Large-Scale Education Reform in General Equilibrium: Regression Discontinuity Evidence from India

A Reply to Roodman (*Journal of Comments and Replications in Economics, 2026)*

Gaurav Khanna*

Journal of Comments and Replications in Economics, Volume 5, 2026-4, DOI: 10.18718/81781.55

Please Cite As: Khanna, G. (2026). Large-Scale Education Reform in General Equilibrium: Regression Discontinuity Evidence from India. A Reply to Roodman (2026). *Journal of Comments and Replications in Economics*, Vol.5 (2026-4). DOI: 10.18718/81781.55

*UCSD's School of Global Policy and Strategy. gakhanna@ucsd.edu

I thank David Roodman for his in-depth reanalysis, and for sharing various versions of his comment. I was happy to share data and code while K23 was still under review, and systematically addressed an evolving set of comments raised over regular emails since January 2022. Here, I respond to a new set of comments shared with me and also the version submitted to a journal.

Received November 16, 2025; Published March 5, 2026.

1 Introduction

Gaurav (2023) (henceforth K23) argues that large-scale investments in education depend on the labor market GE effects. K23 develops a GE model that captures how expansions affect the returns to schooling and estimates the model's parameters leveraging variation from India's large-scale school-building program District Primary Education Program (DPEP). K23 uses various estimation strategies (RDs and Diff-in-Diffs) to show that the school-building raised local education levels and consequently depressed the skill premium. K23 then evaluates the overall welfare consequences of large-scale schooling interventions.

Replications are critical to furthering what we learn, and I am happy to see them becoming more common. Roodman (2023) (R23) provides a re-evaluation of a part of K23. They replicate the main results, confirming K23's findings, and finding no errors in the original analysis code. They raise new critiques. I respond to each point in detail and show that the results in K23 remain conclusively robust.

First, R23 argues that the RD graphs override Stata defaults affecting the appearance of the discontinuity. I show that K23's modifications follow clear and sensible guidelines in the RD literature.

Second, R23 constructs a different dataset than K23, creating a long-run district-level sample, linking multiple sources across decades, dropping certain parent-child districts at an arbitrary cutoff, and changing how the primary outcomes are measured. K23 merges many other outcomes, with varying degrees of district coverage. Somewhere between 18-87 districts (out of 618) are matched differently, of which four are near the RD cutoff. Despite various differences in approach, R23 argues that the education results are marginally insignificant at conventional levels, when using these four districts. I comprehensively evaluate R23's claim. Indian districts endogenously change boundaries, and are split and merged regularly. R23's sample adds noise right at the cutoff attenuating the program-assignment first stage. For simple and conventional re-evaluations of R23's sample (taken as given), K23's results remain consistently robust, despite K23 merging many more databases (given the broader scope of K23).

R23's third point states that the GE effects depend on endogenous quantities. The original K23 paper highlights several times that the returns to schooling (both in partial and general equilibrium) do depend on labor market characteristics. Indeed, that is K23's contribution. Researchers have long treated the returns to education as a parameter, and K23's goal was to show how these returns in GE depend on various labor market quantities. A few of R23's statements here perhaps reflect a misreading of GE models.

R23's final point is that they do not evaluate the Diff-in-Diff as it does not rely on the RD variation. I show that even the Diff-in-Diff is robust to R23's sample and data changes.

I begin by focusing on R23's main concern: the second point of different districts. I then address the other points in detail.

2 Robustness to R23's Sample

First, I discuss R23's second (and main) point. R23 replicates the main results of K23, finding no analysis errors. R23 then recreates part of the estimation sample. This is challenging, as the data span two decades, from seven different sources, and Indian districts are frequently split, merged, and renamed. Importantly, K23 merges many more datasets than R23, as K23's GE analysis is broader in scope than R23's replication. The analysis in K23 includes various additional outcomes (firm-level outcomes, school-level outcomes, test scores, district domestic product, etc.), many of which do not cover all districts, but are necessary for the comprehensive analysis. I show robustness to R23's small-outcome sample.

