Foundations of Econometrics

Exam Notes

Mathias Schindler
Data Science for Decision Making, Class of '22,
Barcelona School of Economics

Dec 14, 2021

Contents

I Maite's Part **3**

1 Formulas to Remember **3**
1.1 Graphical Statistical Tests 4

2 OLS in Finite Sample Theory **6**
2.1 Classical Assumptions 6
2.2 Bits and Pieces 6

3 Statistics and DGP **8**
3.1 Statistics – Not dependent on DGP 8
3.2 Statistics – Dependent on DGP 8

4 Statistics and Units **9**
4.1 Statistics – Unit-dependent 9
4.2 Statistics – Unit-free 9

5 Large Sample Theory / Asymptotics **10**
5.1 Asymptotic Assumptions 10
5.2 Asymptotic OLS Estimator 10
5.3 Asymptotic Statistical Tests 10

6 Non-Spherical Disturbances **12**
6.1 Heteroskedasticity 12
6.2 Clustering 13

7 Endogeneity **14**

II Hanna's Part **15**

8 Key Assumptions **15**

9 Required Data **16**

10 Estimators **17**

11 Difference-in-Differences **19**
11.1 Synthetic Control Groups 20

12 Regression Discontinuity **21**
12.1 Sharp RD 21
12.1.1 Methodology 23
12.1.2 Falsification Checks 24
12.1.3 Limitations of Sharp RD 26
12.2 Fuzzy RD 26
12.2.1 Fuzzy RD as Instrument 27

13 Homeworks **30**
13.1 Assignment 7 30

14 Student Presentations **31**
14.1 My Presentation 31

Part I
Maite's Part

1 Formulas to Remember

OLS Estimator

$$\hat{\beta} = (X'X)^{-1}X'Y$$

OLS Distribution Under classical assumptions (see 2.1):[1]

$$\begin{aligned} & \hat{\beta}|X \sim \mathcal{N}(\beta, \sigma^2(X'X)^{-1}) \\ \Rightarrow \quad & \hat{\beta}_k|X \sim \mathcal{N}(\beta_k, \sigma^2(X'X)^{-1}_{kk}) \end{aligned}$$

Standard Error

$$\text{s.e.}(\hat{\beta}_k) = \sqrt{\hat{\sigma}^2(X'X)^{-1}_{kk}}, \qquad \hat{\sigma}^2 = \frac{SSR}{n-K}$$

Variance Decomposition

$$Var(\hat{\beta}_k|X) = \sigma^2 \times \frac{1}{SST_k} \times \frac{1}{1-R_k^2}$$

R_k^2: From regressing x_k on all other regressors
SST_k: From regressing x_k on all other regressors (or $\sum_{i=1}^n (x_{ik} - \bar{x}_{ik})^2$)

***t*-test** Assumption: Normal errors and strict exogeneity[2]

$$t\text{−statistic} = \frac{\hat{\beta}_k - r}{\text{s.e.}(\hat{\beta}_k)} \underset{\text{under } \mathcal{H}_0}{\sim} t(n-K)$$

r: From $\mathcal{H}_0 : \hat{\beta}_k = r$

***F*-test** Assumption: Normal errors and strict exogeneity[3]

$$F = \frac{[RSSR - SSR]/q}{SSR/[n-K]} \underset{\text{under } \mathcal{H}_0}{\sim} F(q, n-K)$$

$RSSR$: Restricted model's residual sum of squares[4]
q: Number restrictions tested

[1] Unbiased, $\mathbb{E}\{\mathcal{N}(.)\}$, for Gauss-Markov assumptions 1-4. Variance only if assumption 5 as well, $Var\{\mathcal{N}(.)\}$. Assumptions: 2.1.

[2] Otherwise: Asymptotic t-test $\overset{a}{\sim} \mathcal{N}(0,1)$ – see Section 5.3.

[3] Otherwise: Asymptotic F-test $\overset{a}{\sim} \frac{\chi^2(q)}{q}$ or Wald test – see Section 5.3.

[4] Same as main model but without the parameters tested.

***p*-value**

$$p\text{−value} = Pr(|i - \text{statistic}| > |i - \text{critical value}|) \qquad ([5])$$

***R*²**

$$R^2 = \frac{SSE}{SST}, \qquad R^2 = 1 - \frac{SSR}{SST}$$

SSE Explained Sum of Squares

$$SSE = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

SSR Residual Sum of Squares

$$SSR = \sum_{i=1}^{n} \hat{u}_i^2$$

SST Total Sum of Squares

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

1.1 Graphical Statistical Tests

Intuition Assuming $\mathcal{H}_0$, a tested parameter is distributed with most probability mass close to zero. If $\mathcal{H}_0$ were true, the test statistic would be 'close' to the tested value (i.e. 0)

→ Critical values determine what is 'far' and 'close'. Distance measure is unit-free (Mahalanobis).

***t*-test** Figure 1.

***F*-test** Figure 2.

χ^2-test Essentially the same as F-test.
Difference: Doesn't depend on DGP (no need for normality nor i.i.d.).

[5] i: E.g. t, F or χ^2.

