Evaluating the Effects of HEERF Aid to Students

A regression-discontinuity design to estimate the effects of emergency aid on student academic outcomes.

Office of Evaluation Sciences
February 2021

Evaluating the Effects of HEERF Aid to Students

A regression-discontinuity design to estimate the effects of emergency aid on student academic outcomes at institutions that used eligibility thresholds to distribute aid.

What was the challenge?

The Higher Education Emergency Relief Fund (HEERF), created as part of the Coronavirus Aid, Relief, and Economic Security Act (CARES Act) signed into law on March 27, 2020, authorized the Department of Education (ED) to provide $14 billion for institutions of higher education (IHEs), half of which was intended to go directly to students to help offset the costs of COVID-19-related campus disruptions.[1] HEERF resulted in the disbursement of hundreds of thousands of one-time cash grants to students struggling with the adverse effects of the COVID-19 pandemic.

Given the urgency and severity of the COVID-19 pandemic, it was not likely feasible to prospectively embed evidence-building activities into HEERF. However, it may be possible to design, retrospectively, rigorous impact evaluations using available administrative data. This project leverages the fact that IHEs were given wide latitude in how they determined student eligibility for funds and how funds were distributed. Many IHEs used eligibility criteria that sorted students into groups that did and did not qualify for (or qualified for different amounts of) HEERF aid. Researchers can use these criteria, paired with available administrative data, to estimate the effects of HEERF aid to students. Understanding the effects of HEERF aid to students on student academic outcomes can inform the design of future emergency aid efforts and provide guidance to institutions as they consider effective strategies to encourage academic success.

What did we do?

OES designed an impact evaluation to help ED and IHEs answer the question: *What is the effect of providing HEERF aid on short- and medium-term student outcomes?*

Some IHEs chose strategies to determine student eligibility for funds that researchers can use to examine the effects of receiving aid or different amounts of aid. OES identified that many institutions used strict eligibility cutoffs that lend themselves to regression discontinuity (RD) designs—a type of quasi-experimental method—that can be applied with many IHEs.

How is the evaluation designed to build evidence?

An RD design can compare outcomes for people just above and below a threshold if the threshold is set such that eligibility for a program or benefit is essentially random. People near the threshold

[1] Streeter, Michele. 2020. What's in the CARES Act? Higher Education & Student Debt. The Institute for College Access and Success
https://ticas.org/covid-19-2/covid-19-affordability/whats-in-the-cares-act-higher-education-student-debt/

can be assumed to be similar on all characteristics, both observed and unobserved, if they cannot sort themselves to one side of the threshold or the other (for example, by providing specific information to 'game' the system) and the threshold doesn't also determine other differences between people who are eligible and those who are not. The arbitrarily chosen threshold that qualified some individuals but not others for the program allows researchers to observe the effect of the program or benefit on outcomes.

In the case of HEERF, required reports on IHE websites indicate that several institutions determined aid eligibility using Estimated Family Contribution (EFC), which is part of a student's financial aid record if they have submitted a Free Application for Federal Student Aid (FAFSA). Because students filled out the FAFSA before IHEs made HEERF eligibility decisions, and provided that the same EFC threshold was not used for other aid decisions (for example, the Pell Grant), eligibility for HEERF aid can be treated as essentially random for students near the threshold.