Researchers likely use different assumptions to handle district splits and merges. R23's sample construction was not coded in statistical software, so I could not evaluate it in the timeframe provided. Rather than examining how R23 construct their sample, I simply assume R23 made good-faith assumptions, and show robustness using R23's sample.[1]

R23 and K23 differ in the assignment of between 18 (in the ITT) to 87 (in the 2SLS) out of 618 districts. These perhaps stem from several reasons: First, K23 merges many more datasets (for a comprehensive GE analysis) that may have lower district coverage, and other time horizons covering splits and merges. Second, as districts split and merge, R23 picks an arbitrary cutoff to decide which parent-child districts to drop. Third, R23 uses the revised 2001 list of districts for program assignment rather than the original (and more exogenous) list. Finally, R23 redefines the wage variable and recodes the main education variable, for instance, coding non-formal education, adult education, and the total literacy campaign as 0 years of schooling. Putting aside the soundness of these choices, I show that K23's original findings hold using R23's sample.

R23 contends that of the 87 districts that do not match the samples, only the four near the RD cutoff make a difference in the results. So, I test for robustness first, including the four districts, and then the full R23 data. The original K23 paper shows the main results using Sebastian Calonico & Titiunik (2014) (henceforth CCT) and Imbens & Kalyanaraman (2012) (henceforth IK) bandwidth selection procedures, and shows robustness to various other methods (parametric RDs, Bartalotti & Brummet (2017) clustering, difference-in-differences, two-sided bandwidths, etc.). R23 only examines the CCT bandwidth.

I show robustness to R23's re-estimation in multiple ways. Since the IK bandwidth is in all of K23's RD tables, I re-introduce the IK bandwidths. CCT bandwidths are smaller, and so more sensitive to measurement error (e.g., based on district splits) right at the cutoff (as rdrobust uses a triangular kernel, it heavily upweights the districts right at the cutoff). R23 also keeps the bandwidth fixed when first introducing the new districts, but as CCT points out, the bandwidth is endogenous to the sample. So, I also re-estimate the IK bandwidth in all tables when adding data from R23 (henceforth New BW).

Table 1 first reproduces the main results from K23 and R23 for the years of education for the

[1] While I take these changes as given, I have strong reservations about R23's sample, as I point out construction issues below. For instance, R23 claims "no new multi-parent districts were formed in 2001-9", which is incorrect.

Table 1: K23 and R23 Results: Years of Education (Young age group)

	Original		R23 sans 4 districts			With 4 districts		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RD	0.720	0.698	0.656	0.622	0.628	0.297	0.365	0.572
	(0.199)***	(0.174)***	(0.195)***	(0.170)***	(0.173)***	(0.189)	(0.166)**	(0.136)***
Obs	10175	14277	10634	14922	14554	10809	15114	24103
BW	CCT	IK	CCT	IK	New BW	CCT	IK	New BW
BW	0.103	0.134	0.103	0.134	0.130	0.103	0.134	0.231

Notes: First 2 columns are from the original K23 paper. Columns 3-5 are from R23's sample without the 4 districts near the RD cutoff. Columns 6-8 are adding the 4 districts to K23's sample. Except for the first 2 columns, all use R23's recoded variables.

young sample.[2] The first two columns show K23's original results. Columns 3-5 use R23's full sample (including variable recodes, program assignment, and dropped-due-to-parentage districts), but without the four additional districts. The results are qualitatively similar to K23. R23 only shows column 3, but I also include columns 4 (IK) and 5 (new BW). In columns 6-9, I add the four districts to the original sample. The CCT result is now marginally insignificant at conventional levels (p-value=0.116), but the IK and new BW results are still both economically and statistically significant at conventional levels. R23 further argues that including all the R23 sample, the CCT p-value is 0.147.

One concern with the R23 sample is that it may be introducing measurement error in treatment assignment right at the cutoff. The top panel of R23's Figure 2 shows that first-stage treatment assignment falls substantially, and is noisier. Not surprisingly, the outcomes in ITT are attenuated by similar amounts. This noisy assignment could result from various factors, including district splits/merges over time, or slippage in assignment. Fortunately, the RD literature has clear directives for noise and slippage (for instance, donut-hole RDs). Given that CCT bandwidths are small, and all estimation here relies on a triangular kernel that heavily upweights districts right at the cutoff, the noisier assignment in R23's sample may attenuate the effects on education.