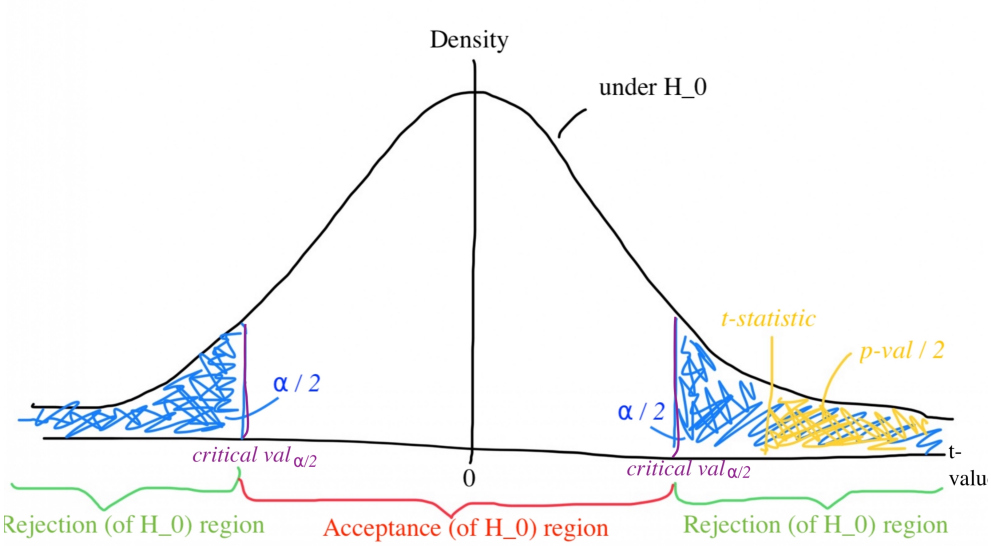

Figure 1: Two-Sided t-Test with Rejection of $\mathcal{H}_0$

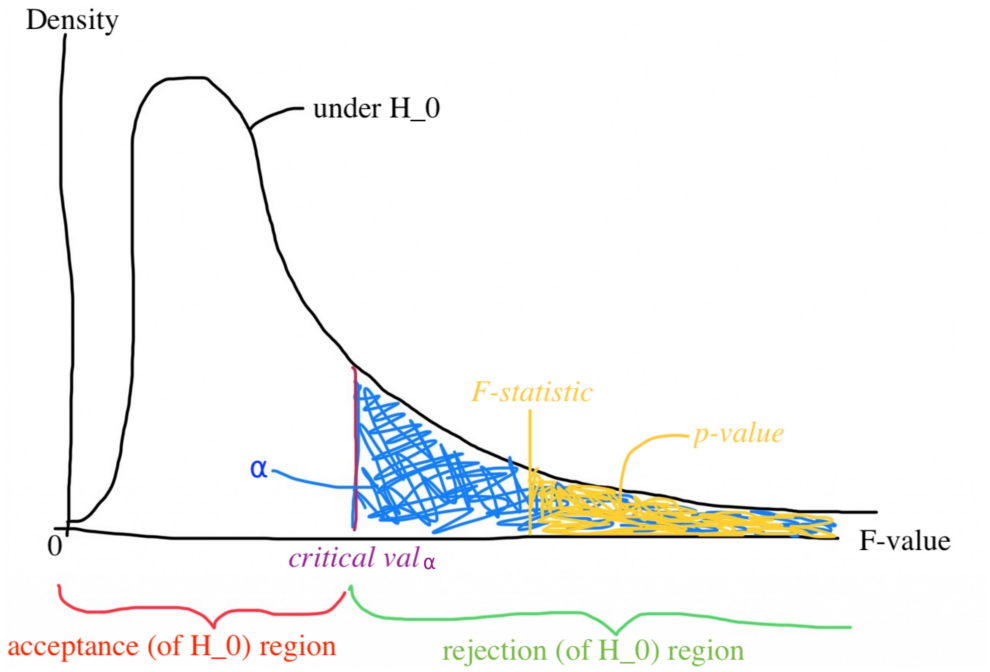

Figure 2: F-Test with Rejection of $\mathcal{H}_0$

2 OLS in Finite Sample Theory

2.1 Classical Assumptions

1. Linearity
2. No multicollinearity
3. Strict Exogeneity: $\mathbb{E}(\varepsilon_i|X) = 0 \ \forall i$
4. Conditional Spheric Disturbances
   - Homoscedasticity (Conditional): $\mathbb{E}(\varepsilon_i^2|X) = \sigma^2 > 0 \ \forall i$
   - Uncorrelated (conditional) disturbances: $\mathbb{E}(\varepsilon_i\varepsilon_j|X) = 0 \ \forall i \neq j$
5. Normality of Conditional Disturbances Vector: $\varepsilon|X \sim \mathcal{N}(.)$

Lingo 1–4 are "Gauss-Markov Assumptions"
1–5 are "Classical Assumptions"

2.2 Bits and Pieces

Sum of Residuals Equal to zero $\Rightarrow$ if constant term is included!

$$\sum_{i=1}^{n} \hat{\varepsilon}_i = 0$$

Intercept

$$\hat{\beta}_0 = \bar{y} \qquad \text{if} \quad y = \beta_0 + \varepsilon \tag{1}$$

Intercept is sample mean if regression w. no regressors.[6]

Variance Inflating Factor Multicollinearity for $VIF >$ 5-10 (small samples).

$$VIF_k = \frac{1}{1 - R_k^2}$$

R_k^2 from auxiliary regression which regresses x_k on all other regressors.

R^2 Interpretation The added explained variation of adding predictors compared to using the sample mean.

[6] I.e. at baseline levels where all $X_i = 0, i = 1, ..., k$

R^2 for No Regressors Model with no regressors then $R^2 = 0$.

Proof.

$$SSE \equiv \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 = \sum_{i=1}^{n} (\overbrace{X}^{=1} \overbrace{\hat{\beta}_0}^{=\bar{y} \text{ cf. } (1)} - \bar{y})^2) = \sum_{i=1}^{n} (\bar{y} - \bar{y})^2 = 0$$

$$\Rightarrow \quad R^2 = \frac{SSE}{SST} = \frac{0}{SST} = 0$$

□

3 Statistics and DGP

3.1 Statistics – Not dependent on DGP

OLS Estimator

$$\hat{\beta} = (X'X)^{-1}X'Y$$

It is a function only of the observed data. Hence, the DGP (distributions) has nothing to do with it. It doesn't depend on the DGP.

R^2

$$R^2 = \frac{SSE}{SST}, \quad R^2 = 1 - \frac{SSR}{SST}$$

SSE, SST, SSR calulated from observed data and OLS estimates. Hence doesn't depend on the DGP.

3.2 Statistics – Dependent on DGP

Standard Error

$$\text{s.e.}(\hat{\beta}_k) \equiv \sqrt{\widehat{Var(\hat{\beta}_k|X)}} = \sqrt{\hat{\sigma}^2(X'X)^{-1}_{kk}} \qquad \text{where} \quad \hat{\sigma}^2 = \frac{SSR}{n-K}$$

so because it depends on $\hat{\sigma}^2$ it depends on the DGP.[7]

***t*-statistic**

$$t\text{-statistic} = \frac{\hat{\beta}_k - r}{\text{s.e.}(\hat{\beta}_k)}$$

which depends on s.e.$(\hat{\beta}_k)$ and hence on the DGP.

***p*-value**

$$p - \text{value} \equiv Pr(|t\text{-statistic}| > |t\text{-critical value}|)$$

where t-statistic depends on s.e.$(\hat{\beta}_k)$ and hence on the DGP.

