Short-term effects of an intervention to increase upper secondary completion

Ingrid Huitfeldt, Lars Kirkebøen and Marte Rønning

(Research Department, Statistics Norway)

25 May 2016

Introduction

- Approximately 30 percent of students who start secondary education does not complete within 5 years (Statistics Norway, 2015).
  - This is high compared to other OECD countries: Norway has a completion rate below the OECD-average among under 25 years olds, and is 3rd last on on-time completion with 57 percent (OECD, 2014).
- The individual and social costs of dropping out are substantial (Lillejord et al, 2015; Falch et al, 2009).
  - Falch et al: PV cost of 900k NOK/drop-out (about 100k EUR)
- Higher completion rates of secondary education is a high priority for Norwegian policy makers
  - "Ny GIV" introduced in 2010. Main objective: Increase completion rates by 6 percentage points, from 69 percent for the 2004 cohort to 75 percent for the 2010 cohort.

What is Ny GIV?

- Completion of upper secondary is significantly lower among students with weak academic performance of elementary school (Falch et. al., 2014; Statistics Norway, 2015).
- Ny GIV: Intensive training to the lowest performing students:
  - The intensive training should start early in the second semester of 10th grade (the final year of lower secondary)
  - Intensive training could replace up to 7.5 hours per week of regular classes
  - Counties received funds for project administration, teachers received some training - otherwise no extra resources
- Students eligible for Ny GIV: 10 percent with lowest midterm grades in each municipality
  - Students already recieving special education considered individually

Implementation of Ny GIV

- Gradually implemented in three waves: 2010/2011, 2011/2012 and 2012/2013.
- Responsibility for implementation was given to the counties (19 in total).
- An invitation letter from the Norwegian Ministry of Education described how the implementation should take place
  - In particular, eligible students
- However, these instructions were interpreted differently by the various counties, but also municipalities, school leaders and even teachers
  - This led to considerable variation in terms of selection of students, course content and implementation of intensive training
  - Described in more detail in comissioned reports, based on register data, surveys, interviews and observation of teaching: Sletten et al (2011), Holen and Lødding (2012, 2013), Helgøy and Homme (2013) and Rønning et al (2013)

This policy evaluation

- Aim: Determine whether training for the weakest performing students have a causal effect on high school completion and other learning outcomes, such as grades.
- This current analysis is the second evaluation of Ny GIV.
  - The first evaulation of Ny GIV (Eielsen, Kirkebøen, Leuven, Raaum and Rønning, 2013) was not able to detect any effects.
- Twofold:
  1. Using the roll-out of the program in a difference-in-differences setup.
     - Whole country included
  2. Using the the "assignment rule" in a RD framework
     - Only Stavanger is included: Stavanger is the only municipality that applied a strict rule when assigning students

Existing literature

- There is some research that studies the effects of intensive training on student achievement, both at upper secondary and high school level
  - Lavy and Schlosser (2005): Extra training has positve effects on poor-performing Israeli high school students
    - The school's average graduation rate increases by aboute 3.3 percentage points.
  - Cortes et al (2014): Doubling the number of mathematics lessons it the first year of high school increases completion for Chicago students
    - Students receiving extra lessons in mathematics about 10 percentage points more likely to complete (completion increased from approximately 60 to 70 percent).
  - Cook et al (2015): Intensive mathematics training provided by tutor increase math results
    - Increase by 0.2-0.3 standard deviations, the proportion who failed in mathematics was halved
    - Evaluated using a randomized controlled experiment

Data

- The data used is primarily registry data at the individual level from Statistics Norway and the Norwegian Directorate of Education
  - Complete coverage (about 60,000 students/year, 2,000-6,000 treated)
  - Student characteristics, incl midterm and final grades, progression through upper secondary, family background
  - Student outcomes observed until 2013/2014, i.e, 1-3 years after treatment
    - Too early to observe actual completion: Grades and early measures of progression correlatate highly in cross section
- Also data on school characteristics and implementation
  - School level register data from the Schools' Information System, e.g. share special needs students
  - Survey data from descriptive evaluations

