 0$ for actually took control (didn't answer the door).
    - Full compliance means $Z_i = D_i$ for all i

How to handle noncompliance

- Two approaches common seen in applied studies.
- **Intent-to-treat** analysis (ITT): just use randomization.
    - Use Z_i as the treatment and analyze as a typical experiment.
    - Downside: can't learn about the effect of actually being canvassed.
- **As-treated** analysis: just use treatment uptake.
    - Act as if D_i was randomly assigned.
    - Not valid if uptake is **correlated** with the outcome.
    - $\rightsquigarrow$ unmeasured confounding between D_i and Y_i and bias.
- Alternative: leverage latent strata of **compliance types**

Setup

- Treatment assignment, $Z_i \in \{0, 1\}$, treatment uptake $D_i \in \{0, 1\}$
- Treatment uptake now affected by assignment: $D_i(z)$
    - $D_i(1) = 1$ if assigned to canvassing, I **would** open my door.
    - $D_i(1) = 0$ if assigned to canvassing, I **would not** open my door.
    - Noncompliance means $D_i(z) \neq z$ for some i.
- Consistency for the observed treatment as usual:

$$D_i = Z_i D_i(1) + (1 - Z_i) D_i(0)$$

- Canvassing is an example of **one-sided noncompliance**.
    - People might refuse treatment when offered ($D_i(1) = 0$)
    - But no one receives treatment if in control ($D_i(0) = 0, \ \forall i$)
    - **Two-sided noncompliance** is when you can refuse to comply with treatment **or** control.

Potential outcomes

- Outcomes might depend on assignment and uptake: $Y_i(z, d)$.
    - $Y_i(1, 1)$: would I vote if I were assigned to canvassing and received it?
- Can only observe two potential outcomes: $Y_i(1, D_i(1))$ and $Y_i(0, D_i(0))$.
    - $Y_i(1, D_i(1))$: potential outcome when assigned canvassing and whatever uptake occurs for unit *i* when assigned to canvassing.
    - $Y_i(1, 1 - D_i(1))$ not possible to ever observe (cross-world or a prior counterfactual)
- Consistency assumption: $Y_i = Y_i(Z_i, D_i(Z_i))$

Some notation

- Let's use 0/1 subscripts for assignment and t/c subscripts for uptake:

$$n_1 = \sum_{i=1}^{n} Z_i \qquad n_0 = \sum_{i=1}^{n} 1 - Z_i \qquad n_t = \sum_{i=1}^{n} D_i \qquad n_c = \sum_{i=1}^{n} 1 - D_i$$

- Average outcomes and uptake in each assignment group:

$$\overline{Y}_1 = \frac{1}{n_1} \sum_{i=1}^{n} Z_i Y_i \qquad \overline{Y}_0 = \frac{1}{n_0} \sum_{i=1}^{n} (1 - Z_i) Y_i$$

$$\overline{D}_1 = \frac{1}{n_1} \sum_{i=1}^{n} Z_i D_i \qquad \overline{D}_0 = \frac{1}{n_0} \sum_{i=1}^{n} (1 - Z_i) D_i$$

- Assumption 1: **randomization** $[\{Y_i(d,z), \forall d,z\}, D_i(1), D_i(0)] \perp\!\!\!\perp Z_i$
  - For observational uses of IV, might condition on some $\mathbf{X}_i$.

ITT effects

- **Intent-to-treat** (ITT) effects are just the ATEs of Z_i

$$\text{ITT}_D = \frac{1}{n}\sum_{i=1}^{n} D_i(1) - D_i(0) \qquad \text{ITT}_Y = \frac{1}{n}\sum_{i=1}^{n} Y_i(1, D_i(1)) - Y_i(0, D_i(0))$$

  - SATE of assignment on treatment uptake and the outcome.
  - If noncompliance is one-sided, then $\text{ITT}_D \geq 0$

- Standard estimators for these quantities:

$$\widehat{\text{ITT}}_D = \overline{D}_1 - \overline{D}_0 \qquad \widehat{\text{ITT}}_Y = \overline{Y}_1 - \overline{Y}_0$$

- Under randomization of Z_i, everything just like Neyman approach.
  - Variances, tests, CIs all standard.
- Problem: ITT_Y is a combination of true effect of D_i and noncompliance.
  - Effect of D_i is maybe more externally valid than Z_i.

