Effects of Early Warning Emails on Student Performance

J. Klenke, T. Massing, N. Reckmann, J. Langerbein, B. Otto, M. Goedicke, C. Hanck

15^{TH} International Conference on Computer Supported Education

Prague, 21-23 April, 2023

Outline

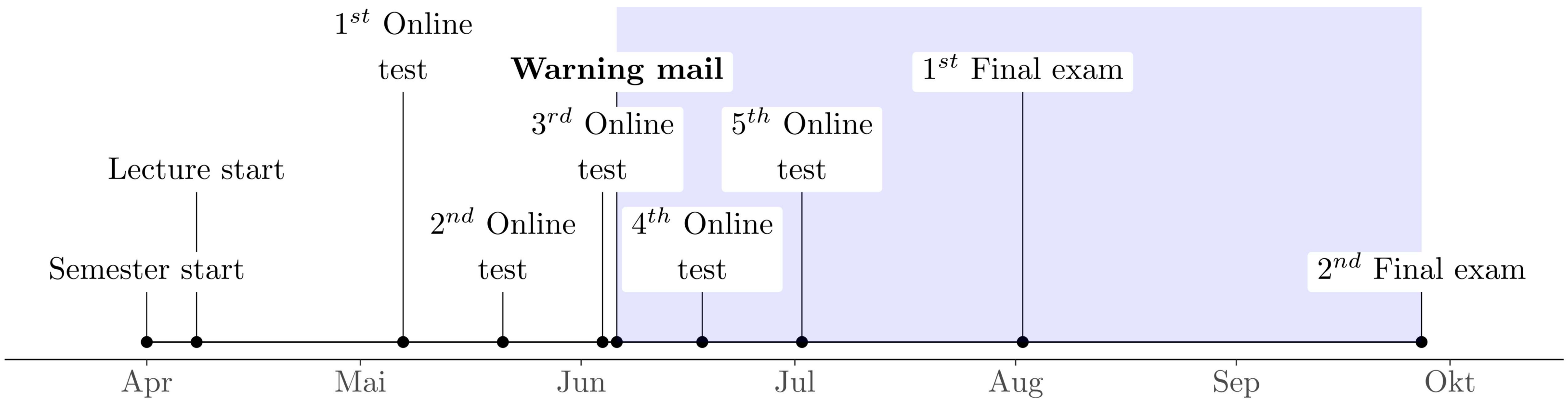
Research Idea and Course Description

- **Research Idea:** students should receive objective and motivating feedback through a warning email
- Analyzed Course: *Inferential Statistics* at the University of Duisburg-Essen
- Compulsory for business and economics
    - Weekly 2-hour lecture
    - Weekly 2-hour exercise
    - Kahoot! games used during classes
    - Homework and 5 online tests on the e-assessment platform JACK
- **802** students at the beginning of the semester
    - **337** students took an exam at the end of the semester

Decision Rule

- Logit model was used to predict students probability to pass the exam based on the first 3 online tests
  - Model was trained with the latest data obtained from previous edition of the same course
- If predicted probability to pass ≤ 0.4 the student got a warning mail

Course Timeline Main Events



Timeline for the key events in the 2019 summer term course Inferential Statistics (treatment cohort).

- The shaded area indicates the period after treatment.
- 57 days between the warning mail and 1^{st} exam
- 113 days between the warning email and 2^{nd} exam