[7] When classical assumptions don't hold then

$$Var(\hat{\beta}_k|X) \neq \sigma^2(X'X)^{-1}_{kk}$$

4 Statistics and Units

4.1 Statistics – Unit-dependent

OLS Estimator $\hat{\beta}_k$
Changing units of regressor k changes $\hat{\beta}_k$. It doesn't change $\hat{\beta}_{\neg k}$.

Standard Errors s.e.$(\hat{\beta}_k)$

Confidence Intervals

$$CI = \{\hat{\beta}_k \pm 1.96 \times \text{s.e.}(\hat{\beta}_\text{k})\}$$

because includes $\hat{\beta}_k, \text{s.e.}(\hat{\beta}_\text{k})$ is unit-dependent.

4.2 Statistics – Unit-free

***t*-test** Based on Mahalanobis distance (unit-free).

***p*-value** Based on t-statistic and hence unit-free,

$$p - \text{value} \equiv Pr(|t\text{-statistic}| > |t\text{-critical value}|)$$

***F*-test** Based on Mahalanobis distance (unit-free).

5 Large Sample Theory / Asymptotics

5.1 Asymptotic Assumptions

1. Linearity
2. Regressors:
   (i) matrix X has full rank with probability 1
   (ii) $Q_{XX} \equiv plim \sum_i \frac{x_i x_i'}{n}$ is positive definite
3. Weak Exogeneity: $\mathbb{E}(\varepsilon_i | x_i) = 0 \ \forall i$
4. Homoscedasticity (Conditional): $\mathbb{E}(\varepsilon_i^2 | x_i) = \sigma^2 \ \forall i$
5. (Data are independent over i)

5.2 Asymptotic OLS Estimator

Convergence Consistent for: (i) Linearity, (iii) Weak exogeneity and (ii) Full-rank matrix X and positive definite Q_{XX}. I.e. assumptions 1-3 in 5.1.

Asymptotic Distribution

$$\sqrt{n}(\hat{\beta}_n - \beta) \overset{a}{\sim} \mathcal{N}(0, \underbrace{\sigma^2 Q_{XX}^{-1}}_{\equiv avar(\hat{\beta})}) \tag{8}$$

5.3 Asymptotic Statistical Tests

Asymptotic *t*-test

$$t\text{-statistic} = \frac{\hat{\beta}_k - r}{\text{s.e.}(\hat{\beta}_k)} \underset{\text{under } \mathcal{H}_0}{\overset{a}{\sim}} \mathcal{N}(0,1)$$

Unnecessary assumptions:[9]

- Strict exogeneity
- Normality of errors

Valid for

- Weak exogeneity
- Linearity
- Large samples

[8] Where $Q_{XX} \equiv plim\left(\frac{X'X}{n}\right)$.
[9] Relative to exact t-test)

Comparison to exact t-test

- $t(n-K)$-distribution fatter tails than $\mathcal{N}(0,1)$
  $\rightarrow$ Larger critical values $\rightarrow$ wider acceptance region $\rightarrow$ larger p-values $\rightarrow$ more conservative inference.

Wald Test .

Advantage Defined asymptotic distribution without:

- (i) Nonnormality of errors
- (ii) Strict exogeneity

Disadvantage F-test converges to Wald test at the limit anyway so no need for it really.

Asymptotic F-test

$$F\text{-statistic} = \frac{\text{Wald test}}{q} \underset{\text{under } \mathcal{H}_0}{\overset{a}{\sim}} \frac{\chi^2(q)}{q}$$

Unnecessary assumptions:[10]

- Strict exogeneity
- Normality of errors

Valid for

- Weak exogeneity
- Linearity
- Large samples

Comparison to exact F-test

- $F(q, n-K)$-distribution fatter tails than $\frac{\chi^2(q)}{q}$
  $\rightarrow$ Larger critical values $\rightarrow$ wider acceptance region $\rightarrow$ larger p-values $\rightarrow$ more conservative inference.

[10] Relative to exact F-test)

6 Non-Spherical Disturbances

6.1 Heteroskedasticity

Definition When

$$\Sigma = \begin{bmatrix} \sigma_1^2 & & \dots & \mathbf{0} \\ & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & \\ \mathbf{0} & \dots & & \sigma_n^2 \end{bmatrix}$$

where n is number of observations

Detection Plot Plot fitted values on x-axis and residuals on y-axis. See Figure 3.

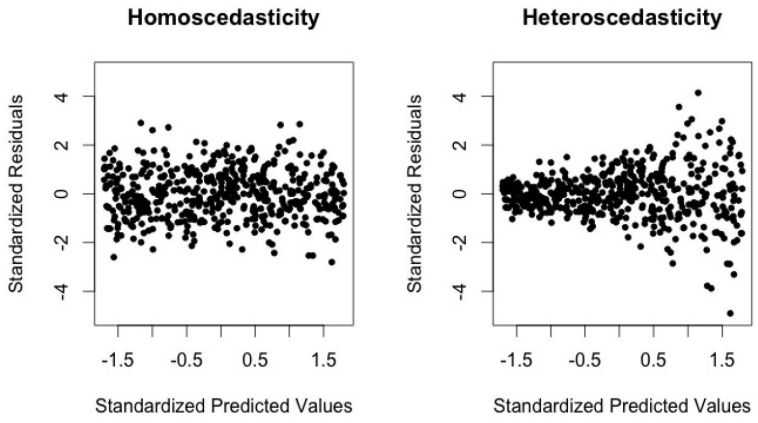

Figure 3: Heteroscedasticity Detection Plot

Breusch-Pagan Test Test whether fitted values $\hat{y}_i$ can explain residuals $\hat{\varepsilon}_i$ (RHS of Fig 3) or not (LHS of Fig 3):

$$\mathcal{H}_0 : \alpha_1 = 0, \quad \mathcal{H}_1 : \neg\mathcal{H}_0$$

for regression: $\hat{\varepsilon}_i = \alpha_0 + \alpha_1 \hat{y}_i + \mu_i$

Null is homoscedastic disturbances against alternative of heteroscedasticity.

6.2 Clustering

Definition When

$$\Sigma = \begin{bmatrix} \Sigma_1 & & \dots & \mathbf{0} \\ & \Sigma_2 & & \vdots \\ \vdots & & \ddots & \\ \mathbf{0} & \dots & & \Sigma_C \end{bmatrix}, \quad \text{with:} \quad \Sigma_c = \begin{bmatrix} \sigma_1^2 & & \dots & \neq \mathbf{0} \\ & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & \\ \neq \mathbf{0} & \dots & & \sigma_{n_c}^2 \end{bmatrix}$$

where C is number of clusters.