2/ Compliance types

Compliance status

- We can stratify units by their **compliance type**.
  - Compliance type is how they would respond to treatment assignment.
  - Basically it's the value of $(D_i(0), D_i(1))$ for any unit.
- Under one-sided noncompliance, there are two types:
  - **Compliers** with $D_i(1) = 1$ and **noncompliers** with $D_i(1) = 0$.
  - Compliers answer the door when assigned to canvassing
  - Noncompliers don't answer the door when assigned to canvassing
  - Everyone has $D_i(0) = 0$, so no noncompliance there.
- Compliance is a function of potential outcomes so it is **pretreatment**!
  - $\rightsquigarrow$ treatment assignment independent of C_i

ITTs among the compliance groups

- Compliance type indicator $C_i \in \{\text{co}, \text{nc}\}$.
    - Number of compliers: $n_{\text{co}} = \sum_{i=1} \mathbf{1}(C_i = \text{co})$.
    - Proportion of compliers: $\pi_{\text{co}} = n_{\text{co}}/n$
    - Same for noncompliers: n_{nc} and π_{nc}
- ITT on uptake directly related to compliance type:

$$\text{ITT}_{D,\text{co}} = \frac{1}{n_{\text{co}}} \sum_{i:C_i=\text{co}} D_i(1) - D_i(0) = 1$$

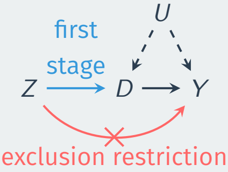
$$\text{ITT}_{D,\text{nc}} = \frac{1}{n_{\text{nc}}} \sum_{i:C_i=\text{nc}} D_i(1) - D_i(0) = 0$$

    - Intuition: no effect of assignment on uptake for noncompliers!
- Implies overall ITT on uptake is equal to the **proportion of compliers**

$$\text{ITT}_D = \pi_{\text{co}}\text{ITT}_{D,\text{co}} + \pi_{\text{nc}}\text{ITT}_{D,\text{nc}} = \pi_{\text{co}}$$

3/ Instrumental variables

Exclusion restriction



- Assumption 2: **first-stage** $\text{ITT}_D = \pi_{\text{co}} \neq 0$
    - At least one person complies with treatment.
- Assumption 3: **exclusion restriction** Z_i only affects Y_i through D_i
    - $Y_i(z, d) = Y_i(z', d)$ for all z, z' and d.
    - Assignment to canvassing only affects turnout through actual canvassing.
    - Not a testable assumption and can't be guaranteed by design.
- Implies that potential outcomes only a function of D_i:

$$Y_i(1) = Y_i(D_i = 1) = Y_i(Z_i = 1, D_i = 1) = Y_i(Z_i = 1, D_i = 1)$$

$$Y_i(0) = Y_i(D_i = 0) = Y_i(Z_i = 1, D_i = 0) = Y_i(Z_i = 1, D_i = 0)$$

Outcome ITTs and compliance types

- We can define the ITTs on the outcome by compliance type as well.
  - $\text{ITT}_{Y,\text{co}}$ effect of assignment on outcome among compliers.
  - $\text{ITT}_{Y,\text{nc}}$ effect of assignment on outcome among noncompliers.
  - Only $\text{ITT}_{Y,\text{co}}$ actually picks up an effect of D_i
- Exclusion restriction has implications for these:
  - Implies that $\text{ITT}_{Y,\text{nc}} = 0$: if D_i doesn't change, Y_i can't change.
  - Implies that $\text{ITT}_{Y,\text{co}}$ is due entirely to treatment uptake.
- Allows us to connect the ITT on the outcome to compliance groups:

$$\text{ITT}_Y = \pi_{\text{co}}\text{ITT}_{Y,\text{co}} + \pi_{\text{nc}}\text{ITT}_{Y,\text{nc}} = \text{ITT}_D\text{ITT}_{Y,\text{co}}$$