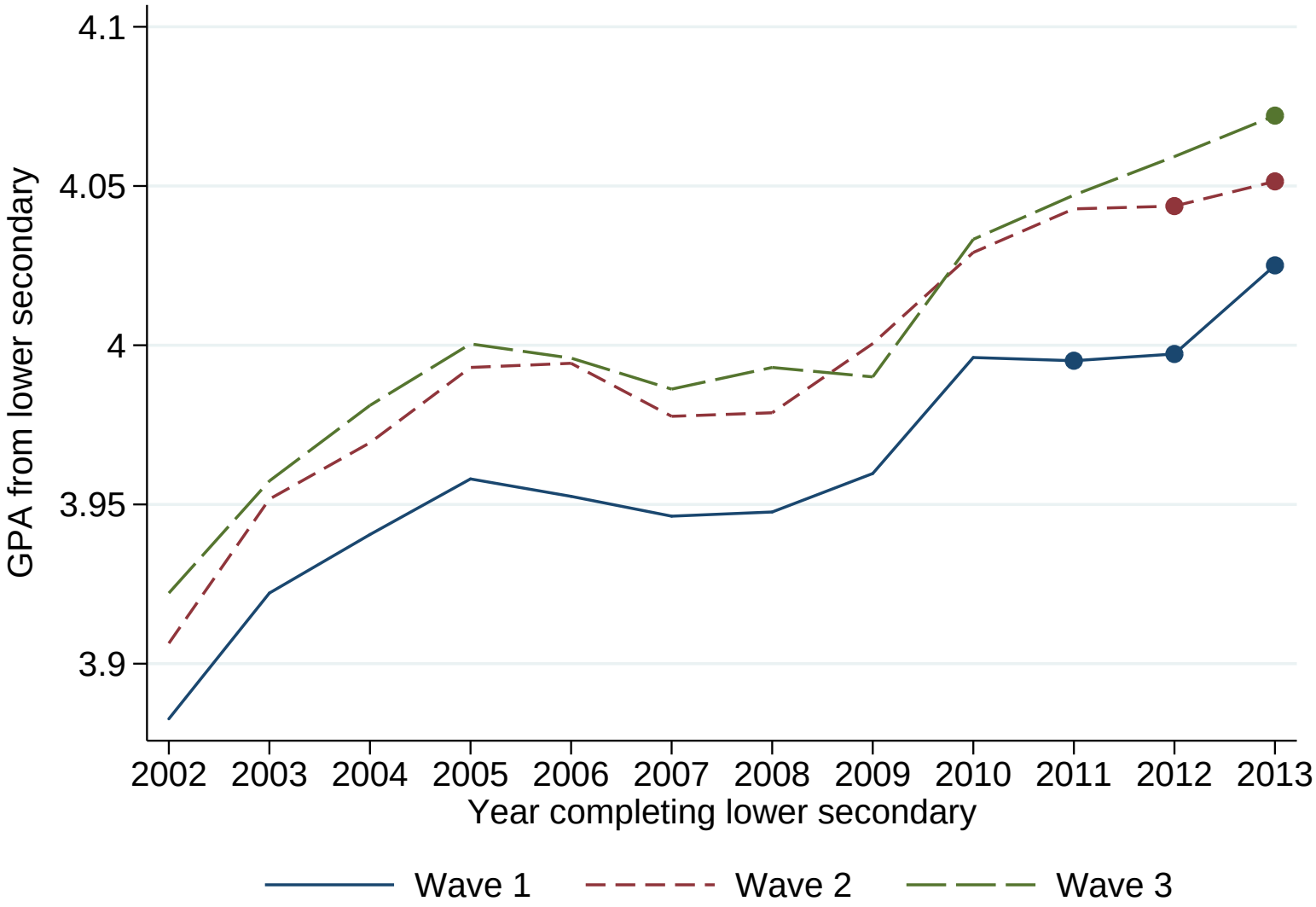
Data: Outcome variables

	2010/11	2011/12	2012/13
Intensive training	0,031	0,075	0,093
GPA 10th grade	4,029	4,032	4,044
Exam score 10th grade	3,456	3,406	3,393
Completion of high school			
1st year, on time	0,794	0,800	0,818
2nd year on time	0,714	0,728	.
3rd year, on time	0,483		
Enrolled in high school			
2nd year, on time	0,811	0,812	0,836
2nd year, one year delayed	0,868	0,869	.