Magnitude Clustered s.e.'s are usualy always larger. Default standard errors fail to include the positive covariances between disturbances within each given cluster.

Larger clustered standard errors $\rightarrow$ smaller t-values.

Consistency of Estimator Cluster-Robust Variance Matrix Estimator is consistent for $C \rightarrow \infty$. In practice: Have $C > 50$ clusters.

If Clusters$<$ 50 Can use bootstrapping.

7 Endogeneity

Confounders or Omitted Variables Bias.
Solution: Use treatment effect models.

Simultaneity

Measurement Error In Regressor (Only)
If

$$x^{\text{observed}} = x^{\text{true}} + v_i$$

where v_i is error term with mean 0 and some variance. Then slope biases estimate towards zero (attenuation bias). As in Figure 4; red line with measurement error, green line without.

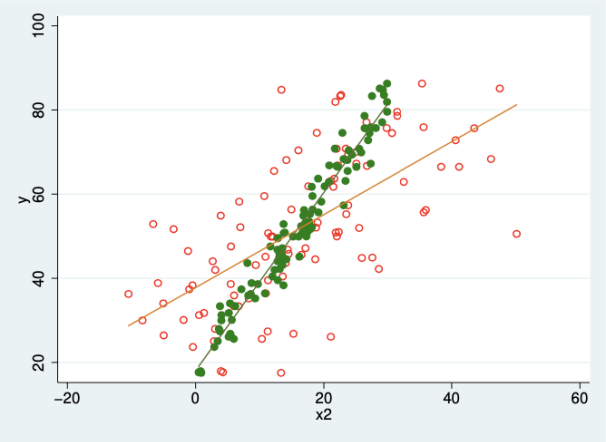

Figure 4: Measurement Error in Regressor

In Outcome (Only)
If unsystematic measurement error (i.e. mean zero). Then parameter estimates will be the same as without error. However, standard errors will be larger.

No Strict/Weak Exogeneity

$$\mathbb{E}(\varepsilon_i|X) \neq 0 \quad \forall i \qquad \text{(Strict)}$$
$$\mathbb{E}(\varepsilon_i|x_i) \neq 0 \quad \forall i \qquad \text{(Weak)}$$

$\Rightarrow$ OLS estimator is biased: $\mathbb{E}(\hat{\beta}|X) \neq \beta$
$\rightarrow$ Exact tests are invalid

$\Rightarrow$ OLS estimator is inconsistent: $plim(\hat{\beta}) \neq \beta$
$\rightarrow$ Asymptotic tests are invalid

Part II
Hanna's Part

8 Key Assumptions

Randomized Experiments Unconditional independence of treatment for potential outcomes

$$Y_{i0}, Y_{i1} \perp\!\!\!\perp D_i$$

Matching Conditional Independence. Potential outcomes not affected by anything other than treatment once controlled for characteristics:

$$Y_{i0}, Y_{i1} \perp\!\!\!\perp D_i | X_i$$

Difference-in-Differences Parallel Trend: In absence of treatment, outcomes for groups would have grown in parallel (or stayed the same).

$$\mathbb{E}(Y_{i0}|D_i = 1, t = 1) - \mathbb{E}(Y_{i0}|D_i = 1, t = 0) = \mathbb{E}(Y_{i0}|D_i = 0, t = 1) - \mathbb{E}(Y_{i0}|D_i = 0, t = 0)$$

Regression Discontinuity Individuals close to the cutoff (right under and right over) are statistically (almost) identical.

Instrumental Variables

- Relevance: $Cov(Z_i, X_i) \neq 0$
- Exclusion: $Z_i \to X_i \to Y_i$ and $Z_i \nrightarrow Y_i$
- Monotonicity:[11] No defiers in sample

[11] For heterogenuous treatment effects.

9 Required Data

Randomized Experiments

- Outcomes $Y(t = 0)$, pre-treatment
- Outcomes $Y(t = 1)$, post-treatment
- Indicator, control or treatment group

Matching

- Outcomes $Y(t = 1)$, post-treatment
- Indicator, control or treatment group
- Characteristics, X

Difference-in-Differences

- Outcomes $Y(t = 0)$, pre-treatment
- Outcomes $Y(t = 1)$, post-treatment
- Indicator, control or treatment group
- Characteristics, X

Regression Discontinuity

- Outcomes $Y(t = 0)$, pre-treatment
- Outcomes $Y(t = 1)$, post-treatment
- Forcing variable, Q
- Eligibility cutoff, D_i
- Characteristics, X

Instrumental Variables

- Outcomes $Y(t = 0)$, pre-treatment
- Outcomes $Y(t = 1)$, post-treatment
- Indicator, control or treatment group
- Characteristics, X

10 Estimators

Randomized Experiments

$$\hat{\alpha} \equiv \bar{Y}_{treat} - \bar{Y}_{control} = \alpha_{\text{ATE}} = \alpha_{\text{ATT}}$$

Matching When $X_i \approx X_i^M \Rightarrow Y_i(0) \approx Y_i^M(0)$ where $Y_i(0)$ is unobserved.

$$\alpha_{ATE} = \sum_{i=1}^{N} [\bar{Y}^j(1) - \bar{Y}^j(0)] \frac{N^j}{N}$$

So weight by number in category j as proportion of total N

$$\alpha_{ATT} = \sum_{i=1}^{N} [\bar{Y}^j(1) - \bar{Y}^j(0)] \frac{N^j_{treat}}{N_{treat}}$$

So weight by number of treated in category j as proportion of total treated N_{treat}

Difference-in-Differences

$$\begin{aligned}\alpha_{DiD} = &\underbrace{\{\mathbb{E}(Y|D_i = 1, t = 1) - \mathbb{E}(Y|D_i = 1, t = 0)\}}_{\text{=Effect on treatment group}} \\ &- \underbrace{\{\mathbb{E}(Y|D_i = 0, t = 1) - \mathbb{E}(Y|D_i = 0, t = 0)\}}_{\text{=Effect on control group}}\end{aligned}$$

Whether ATE or ATT not mentioned anywhere.

Regression Discontinuity

$$\begin{aligned}\alpha_{\text{RD}} &\equiv \mathbb{E}(Y_1|X = c) - \mathbb{E}(Y_0|X = c) &(2)\\ &= \lim_{k \underset{\text{from left}}{\to} c} \mathbb{E}(Y|X = k) - \lim_{k \underset{\text{from right}}{\to} c} \mathbb{E}(Y|X = k) &(3)\end{aligned}$$

- Sharp RD: ATT (around cutoff) – all right to cutoff are treated
- Fuzzy RD: LATE (around cutoff), or/also ITT[12] – effect of eligibility of treatment.

