

NBER WORKING PAPER SERIES

WHY HIGH-ORDER POLYNOMIALS SHOULD NOT BE USED IN REGRESSION
DISCONTINUITY DESIGNS

Andrew Gelman
Guido Imbens

Working Paper 20405
<http://www.nber.org/papers/w20405>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2014

We thank Jennifer Hill and Joseph Cummins for helpful comments and the National Science Foundation and Institute for Education Sciences for partial support of this work. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at <http://www.nber.org/papers/w20405.ack>

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2014 by Andrew Gelman and Guido Imbens. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Why High-order Polynomials Should not be Used in Regression Discontinuity Designs

Andrew Gelman and Guido Imbens

NBER Working Paper No. 20405

August 2014

JEL No. C01,C1

ABSTRACT

It is common in regression discontinuity analysis to control for high order (third, fourth, or higher) polynomials of the forcing variable. We argue that estimators for causal effects based on such methods can be misleading, and we recommend researchers do not use them, and instead use estimators based on local linear or quadratic polynomials or other smooth functions.

Andrew Gelman
Department of Statistics
and Department of Political Science
Columbia University
New York, NY 10027
gelman@stat.columbia.edu

Guido Imbens
Graduate School of Business
Stanford University
655 Knight Way
Stanford, CA 94305
and NBER
Imbens@stanford.edu

Why High-Order Polynomials Should Not be Used in Regression Discontinuity Designs*

Andrew Gelman[†] Guido Imbens[‡]

2 Aug 2014

Abstract

It is common in regression discontinuity analysis to control for high order (third, fourth, or higher) polynomials of the forcing variable. We argue that estimators for causal effects based on such methods can be misleading, and we recommend researchers do not use them, and instead use estimators based on local linear or quadratic polynomials or other smooth functions.

Keywords: identification, policy analysis, polynomial regression, regression discontinuity, uncertainty

1. Introduction

1.1. Controlling for the forcing variable in regression discontinuity analysis

Causal inference is central to science, and identification strategies in observational studies are central to causal inference in aspects of social and natural sciences when experimentation is not possible. Regression discontinuity designs (RDD) are a longstanding (going back to Thistlewaite and Campbell, 1960), and recently popular, way to get credible causal estimates when applicable. But implementations of regression discontinuity inference vary considerably in the literature, with many researchers controlling for high-degree polynomials of the underlying continuous forcing variable. In this note we make the case that high-order polynomial regressions have poor properties and argue that they should not be used in these settings.

There are three, somewhat related, reasons why we think that high-order polynomial regressions are a poor choice in regression discontinuity analysis:

1. An estimate based on a polynomial regression, with or without trimming, can be interpreted as the difference between a weighted average of the outcomes for the treated and a weighted average for the controls. Given the choice of estimator, the weights depend only on the threshold and the values of the forcing variable, not on the values for the outcomes. One can, and should in applications, inspect these weights. We find that doing so in some applications suggests that the weights implied by higher-order polynomial regressions are not attractive compared to those for local linear regressions.

*We thank Jennifer Hill and Joseph Cummins for helpful comments and the National Science Foundation and Institute for Education Sciences for partial support of this work

[†]Department of Statistics, Columbia University, New York.

[‡]Graduate School of Business, Stanford University, and NBER.

2. Results based on high order polynomial regressions are sensitive to the order of the polynomial. Moreover, we do not have good methods for choosing that order in a way that is optimal for the objective of a good estimator for the causal effect of interest. Often researchers choose the order by optimizing some global goodness of fit measure, but that is not closely related to the research objective of causal inference.
3. Inference based on high-order polynomials is often poor. Specifically, confidence intervals based on such regressions, taking them as accurate approximations to the regression function, are often misleading. Even if there is no discontinuity in the regression function, high-order polynomial regressions often lead to confidence intervals that fail to include zero with probability substantially higher than the nominal Type 1 error rate.