Data: Control variables

	Wave 1	Wave 2	Wave 3
Individual level			
Score standardized test, 8th grade	-0,030	-0,003	-0,013
Father's education	4,211	4,180	4,130
Immigrant	0,062	0,048	0,041
Norwegian-born of immigrant parents	0,066	0,024	0,018
School level			
Share of pupils with special needs	0,106	0,113	0,136
Teacher hours/student hours	0,053	0,062	0,077
Share of teachers withouth a certificate	0,031	0,040	0,051
# municipalities	50	183	211
# schools	206	392	418
# students	68,625	95,862	68,408

Graphical illustration of DiD: Trends in GPA across waves



Research design - DiD

- Outcome of student i in school j and graduation cohort t:

$$y_{ijt} = \alpha_j + \delta_t + \gamma T_{jt} + x_{ijt}\beta + \epsilon_{ijt}$$

- Treatment (T) vary by wave*year
  - Control for wave and year fixed effects - assume common trends across waves in absence of treatment
    - Supported by joint tests in fully interacted model ($y_{ijt} = \alpha_{jt} + x_{ijt}\beta + \epsilon_{ijt}$) estimated on pre-reform years
  - $x\beta$ irrelevant for identification, can reduce residual variation and improve precision
  - Cluster standard errors at municipality level (municipality responsible for primary and lower sec schooling)

DiD: Baseline results

	Coeff	Se.
1st stage: Intensive training	0,104**	(0,0086)
GPA 10th grade	-0,013	(0,0080)
Exam score 10th grade	0,001	(0,0132)
Completion of high school		
1st year, on time	-0,002	(0,0035)
2nd year on time	0,004	(0,0059)
3rd year, on time	-0,005	(0,0071)
Enrolled in high school		
2nd year, on time	0,005	(0,0034)
2nd year, one year delayed	0,004	(0,0030)