Literature on Warning Systems in Education

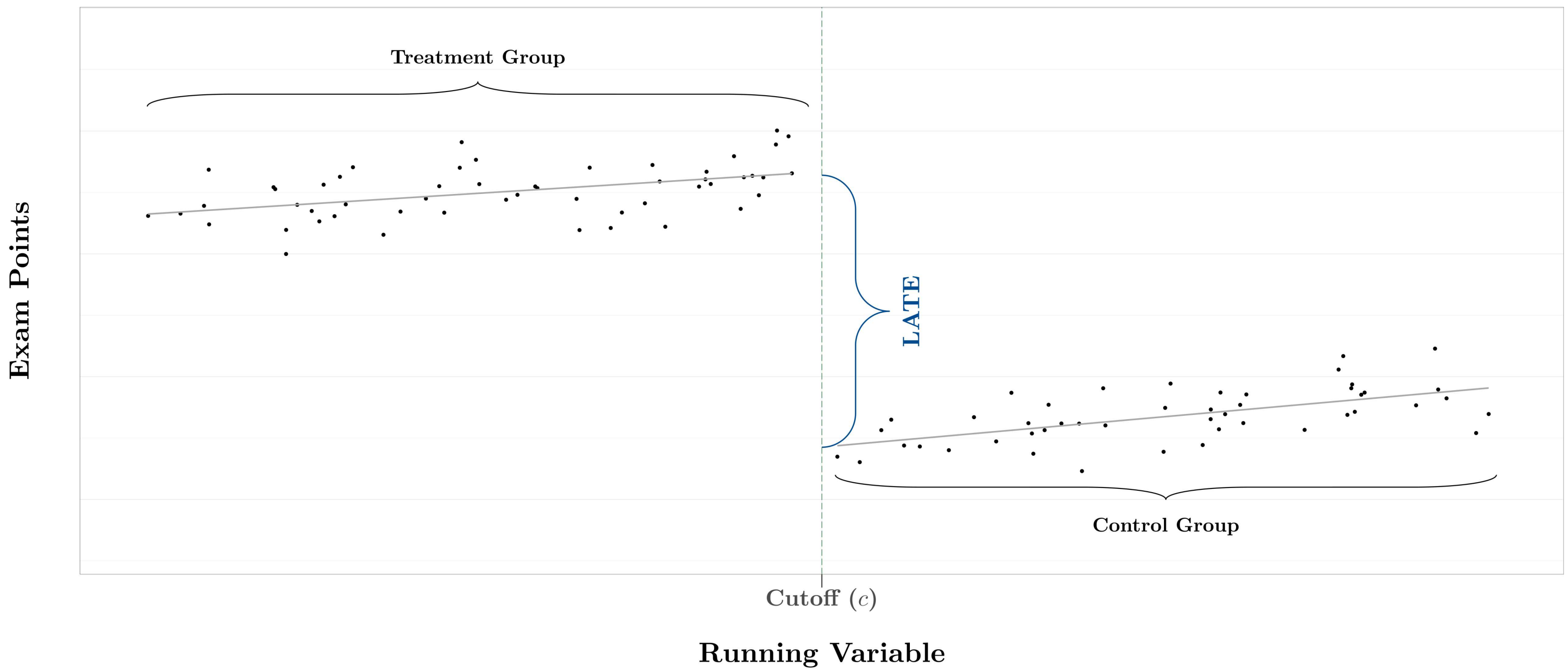
- Arnold and Pistilli (2012) investigated the effect of the signal light system at the Purdue University and found a positive effect on student grades
- Bañeres, Rodríguez, Guerrero-Roldán, and Karadeniz (2020) implemented an early warning system but did not analysed the effect on students' performance
- Şahin and Yurdugül (2019) invented an *Intelligent Intervention System* where for each assessment the students get feedback. Students emphasized the usefulness of the system.
- Mac Iver, Stein, Davis, Balfanz, and Fox (2019) could not find an effect from their early waning system in the ninth grade.
- Edmunds and Tancock (2002) analyzed the effects of incentives on third and four-graders' reading motivation and did not find an effect.