In Table 2, I examine the robustness of R23's sample, focusing on the two samples that R23 claims have weak results: (a) the K23 sample, adding the four districts, and (b) the full R23 sample. Both always include R23's several analytical choices, including variable re-codes, arbitrarily dropping parentage, etc.

First, given the tight bandwidth and the triangular kernel sensitivity to noise at the cutoff, in Panel A, I drop only the one closest district (out of 618) to the cutoff, and all results are economically and statistically significant at conventional levels. Interestingly, the 1st district just below the cutoff (Tiruvanamalai) was newly created just before treatment assignment, and the 1st district just above the cutoff (Champawat) was newly created after treatment but before the outcomes were measured. This introduces noise: Champawat was formed from two districts, so we do not have a clean running

[2] I focus on the years of education, as that is the main first step in the analysis. Average wages are allowed to theoretically be positive, zero, or negative, based on the GE effects. Appendix Table 4 shows re-estimations of the wage consequences show results similar to K23.

Table 2: R23 Sample: Robustness Checks (Years of Education for Young)

Panel A	Without 1 nearest district to cutoff					
	With 4 districts			Full R23 Sample		
RD Estimate	0.422	0.458	0.458	0.398	0.416	0.428
	(0.191)**	(0.168)***	(0.169)***	(0.190)**	(0.166)**	(0.184)**
Observations	10719	15024	14767	10807	15095	11957
BW Type	CCT	IK	New BW	CCT	IK	New BW
BW	0.103	0.134	0.132	0.103	0.134	0.109
Panel B	Donut hole					
	With 4 districts			Full R23 Sample		
RD Estimate	0.951	0.861	0.865	0.921	0.812	0.810
	(0.203)***	(0.176)***	(0.165)***	(0.201)***	(0.175)***	(0.167)***
Observations	10396	14701	17483	10482	14770	15917
BW Type	CCT	IK	New BW	CCT	IK	New BW
BW	0.103	0.134	0.161	0.103	0.134	0.154
Panel C	No split-districts in donut					
	With 4 districts			Full R23 Sample		
RD Estimate	0.423	0.464	0.464	0.399	0.423	0.406
	(0.190)**	(0.167)***	(0.167)***	(0.189)**	(0.166)**	(0.188)**
Observations	10755	15060	15060	10843	15131	10843
BW Type	CCT	IK	New BW	CCT	IK	New BW
BW	0.103	0.134	0.134	0.103	0.134	0.104
Panel D	No split-districts in donut + no Cuddalore					
	With 4 districts			Full R23 Sample		
RD Estimate	0.524	0.545	0.547	0.501	0.503	0.522
	(0.192)***	(0.169)***	(0.171)***	(0.191)***	(0.167)***	(0.180)***
Observations	10668	14973	14595	10756	15044	12918
BW Type	CCT	IK	New BW	CCT	IK	New BW
BW	0.103	0.134	0.130	0.103	0.134	0.115
Panel E	All Districts: Uniform Kernel					
	With 4 districts			Full R23 Sample		
RD Estimate	0.597	0.420	0.503	0.565	0.361	0.509
	(0.177)***	(0.157)***	(0.138)***	(0.175)***	(0.156)**	(0.202)**
Observations	10809	15114	20039	10898	15186	8274
BW Type	CCT	IK	New BW	CCT	IK	New BW
BW	0.103	0.134	0.182	0.103	0.134	0.0750

Notes: Outcome is years of education (R23's recoded version). The first three columns add the 4 districts, whereas the last 3 columns are the full R23 sample. All other R23 sample changes (recoding of education and wages, dropping certain parent-child district merges, etc.) are incorporated fully.

variable value for it; and since the boundaries are endogenous, it is likely not a simple weighted average of its parent district's running variable values. Further, subsequent outcome sampling is

at the 'newly-formed' district level, rather than a weighted average of the parent districts. Even dropping only one of the districts (either above or below the cutoff) similarly shows all results are economically and statistically significant. This may be expected as the results were always robust to the slightly larger IK bandwidth, and the smallest possible donut hole makes the small-BW CCT results significant at conventional levels.