Instrumental Variables

$$\alpha_{\text{IV}}^{\text{homo}} \equiv \frac{\mathbb{E}(Y_i|Z_i = 1) - \mathbb{E}(Y_i|Z_i = 0)}{\mathbb{E}(D_i|Z_i = 1) - \mathbb{E}(D_i|Z_i = 0)}$$

- Homogenuous Effects: ATE

[12] Intent-to-Treat

- Heterogenuous Effects:
  - No defiers: LATE – Effect for compliers
  - No defiers and always-takers: LATE=ATT – only treated units are the compliers
  - No defiers, always-takers, never-takers: LATE=ATT=ATE – sample is only compliers

11 Difference-in-Differences

Key Assumption In absence of treatment, outcomes for groups would have grown in parallel (or stayed the same), i.e. Parallel Trend.

Estimator

$$\begin{aligned}\delta_{DiD} = &\underbrace{\{\mathbb{E}(Y|D_i = 1, t = 1) - \mathbb{E}(Y|D_i = 1, t = 0)\}}_{\text{=Effect on treatment group}} \\ &- \underbrace{\{\mathbb{E}(Y|D_i = 0, t = 1) - \mathbb{E}(Y|D_i = 0, t = 0)\}}_{\text{=Effect on control group}}\end{aligned}$$

Graphical For binary treatment $D_i \in \{0, 1\}$ and two-time period $t \in \{0, 1\}$ see Figure 5.

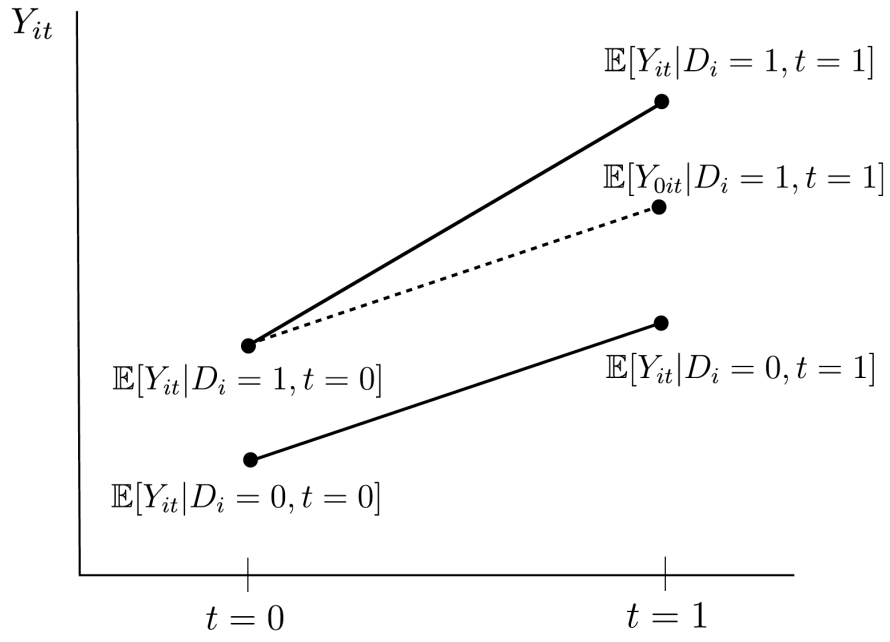

Figure 5: Difference-in-Differences in Binary Case

Regression Model

$$Y = \beta_0 + \delta_{DiD}(D \times t) + \beta_1 D + \beta_2 t + \varepsilon$$

11.1 Synthetic Control Groups

Basically used by taking multiple control groups and doing a weighted average with weights chose so as to mimic the treatment group optimally as in Figure 6.

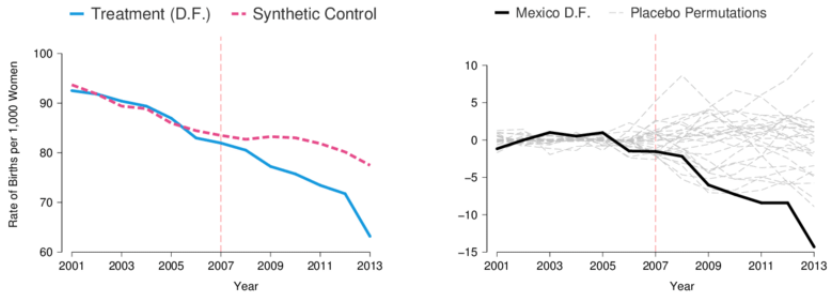

Figure 6: Difference-in-Differences with Synethetic Control Group

The idea is similar to matching and estimates the weights such that

$$t > T_0: \quad Y_{1t}(0) = \sum_{j=2}^{J} \omega_j Y_{jt}, \quad j = 1, ..., J$$

where $Y_{1t}(0)$ is the estimated counterfactual after time of treatment $t > T_0$, $j = 2, ..., J$ are the control groups (i.e. $j = 1$ is the treatment group), i.e. the grey lines in RHS of Figure 6, and ω_j are the estimated weights.

The weights should ideally satisfy that

$$t \leq T_0: \quad Y_{1t} = \sum_{j=2}^{J} \omega_j Y_{jt}, \quad j = 1, ..., J$$

Such that before treatment, $t \leq T_0$, the treatment group Y_{1t} and synthetic control group, $\sum_{j=2}^{J} \omega_j Y_{jt}$ are the same.

12 Regression Discontinuity

Key Assumption Individuals close to the cutoff (right under and right over) are statistically (almost) identical.

External Validty Iff the individuals around the cutoff (right under and right over) are representative of the whole population.

12.1 Sharp RD

Internal Validty Estimated effect for the individuals close around the cutoff.