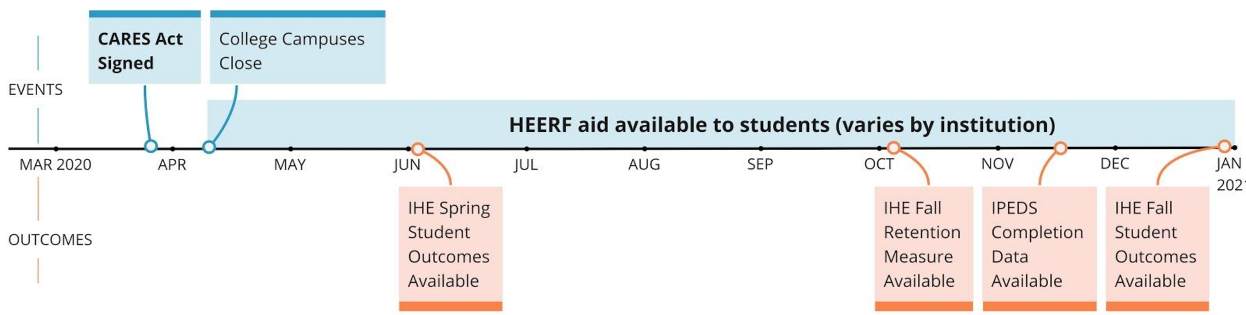

Figure 1: Timeline of HEERF-related events and observation of student outcomes.

Description of data required:

- *Eligibility threshold*: A student-level indicator of eligibility for HEERF aid, such as EFC. For a regression discontinuity design to work with EFC as the running variable, it must be true that students just above and below the EFC threshold look the same both on observable characteristics (e.g., gender, race, age, other support received) and characteristics not captured in administrative data (e.g., motivation, career aspirations).
- *Amount of HEERF aid received*: Student-level administrative data on the amount of aid received and date of disbursement. These data are usually maintained by the institution's financial aid or bursar's office.
- *Primary student-level outcomes*, usually collected in institution registration data:
    - *Retention:* Binary indicator of continued enrollment (or withdrawal) in the Spring 2020 term.
    - *Credits earned:* Number of credits successfully earned in the Spring 2020 term.
    - *Enrollment intensity:* Registered credits as a proportion of full-time enrollment status.
    - *Fall 2020 enrollment:* Binary indicator of enrollment in courses in the Fall 2020 term.
    - *Time to degree:* Number of academic terms from post-secondary entry to graduation.
- *Student-level financial aid package data*: Information for each student on other financial aid and grants received, tuition and fees associated with Spring and Fall 2020 enrollment, and student account balances at the time of aid disbursement.

- *Individual student characteristics*: Race, gender, socioeconomic status, first generation status, and other individual characteristics collected by the institution.

Time frame for observing outcomes and data availability: Analysis of Spring 2020 student-level outcomes could likely begin immediately, with analysis of Fall 2020 outcomes potentially available in early 2021. Student-level outcomes are likely to be available shortly after the semester for which outcomes will be observed. Obtaining access to student-level data may require data sharing agreements and custom pulls of data by an institution's bursar's and registrar's office, which could create delays in conducting the analyses.

Key requirements to apply this method

1. *Threshold Fidelity:*
    a. There is a continuous variable used to determine the allocation of HEERF aid and a threshold where the probability of students receiving HEERF aid is discontinuous, meaning significantly higher on one side of the threshold than the other.
    b. Students' position above or below threshold is as good as random, meaning it is not likely that students are able to 'game the system' and position themselves just above the threshold to receive a grant.
    c. No other aid programs or sources of support that may affect student outcomes utilizes the same cutoff. For example, if the cutoff for HEERF aid is the same as for receipt of Pell Grants, any differences in student outcomes on either side of the cutoff could not be attributed to HEERF aid independent of Pell Grant receipt.

2. *Sufficient sample size to detect policy-relevant effect sizes*:
    a. Based on a simulated analysis of data from a large public flagship university that enrolled 30,000 students in 2015-16, an evaluation could detect a 5 percentage point change in the likelihood of re-enrollment and graduation with **approximately 3,000 students** near the EFC threshold.[2] This implies a sample of approximately 3,000 students with an EFC between $4,000 and $6,000 in order to detect a meaningful effect with a high degree of confidence.

How estimates are generated: The effect of receiving aid can be estimated in most cases using an ordinary least squares (OLS) regression of the outcome of interest on an indicator of receipt of aid and a function of distance from the threshold. This regression would yield unbiased estimates of the local average treatment effect of HEERF aid on student outcomes, meaning the effect among students close to the threshold. A thorough examination of the effect may include tests of whether the estimates are sensitive to changes in the bandwidth (how far away from the threshold you are willing to include data) and functional form (i.e., linear vs. quadratic).