Yet, dropping just 1 district may not be the conventional donut hole approach, as the convention is to pick a range on the running variable. In Panel B, I create a tight donut hole of ± 0.004 of the running variable (the cutoff is 0.393). The results across all specifications and R23 samples are economically and statistically significant. Some of the districts in the donut hole may result from endogenous splits and merges, perhaps not reflecting their 'true' running-variable value and adding noise near the cutoff.[3]

In Panel C, I only drop the three districts within the tight donut-hole that were newly created between 1991 (when the running variable was measured) and 2001.[4] To further reduce noise, we may also want to drop districts created between 2001 and 2009 (when the outcome was measured), but I could not find 2009 district names in R23's data, so Panel C estimates are conservative.[5] I keep all other districts in the donut, and the final sample includes the 4 additional districts. The results are economically and statistically significant.

So far, I always include the four districts that R23 contests. When inquiring about the four districts near the cutoff on Jan 4, 2023, the R23 author says "I know there's a lot of complexity because districts split and merge over time. But it looks to me like these four weren't affected by that." But at least one of the four (Cuddalore in Tamil Nadu) was carved out in the middle of the sample period in 1997.[6] In Panel D, I exclude Cuddalore as well, and the results are even stronger.

Finally, part of the sensitivity to measurement error right at the cutoff may be because the estimations so far use a triangle kernel that heavily weights the districts right at the cutoff, with steadily falling weights as one moves away from the cutoff. In Panel E, I keep all the districts (no donut holes, nor dropping splits/merges) and simply use a more conventional uniform kernel, which weights all districts in the bandwidth equally. The results are economically and statistically significant. Using an Epanechnikov kernel, which is similar to a triangle, but less sharp, also produces results significant at conventional levels.

Since the very tight bandwidth with a triangular kernel is sensitive to noise at the cutoff, we may manually expand the bandwidth to test for robustness. In Figure 1, I keep the sensitive-to-noise triangular kernel (and the full data sample), but manually change the bandwidths, starting just

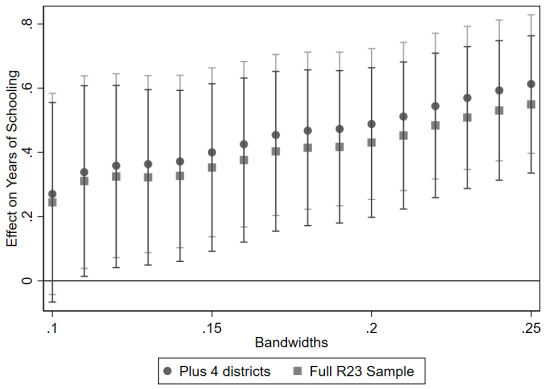
[3] A rich part of a district may be carved out, and later the outcome is measured for this new rich district, but the running variable value puts it near the cutoff.

[4] Many districts underwent a unique major redesign when Uttaranchal was created in 1997.

[5] R23 claims, "no new multi-parent districts were formed in 2001-09." This is incorrect, and concerns me about how the R23 sample is formed. Here are just a few (non-exhaustive list of) examples contradicting R23's claim: In 2008, Pratapgarh was created from various parts of Banswara, Chittorgarh, and Udaipur. In 2004, Baksa was created from parts of Barpeta, Nalbari, and Kamrup. In 2006, Samba was derived from parts of Jammu and Kathua. In 2008, Tiruppur district was created from parts of Coimbatore and Erode. In 2004, Udalguri was carved out from Darrang and Sonitpur. In 2006, Ramban was formed from parts of Doda and Udhampur. In 2004, Chirang was created from parts of Barpeta, Bongaigaon, and Kokrajhar. In 2006, Mohali was formed from parts of Patiala and Rupnagar.

[6] Two of the four were split or formed just before the running variable was measured.

Figure 1: Robustness to BWs (R23 sample)



Notes: RDrobust estimates to manually varying bandwidths, on R23's sample.

below the CCT bandwidth. All other bandwidths show economically and statistically significant results.

