Examining the effectiveness of a bridging programme using a regression discontinuity design

Michael Murray

(University of KwaZulu-Natal)

Outline

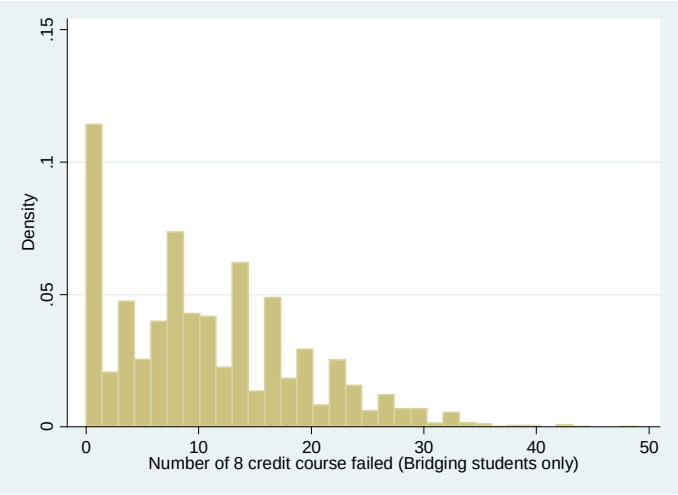
- Problem formulation
- A naïve difference in means estimator
- Selection on observable covariates
- Regression-based adjustment procedure
- Matching on a Propensity score
- Selection on unobservable covariates
- Instrumental variables
- Model the endogenous treatment effect (Heckit)
- Regression Discontinuity design-work in progress

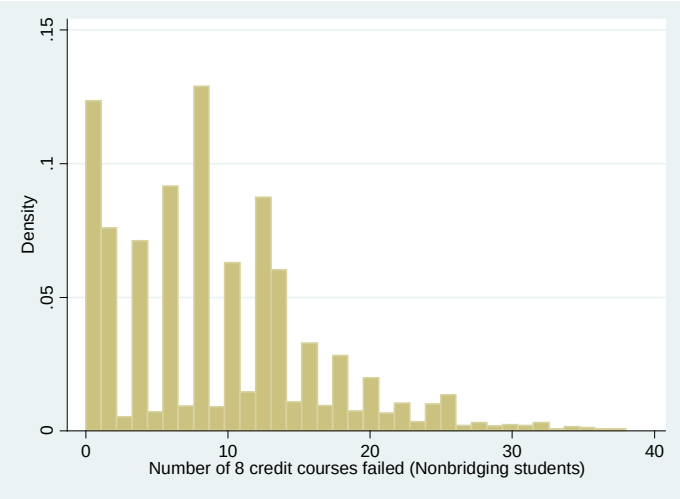
Faculty of Science at UKZN

- Student must come from a disadvantaged school background
- BSc Augmented 4-year program- Regular lectures augmented by additional lectures.
- Four modules instead of eight taken in first year plus two Scientific Communication modules.
- Entry requirement:- At least 28 Matric points (2007-2009) At least 22 points (excluding Life orientation) 2010+
- BSc Foundation program- Earn 32 credit points that may be carried forward to their mainstream qualification.
- Entry requirement:- At least 20 Matric points (2007-2009) At least 16 points (excluding Life orientation) 2010+
- Main stream programs
- M-stream - At least 34 Matric points2007-2009 At least 30 points (excluding Life orientation) 2010+
- LES-stream - At least 34 Matric points2007-2009 At least 28 points (excluding Life orientation) 2010+
- 2008 Higher/Standard/ lower grade replaced with a single level only

Response vs treatment variable

Y=number of eight credit courses failed for the <u>first time</u>





Treatment	E(Y)	sd(Y)	Number of students
Bridged (T=1)	10.77325	8.315843	2527
Not bridged (T=0)	9.212121	7.119832	3465

Bridged students fail more courses=> bridging not successful !!!

Heart transplant example

- Surgeon A: average survival time is 2.5years
- Surgeon B: average survival time is 4.5years
- => Surgeon B `better'
- However, surgeon B only operates on patients below the age of 35
- Need to adjust survival times for the age of the patients
- What about other confounding covariates?