Interpretation of estimates: Estimating the equation would reveal causal estimates of the local average treatment effect of HEERF aid on student outcomes for the set of students near the threshold. If the threshold identifies students who received HEERF aid versus those that did not,

[2] In these estimates, we controlled for standard demographic characteristics and other elements of students' financial aid packages, which increased precision.

the estimates indicate how receiving aid affects the outcomes of interest relative to not receiving any aid. If the threshold identifies students who received a higher versus lower amount of aid, the estimates indicate the effect on outcomes of receiving more aid relative to less.

Limitations:

Some Disbursement Strategies are Not Appropriate for an RD: A regression discontinuity design is only appropriate if the disbursement strategy relies on an arbitrary cut off where students' placement above and below an eligibility threshold is as good as random. Disbursement strategies that may not satisfy randomness include those that require applications or documentation of expenses from students.

Limited External Validity: The local average treatment effects only capture the effects of HEERF on the subset of students close to the threshold. If the threshold is based on family income, the findings may only generalize to the subset of low-income students who completed the FAFSA and are likely to be eligible for need-based financial aid and HEERF aid. Also, effects may not generalize across institutions when thresholds and aid amounts differ between institutions.

Sample Size and Statistical Power: A sufficient number of students near the threshold is necessary to generate precise estimates of the effects. If an insufficient number of students are close to the threshold it may be difficult to detect effects with enough precision to be useful for decision making. Moreover, a larger bandwidth around the threshold may reduce validity if students further from the threshold are systematically different from those near the threshold.

What did we learn?

HEERF aid was passed, implemented, and distributed to students over a short period of time in Spring 2020, which precluded embedding evidence building into the aid program. But opportunities exist to build evidence using quasi-experimental methods, including a regression-discontinuity design.

The RD design can provide evidence of the effect of one-time emergency aid on student outcomes. Little evidence is currently available to indicate whether emergency aid helps students continue their education during campus disruptions, especially for a disruption of the magnitude of the COVID-19 pandemic.[3] Evaluations using an RD design can indicate how much emergency aid programs change student outcomes.

Building robust evidence may benefit from coordinated evaluations with multiple IHEs. Because institutions varied in the amounts of aid awards, how eligible students were identified, and the thresholds used for eligibility, results from an evaluation at any one institution may not be generalizable to other institutions. Developing a coordinated effort—for example, with a consortium of IHEs—to conduct RD evaluations at multiple institutions may yield a fuller picture of how HEERF aid affects student academic outcomes.

[3] Evans, William N., Melissa S. Kearney, Brendan Perry, and James X. Sullivan. 2019. A Cautionary Tale on Emergency Student Aid. *American Economic Review*. https://www.aeaweb.org/articles?id=10.1257/pandp.20191014
Dynarski, Sue and Judith Scott-Clayton. 2013. Financial Aid Policy: Lessons from Research. https://www.edpolicy.umich.edu/files/wp-dynarski-scott-clayton-financial-aid-policy-2013.pdf

Technical Annex

The following sections elaborate on each step in the process of scoping, validating, and estimating a regression discontinuity design. To use this evaluation strategy, the evaluator needs to have some knowledge of linear and non-linear regression and inferential statistics.

Step 1: Identify a threshold

Example: Expected Family Contribution (EFC)

Many institutions of higher education used EFC from the previous year's FAFSA application to determine who received aid and the amount of aid students received.[4] For example, a hypothetical institution could have disbursed their HEERF aid based on EFC in the following manner:

Contact Order	EFC	Pell Recipient	Grant Amount
First	$0	Yes—Full Pell	$1,250
Second	$1-5,576	Yes—Partial Pell	$500
Third	None-Specified. Sect. 484 Eligible	No	$300
Fourth	All Title IV Eligible Students	No	$300

In this type of distribution model, evaluators could use the zero EFC threshold to estimate the local average treatment effect[5] of receiving an additional $750 in grant aid on student outcomes. Alternative running variables could include GPA, if there was an academic eligibility component, or any other continuous characteristic of the student that influenced the probability that the student received HEERF aid.