Finally, if the CCT is sensitive to noise at the cutoff, we could stick to more conventional RD estimation techniques and estimate a parametric RD around the cutoff (Imbens & Lemieux, 2008). In Table 3, I estimate a local linear regression on each side of the cutoff (the slopes vary on either side), and show robustness to the R23 samples, and bandwidths that span both the CCT and IK bandwidths.[7]

Table 3: Parametric RD (and with weights)

	Unweighted		R23 Weights		Unweighted		R23 Weights	
	+4 dist	Full R23	+4 dist	Full R23	+4 dist	Full R23	+4 dist	Full R23
	0.552	0.527	0.617	0.582	0.559	0.500	0.416	0.336
RD	(0.186)***	(0.184)***	(0.181)***	(0.179)***	(0.157)***	(0.155)***	(0.152)***	(0.151)**
Obs	10,672	10,757	10,672	10,757	16,257	16,287	16,257	16,287
BW	0.100	0.100	0.100	0.100	0.150	0.150	0.150	0.150

Notes: Local linear regressions around RD cutoff. The 'R23 Weights' column includes R23's weights.

R23 raises a few other minor points within the same section. First, R23 reports the robust coefficient and standard errors from CCT. As CCT make clear, the robust coefficient is not meant to be reported (in fact, Stata hides it), as it is not an RD estimate. It is only meant as a reference, as the robust SE is not centered around the conventional estimate. Bias correction (perhaps rarely implemented) is also likely to be more sensitive to noise at the cutoff as it fits a higher-order polynomial close to the cutoff.[8] Second, R23 clusters SEs at values of the running variable, without changing the

[7] Many applied papers use CCT simply to compute an optimal bandwidth.

[8] A list of the top cited papers using CCT are here: https://drive.google.com/file/d/1CVz3ZnuFa7qngET0QQd2uIAv0LUqd-P9/view?usp=sharing. Many just use CCT for bandwidth selection.

bandwidth. While CCT does not allow for clustering, the new recently updated packages from the authors do allow for it (for their MSE and CER bandwidths). But the authors make clear that when clustering, we must re-estimate the bandwidths, and so unlike the conventional case, clustering will also change the point estimate (and the sample). In K23, I already show robustness to clustering using the new packages from these authors, and also to using Bartalotti & Brummet (2017) (which allows me to keep the same bandwidths), and finally also to collapsing the data to district-age cells (imposing an intra-cluster correlation=1).[9] Third, R23 introduces sample weights, modifies them, and then applies them to the CCT estimation using the 'weights' command. But the CCT weights command is meant for RD estimation (and these weights are multiplied by the triangular or uniform kernel, weighting observations differently). While I am uncomfortable about imposing weights in RDs in general, in Table 3, I include weights for the parametric RD, and show robustness to samples and bandwidths.

Overall, I conclude my results are robust to R23's sample changes. Below, I address the other points.

3 Other Points

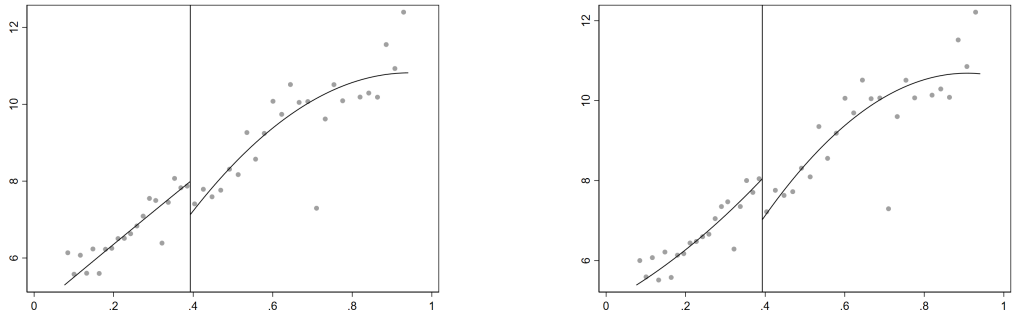
3.1 RD Plots

R23 argues that the RD graphs override Stata defaults affecting the appearance of the discontinuity. I argue K23 made correct choices in this context. R23 (figure 2) shows that when re-scaling the y-axis to include white space above and below, the main part of the graph is obviously shrunk, and it naturally follows that the appearance of a discontinuity is reduced. Then, even though the variation is at the district level, R23 plots the unbinned raw data at the individual level such that there is wide variation. R23 then plots a 4th-order polynomial fit on each side of the discontinuity, which the literature discourages us from doing (Gelman & Imbens, 2018). R23 use Stata defaults without questioning them, but K23 takes careful consideration of the variation before using Stata defaults.