LATE

- Under the exclusion restriction, $\text{ITT}_{Y,\text{co}}$ is the effect of treatment receipt:

$$\begin{aligned}\text{ITT}_{Y,\text{co}} &= \frac{1}{n_{\text{co}}} \sum_{i:C_i=\text{co}} Y_i(1, D_i(1)) - Y_i(0, D_i(0)) \\ &= \frac{1}{n_{\text{co}}} \sum_{i:C_i=\text{co}} Y_i(D_i = 1) - Y_i(D_i = 0) = \tau_{\text{LATE}}\end{aligned}$$

- This quantity is the **local ATE** (LATE), local to compliers.
  - It's a conditional ATE, where we condition on being a complier.
  - Also called the **complier average causal effect** (CACE).
- **LATE Theorem** under one-sided noncompliance, exclusion restriction, first-stage, and randomization:

$$\tau_{\text{LATE}} = \text{ITT}_{Y,\text{co}} = \frac{\text{ITT}_Y}{\text{ITT}_D}$$

Wald estimator

- **Wald** or **instrumental variables estimator** for the LATE:

$$\widehat{\tau}_{iv} = \frac{\widehat{\text{ITT}}_Y}{\widehat{\text{ITT}}_D}$$

  - Ratio of the two unbiased ITT estimators.
  - Not unbiased, but it is **consistent** for τ_{LATE}.
- Equivalent to the **two-stage least squares** estimator:
  - Regress D_i on Z_i and get fitted values $\widehat{D}_i$
  - Regress Y_i on $\widehat{D}_i$
- We can use the delta method to find the (superpopulation) variance:

$$\mathbb{V}[\widehat{\tau}_{iv}] = \frac{1}{\text{ITT}_D^2}\mathbb{V}\left[\widehat{\text{ITT}}_Y\right] + \frac{\text{ITT}_Y^2}{\text{ITT}_D^4}\mathbb{V}\left[\widehat{\text{ITT}}_D\right] - 2\frac{\text{ITT}_Y}{\text{ITT}_D^3}\text{cov}\left[\widehat{\text{ITT}}_Y, \widehat{\text{ITT}}_D\right]$$

4/ Two-sided noncompliance

Two-sided noncompliance

- Two-sided noncompliance: those in control can select into treatment.
- **Encouragement design**: randomly assign an encouragement of some treatment.
    - Some may refuse encouragement and opt to not take treatment.
    - Some may take treatment even without encouragement.
- Z_i is the encouragement and D_i is the treatment.
- No change in estimation, just different identification assumptions.

Compliance types

- Four compliance types (or **principal strata**) in this setting:
    - Complier $D_i(1) = 1$ and $D_i(0) = 0$
    - Always-taker $D_i(1) = D_i(0) = 1$
    - Never-taker $D_i(1) = D_i(0) = 0$
    - Defier $D_i(1) = 0$ and $D_i(1) = 1$
- Connections between observed data and compliance types:

	$Z_i = 0$	$Z_i = 1$
$D_i = 0$	Never-taker or Complier	Never-taker or Defier
$D_i = 1$	Always-taker or Defier	Always-taker or Complier

- Let π_{co}, π_{at}, π_{nt}, and π_{de} be the proportions of each type.
- ITT effects on D_i are more murky: $\text{ITT}_D = \pi_{\text{co}} - \pi_{\text{de}}$
    - Defiers really make things messy!

Instrumental variables assumptions

- Canonical IV assumptions for Z_i to be a valid instrument:
  1. Randomization of Z_i
  2. Presence of some compliers $\pi_{\text{co}} \neq 0$ (first-stage)
  3. Exclusion restriction $Y_i(z, d) = Y_i(z', d)$
  4. **Monotonicity**: $D_i(1) \geq D_i(0)$ for all i (no defiers)
- Implies ITT effect on treatment equals proportion compliers: $\text{ITT}_D = \pi_{\text{co}}$
- Implies ITT for the outcome has the same interpretation:

$$
\begin{aligned}
\text{ITT}_Y &= \text{ITT}_{Y,\text{co}}\pi_{\text{co}} + \underbrace{\text{ITT}_{Y,\text{at}}}_{=0 \text{ (ER)}}\pi_{\text{at}} + \underbrace{\text{ITT}_{Y,\text{nt}}}_{=0 \text{ (ER)}}\pi_{\text{nt}} + \text{ITT}_{Y,\text{de}}\underbrace{\pi_{\text{de}}}_{=0 \text{ (mono)}} \\
&= \text{ITT}_{\text{co}}\pi_{\text{co}}
\end{aligned}
$$

- $\rightsquigarrow$ same identification result: $\tau_{\text{LATE}} = \text{ITT}_Y / \text{ITT}_D$

Is the LATE useful?