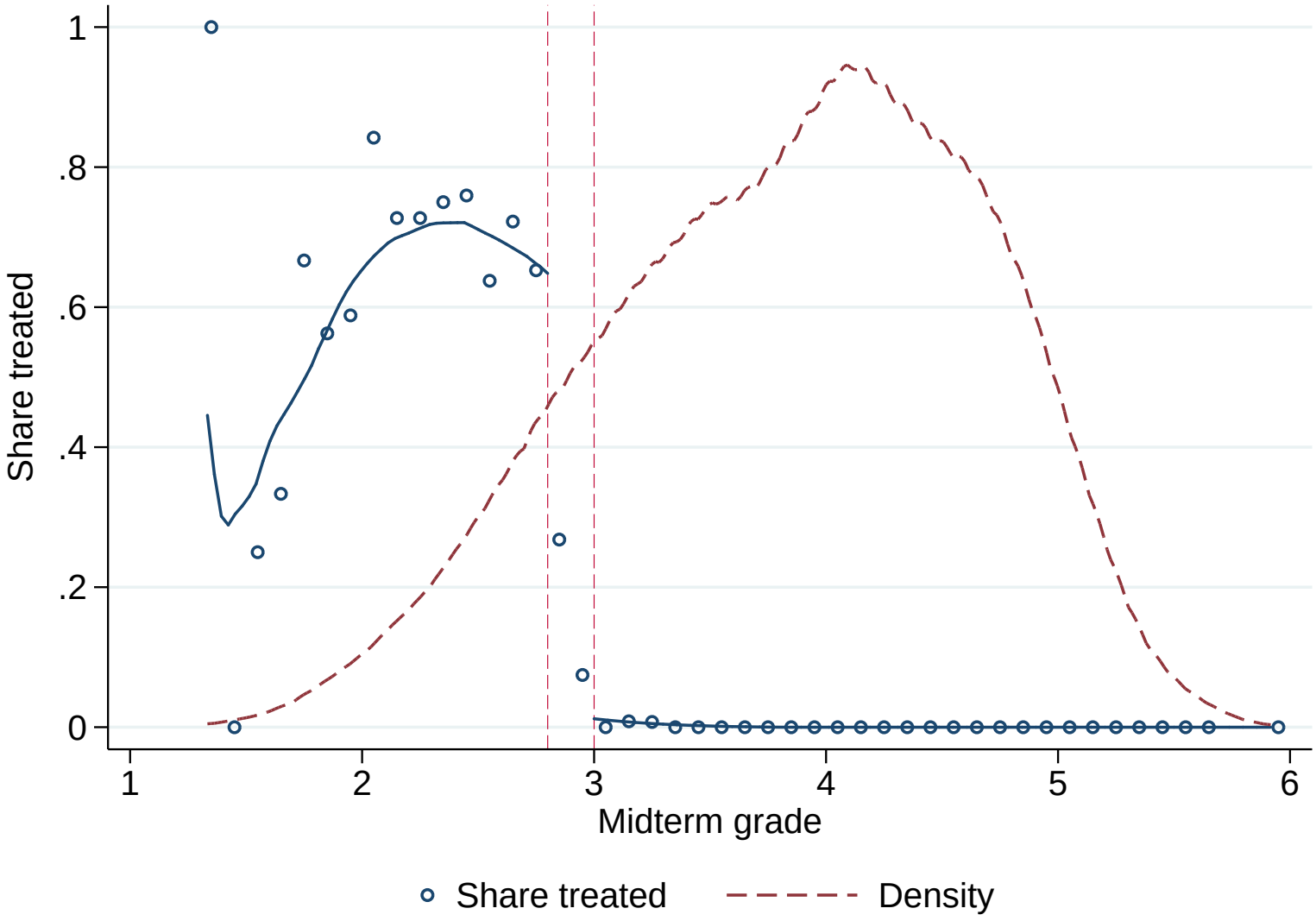
DiD results - cont'd

- Share of pupils treated increases to about 10 percent
- No other differences are statistically significant.
  - Large data set (12 000 treated students): We can exclude relatively small effects. E.g., it is unlikely that enrollment in 2nd year has increased by more than about 1 percentage point
  - This is a minor change compared to the objective of Ny GIV
- The estimates give us the average ITT of introducing Ny GIV
  - Only 10 percent of students participate: We cannot rule out a direct effect of up to about 10 percentage points on enrollment in 2nd year
  - We cannot distinguish between effects on students who participate (direct effect) and other students in treated schools (spill-overs).
    - Spill-overs may arise e.g. because of teacher training, changed group composition and reallocation of teachers
  - Heterogeneous implementation: May have succeeded in some schools?

Design of intervention - suitable for RD?

- Intensive training explicitly targeted towards the students in the lowest (municipality) decile in the distribution of midterm grades in tenth grade
  - In general, no clear discontinuity in participation rate (Eielsen et al, 2013)
  - However, one municipality (Stavanger) assigned students to intensive training according to the rule
- Thus, use Stavanger for RD analysis
  - Compare students just below the cutoff to students just above the cutoff: Apart from participation in extra training programs, students in a small interval on both sides of this cutoff are expected to be similar
  - Limited sample: 400 treated students
  - Different effect than DiD: Difference between effect and spill-overs on marginal (un)treated students
    - Relevant for policy: Effect of marginal expansion/contraction of program