- The literature on the effects of warning system is inconclusive
- Many studies analyzed the sytstem with questionnaires

➔ We try to measure the effect directly on students' performance

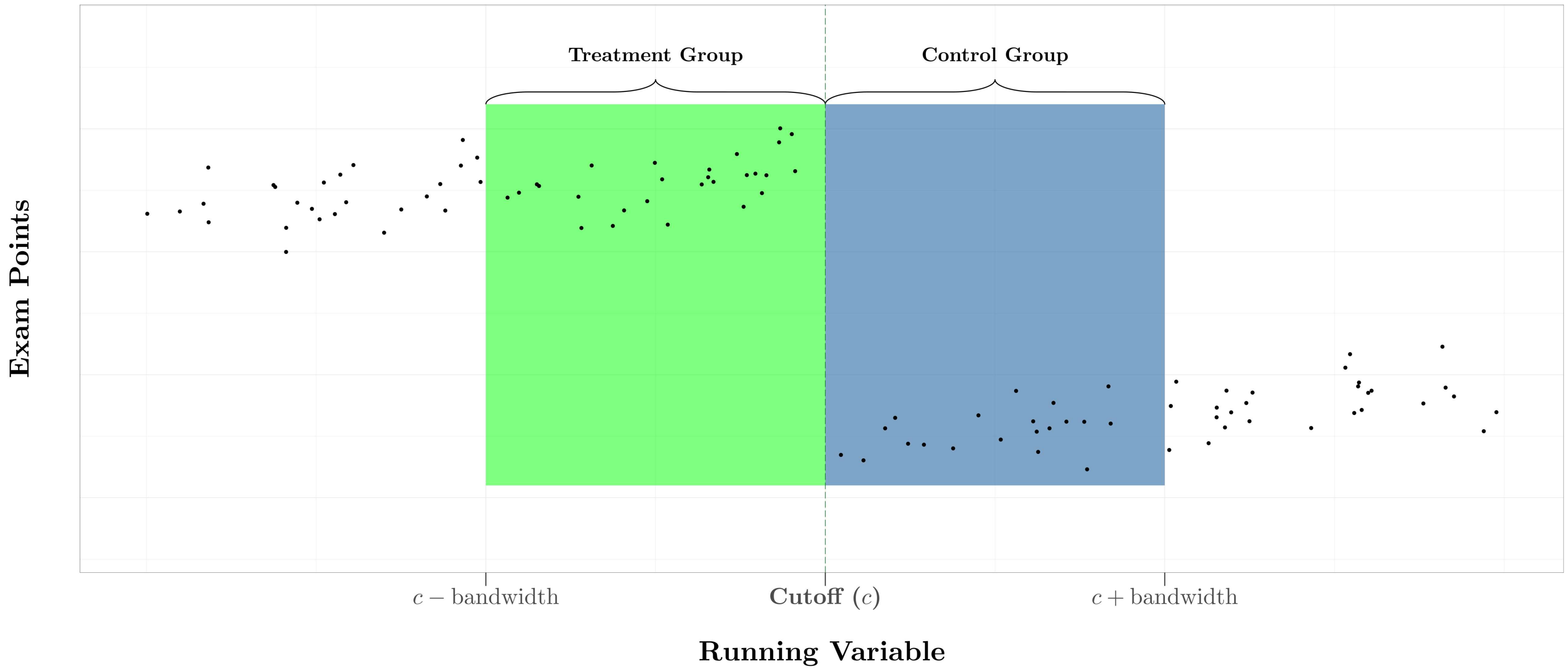
RDD Toy Example – I

Parametric Estimation



RDD Toy Example – II

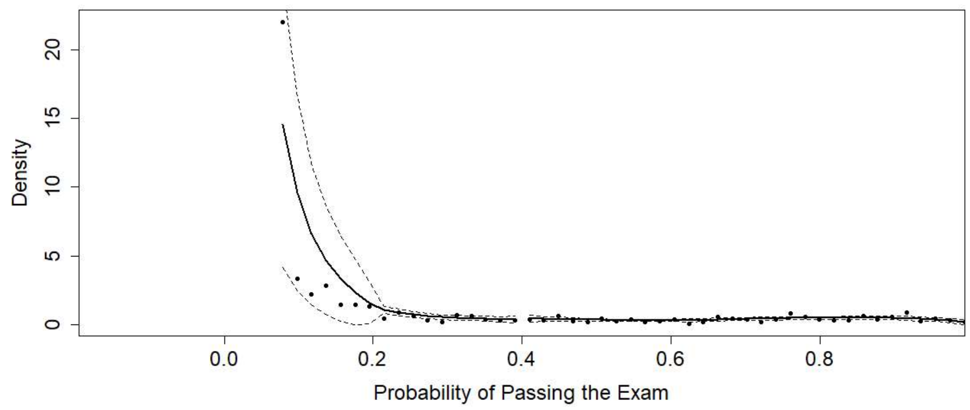
Non-parametric Estimation



- We used the data-driven approach by Imbens and Kalyanaraman (2009) to determine the bandwidth

Model Assumptions

- The running variable W must not have a jump around the cutoff in the density function.