Step 2: Check the Fidelity of the Threshold

To approximate a randomized evaluation, evaluators have to prove that the threshold identifies exogenous variation that makes the placement of students along the threshold as good as random. In order for the RD approach to be valid, there are a number of conditions including:

1. Sufficient sample size close to the threshold (and within the analytical bandwidth) to estimate the local average treatment effects with enough statistical power to identify policy relevant effects;
2. Evidence that the assignment rule was followed with a high degree of fidelity;

[4] This application may not have reflected the current financial hardships students were facing in the COVID-19 pandemic, so many colleges also allowed students to complete an application for HEERF aid.
[5] The local average treatment effect is the effect of HEERF Aid on students who *received* the aid who have an EFC *near* this threshold.

3. Evidence that there was no manipulation at the threshold (i.e. students weren't able to preemptively take action to strategically position themselves on one side of the threshold);
4. Evidence that students just above the threshold are similar to those just below the threshold in terms of observable characteristics; and
5. Evidence on whether the threshold coincides with other financial aid policies

To check the first and second condition, evaluators could create a plot like Figure 2 below, which shows that HEERF Aid was distributed to all students below the 70th percentile of the EFC distribution, and not to any students above the threshold. In reality, most of the time there is not a sharp jump from 0 percent likelihood of receiving aid on one side of the threshold to 100 percent likelihood on the other side of the threshold. This graph can also be utilized to examine the density surrounding the threshold, as a preliminary examination of sample size.

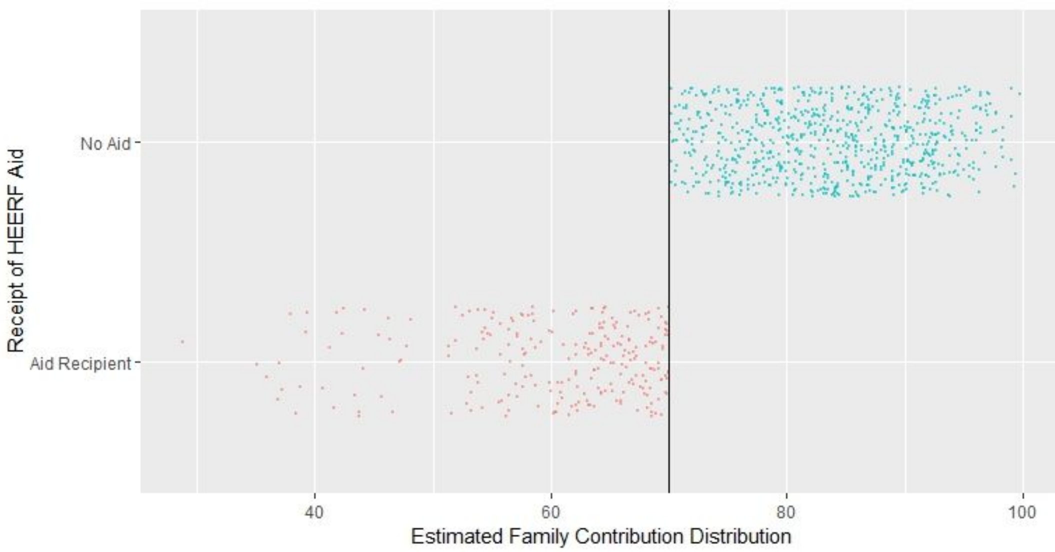

Figure 2. Visualizing the fidelity of the assignment rule. High fidelity to assignment is evident when all or most students with an EFC below a threshold (a 70th percentile cutoff in this example) receive aid (the red dots), and all or most students with an EFC above the threshold do not receive aid (the blue dots).

Next, evaluators would want to check whether there was any manipulation at the threshold, meaning students were able to take action to change their position relative to the threshold. To check this visually, evaluators would create a graph to examine whether there appears to be stacking of students just above or below the EFC threshold (Figure 3). In the particular case where evaluators use the EFC threshold, it is highly unlikely that students were able to manipulate their position relative to the threshold because the FAFSA was filled out in the prior year. Therefore, students were unaware at the time they filled out the FAFSA that HEERF aid would be allocated.