Program participation in Stavanger



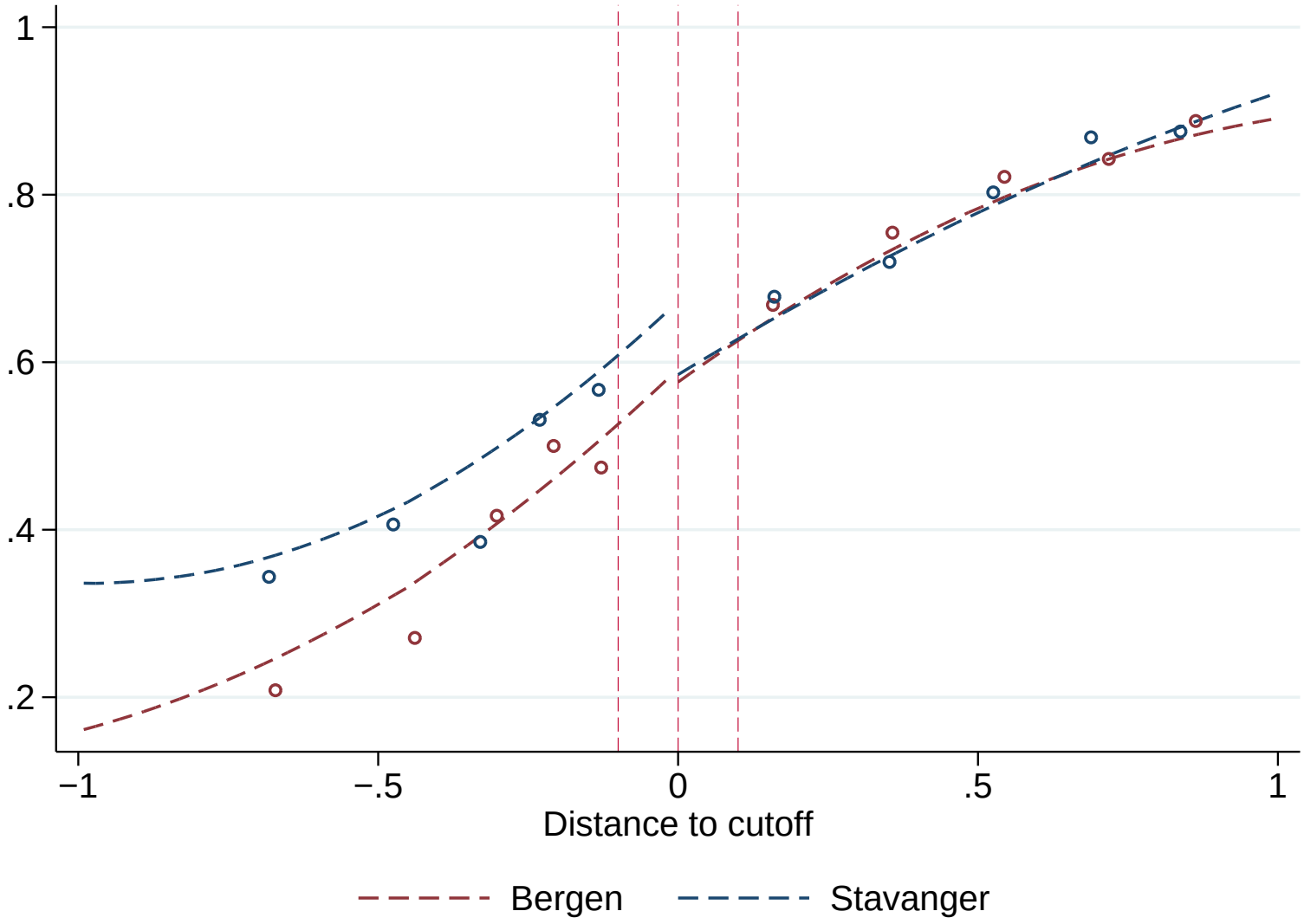
Research design - RD

- Estimates variations of $y = f(z) + \gamma D + x\beta + \epsilon$
  - $D = 1(z \leq 0)$, donut around cutoff, local quadratic regression and triangular weights in all specifications
- Basic RD design: Estimate discontinuity at cutoff:

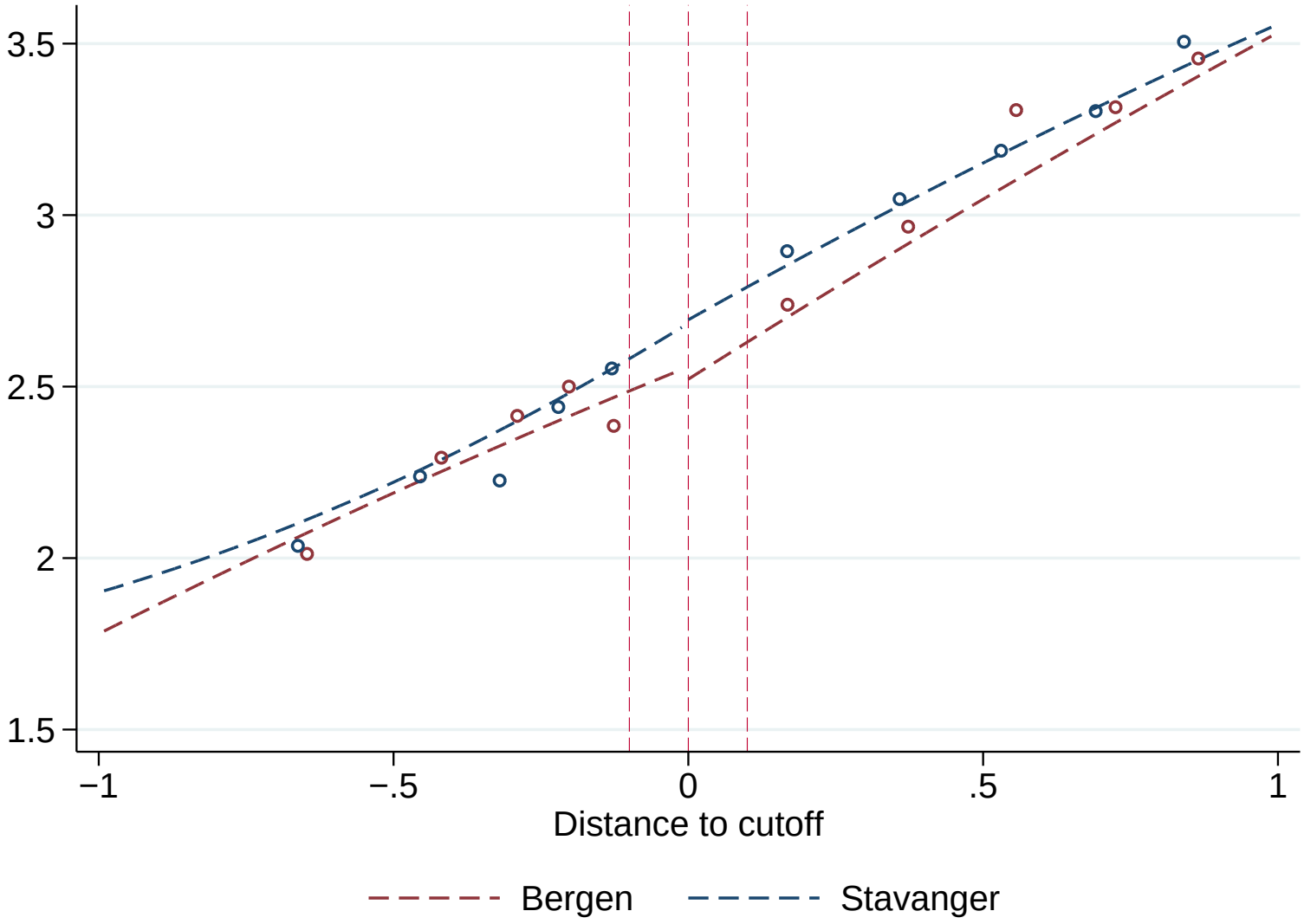
$$f(z) = Dz\alpha_0 + (1 - D)z\alpha_1, \qquad z\alpha_j = \alpha_j^1 z + \alpha_j^2 z^2$$

- Extensions with untreated control group (C), i.e. with no change in treatment status at $z = 0$
  - DiRD: Difference in discontinuity at cut-off: $f(z) = d_C(Dz\alpha_0^C + (1-D)z\alpha_1^C) + (1-d_C)(Dz\alpha_0 + (1-D)z\alpha_1) + \eta_1 d_C + \eta_2 d_C D$
  - DiDS: Assumes equal relationship $f(z)$ for treatment and control group: $f(z) = (d_0 z\alpha_0 + d_1 z\alpha_1) + \eta_1 d_C + \eta_2 d_C D$
  - DiRD nests DiDS, can test equality of $f(z)$. Not rejected, thus use DiDS to get precision.