Property Treatment is deterministic over and below the cutoff such that

$$Pr(D_i|g+\epsilon) = 1$$
$$Pr(D_i|g-\epsilon) = 0$$

where g is the cutoff and $\epsilon > 0$ is some threshold.

$$Pr(D_i|g+\epsilon) \in (0,1)$$
$$Pr(D_i|g-\epsilon) = 0$$

Required Variables

- Outcome Y
  Example: Earnings
- Forcing/running variable X that determines treatment
  - Continuous, i.e. $X \in \mathbb{R}$
  - Well-defined single cutoff determines eligibility of treatment $c \in X$

  Example: SAT scores
- Binary treatment $D_i\{0,1\}$
  Example: Scholarship

Assumption 1: Deterministic

$$D_i = \mathbb{1}(X_i > c)\ \forall i$$

Assumption 2: No Compliance Issue Individual cannot manipulate their X_i

Estimator

$$\alpha_{\text{RD}} \equiv \lim_{\substack{k \to c \\ \text{from left}}} \mathbb{E}(Y|X=k) - \lim_{\substack{k \to c \\ \text{from right}}} \mathbb{E}(Y|X=k) \tag{4}$$

$$= \mathbb{E}(Y_1|X=c) - \mathbb{E}(Y_0|X=c) \tag{5}$$

where we do not observe the 2nd term in (5).[13] Because of continuity the 2nd term in (4) collapses to $\mathbb{E}(Y_0|X=c)$ in the limit.

Estimated Effect Regression Discontinuity establishes the Average Treatment Effect on the Treated (ATT) – but only around the cutoff.[14]

Regression Regress (with a restricted sample with h around the cutoff and X_i normalized to zero at cutoff)[15]

$$Y = \beta_0 + \beta_1 D_i + \beta_2(X_i - c) + \beta_3[D_i \times (X_i - c)] + \varepsilon \tag{6}$$

where β_1 is then the Sharp RD-effect.

Extended Regressions (6) can be extended such that polynomial terms and controls are added:

$$\begin{aligned} Y =& \beta_0 + \beta_1 D_i + \beta_2(X_i - c) + \beta_3 D_i \times (X_i - c) + \beta_4(X_i - c)^2 + \beta_5 D_i \times (X_i - c)^2 \\ &+ \cdots + \beta_k(X_i - c)^k + \beta_{k+1} D_i \times (X_i - c)^k \\ &+ \gamma_1 Z_1 + \gamma_2 Z_2 + \cdots + \gamma_q Z_q + +\varepsilon \end{aligned} \tag{7}$$

NB: Estimation of β_1 should not be very sensitive to added polynomials and controls. Estimation of (6) should always be reported before (7).

Graphical RD Figure 7.

[13] Potential outcome of nontreatment at the cutoff, $\mathbb{E}(Y_0|X=c)$,is unobserved by construction because of deterministic assignment of treatment.

[14] https://dimewiki.worldbank.org/Regression_Discontinuity#Fuzzy_RDD.

[15] I think the interaction term $[D_i \times (X_i - c)]$ allows the slopes to differ on either side of the cutoff.

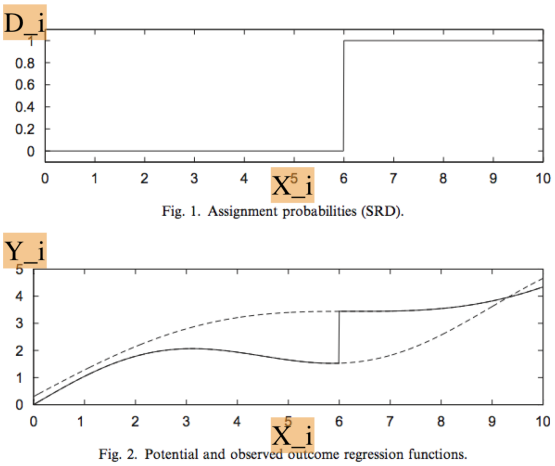

Figure 7: Graphical Sharp RD

12.1.1 Methodology

1. Transform forcing variable X to be zero at cutoff c.
2. Restrict sample to h away from cutoff.
   h chosen by the researcher. Different h-restrictions depicted in Figure 8.
3. Regress (6) and potentially add controls Z and polynomials.

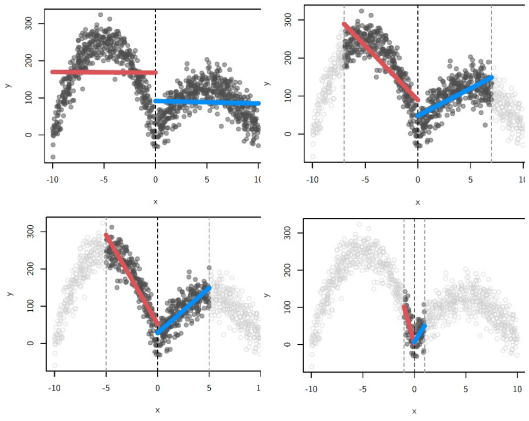

Figure 8: Restricted Samples for RD with Different h

12.1.2 Falsification Checks

Various things can go wrong so that detected discontinuity is not one in fact.

- Spurious Jump: Nonlinearity in y and x should not be mistaken for discontinuity (see Figure 9).
  - Test: Regress for different restricted samples with different restriction bandwidth h, e.g. for each sample in Figure 8. Estimates should be robust to this.

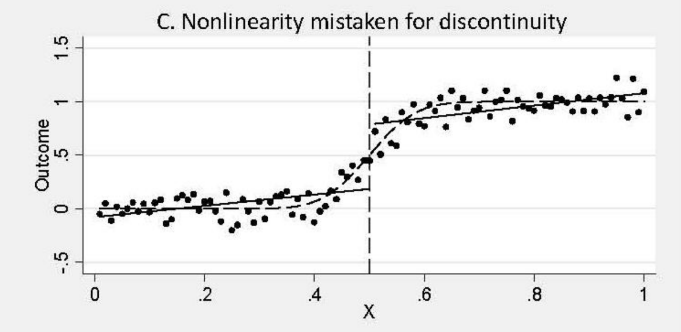

Figure 9: Misspecified Discontinuity

- Placebo Test: Effect should only be at cutoff not anywhere else.
  - Test: Regress (6) for some placebo threshold $c^* \neq c$. Here the estimator β_1 should be insignificant (zero).
- Ignorability: No other covariates than the outcome should 'move' at cutoff.
  - Test: Check for correlations between correlations between various controls Z and forcing variable X as in Figure 10. (The effect studied is certainty of election of Democrat on liberal policy-voting.)
- Sorting/Bunching: If subjects can manipulate their forcing variable $X \rightarrow$ Violation of 'No Compliance Issue'-assumption.
  - Solution: Plot densities of the forcing variable X. Should be smooth around cutoff.
  - Example: Poverty index used to determine eligibility for national aid. Cutoff publicly available in late 1997. Figure 11: After 1998, Colombia finds bunching to the low side of poverty index, i.e. more national aid. Figure 12 in general.