K23 bins plots to better examine the data following clear guidelines from the writers of the RD package (Matias D. Cattaneo & Titiunik, 2019). For instance, they begin with an example of noisy scatters, but say that "A more useful approach is to aggregate or "smooth" the data before plotting," so as to better examine the data variation. They walk through various types of binning, and state that "In sum, bins can be chosen in many different ways. Which method of implementation is most appropriate depends on the researcher's particular goal, for example, illustrating/testing for the overall functional form versus showing the variability of the data." In Figure 2, I increase the number of bins and plot the RD figures, again showing the discontinuity.

[9] I also show nnclustering, rather than clustering at the running variable (as R23 does) given the mass points. As the Causal Inference mix tape says, citing the latest literature: "extensive theoretical and simulation-based evidence that clustering on the running variable is perhaps one of the worst approaches you could take. In fact, clustering on the running variable can actually be substantially worse than heteroskedastic-robust standard errors.....But whatever you do, don't cluster on the running variable, as that is nearly an unambiguously bad idea." https://mixtape.scunning.com/06-regression_discontinuity.

Figure 2: Years of Education for Young (R23 Samples with donuts)



(a) +4 Districts (b) Full R23 Sample

Notes: RDplot with donut hold sample, on R23 sample. Outcome is years of education (R23 recode).

3.2 GE effects

R23's main contentions on the model are that (a) the section is 'somewhat complicated,' and that (b) the GE effects depend on endogenous quantities. On (a), I did my best to explain it to a wide audience, even those unfamiliar with GE models.

I believe that (b) may stem from misunderstandings of the underlying economics.[10] As I say in the paper and emails, the fact that the returns (in general or partial equilibrium) depend on endogenous quantities is, in many ways, the point of the paper.

In K23, when I first define the returns to education, I state: "*This highlights an important fact: the returns to skill are not an exogenous parameter, but rather an endogenous variable that depends on local labor market conditions*" (then I list the various labor market factors, including the aggregate and cohort-specific skill distributions). And say: "*In regions that have relatively more skilled workers, the returns to acquiring skill will be relatively lower, whereas, for regions with more skill-biased capital, the returns are higher.*" And: "*As such, the GE effects are for a given change in the skill distribution.*" Then, as around equations 15 and 16, I show how the GE effects depend on the change in the skill distributions.

Perhaps best explained in my 3/8/22 email to R23: "*I think one point I try to make in my paper is that while most others treat the returns to education as a 'fixed exogenous economic parameter', the point I'm trying to make is that it's an 'endogenous quantity'. So conceptually the returns are actually more like 'the number of workers' rather than a 'risk aversion parameter'. Sorry if this wasn't clear. The idea is that in different settings the quantities of skilled and unskilled labor by cohort are different, and so the endogenous returns are different. And I show what determines why they're different. This means the returns in India in 2007 are different from the returns in India in 2015, and from Mexico in 2007.*

[10] First, I think it important to clarify the distinction between two concepts that R23 conflates: (i) the returns to education in general equilibrium, and (ii) the general equilibrium effect on the returns to education: the latter is quite simply, the difference between the general and partial equilibrium returns. I think, perhaps, R23 may mean 'the returns in GE', sometimes when they say the 'GE effects.'

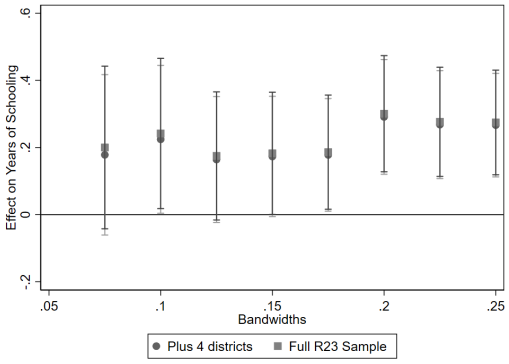
So rather than use the estimates to make decisions on where to spend, I was hoping researchers would try to use the method I outline to estimate returns in other contexts."