What we would like to observe

- Potential outcomes
- $Y_i(0)$ =response variable if student not bridged
- $Y_i(1)$ =response variable if same student bridged
- We could then calculate for each student the following treatment effect due to bridging $\boldsymbol{\Delta}_i$=Y_i (1)- $Y_i(0)$

- What we actually observe:-
- T_i =1 if bridged, T_i =0 if not bridged
- Y_i=(1-T_i) $Y_i(0)$+T_i $Y_i(1)$ = Total number of 8 credit courses failed for the first time
- Some additional covariates X_i
- Because we observe a random sample $\{Y_i, T_i, X_i$; i=1…N$\}$ we can identify E($Y_i|\ T_i = 1, X_i$) and E($Y_i|\ T_i = 0, X_i$) either parametrically (through a regression function) or non-parametrically using it's sample analogue

What we would like to estimate

- Average treatment effects that may be of interest?

$$\text{ATE} \equiv E\{Y_i(1) - Y_i(0)\}$$

- Measures the average effect of treatment on an individual that is being drawn randomly from the overall population– it may include in it's value individuals who may never be eligible for treatment. For our dataset ATE>0 not a problem! Because bridged students have a lower Matric Point score one would expect bridged students to fail more courses on average than their non-bridged counterparts.

$$\text{ATT} \equiv E\{Y_i(1) - Y_i(0) \mid T_i = 1\}$$

- Measures the average effect of treatment on an individual that is being drawn randomly from the treated subpopulation. ATT<0 implies that amongst the subpopulation of treated (bridging) students, they would fail less courses than they would if they had not been treated (bridged) => bridging program is a success

- When does Naïve difference in means estimator= ATE=ATT?

Correcting for sample selection bias

- Random experiment:- Being able to randomly assign treatment (T) to individuals will ensure that we have

  Naïve estimator= ATE=ATT

- In our context however we have an element of self selection into a particular treatment taking place – called an observational study

Selection based on observables

- Can we find enough observable covariates X so that the following condition becomes valid for our dataset when we condition (control) for X?

$$\Pr[T_i \mid Y_i(0), Y_i(1), X_i] = \Pr(T_i \mid X_i) \text{ CIA}$$

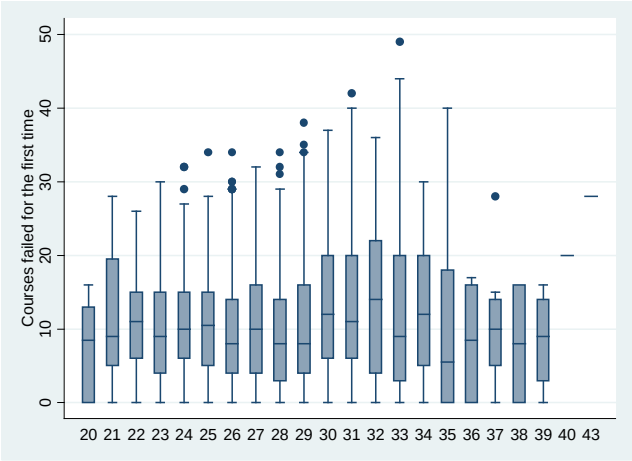
- If Yes-> use a regression adjustment method or a propensity score based matching method
- If No-> use instrumental variables. These are variables, observable to the researcher which help to determine the treatment status (T) but not the outcome variable (Y) . Difficult to find!!!
- Or else specify functional forms for the treatment selection and response variable and attempt to model any correlation that may occur in the distribution of the disturbance terms associated with these two processes . Heckman's procedure-- functional form assumptions may be hard to justify.

Observable covariates (X)

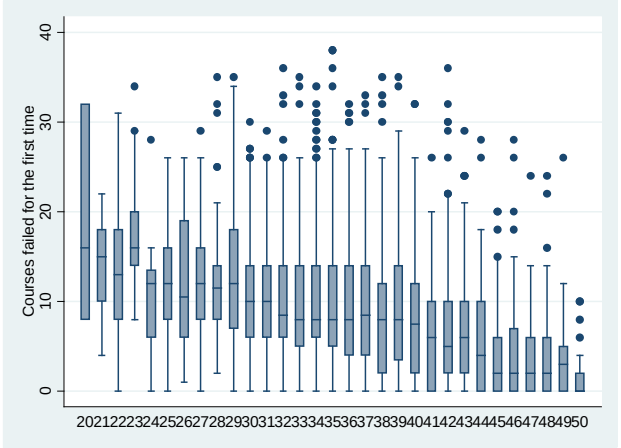
- Male (0/1), African(0/1), Residence(0/1),
- OBE indicator Senior Certificate (Higher/Standard/Lower grade) replaced in 2008 with National Senior Certificate(Single grade)(0=pre 2008,1=Single grade matric)
- Matric points:- Total matric point count
- Financial aid:- Amount of funding received