- The LATE is a unknown subset of the data.
    - Treated units are a mix of always takers and compliers.
    - Control units are a mix of never takers and compliers.
- Without further assumptions, $\tau_{\text{LATE}} \neq \tau$.
- Complier group depends on the instrument $\rightsquigarrow$ different IVs will lead to different identified estimands.
- But we cannot do any better in terms of point estimation without more assumptions.
    - Alternative: bound the ATE?

5/ Basic two-stage least squares

TSLS

- **Two stage least squares** (TSLS) is the classical approach to IV.
- Basic idea is to assume two constant effects linear models:

$$Y_i = \alpha + \tau D_i + \varepsilon_i$$
$$D_i = \delta + \gamma Z_i + \eta_i$$

- Here the treatment D_i is **endogenous** so $\mathbb{E}[\varepsilon_i \mid D_i] \neq 0$
- But we have an **instrument** Z_i that is exogenous $\mathbb{E}[\varepsilon_i \mid Z_i] = 0$
    - It also is exogenous for treatment, so $\mathbb{E}[\eta_i \mid Z_i] = 0$.
- This implies the following CEF form for Y_i conditional on Z_i:

$$\mathbb{E}[Y_i \mid Z_i] = \alpha + \tau \mathbb{E}[D_i \mid Z_i] = \alpha + \tau \cdot (\gamma Z_i)$$

TSLS estimands

- Under the model, we have the following CEF: $\mathbb{E}[Y_i \mid Z_i] = \alpha + \tau \cdot (\gamma Z_i)$
    - $\rightsquigarrow$ a regression of Y_i on γZ_i would have τ as the slope.
- If the CEF is linear, we have this simple relationship slopes:

$$\mathbb{E}[D_i \mid Z_i] = \delta + \gamma Z_i \quad \rightsquigarrow \quad \gamma = \frac{\text{cov}(D_i, Z_i)}{\mathbb{V}[Z_i]}$$

- Applying this to above CEF we have:

$$\tau = \frac{\text{cov}(Y_i, \gamma Z_i)}{\mathbb{V}[\gamma Z_i]} = \frac{\text{cov}(Y_i, Z_i)}{\gamma \mathbb{V}[Z_i]} = \frac{\text{cov}(Y_i, Z_i)}{\text{cov}(D_i, Z_i)}$$

- TSLS estimator:
    - Estimate $\widehat{\gamma}$ from regression of treatment D_i on instrument Z_i
    - Estimate $\widehat{\tau}_{2SLS}$ as the slope of a regression of Y_i on $\widehat{\gamma} Z_i$
    - Under this model, $\widehat{\tau}_{2SLS} \overset{p}{\rightarrow} \tau$ (but don't use SEs from second stage)

Binary treatment and instrument

- Under binary treatment/instrument, TSLS estimand is the LATE:

$$\tau = \frac{\text{cov}(Y_i, Z_i)}{\text{cov}(D_i, Z_i)} = \frac{\mathbb{E}[Y_i \mid Z_i = 1] - \mathbb{E}[Y_i \mid Z_i = 0]}{\mathbb{E}[D_i \mid Z_i = 1] - \mathbb{E}[Y_i \mid D_i = 0]} = \frac{\text{ITT}_Y}{\text{ITT}_D} = \tau_{\text{LATE}}$$

- And the TSLS estimator is the Wald estimator:

$$\widehat{\tau}_{2SLS} = \frac{\widehat{\text{cov}}(Y_i, Z_i)}{\widehat{\text{cov}}(D_i, Z_i)} = \frac{\overline{Y}_1 - \overline{Y}_0}{\overline{D}_1 - \overline{D}_0} = \frac{\widehat{\text{ITT}}_Y}{\widehat{\text{ITT}}_D} = \widehat{\tau}_{iv}$$