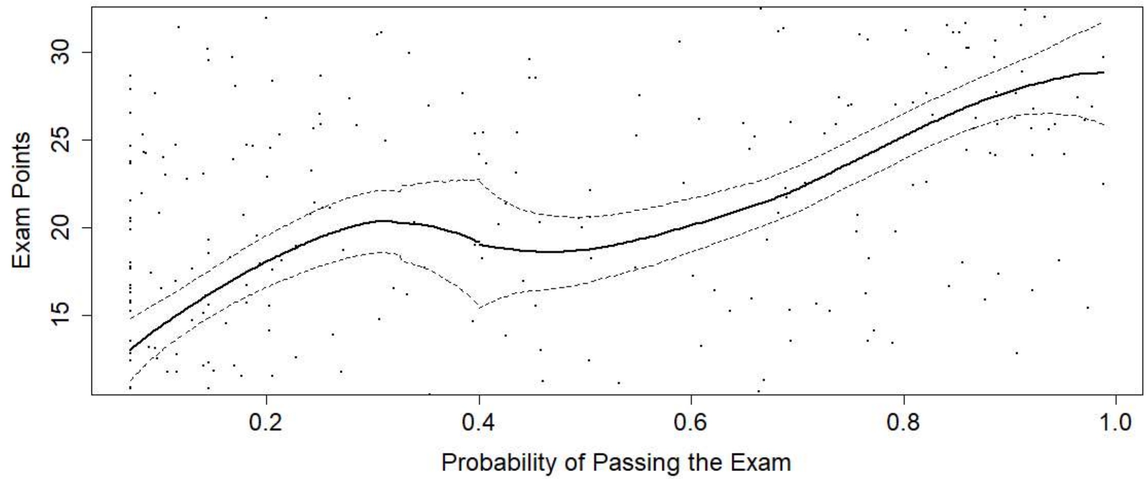


Graphical illustration of the McCrary sorting test.

- There is no jump in the density around the cutoff point of 0.4.
- p-value: 0.509
- The incentive to manipulate the treatment is quiet low.

- Also standard IV estimation assumptions must hold

Empirical Results – I



Graphical illustration of the RDD model.

Estimate

- LATE: 0.193
  - SE: (4.889)
  - p-value: 0.968
- Bandwidth: 0.255
- N: 126

Empirical Results – II

- The LATE estimate is positive but not significant
  - An estimate of 0.193 means that students who received the warning email achieved 0.193 points more than comparable students who did not
  - Compared to the 60-point exam, the effect size seems limited
- Bandwidth of 0.255
  - Only students with a predicted probability 0.4 (cutoff) ± 0.255 (bandwidth), are included in the analysis.
- This leads to the effective sample size of 126 students.

Discussion – I

- Our RDD results do not provide evidence that the warning email has a significant effect on students' results (or behavior)
- The variance around the cutoff is rather high, which compromises the detection of an effect
- Many individuals are not included in the final analysis for several reasons
  - Students dropping the course
  - Students far away from the cutoff are not providing much information for the model

→ Thus precise estimation of the treatment becomes more difficult

Discussion – II

- Students get also feedback through their online tests
- The warning may also lead weak students to postpone participation to a later semester
    - The cost to postone exams are in our program quiet low
- The objective feedback and motivation from one warning email is rather small

Further Research

- The effect on the dropout rate from such warning emails or systems requires further attention
- An automatic repeatedly feedback system could possibly have a greater impact on students motivation
    - A detailed recurring feedback could also used to guide students

> We see the open and transparent communication of the student’s performance to the students as a positive aspect of the system.