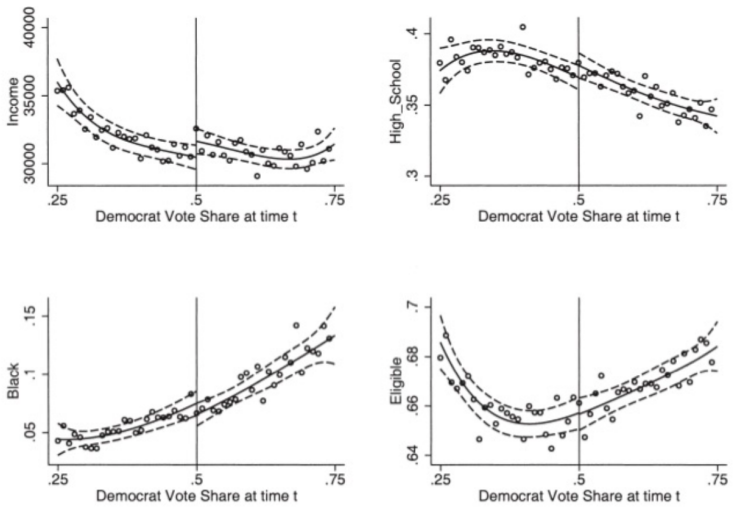

Figure 10: Ignorability Falsification Check

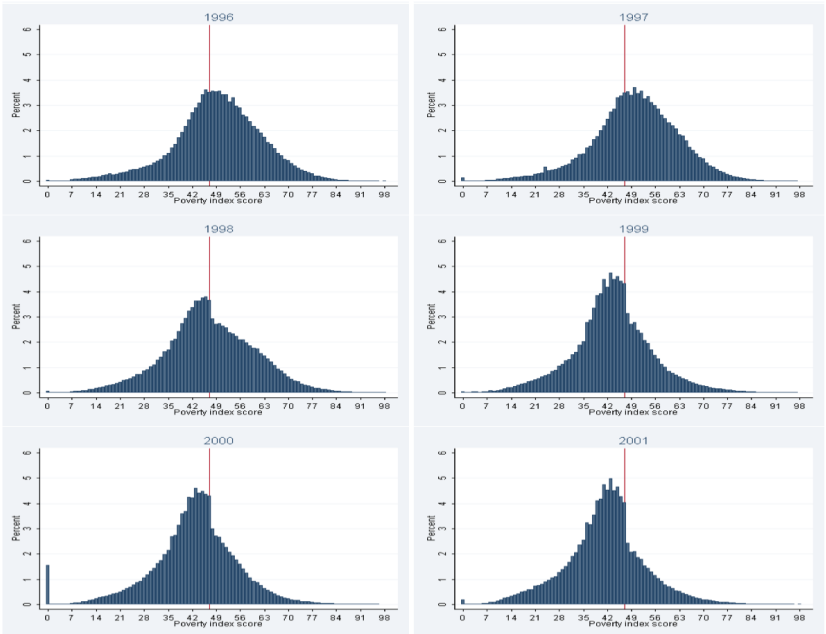

Figure 11: Sorting/Bunching Falsification Check for Colombia

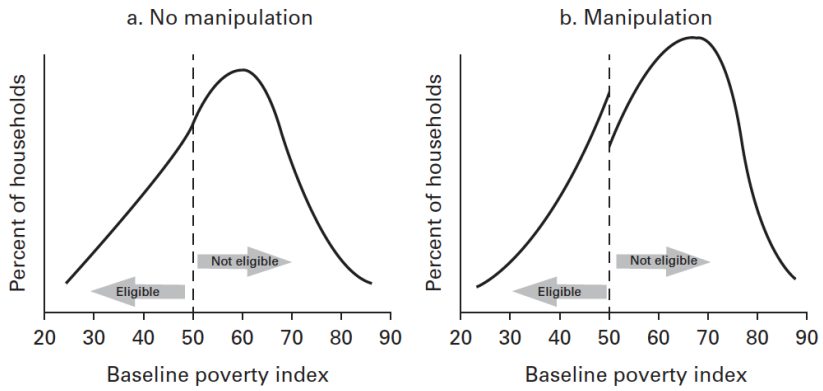

Figure 12: Sorting/Bunching Falsification Check in General

12.1.3 Limitations of Sharp RD

No ignorability of treatment if

- People sort manipulate their eligibility X_i and hence self-select into treatment. Violation of 'No Compliance Issue'-assumption.
- If other things change that cause the discontinuity at the same cutoff as treatment.

12.2 Fuzzy RD

Property Now, probability of treatment is not 0 or 1 anymore, but there is still a discontinuity in probability for treatment as in Figure 13 (the upper part). Hence we need it to hold

$$\lim_{k \underset{\text{from left}}{\to} c} Pr(D_i = 1|X = k) \neq \lim_{k \underset{\text{from right}}{\to} c} Pr(D_i = 1|X = k)$$

Verbally, probability of participation in program, $D_i = 1$, can depend on different X-values but for some value c the probability structurally increases.

Application Typically useful, when some program encourages program participation after some threshold c but doesn't enforce it.

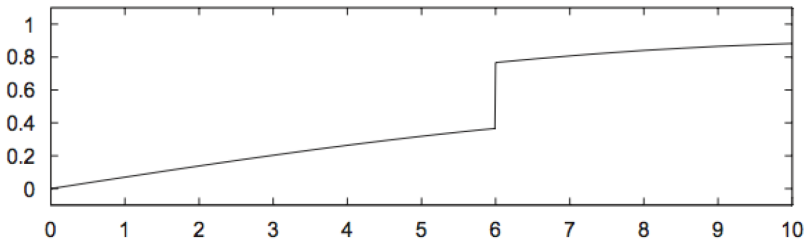

Fig. 3. Assignment probabilities (FRD).

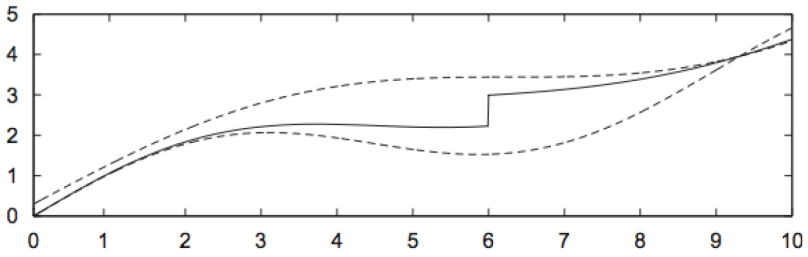

Fig. 4. Potential and observed outcome regression (FRD).