Fit a Poisson regression model

- Assume $Y_i \sim \text{Poisson}(\lambda_i)$ with
- $\lambda_i \equiv \text{E}(Y_i|X_i, T_i, e_i) = \exp(X_i\beta + \delta T_i + e_i)$
- Which covariates do we include?
- Analysis only appropriate if we have included enough observable covariates X so that the CIA condition becomes valid for our dataset when we condition (control) for X?



MatPts for bridged students



MatPts for non-bridged students

Downward sloping for non-bridged students only

Have we included enough covariates?

Parameter Estimates	Model 1 * denotes significant at 5% level	Model 2
bridged	0.0499*	-0.1297*
obe	-0.1570*	-0.2417*
male	0.0391*	0.0258
african	0.2323*	0.1459*
residence	0.1844*	0.1885*
finaid	-0.00004*	-0.00004*
intercept	2.1779*	3.2711*
Matpts		-0.0280*

	Coefficient estimate	Robust std error	95% Confidence Interval
ATE	0.4647	0.5332	[-0.580, 1.509]
ATT	-2.666	0.3365	[-3.326,-2.006]

Have we adjusted for enough covariates X so that CIA becomes valid

Why is the OBE sign negative? – mark inflation

Matching based on a propensity score

- The assignment to treatment probabilities p_i= Pr(T_i | X_i) are called propensity scores
- These methods attempt to compute an average treatment effect by matching treated observations with nontreated (controls) observations whose covariate profiles (propensity scores) are very similar.
- In order to be justified, enough observables X must have been included in the propensity score calculation so that the CIA condition becomes valid

$Y_i(0), Y_i(1) \perp T_i \mid X_i \Leftrightarrow$ Pr[$T_i \mid Y_i(0), Y_i(1), X_i$]= Pr(T_i | X_i)

CIA

```
. teffects ipwra  (failed8 obe MatPts matsq male african Resin1stYoS FinAidin1stYoS, poisson)(brid
> ged MatPts, probit), pstolerance(1e-10)
```

IPRWA method	Coefficient estimate	Robust std error	95% Confidence Interval
ATE	0.7524	0.2266	[0.3083,1.1966]
ATT	-0.3107	0.1207	[-.5474,-0.0741]

Doubly robust (IPRA) estimators combine the outcome based modelling strategy of RA with the treatment based modelling strategy of IPW. These estimators have a remarkable property: although they require us to build two models, we only need to specify one of the two models correctly in order to obtain correct estimates of the treatment effect.

Regression versus matching

- Regression adjustment: Assumes
- E($Y_i(0)|\ X_i,\ T_i) = \beta_0 + \ X_i\beta_1$
- E($Y_i(1)|\ X_i,\ T_i) = \beta_0 + \delta + \ X_i\beta_1$
- CIA and linearity assumed
- Matching:- Assumes
- E($Y_i(0)|\ X_i,\ T_i$) =E($Y_i(0)|\ X_i$)
- E($Y_i(1)|\ X_i,\ T_i$) =E($Y_i(1)|\ X_i$)
- CIA assumed but no parametric form for conditional mean given

- Matching may not be possible in all cells- because non-matched observations will have to be discarded the reduced sample size may affect the standard error associated with one's treatment effect estimator.
- Matching being done ***without*** any direct reference being made to an outcome variable Y_i
- At the cost of making a parametric assumption about the form of E($Y_i|\ X_i, T_i$) all data points can be used in OLS regression which may lead to a smaller standard error being associated with one's OLS based treatment effect estimator