References

Angrist, J. D., G. Imbens, and D. B. Rubin (1996). "Identification of Causal Effects Using Instrumental Variables". In: *Journal of the American Statistical Association* 91.434. Publisher: Taylor & Francis, pp. 444-455.

Arnold, K. E. and M. Pistilli (2012). "Course signals at Purdue: using learning analytics to increase student success". Eng. In: *ACM International Conference Proceeding Series*. LAK '12. ACM, pp. 267-270.

Bañeres, D., M. E. Rodríguez, A. E. Guerrero-Roldán, et al. (2020). "An Early Warning System to Detect At-Risk Students in Online Higher Education". In: *Applied Sciences* 10.13, p. 4427.

Edmunds, K. and S. M. Tancock (2002). "Incentives: The effects on the reading motivation of fourth-grade students". In: *Reading Research and Instruction* 42.2, pp. 17-37.

Huntington-Klein, N. (2022). *The effect : an introduction to research design and causality*. First edition. Chapman & Hall book. Boca Raton ; London ; New York: CRC Press, Taylor & Francis Group.

Imbens, G. and K. Kalyanaraman (2009). "Optimal Bandwidth Choice for the Regression Discontinuity Estimator". In: *National Bureau of Economic Research* 1.14726.

Mac Iver, M. A., M. L. Stein, M. H. Davis, et al. (2019). "An Efficacy Study of a Ninth-Grade Early Warning Indicator Intervention". In: *Journal of Research on Educational Effectiveness* 12.3, pp. 363-390.

Şahin, M. and H. Yurdugül (2019). "An intervention engine design and development based on learning analytics: the intelligent intervention system (In 2 S)". In: *Smart Learning Environments* 6.1, p. 18.

Appendix: Regression Discontinuity Design (RDD) – I

- The treatment is **not** randomly assigned and therefore methods like OLS are not suitable
  - Treatment is a function of the predicted probability to pass the exam
- Consider the following **sharp** RDD representation Huntington-Klein (2022):

$$Y_i = \beta_0 + \alpha T_i + \beta W_i + u_i$$

- W_i denotes the predicted probability to pass the final exam
- T_i indicates if a student received a mail
  - $T_i = 1[W_i \leq c]$, with $c = 0.4$
- α denotes the treatment effect
- u_i denotes the error term

! This design is not suitable for our analysis as our groups are not perfectly separated

Appendix: Regression Discontinuity Design (RDD) – II

Appendix: Fuzzy RDD

- **Fuzzy** RDD allows to analyse a treatment in a setting where the two groups are not perfectly separated
  - Only the likelihood of receiving the treatment needs to *change*
- The effect is estimated through an instrumental variable estimation Angrist, Imbens, and Rubin (1996) where in the first stage the $\widehat{T}_i$ are estimated which then are inserted in the second stage
- First Stage:

$$T_i = \gamma_0 + \gamma_i Z_i + \gamma_2 W_i + \nu_i$$

- Second Stage:

$$Y_i = \beta_0 + \alpha \widehat{T}_i + \delta_1 W_i + \beta X_i + u_i$$

Appendix: Regression Discontinuity Design (RDD) – III

- RDD compares the individuals around the cutoff to estimate the effect
- **Main Assumption:** Individuals around the cutoff are compareable and only differ in the treatment assignment
  - The estimate is called Local Average Treatment Effect (**LATE**)
- For both methods, sharp and fuzzy,the estimation can be either parametric or non-parametric
  - Parametric estimation
    - Uses the whole sample size but (many) more parameters
    - Individuals away from the cutoff are less relevant for the estimation of the effect
  - Non-parametric estimation
    - Only *comparable* (near the cutoff) individuals are used for the analysis
      - Decision whether to include an individual depends on the running variable W
      - The groups are determined by the the data-driven approach of Imbens and Kalyanaraman (2009)
        - With an F-test the bandwidth is determined
        - $c \pm bandwidth$ are the two groups