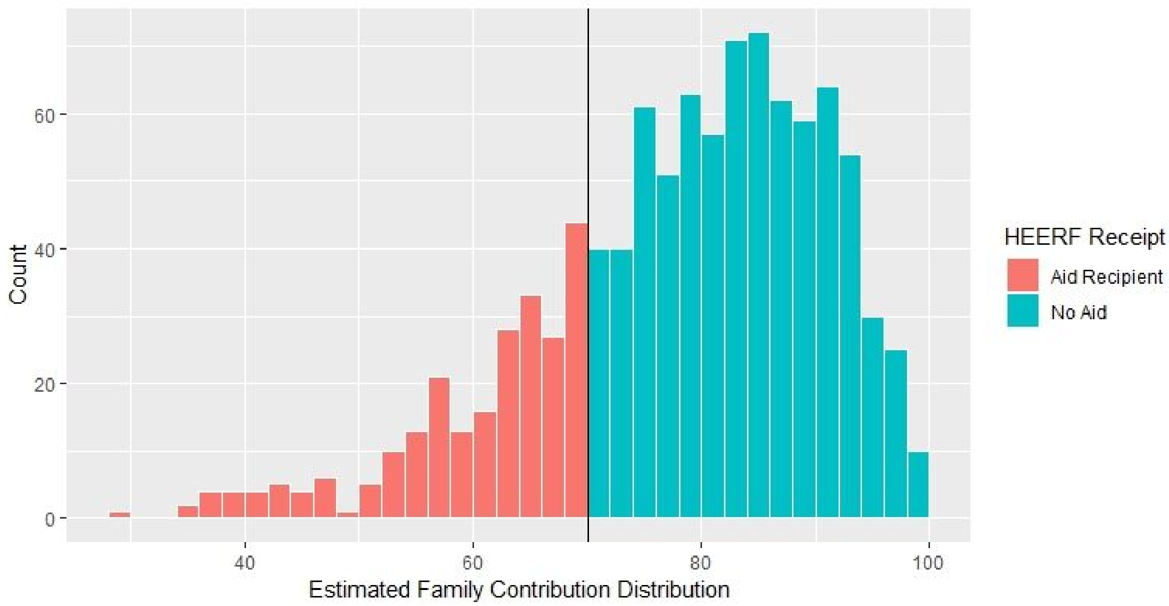

Figure 3. Density of students around the threshold to check for manipulation. An unusually tall red bar immediately below the threshold and an unusually low blue bar immediately above the threshold would indicate the potential for students near the threshold to alter their position to become eligible for aid.

It is possible that because EFC thresholds also determine other aid, such as the Federal Pell Grant, students may have engaged in manipulation. In this case, we predict this to be highly unlikely because of the opaque nature of EFC calculations, which many students would not fully understand when filling out the FAFSA. Regardless, evaluators would want to run a manipulation check, commonly known as the McCrary test[6] and visualize the results in a graph (Figure 4).

[6]McCrary, Justin. 2008. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics* 142, no. 2: 698-714.