Selection based on unobservables

- CIA condition no longer holds true
- Probit-2SLS method- Find a suitable instrumental variable (Z)
- Z_i must be correlated with T_i
- Z_i must not be correlated with the error term e_i helping to generate Y_i thus Z_i affects Y_i only through T_i

$$Z=1_{\{MatPts>34\}}$$

- Heckman switching regression – Model the endogenous treatment selection process explicitly- to be used if we cannot find a suitable instrumental variable Z

Homogeneous treatment effect

$Y_i = \alpha_0 + \tau T_i + X_i\beta_0 + e_{i0}$ with $E(e_{i0} \mid T_i, X_i, Z_i) \neq 0$

Instrumental variable:- $Z = 1_{\{MatPts > 34\}}$ =rvar

bridged	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
rvar	-.5184594	.0875356	-5.92	0.000	-.690026	-.3468929
MatPts	-.2003465	.0084727	-23.65	0.000	-.2169526	-.1837403
obe	-.1726783	.059531	-2.90	0.004	-.2893569	-.0559998
male	.000683	.0475314	0.01	0.989	-.0924768	.0938428
african	1.588862	.0960582	16.54	0.000	1.400592	1.777133
FinAidin1st~S	4.40e-06	1.55e-06	2.84	0.004	1.37e-06	7.43e-06
Resin1stYoS	.3354089	.0560773	5.98	0.000	.2254994	.4453184
_cons	4.693624	.2730262	17.19	0.000	4.158502	5.228746

failed8	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
bridged	-4.426283	.6011115	-7.36	0.000	-5.604682	-3.247885
MatPts	-.4036681	.0334585	-12.06	0.000	-.469259	-.3380772
obe	-2.599681	.2430414	-10.70	0.000	-3.076131	-2.123231
male	.1959381	.1975519	0.99	0.321	-.1913357	.583212
african	1.976758	.3199458	6.18	0.000	1.349547	2.603969
FinAidin1st~S	-.0000326	7.12e-06	-4.57	0.000	-.0000465	-.0000186
Resin1stYoS	2.090235	.2685543	7.78	0.000	1.563771	2.6167
_cons	24.81022	1.329526	18.66	0.000	22.20386	27.41658

	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf.	Interval]
atet	-4.426283	.5768365	-7.67	0.000	-5.556862	-3.295705