Figure 13: Fuzzy RD

12.2.1 Fuzzy RD as Instrument

It is now the case that the forcing variable X acts as an instrument.

Example [16] Research question: Effect of free tutoring program D_i on exit exam scores Y_i.

- If entry exam score $X \leq 70$ students can enroll in free tutoring program

Students are not obligated to enroll.

From Figure 14, more likely to tutor program if $X \leq 70$ but also some with $X > 70$ get tutor. Hence, fuzzy RD.

Conditions for Fuzzy RD (IV):

1. Relevance: $Cov(X_i, D_i) \neq 0$
2. Exclusion: $X_i \nrightarrow Y_i$ and $X_i \rightarrow D_i \rightarrow Y_i$, where $\rightarrow$ denotes causal link.
3. Exogeneity: No shared confounders of Y_i, D_i and X_i
4. Monotonicity

[16] From: https://evalf20.classes.andrewheiss.com/example/rdd-fuzzy/

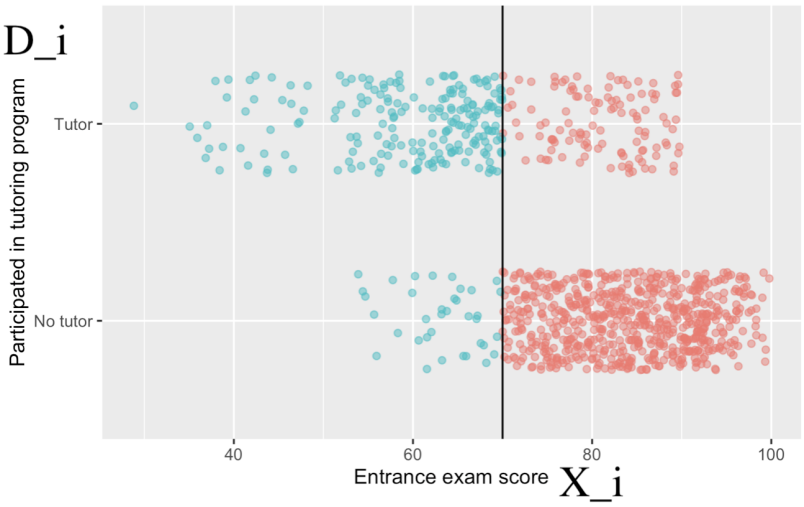

Figure 14: Fuzzy RD example

Condition 1 is verified by Figure 14. Condition 2 is satisfied because entrance exam scores X_i do not causally determine exit exam scores Y_i (only through the tutoring program D_i). Condition 3 is debatable (I think).

Estimator

$$\alpha_{\text{FuzzyRD}} \equiv \frac{\lim_{k \underset{\text{from left}}{\to} c} \mathbb{E}(Y|X=k) - \lim_{k \underset{\text{from right}}{\to} c} \mathbb{E}(Y|X=k)}{\lim_{k \underset{\text{from left}}{\to} c} \mathbb{E}(D_i|X=k) - \lim_{k \underset{\text{from right}}{\to} c} \mathbb{E}(D_i|X=k)}$$

Interpretation of Estimator Because α_{FuzzyRD} is a LATE then it only is for the compliers. So it is only for the individuals that took the tutor when their entry score was ≤ 70 but would not have if it was > 70. Not those that took the tutor and either case and not those that never would.

Internal Validity The effect measured is now only around the cutoff *and* only among the compliers.

Regression Two-stage least squares

$$D_i = \gamma_0 + \gamma_1 \mathbb{1}(\text{below cutoff}) + \gamma_2 X_i + \gamma_1 (D_i \times X_i) + \mu \qquad \text{(First Stage)}$$
$$Y_i = \beta_0 + \beta_1 \hat{D}_i + \beta_2 X_i + \beta_1 (D_i \times X_i) + \varepsilon \qquad \text{(Second Stage)}$$

where $\hat{D}_i$ are then the fitted values from the First Stage.

Robustness Checks All the checks as described for Sharp RD in Section 12.1.2 should/could also be carried out for Fuzzy RD. Also addition of polynomials and controls in (First Stage) and (Second Stage).

13 Homeworks

13.1 Assignment 7

- Outcome Y_i: Nominal earnings
- Running variable X_i: Age left education
- Treatment (cutoff) D_i: Dropout age 15 when aged 14 (year 1947)

Significance of Jumps Check significance of γ_1, δ_1 in regressions:
Regress (first stage)

$$X = \gamma_0 + \gamma_1 D_i + controls + \varepsilon \qquad \text{(First Stage)}$$

Regress (reduced form)

$$Y_i = \delta_0 + \delta_1 D_i + controls + \phi \qquad \text{(Second Stage)}$$

Fuzzy RD, OLS .
Regress OLS (cols 1-3, table 2)

$$Y_i = \alpha_0 + \alpha_1 X_i + controls + \mu$$

Fuzzy RD (cols 4-6, table 2)

$$Y_i = \beta_0 + \beta_1 \hat{X}_i + controls + \omega_1 \qquad \text{(Second Stage)}$$
$$X_i = \psi_0 + \psi_1 D_i + controls + \omega_2 \qquad \text{(First Stage)}$$

14 Student Presentations

14.1 My Presentation

Research Question "Does Danish proficiency for immigrant children (aged 0-13 at arrival) in Denmark causally impact probability of obtaining university degree?"

Data Danish registry data: Population registry, education registry, parents' residence permits, and many more.

Method Instrumented difference in difference (DiD-IV) design:

- OLS: Danish proficiency on university completion would be endogenuous
  - E.g. innate ability
- IV: $age{-}at{-}arrival$ (instrument) would not survive exogeneity
  - E.g. knowledge of culture, educational grading system, etc.
- IV-DiD: $\{(age{-}at{-}arrival) \times (non{-}Germanic\ speaking\ country)\}$ could solve this
  - Both encounter all the same things except different costs of learning Danish

Findings Statistically significant premiums for Danish-speakers of around 40 pct. points. Robust to various specifications.

Limitations .
· Changing compositions within groups over time?
· Proxy for Danish proficiency (GPA in Danish-subject in primary school) accurate?