- $\rightsquigarrow$ constant effects model not required for TSLS in this setting.
- But we need constant effects when we add covariates:

$$Y_i = \alpha + \tau D_i + \mathbf{X}_i'\beta_y + \varepsilon_i$$
$$D_i = \delta + \gamma Z_i + \mathbf{X}_i'\beta_d + \eta_i$$

  - Otherwise, τ is an odd weighted function of causal effects and $\tau \neq \tau_{\text{LATE}}$

Weak instruments

- IV is unstable when instrument weakly affects treatment $\text{cov}(D_i, Z_i) \approx 0$.
- **Example** completely irrelevant instrument:

$$\begin{aligned} Y_i &= \tau D_i + \varepsilon_i & \mathbb{E}[\varepsilon_i \mid D_i] \neq 0 \\ D_i &= 0 \times Z_i + \eta_i & \mathbb{E}[\varepsilon_i \mid Z_i] = \mathbb{E}[\eta_i \mid Z_i] = 0 \end{aligned}$$

  - Note that we only assume mean independence, so $\text{cov}(D_i, Z_i)$ could be nonzero.
- We can write the bias of the Wald estimator as:

$$\widehat{\tau}_{iv} - \tau = \frac{\widehat{\text{cov}}(\tau D_i + \varepsilon_i, Z_i)}{\widehat{\text{cov}}(D_i, Z_i)} - \tau = \frac{\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i Z_i}{\frac{1}{n}\sum_{i=1}^{n} \eta_i Z_i} \xrightarrow{d} \underbrace{\frac{\text{cov}(\varepsilon_i, \eta_i)}{\mathbb{V}[\varepsilon_i]}}_{\text{bias}} + \underbrace{W_i}_{\text{Cauchy r.v.}}$$

- Inconsistent and asymptotically heavy tails (bc of Cauchy)
  - When $Z \rightarrow D$ effect is small but non-zero we see similar behavior.

What to do about weak instruments?

- Detecting weak instruments:
  - F-test on instruments (excluded from second stage): $H_0 : \gamma = 0$.
  - Rule of thumb: bias is small when F-stat ≥ 10 (Stock & Yogo, 2005)
  - Correct coverage may require cutoff $F \geq 104.7$ (Lee et al, 2020)
  - The latter is a worst-case, typical data maybe ok with 10 cutoff
- Anderson-Rubin (1949) test (simplified setting, binary Z/D)
  - $H_0 : \tau = \tau_0$ equivalent to $H_0 : \text{ITT}_Y - \text{ITT}_D \cdot \tau_0 = 0$
  - Under the null, asymptotically we have

$$g(\tau_0) = \widehat{\text{ITT}}_Y - \widehat{\text{ITT}}_D \tau_0 \sim N(0, \Omega(\tau_0))$$
$$\Omega(\tau_0) = \mathbb{V}[\widehat{\text{ITT}}_Y] + \tau_0^2 \mathbb{V}[\widehat{\text{ITT}}_D] - 2\tau_0 \text{cov}(\widehat{\text{ITT}}_Y, \widehat{\text{ITT}}_D)$$

  - AR test statistic: $g(\tau_0)^2 / \Omega(\tau_0) \sim \chi^2$ no matter first-stage effect.
  - Can invert (analytically!) to get confidence intervals

Multi-valued treatments

- Generalization of these ideas:
    - Multi-valued treatment: $D_i \in \{0, 1, \dots, K-1\}$
    - Binary instrument: $Z_i \in \{0, 1\}$
- Assumptions:
    - Randomization: $[\{Y_i(d, z), \forall d, z\}, D_i(1), D_i(0)] \perp\!\!\!\perp Z_i$
    - Monotonicity: $D_i(1) \geq D_i(0)$ (instrument only increases treatment)
    - Exclusion restriction: $Y_i(1, d) = Y_i(0, d)$ for all $d = 0, 1, \dots, K-1$
- Can't identify the proportion of all compliance types here.
- Example: $K = 3 \rightsquigarrow 9$ principal strata
    - Affected: $(D_i(0), D_i(1)) \in \{(0, 1), (0, 2), (1, 2)\}$
    - Unaffected: $(D_i(0), D_i(1)) \in \{(0, 0), (1, 1), (2, 2)\}$
    - Negatively affected: $(D_i(0), D_i(1)) \in \{(1, 0), (2, 0), (2, 1)\}$
    - Last ruled out by monotonicity.
    - 5 unknowns and 4 knowns under monotonicity.