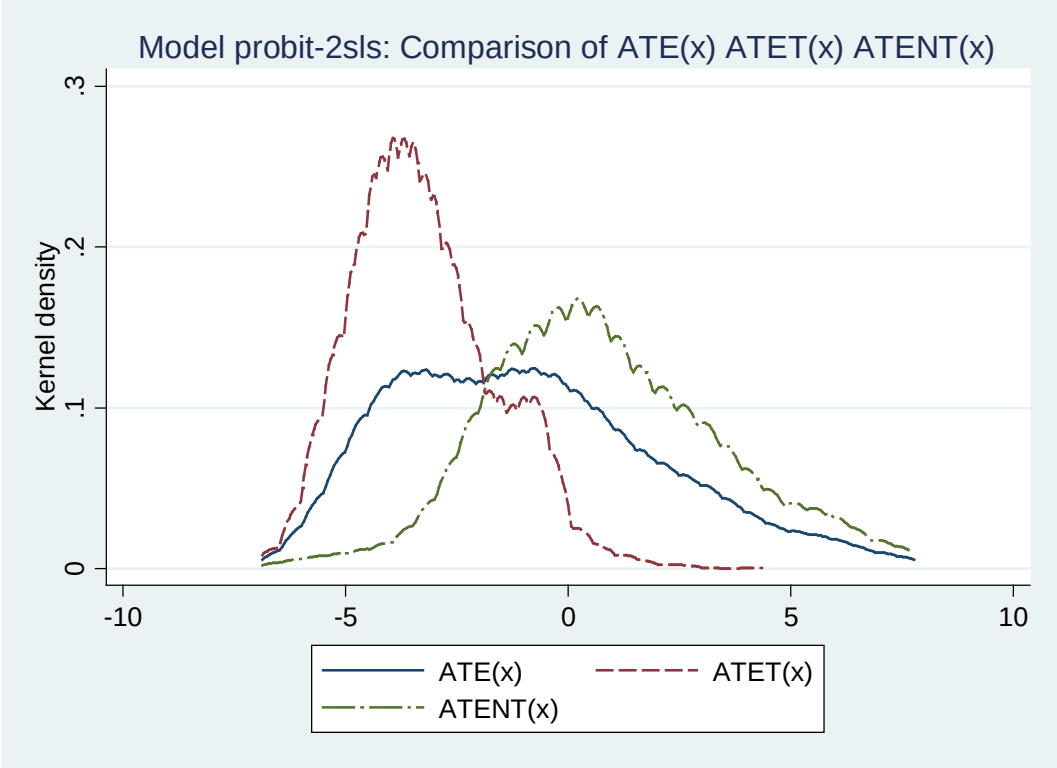
Heterogenous treatment effect

$$Y_i = \alpha_0 + \tau\ T_i + X_i\beta_0 + T_i\,(X_i - \mu_x)\,\delta + e_{i0} + T_i\,(e_{i1} - e_{i0})$$

bridged	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
rvar	-.5184594	.0875356	-5.92	0.000	-.690026	-.3468929
MatPts	-.2003465	.0084727	-23.65	0.000	-.2169526	-.1837403
obe	-.1726783	.059531	-2.90	0.004	-.2893569	-.0559998
male	.000683	.0475314	0.01	0.989	-.0924768	.0938428
african	1.588862	.0960582	16.54	0.000	1.400592	1.777133
FinAidin1st~S	4.40e-06	1.55e-06	2.84	0.004	1.37e-06	7.43e-06
Resin1stYoS	.3354089	.0560773	5.98	0.000	.2254994	.4453184
_cons	4.693624	.2730262	17.19	0.000	4.158502	5.228746

failed8	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
bridged	-.739272	.7791016	-0.95	0.343	-2.266596	.788052
_ws_MatPts	.4984571	.0727609	6.85	0.000	.355819	.6410951
MatPts	-.4314577	.0332925	-12.96	0.000	-.4967232	-.3661923
obe	-2.434555	.2399184	-10.15	0.000	-2.904882	-1.964227
male	.1966549	.194169	1.01	0.311	-.1839873	.5772971
african	1.170286	.333561	3.51	0.000	.5163847	1.824188
FinAidin1st~S	-.0000354	7.01e-06	-5.06	0.000	-.0000492	-.0000217
Resin1stYoS	1.788544	.2673283	6.69	0.000	1.264483	2.312606
_cons	25.7378	1.318871	19.52	0.000	23.15233	28.32327

	Observed Coef.	Bootstrap Std. Err.	z	P>|z|	Normal-based [95% Conf.	Interval]
atet	-3.116156	.5888043	-5.29	0.000	-4.270192	-1.962121



Heckit treatment effects model-

- Response model : Y_i~Poisson(λ_i)
- with
  $\lambda_i \equiv \mathrm{E}(Y_i \mid X_i, T_i, e_i) = \exp(X_i\beta + \delta T_i + e_i)$
- Treatment model :
- T_i=1 if $X_i\gamma + u_i > 0$
- T_i=0 elsewhere
- Correlation structure
- $\begin{pmatrix} e_i \\ u_i \end{pmatrix} \sim MVN \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \rho\sigma \\ \rho\sigma & 1 \end{pmatrix}$
- CIA (i.e. no selection on unobservables) => ρ=0 in the above model formulation

Poisson regression with endogenous treatment — Number of obs = 4923
(24 quadrature points) — Wald chi2(7) = 336.01
Log pseudolikelihood = -18256.469 — Prob > chi2 = 0.0000

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
failed8						
MatPts	.185584	.0293576	6.32	0.000	.1280441	.2431239
matsq	-.0033836	.0004168	-8.12	0.000	-.0042006	-.0025665
male	.0379716	.0393998	0.96	0.335	-.0392505	.1151937
african	.1122926	.0468642	2.40	0.017	.0204406	.2041447
Resin1stYoS	.2246871	.0392917	5.72	0.000	.1476768	.3016974
FinAidin1stYoS	-8.07e-06	1.53e-06	-5.29	0.000	-.0000111	-5.08e-06
1.bridged	-.2653756	.1097736	-2.42	0.016	-.4805278	-.0502233
_cons	-.3478879	.5477618	-0.64	0.525	-1.421481	.7257055
bridged						
Quintile						
2	-.0958927	.1061425	-0.90	0.366	-.3039281	.1121427
3	.042367	.0898388	0.47	0.637	-.1337139	.2184479
4	-.2385257	.0907536	-2.63	0.009	-.4163994	-.060652
5	-1.313806	.0934206	-14.06	0.000	-1.496907	-1.130705
MatPts	-.2556145	.0087253	-29.30	0.000	-.2727157	-.2385133
_cons	8.153117	.2797331	29.15	0.000	7.60485	8.701384
/athrho	.0482742	.0754449	0.64	0.522	-.0995952	.1961436
/lnsigma	-.1632798	.0175095	-9.33	0.000	-.1975978	-.1289619
rho	.0482367	.0752694			-.0992672	.1936663
sigma	.8493535	.0148717			.8206999	.8790075