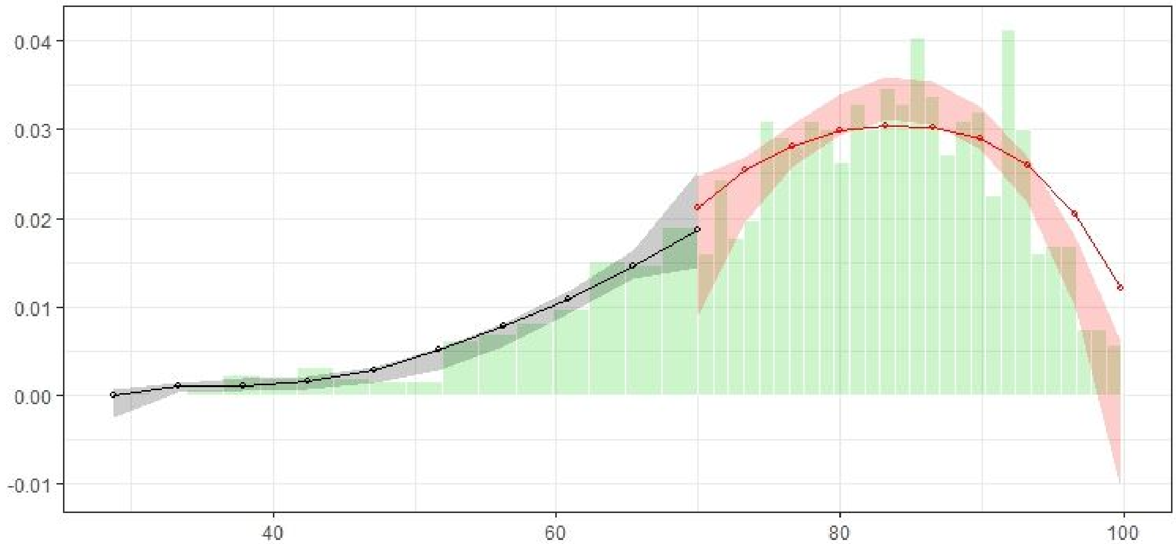

Figure 4. Example visual representation of the McCrary test for manipulation at the threshold. This test shows that the difference in the density on either side of the threshold (where the grey and red lines meet) is not significantly different from zero.

Evaluators would want to examine whether there are differences in observable characteristics (e.g. gender, race/ethnicity, age, first-generation status, etc.) between the students just above and below the threshold. To investigate the potential for differential student composition in observable student characteristics at the EFC or first-come first-served threshold, evaluators should estimate the probability of a student identifying, for instance, as white or as female as a function of their distance from the threshold and whether they are above or below the threshold.

Finally, it is important to identify whether the threshold coincides with other policies. One could imagine, for instance, that the EFC threshold for HEERF could be the same as the cutoff for the Pell Grant disbursement, which would mean students above the threshold are experiencing multiple treatments at one time, making it more difficult to isolate the causal effect of HEERF. In this case, evaluators could implement a difference-in-discontinuities design where changes over time are used to isolate the local average treatment effect of HEERF, while accounting for Pell Grants.

Step 3: Estimate the Local Average Treatment Effect

Part 1: Determine whether to estimate a sharp or fuzzy RD design

Sharp Regression Discontinuity Model

If there is perfect compliance at the threshold (i.e. the probability of receiving HEERF aid goes from 0% to 100%), evaluators should estimate a sharp RD. For colleges that automatically disbursed aid to students on the basis of EFC, there is a possibility of perfect compliance (as shown in Figure 1) which would allow for the estimation of a sharp RD.

$$y_i = \alpha + f(Dist_i) + \delta Below_i + \beta X_i + \varepsilon_i$$

In the equation above, y_i is an outcome of interest, $Below_i$ is a binary indicator of whether the student's EFC is above or below the threshold, $f(Dist_i)$ is a function of the difference between student *i*'s EFC and the threshold, and X_i is a vector of student level characteristics (i.e. race/ethnicity, gender, age, other financial aid awards). A visual example of results from an RD regression is shown in Figure 5.