I believe R23's 'revised' method for estimating GE effects is problematic.[11] I believe their equation (6) and revised method point 7, may be an inadvertent error, as the β terms should cancel out on each side of the equation. It is always worth discussing alternative methods, but I would encourage the use of the method in K23 for now, as it was developed after numerous discussions with seminar audiences, theses advisers, referees, and editors.[12]

3.3 Diff-in-Diffs and Diff-in-Discontinuities

R23's final point is that they do not evaluate the Diff-in-diffs as it does not rely on exogenous variation. In K23, I show tests of pre-trends. One way to evaluate the diff-in-diff is to restrict it to the bandwidth around the RD cutoff, and perform a difference-in-discontinuities. I include age and district fixed effects, where the treated group is young cohorts in DPEP districts, and show the difference-in-discontinuity results for the R23 sample in Figure 3, for various RD bandwidths (errors at the district level).[13]

Figure 3: Difference-in-Discontinuities



Difference-in-discontinuities on R23 sample. Outcome is years of education (R23 recode). Treated group is young in DPEP districts. Age and district fixed effects included. Errors clustered at the district level.

[11] In K23, the GE effects are cleanly estimated by seeing how the skill premium changes at the RD cutoff, rather than depending on the shares of workers. But it looks like R23 wants to measure the GE effects by measuring returns in partial and general equilibrium, separately, making it depend on what R23 calls 'endogenous quantities.'

[12] From what I understand from R23, they feel hesitant to use K23's method as they think it indirect, and solves for elasticities of substitution (σ_A and σ_E). But, as I make clear in K23, solving for these elasticities is not necessary; it's merely an added bonus of the chosen approach. That is the beauty of working with sufficient statistics: "To estimate returns and the GE effects I will not need to estimate every economic parameter (like σ_A and σ_E)....I only need to identify a combination of economic parameters rather than every primitive. This is less demanding of the data, while allowing for clean identification."

[13] R23's data had deleted certain age groups, so I could not re-evaluate K23's original diff-in-diff.

4 Conclusion

I address all of R23's points, and show that K23's results are robust to R23's new sample, and that certain decisions made by R23 are not appropriate for the context.[14] Putting aside the various modifications R23 make to the sample, I show robustness to R23 using best practice methods, and answer all their ancillary questions. I continue to encourage careful well-founded re-analyses, as there is much to learn about the robustness of salient findings.

[14] R23 concludes their introduction by saying the lack of effects may be expected, as it's not obvious a 17.5-20% spending boost for 5-7 years would generate robustly detectable impacts 10-15 years later. This is perhaps an inadvertent misinterpretation: even though the outcome survey is from 10-15 years later, K23 measures education decisions for those who were school-aged contemporaneous to when the spending occurred. And the initial spending (at that time, the largest donor-led program) was to build schools, which continued to be funded years later. DPEP's impact on education was independently shown in other excellent papers using both the RD (e.g., Sunder (2018); Madhuri Agarwal & Javadekar (2023)) and Diff-in-diff (e.g., Azam & Saing (2016)). The author's goal was to evaluate whether schooling investments were fruitful, and I worry R23's conclusions rely on misunderstandings. Footnote 5 of K23 documents other papers from around the world that may argue otherwise.

Bibliography

Azam, M. & Saing, C. H. (**2016**). "Assessing the Impact of District Primary Education Program in India". *Review of Development Economics*, 21(4): 1113–1131. DOI: 10.1111/rode.12281.

Bartalotti, O. & Brummet, Q. (**2017**). "Regression Discontinuity Designs with Clustered Data". DOI: 10.1108/S0731-905320170000038017.

Gaurav, K. (**2023**). "Large-Scale Education Reform in General Equilibrium: Regression Discontinuity Evidence from India". *Journal of Political Economy*, 131(2). DOI: 10.1086/721619.

Gelman, A. & Imbens, G. (**2018**). "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs". *Journal of Business Economic Statistics*, 37(3): 447–456. DOI: 10.1080/07350015.2017.1366909.