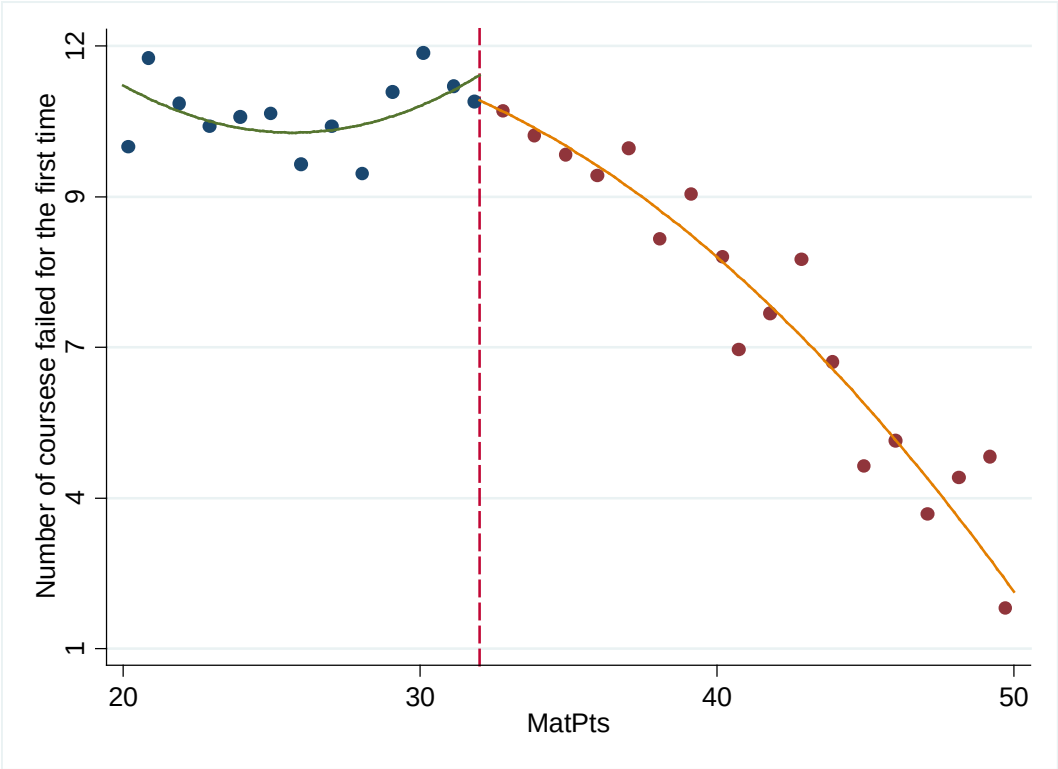
Wald test of indep. eqns. (rho = 0): chi2(1) = 0.41 Prob > chi2 = 0.5223

Wald test indicates that we can fail to reject the null hypothesis of no correlation between the treatment errors and the outcome errors=> CIA valid

