Stable Adjustments in Regression Discontinuity

A Reply to Albada (*Journal of Comments and Replications in Economics*, 2025)

Andrew Gelman* Guido Imbens †

Journal of Comments and Replications in Economics, Volume 4, 2025-12, DOI: *10.18718/81781.51*

Please Cite As: Gelman, A., and Imbens, G. (2025). Stable Adjustments in Regression Discontinuity. A Reply to Albada (*JCRE*, 2025). *Journal of Comments and Replications in Economics*, Vol.4 (2025-12). DOI: 10.18718/81781.51

*Department of Statistics and Department of Political Science, Columbia University, New York. gelman@stat.columbia.edu
†Graduate School of Business, Stanford University. imbens@stanford.edu

Received January 24, 2026; Published January 28, 2026.
©Author(s) 2025. Licensed under the Creative Commons License - Attribution 4.0 International (CC BY 4.0).

Although it originated in psychology in the 1960s (Thistlethwaite and Campbell, 1960), since the early 2000s regression discontinuity (RD) analysis has become a popular identification strategy in economics, political science, and other fields where there is interest in causal inference from natural experiments. Because the running or forcing variable is often highly correlated with the outcome variable, inference from RD can be highly sensitive to the model to adjust for the running variable. Out of concern for bias due to model misspecification, researchers have sometimes been encouraged to go beyond simple linear adjustments and employ flexible functional forms such as splines and fifth or even higher degree polynomials rather than (Lee, and Thomas Lemieux, 2010). The strategy of using flexible regression models in the adjustment appears conservative—fitting the data rather than assuming linearity—but has paradoxically led to a profusion of implausible but statistically significant findings that are explainable as chance variation. In Gelman, and Zelizer (2015) and Gelman and Imbens (2019), we explored how RD adjusting for high-order polynomials can obtain statistical significance at a high rate even in absence of any true effect. In a new paper, Albada (2025) shows how this problem can arise even with local linear and quadratic regressions. They summarize that such results "are contingent on specific contexts and interpretations."

We agree.

We offer two additional pieces of advice when analyzing data from regression discontinuity designs. First, instead of doing local linear or local polynomial regression, it is possible to directly optimize the weights to minimize worst case bias (Armstrong and Kolesár, 2018, Ghosh et al., 2025). Second, one should remember that these are observational studies and to do the usual recommended thing with observational studies, which is to adjust for important pre-treatment variables.

In good RD studies, the running variable is typically a strong predictor of the outcome, and often chosen for precisely that reason. For example in educational studies of the effect of being held back a grade on subsequent performance, RD analyses may exploit the fact that test scores are used to determine whether students are held back or not. The reason test scores are used as the forcing variable is precisely because they are good predictors of subsequent performance. In such cases, adjusting for the forcing variable can do most of the work. In many bad RD studies, there's no good theoretical or empirical reason to think that the forcing variable will be a strong predictor of the outcome, and it is not generally a good idea to adjusting for only one variable in an uncontrolled observational study.

One problem here is that the literature and the textbook treatments of RD tend to oscillate between specific examples and abstract principles, rarely if ever explaining that the substantive context should inform what to adjust for. In problems where the forcing variable captures all important imbalance, we can choose functional forms to avoid the problem noted in our earlier work, that a noisy adjustment can yield unstable and overconfident estimates of the effect at the discontinuity. And in problem with imbalance in many dimensions, no functional form will work if you are only adjusting for a single variable.

Gelman, and Zelizer (2015), Gelman and Imbens (2019), and Albada (2025) focus on the functional form used for the dependence on the forcing variable in the discontinuity regression, but in many problems it's more important to consider other adjustment variables. One example is in Barfort et al. (2021), discussed by Gelman (2022), where the data were on political candidates, the outcome variable was remaining years of life, and a key predictor—not included in the original published regression analysis!—was current age. The estimated coefficient for age in this example was, unsurprisingly, approximately -1.0, and including this predictor increased the R2 of the regression from 1% to 35%. Predicting remaining lifespan without accounting for age makes little sense but can be understood as an example of what is common practice in the RD literature, adjusting for the forcing variable and nothing else in an RD analysis. We have seen too many regression discontinuity analyses that give implausible answers because the jump at the discontinuity cancels or magnifies a sharp jump in the other direction in the fitted curve. When one looks at the regression discontinuity analyses that work (in the sense of giving answers that make sense), the fitted curve is smooth and the relation between the running variable and the outcome makes substantive sense. It makes sense that previous test scores should be a strong linear predictor of future test scores, which bolsters our confidence in, for example, the conclusions of Chay et al. (2005). In contrast, there is no particular reason to expect any consistent relationship between distance north of a river and life expectancy as modeled by Chen et al. (2013), or between vote margin and remaining years of life as modeled by Barfort et al. (2021). In general we are more skeptical of such analysis in which limited effort was taken to adjust for differences between treatment and control groups.