Imbens, G. & Kalyanaraman, K. (**2012**). "Optimal Bandwidth Choice for the Regression Discontinuity Estimator". *The Review of Economic Studies*, 79(3): 933–959. URL: https://www.jstor.org/stable/23261375.

Imbens, G. & Lemieux, T. (**2008**). "Regression Discontinuity Designs: A Guide to Practice". *Journal of Econometrics*, 142(2): 615–635. DOI: 10.1016/j.jeconom.2007.05.001.

Madhuri Agarwal, V. B. & Javadekar, S. (**2023**). "Marrying Young: The Surprising Effect of Education". *SSRN Electronic Journal*. DOI: 10.2139/ssrn.4010142.

Matias D. Cattaneo, N. I. & Titiunik, R. (**2019**). "A Practical Introduction to Regression Discontinuity Designs: Foundations". *Elements in Quantitative and Computational Methods for the Social Sciences*. DOI: 10.1017/9781108684606.

Roodman, D. (**2023**). "Comment: Large-Scale Education Reform in General Equilibrium: Regression Discontinuity Evidence from India". *I4R DP No 70*. URL: https://www.rwi-essen.de/fileadmin/user$_upload/RWI/Publikationen/I4R_Discussion_Paper_Series/070_I4R_Roodman.pdf$.

Sebastian Calonico, M. D. C. & Titiunik, R. (**2014**). "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs". *Econometrica*, 82(6): 2295–2326. DOI: 10.3982/ECTA11757.

Sunder, N. (**2018**). "Parents' Schooling and Intergenerational Human Capital: Evidence from India". URL: https://papers.ssrn.com/sol3/papers.cfm?abstract$_id = 3064665$.

Table 4: RD estimates for Log(Weekly Earnings) – Young Individuals

Panel A	Log(Wages) – R23 definition							
	Original		R23 sans 4 districts			With 4 districts		
RD	0.112	0.144	0.0474	0.0770	0.0747	0.0395	0.0734	0.160
	(0.0312)***	(0.0270)***	(0.0282)*	(0.0245)***	(0.0248)***	(0.0265)	(0.0232)***	(0.0188)***
Obs	10175	14277	10634	14922	14554	10809	15114	24099
BW	CCT	IK	CCT	IK	New BW	CCT	IK	New BW
BW	0.103	0.134	0.103	0.134	0.130	0.103	0.134	0.231

Panel B	Parametric RD							
	Unweighted		Weighted		Unweighted		Weighted	
	+4 dist	R23 sample	+4 dist	R23 sample	+4 dist	R23 sample	+4 dist	R23 sample
RD	0.104	0.0948	0.194	0.174	0.121	0.106	0.176	0.153
	(0.0265)***	(0.0269)***	(0.0244)***	(0.0249)***	(0.0222)***	(0.0225)***	(0.0205)***	(0.0209)***
Obs	10,672	10,757	10,672	10,757	16,257	16,287	16,257	16,287
BW	Para	Para	Para	Para	Para	Para	Para	Para
BW	0.100	0.100	0.100	0.100	0.150	0.150	0.150	0.150

Panel C	Donut hole					
	With 4 districts			Full R23 Sample		
RD	0.115	0.134	0.147	0.104	0.122	0.128
	(0.0286)***	(0.0246)***	(0.0230)***	(0.0289)***	(0.0250)***	(0.0238)***
Obs	10396	14701	17483	10482	14770	15917
BW	CCT	IK	New BW	CCT	IK	New BW
BW	0.103	0.134	0.161	0.103	0.134	0.154

Panel D	All Districts: Uniform Kernel					
	With 4 districts			Full R23 Sample		
RD	0.102	0.106	0.185	0.0933	0.0887	0.0361
	(0.0246)***	(0.0218)***	(0.0189)***	(0.0250)***	(0.0221)***	(0.0288)
Obs	10809	15114	20039	10898	15186	8274
BW	CCT	IK	New BW	CCT	IK	New BW
BW	0.103	0.134	0.182	0.103	0.134	0.0750

Notes: Outcome is R23's recode of the Log(weekly earnings) for the young variable. Estimates include all other R23 sample construction changes.