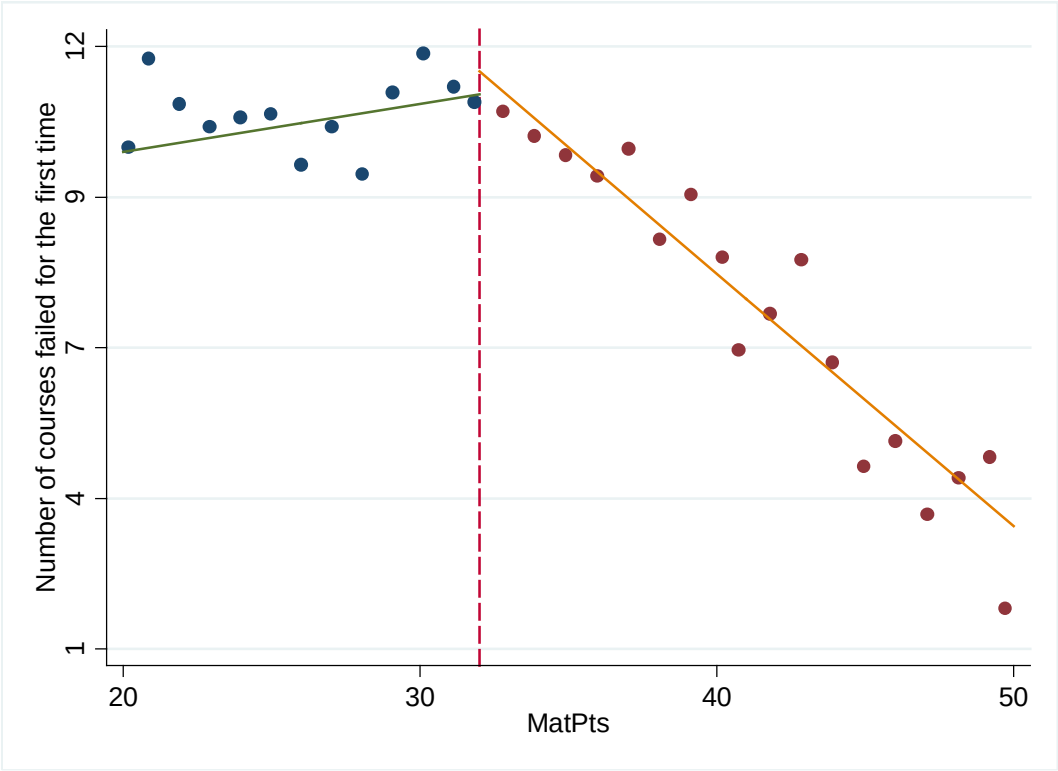
ATT

	Margin	Unconditional Std. Err.	z	P>|z|	[95% Conf.	Interval]
_cons	-3.356659	1.525163	-2.20	0.028	-6.345923	-.3673943

The Regression Discontinuity Design



Regression Discontinuity Design



- We observe treatment status T_i of unit *i*
- We observe covariates (X_i , Z_i) known not to have been affected by treatment.
- KEY idea: assignment to treatment is determined by the value of X_i being smaller or larger than a given threshold. $T_i = 1_{\{X_i \geq c\}}$
- - deterministic treatment assignment ! SHARP regression discontinuity (SRD)
- - probabilistic treatment asignment ! FUZZY regression discontinuity (FRD)
- X_i is known as the forcing variable (or running variable)

Fuzzy RD

```
. rdrobust failed8 rvar, c(0) fuzzy(bridged) all
```

Sharp RD Estimates using Local Polynomial Regression.

Cutoff c = 0	Left of c	Right of c
Number of obs	646	670
Order Loc. Poly. (p)	1	1
Order Bias (q)	2	2
BW Loc. Poly. (h)	2.189	2.189
BW Bias (b)	2.928	2.928
rho (h/b)	0.747	0.747

Number of obs =	5014
NN Matches =	3
BW Type =	CCT
Kernel Type =	Triangular

Structural Estimates. Outcome: failed8. Running Variable: rvar. Instrument: bridged.

Method	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Conventional	.99175	3.8508	0.2575	0.797	-6.55562	8.53911
Robust	-	-	0.6356	0.525	-3.52169	6.90198

First-Stage Estimates. Outcome: bridged. Running Variable: rvar.

Method	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Conventional	-.29808	.06404	-4.6542	0.000	-.423602	-.172552
Robust	-	-	-6.4616	0.000	-.360838	-.192881

All Structural Estimates.

Method	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Conventional	.99175	3.8508	0.2575	0.797	-6.55562	8.53911
Bias-Corrected	1.6901	3.8508	0.4389	0.661	-5.85722	9.23751
Robust	1.6901	2.6591	0.6356	0.525	-3.52169	6.90198