Graphical illustration of RD: Share completing first year



Graphical illustration of RD: Exam score



RD: Baseline results

	RD	DiRD	DiDS
1st stage: Intensive training	0,565**	0,565**	0,692**
	(0,090)	(0,090)	(0,023)
Balancing: Index of background char	0,006	-0,017	-0,023**
	(0,023)	(0,035)	(0,009)
Exam 10th grade	0,002	-0,033	-0,067
	(0,192)	(0,294)	(0,076)
Completion of 1st year, on time	0,103	0,072	0,096**
	(0,105)	(0,158)	(0,041)
Completion of 2nd year, on time	0,094	-0,098	0,066
	(0,130)	(0,175)	(0,044)
Enrolled 2nd year, on time	0,200	0,138	0,088**
	(0,111)	(0,159)	(0,042)

Interpretion of RD results

- No effects on GPA or exam score.
- However, some evidence of effect on completion and enrollment:
  - Only significant for DiDS, but consistent estimate across techniques
  - ITT estimate: The share completing 1st year/enrolling 2nd year (just) below the cutoff is 9 percentage points higher in Stavanger than in Bergen.
    - Corresponds to ATT of about 13 percentage points
  - We find no clear effets on later completion of/enrollment in later grades
    - May be due to few available cohorts/little power

Interpretion of RD results (cont'd)

- RD effect: Marginal direct effect - spill-overs
  - Similar results above cutoff and parallel below cutoff in control and treated groups: Suggests homogeneous effect in Stavanger
  - Caveat: Higher-performing Stavanger student could have performed better in absence of program
- Study several outcomes and specifications: Significant results by chance?
  - Effect on completion/enrollment not significant if Bonferroni adjusted, but more low p-values than from a uniform distribution
  - P-values (roughly) follow uniform distribution for effects on grades

Cost benefit

- Is an effect of 9 percentage points important?
  - ITT effect of 9 percentage points on target students - about 1 percentage point on all students
  - Smaller than objective (6 percentage points)
  - Still, potentially, yes
- Improving national upper secondary completion rates rates by 1 percentage point could provide a benefit of >500M NOK (>50M EUR) for each cohort (cf Falch et al, 2009)
- Reported cost of Ny GIV is 726M NOK (treating 12,000 students)
  - Includes phasing in, training of teachers and evaluation, cost of continuation likely to be lower.
- Thus, a long-term completion effect similar to the short-term effect in Stavanger will be cost effective

Differences in implementation

- School owner (municipality) more involved in Stavanger than rest of country
  - "School owners have prepared instructions/guidelines": 90 vs 35 percent
  - "School owners coordinate networks for exchange of experience": 90 vs 50 percent
  - Also directly involved in selection of students
- Other characterics similar
  - E.g. group size, teacher qualifications, teaching in/ouside regular classes

Conclusion

- Decent quasi-experimental varition, combined with process evaluation
  - But: Unclear implementation complicates evaluation
- DiD estimates: No effects of introduction of program
  - Unlikely to find significant effects further on
  - Can rule out effects smaller than objective of program, but may be effects on target group
- RD estimates: Indications of some positive effects on completion/enrollment in one municipality
  - Compared to control group: Seems to be a general effect on low-performing students
  - Hard to generalize results: How and why is implementation different in Stavanger?

Implications for policy

- Overall, program has been unsuccessful at improving average school performance
  - But may still contain elements that can work well, and provide benefits for individuals and society
- Implementation is important - for effect and effect estimates
- Difficult for Ministry of education to micromanage implementation
- Supports decentralized approach:
  - Provide funding to counties (which are responsible for upper secondary schooling) and researchers, given that the suggested interventions:
    - Can be expected to be effective (based on previous research or experiences)
    - Are considered policy-relevant (i.e., may be reproduced by other schools/counties)
    - Provide a clear plan for evaluation, preferably involving RCT

Current policy and our new project

- Ministry currently funds six large-scale RCTs:
  - 2 reducing class size in grades 1-4
  - 4 targeting increased completion of upper secondary
- Our project: Intensive math train for low-performing students in year 8 and 11
  - Close collaboration with municipality of Oslo
  - Randomize 48 lower secondary schools, 17 upper sec.
    - Outcomes from adm data: Test scores, completion, survey data?
  - Research collaboration:
    - Process evaluation
    - Didactics - content of intervention