Based on these arguments we recommend that researchers not use such methods, and instead control for local linear or quadratic polynomials or other smooth functions.

1.2. Theoretical framework

Regression discontinuity analysis has enjoyed a renaissance in social science, especially in economics; Imbens and Lemieux (2008), Van Der Klaauw (2013), Lee and Lemieux (2010) and DiNardo and Lee (2010) provide recent reviews.

Regression discontinuity analyses is used to estimate the causal effect of a binary treatment on some outcome. Let $(Y_i(0), Y_i(1))$ denoted the pair of potential outcomes for unit i , and let $W_i \in \{0, 1\}$ denote the treatment. The realized outcome is $Y_i^{\text{obs}} = Y_i(W_i)$. Although the same issues arise in fuzzy regression discontinuity designs, for ease of exposition we focus on the sharp case where the treatment is a deterministic function of a pretreatment forcing variable X_i :

$$W_i = \mathbf{1}_{X_i \geq 0}.$$

Define

$$\tau(x) = \mathbb{E}[Y_i(1) - Y_i(0)|X_i = x].$$

Regression discontinuity methods focus on estimating the average effect of the treatment at the threshold (equal to zero here):

$$\tau = \tau(0).$$

Under some conditions, mainly smoothness of the conditional expectations of the potential outcomes as a function of the forcing variable, this average effect can be estimated as the discontinuity in the conditional expectation of Y_i^{obs} as a function of the forcing variable, at the threshold:

$$\tau = \lim_{x \downarrow 0} \mathbb{E}[Y_i|X_i = x] - \lim_{x \uparrow 0} \mathbb{E}[Y_i|X_i = x].$$

The question is how to estimate the two limits of the regression function at the threshold:

$$\mu_+ = \lim_{x \downarrow 0} \mathbb{E}[Y_i|X_i = x], \quad \text{and} \quad \mu_- = \lim_{x \uparrow 0} \mathbb{E}[Y_i|X_i = x].$$

We focus in this note on two approaches researchers have commonly taken to estimating μ_+ and μ_- . Typically researchers are not confident that the two conditional means $\mu_+(x) =$

$\mathbb{E}[Y_i^{\text{obs}}|X_i = x, x > 0]$ and $\mu_-(x) = \mathbb{E}[Y_i^{\text{obs}}|X_i = x, x < 0]$ can be well approximated by a global linear function. One approach researchers have taken is to use a global high-order polynomial approach. In this approach, researchers choose some integer K , and estimate the regression function,

$$Y_i^{\text{obs}} = \sum_{k=0}^K X_i^k \beta_{+j} + \varepsilon_{+i},$$

separately on the units with values for $X_i \geq 0$ and units with values for $X_i < 0$. The discontinuity in the value of the regression function at zero is then estimated as,

$$\hat{\tau} = \hat{\mu}_+ - \hat{\mu}_- = \hat{\beta}_{+0} - \hat{\beta}_{-0} - .$$

In practice, researchers often use cubic or higher order polynomials, often using statistical information criteria or cross-validation to determine choose the degree K of the polynomial.

The second commonly-used approach is local linear or sometimes local quadratic approximation. In that case researchers discard the units with X_i more than some bandwidth h away from the threshold and estimate a linear or quadratic function on the remaining units; see Hahn, Todd, and Van Der Klaauw (2001) and Porter (2003). Imbens and Kalyanaraman (2012) suggest a data driven way for choosing the bandwidth in connection with a local linear specification. Calonico, Cattaneo, and Titiunik (2014) suggest using a quadratic specification.

The main point of the current paper is that we think the approach based on high order global polynomial approximations should not be used, and that instead inference based on local low order polynomials is to be preferred. In the next three sections we discuss three arguments in support of this position and illustrate these in the context of some applications.

2. Issue 1: Noisy weights

Our first argument against using global high-order polynomial methods focuses on the interpretation of linear estimators for the causal estimand as weighting averages. More precisely, these estimators can be written as the difference between the weighted averages of the outcomes for the treated and controls, with the weights a function of the forcing variable. We show that for global polynomial methods these weights can have unattractive properties.