TSLS with multivalued treatments

- Let $C_i = jk$ be an indicator for compliance type $D_i(1) = j$ and $D_i(0) = k$.
    - People that are moved from k to j by the instrument.
    - Let $\rho_{jk} = \mathbb{P}(D_i(1) = j, D_i(0) = k)$ be the strata size.
- We can show that the 2SLS estimator converges to:

$$\widehat{\tau}_{2SLS} \overset{p}{\to} \sum_{k=0}^{K-1} \sum_{j=k+1}^{K-1} \omega_{jk} \mathbb{E}\left(\frac{Y_i(1) - Y_i(0)}{j - k} \mid C_i = jk \right)$$

$$\omega_{jk} = \frac{(j-k)\rho_{jk}}{\sum_{s=0}^{K-1} \sum_{t=s+1}^{K-1} (s-t)\rho_{st}}$$

- Intuition: a weighted average of effects per dose for each affected type.
    - Weights are proportional to size of the strata and how big the effect of the instrument is for that strata.
    - If instrument can only increase by 1 dose, then simplifies to weighted average of principal strata effects.

6/ General two-stage least squares

General 2SLS

- Linear model for each i:

$$Y_i = \mathbf{X}_i'\beta + \varepsilon_i$$

  - $\mathbf{X}_i$ is $k \times 1$ now includes D_i and any pretreatment covariates.
  - Parts of $\mathbf{X}_i$ are endogenous so that $\mathbb{E}[\varepsilon_i \mid \mathbf{X}_i] \neq 0$
- Instruments $\mathbf{Z}_i$ that is $\ell \times 1$ vector such that $\mathbb{E}[\varepsilon_i \mid \mathbf{Z}_i] = 0$.
  - $\mathbf{Z}_i$ might include exogenous/pretreatment variables from $\mathbf{X}_i$ as well.
  - Rank condition: $\mathbb{E}[\mathbf{Z}_i\mathbf{Z}_i']$ and $\mathbb{E}[\mathbf{X}_i\mathbf{Z}_i']$ have full rank.
- Identification:
  - $k = \ell$: just-identified.
  - $k < \ell$: over-identified (can test the exclusion restriction, kinda)
  - $k > \ell$: unidentified (fails rank condition)

Nasty Matrix Algebra

- Projection matrix projects values of $\mathbf{X}_i$ onto $\mathbf{Z}_i$:

$$\begin{aligned} \mathbf{\Pi} &= (\mathbb{E}[\mathbf{Z}_i \mathbf{Z}_i'])^{-1} \mathbb{E}[\mathbf{Z}_i \mathbf{X}_i'] && \text{(projection matrix)} \\ \tilde{\mathbf{X}}_i &= \mathbf{\Pi}' \mathbf{Z}_i && \text{(projected values)} \end{aligned}$$

- To derive the 2SLS estimator, take the fitted values, $\mathbf{\Pi}'\mathbf{Z_i}$ and multiply both sides of the outcome equation by them:

$$\begin{aligned} Y_i &= \mathbf{X}_i'\beta + \varepsilon_i \\ \mathbf{\Pi}'\mathbf{Z}_i Y_i &= \mathbf{\Pi}'\mathbf{Z}_i \mathbf{X}_i'\beta + \mathbf{\Pi}'\mathbf{Z}_i \varepsilon_i \\ \mathbb{E}[\mathbf{\Pi}'\mathbf{Z}_i Y_i] &= \mathbb{E}[\mathbf{\Pi}'\mathbf{Z}_i \mathbf{X}_i']\beta + \mathbb{E}[\mathbf{\Pi}'\mathbf{Z}_i \varepsilon_i] \\ \mathbb{E}[\mathbf{\Pi}'\mathbf{Z}_i Y_i] &= \mathbb{E}[\mathbf{\Pi}'\mathbf{Z}_i \mathbf{X}_i']\beta + \mathbf{\Pi}' \mathbb{E}[\mathbf{Z}_i \varepsilon_i] \\ \mathbb{E}[\mathbf{\Pi}'\mathbf{Z}_i Y_i] &= \mathbb{E}[\mathbf{\Pi}'\mathbf{Z}_i \mathbf{X}_i']\beta \\ \mathbb{E}[\tilde{\mathbf{X}}_i Y_i] &= \mathbb{E}[\tilde{\mathbf{X}}_i \mathbf{X}_i']\beta \\ \beta &= (\mathbb{E}[\tilde{\mathbf{X}}_i \mathbf{X}_i'])^{-1} \mathbb{E}[\tilde{\mathbf{X}}_i Y_i] \end{aligned}$$