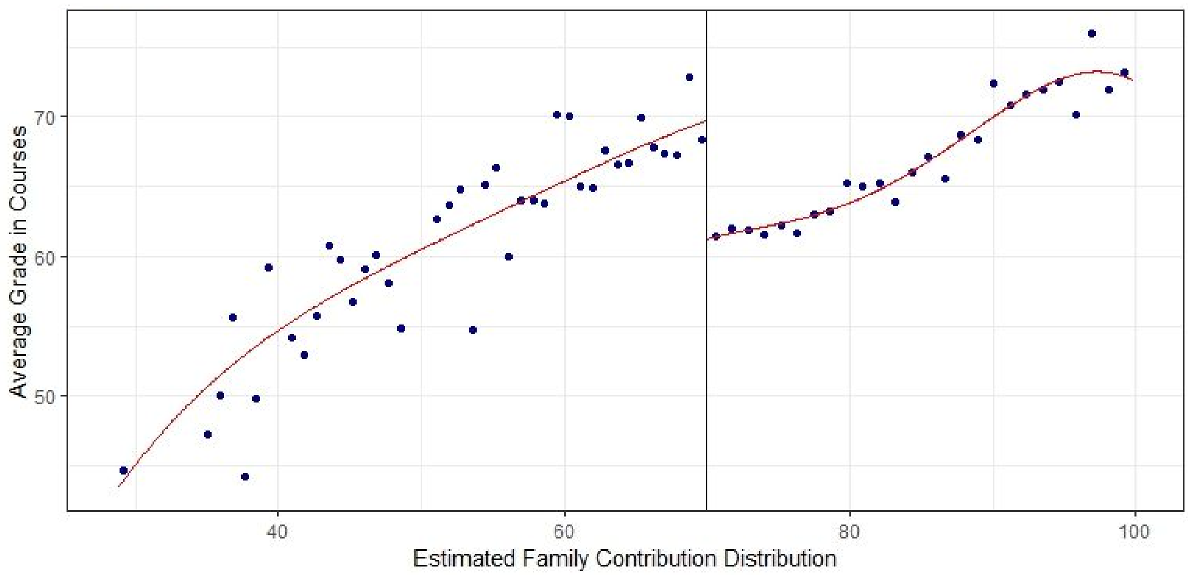

Figure 5: Example of regression discontinuity results. At the RD threshold (70th percentile), average grades are significantly higher for students who receive aid (to the left of the vertical line) compared with those who don't receive aid (to the right of the line). The difference between these lines is the Local Average Treatment Effect (LATE).

In estimating this equation, a key issue for evaluators is how many students to include above and below the threshold (i.e. the bandwidth). It is standard practice to estimate data-driven bandwidths and kernels using the algorithm developed in this rdrobust package. There are multiple types of bandwidths, and the selection of bandwidth is an important step in the analysis, given that this affects which students are included and excluded from the estimation. In the robustness section below, we expand on the different potential bandwidth estimation strategies that can be implemented to test the sensitivity of estimates

Fuzzy Regression Discontinuity Model

Given that many institutions allocate HEERF aid both based on a threshold and based on applications for students whose EFC is missing or needs to be updated, it is likely that evaluators will need to use a fuzzy RD, rather than a sharp RD. Evaluators would implement a two-stage least squares model (2SLS) where HEERF aid recipient (T_i) is predicted in the following first-stage model:

$$T_i = \alpha + f(Dist_i) + \delta Below_i + \beta X_i + \varepsilon_i$$

as a function of the distance from the threshold ($f(Dist_i)$), an indicator of scoring above the threshold ($Below_i$), a vector of student characteristics (X_i), and an idiosyncratic error term (ε_i). The predictions for T in the first-stage equation above are denoted as $\check{T}$ in the second stage equation presented below. In the second stage model below, ϕ represents the local average treatment effect (LATE) of HEERF aid on student outcomes.

$$Y_i = \alpha + f(Dist_i) + \varphi\check{T} + \beta X_i + \varepsilon_i$$

A key issue in estimation is the type of bandwidth and the type of function utilized to estimate the local average treatment effect. A common approach to the estimation of bandwidths is an algorithm[7] This LATE represents the causal impact of HEERF aid on student outcomes for those students near the threshold for whom scoring above the threshold would have resulted in receiving HEERF aid.

Part 2: Robustness Checks

1. Estimate the local average treatment effect using various functional forms
2. Test whether the results are robust to multiple bandwidths

It is standard practice to estimate both a linear and a quadratic functional form to test whether the estimates are sensitive to changes in functional form. Additionally, evaluators should also utilize multiple data-driven bandwidths to further probe whether the estimates are robust across specifications.

[7] See Calonico, S., M. D. Cattaneo, and R. Titiunik. 2015. Optimal Data-Driven Regression Discontinuity Plots. Journal of the American Statistical Association 110(512): 1753-1769.