We have seen lots of poorly conducted regression discontinuity analyses where a common theme was that they adjusted for the running variable (however defined) and nothing else, just following bad practice from the standpoint of observational studies. Those fifth-degree polynomials are particularly troublesome, but we would not want this technical discussion to distract from the larger problem that many analysts seem to operate on the principle that if you can find a discontinuity in treatment assignment, then you can just regress on a corresponding continuous variable and then ignore all the usual concerns and strategies associated with observational studies.

A discontinuity analysis is an observational study. The forcing variable is important, but it is not the only thing in town. The big mistakes seem to come from: (a) unregularized regression on the forcing variable which randomly give you wild jumpy curves that pollute the estimate of the discontinuity, (b) not adjusting for other important pre-treatment predictors, and (c) taking statistically significant estimates and treating them as meaningful, without looking at the model that's been fit.

As with any observational study, if you have a really good pre-treatment predictor, you might be able to get away with just adjusting for that and nothing else, but in general you need to be concerned with balance on all pre-treatment predictors. When there's lack of overlap, the form of the regression function can be important. Albada (2025) makes the valuable point that these issues are not restricted to polynomials. Adjusting for the running variable alone, no matter what the functional form, does not in general correct for imbalance in other variables.

References

Albada, M. R. (2025). "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs. A Comment on Gelman and Imbens (Journal of Business & Economic Statistics, 2019).” *Journal of Comments and Replications in Economics*, vol. 4, no. 11. DOI: 10.18718/81781.50.

Armstrong, T. B., & Kolesár, M. (2018). “Optimal Inference in a Class of Regression Models.” *Econometrica*, vol. 86: 655-683. DOI: 10.3982/ECTA14434.

Barfort, S., Klemmensen, R., & Larsen, E. G. (2021). “Longevity Returns to Political Office.” *Political Science Research and Methods*, vol. 9: 658-664. DOI: 10.1017/psrm.2019.63.

Chay, K. Y., McEwan, P. J., & Urquiola, M. (2005). “The Central Role of Noise in Evaluating Interventions that Use Test Scores to Rank Schools.” *American Economic Review*, vol. 95: 1237-1258. DOI: 10.1257/0002828054825529.

Chen, Y., Ebenstein, A., Greenstone, M., et al. (2013). “Evidence on the Impact of Sustained Exposure to Air Pollution on Life Expectancy from China’s Huai River Policy.” *Proceedings of the National Academy of Sciences* , vol. 110: 12936-12941. DOI: 10.1073/pnas.1300018110.

Gelman, A. (2022). “Criticism as Asynchronous Collaboration: An Example from Social Science Research.” *Stat*, vol. 11: e464. DOI: 10.1002/sta4.464.

Gelman, A., & Imbens, G. (2019). “Why High-order Polynomials should not be used in Regression Discontinuity Designs.” *Journal of Business and Economic Statistics*, vol. 37: 447-456. DOI: 10.1080/07350015.2017.1366909.

Gelman, A., & Zelizer, A. (2015). “Evidence on the Deleterious Impact of Sustained Use of Polynomial Regression on Causal Inference.” *Research and Politics*, vol. 2: 1-7. DOI: 10.1177/2053168015569830.

Ghosh, A., Imbens, G., & Wager, S. (2025). “PLRD: Partially Linear Regression Discontinuity Inference.” URL: https://arxiv.org/abs/2503.09907/.

Lee, D. S., & Lemieux, T. (2010). “Regression Discontinuity Designs in Economics.” *Journal of Economic Literature*, vol. 48: 281-355. DOI: 10.1257/jel.48.2.281.

Thistlethwaite, D. L., & Campbell, D. T. (1960). “Regression-discontinuity Analysis: An Alternative to the ex post facto Experiment.” *Journal of Educational Psychology* , vol. 51: 309-317. DOI: 10.1146/10.1037/h0044319.