How to estimate the parameters

- Collect $\mathbf{X}_i$ into a $n \times k$ matrix $\mathbb{X} = (\mathbf{X}_1', \ldots, \mathbf{X}_n')$
- Collect $\mathbf{Z}_i$ into a $n \times \ell$ matrix $\mathbb{Z} = (\mathbf{Z}_1', \ldots, \mathbf{Z}_n')$
- In-sample projection matrix produces fitted values: $\widehat{\mathbb{X}} = \mathbb{Z}(\mathbb{Z}'\mathbb{Z})^{-1}\mathbb{Z}'\mathbb{X}$
    - Fitted values of regression of $\mathbb{X}$ on $\mathbb{Z}$.
    - Matrix party trick: $\mathbb{X}'\mathbb{Z}/n = (1/n)\sum_i^n \mathbf{X}_i\mathbf{Z}_i' \overset{p}{\to} \mathbb{E}[\mathbf{X}_i\mathbf{Z}_i']$.
- Take the population formula for the parameters:

$$\beta = (\mathbb{E}[\tilde{\mathbf{X}}_i\mathbf{X}_i'])^{-1}\mathbb{E}[\tilde{\mathbf{X}}_i Y_i]$$

- And plug in the sample values (the n cancels out):

$$\hat{\beta}_{2SLS} = (\widehat{\mathbb{X}}'\mathbb{X})^{-1}\widehat{\mathbb{X}}'\mathbf{y} \overset{p}{\to} \beta$$

- This is how R/Stata estimates the 2SLS parameters

Asymptotic variance for 2SLS

- We can write the centered, normalized TSLS estimator as:

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta) = \underbrace{\left(n^{-1}\sum_i \widehat{\mathbf{X}}_i\widehat{\mathbf{X}}_i'\right)^{-1}}_{\xrightarrow{p}(\mathbb{E}[\widehat{\mathbf{X}}_i\widehat{\mathbf{X}}_i'])^{-1}} \underbrace{\left(n^{-1/2}\sum_i \widehat{\mathbf{X}}_i\varepsilon_i\right)}_{\xrightarrow{d}N(0,\mathbb{E}[\widehat{\mathbf{X}}_i'\varepsilon_i'\varepsilon_i\widehat{\mathbf{X}}_i])}$$

- Thus, we have that $\sqrt{n}(\hat{\beta}_{2SLS} - \beta)$ has asymptotic variance:

$$(\mathbb{E}[\widehat{\mathbf{X}}_i\widehat{\mathbf{X}}_i'])^{-1}\mathbb{E}[\widehat{\mathbf{X}}_i'\varepsilon_i'\varepsilon_i\widehat{\mathbf{X}}_i](\mathbb{E}[\widehat{\mathbf{X}}_i\widehat{\mathbf{X}}_i'])^{-1}$$

- **Robust 2SLS variance estimator** with residuals $\hat{u}_i = Y_i - \mathbf{X}_i'\hat{\beta}$:

$$\widehat{\text{var}}(\hat{\beta}_{2SLS}) = (\widehat{\mathbb{X}}'\widehat{\mathbb{X}})^{-1}\left(\sum_i \hat{u}_i^2\widehat{\mathbf{X}}_i\widehat{\mathbf{X}}_i'\right)(\widehat{\mathbb{X}}'\widehat{\mathbb{X}})^{-1}$$

- HC2, clutering, and autocorrelation versions exist