2.1. The weighted average representation of polynomial regressions

The starting point is that polynomial regressions, whether global or local, lead to estimators for μ_+ and μ_- that can be written as weighted averages. Focusing on $\hat{\mu}_+$, the estimator for μ_+ , we can write $\hat{\mu}_+$ as a weighted average of outcomes for units with $X_i \geq 0$:

$$\hat{\mu}_+ = \frac{1}{N_+} \sum_{i:X_i \geq 0} \omega_i Y_i^{\text{obs}},$$

where the weights ω_i have been normalized to have a mean of 1 over all units with a value of the forcing variable exceeding zero. The weights are an estimator-specific function of the

full set of values X_1, \dots, X_N for the forcing variable that does not depend on the outcome values $Y_1^{\text{obs}}, \dots, Y_N^{\text{obs}}$. Hence we can write the weights as

$$\omega_i = \omega(X_1, \dots, X_N).$$

The various estimators differ in the way the weights depend on value of the forcing variable. Moreover, we can inspect, for a given estimator, the functional form for the weights. Suppose we estimate a K -th order polynomial approximation using all units with X_i less than the bandwidth h (where h can be ∞ so that this includes global polynomial regressions). Then the weight for unit i in the estimation of μ_+ , $\hat{\mu}_+ = \sum_{i:X_i \geq 0} \omega_i Y_i^{\text{obs}} / N_+$, is

$$\omega_i = \mathbf{1}_{0 \leq X_i < h} \cdot e'_{K1} \left(\sum_{j:0 \leq X_j < h} \begin{pmatrix} 1 & X_j & \dots & X_j^K \\ X_j & X_j^2 & \dots & X_j^{K+1} \\ \vdots & \ddots & & \vdots \\ X_j^K & X_j^{K+1} & \dots & X_j^{2K} \end{pmatrix} \right)^{-1} \begin{pmatrix} 1 \\ X_i \\ \vdots \\ X_i^K \end{pmatrix},$$

where e_{K1} is the K -component column vector with all elements other than the first equal to 0, and the first element equal to 1.

There are two important features of these weights. First, the values of the weights have nothing to do with the actual shape of the regression function, whether it is constant, linear, or anything else. Second, one can inspect these weights based on the values of the forcing variable in the sample, and compare them for different estimators. In particular we can compare, before seeing the outcome data, the weights for different values of the bandwidth h and the order of the polynomial K .

2.2. Example: Matsudaira data

To illustrate, we inspect the weights for various estimators for an analysis by Matsudaira (2008) of the effect of a remedial summer program on subsequent academic achievement. Students were required to participate in the summer program if they score below a threshold on either a mathematics or a reading test, although not all students did so, making this a fuzzy regression discontinuity design. We focus here on the discontinuity in the outcome variable, which can be interpreted as an intention-to-treat estimate. There are 68,798 students in the sample. The forcing variable is the minimum of the mathematics and reading test scores normalized so that the threshold equals 0. The outcome we look at here is the subsequent mathematics score. There are 22,892 students with the minimum of the test scores below the threshold, and 45,906 with a test score above.

In this section we discuss estimation of μ_+ only. Estimation of μ_- raises the same issues. We look at weights for various estimators, without using the outcome data. First we consider global polynomials up to sixth degree. Next we consider local linear methods. The bandwidth for the local linear regression is 27.6, calculated using the Imbens and Kalyanaraman (2012) bandwidth selector. This leaves 22,892 individuals whose value for the forcing variable is positive and less than 27.6, out of the 45,906 with positive values for the forcing variable. We estimate the local linear regression using a triangular kernel.

Figure 1 and Table 2.2 present some of the results relevant for the discussion on the weights. Figure 1a gives the weights for the six global polynomial regressions. Figure 1b

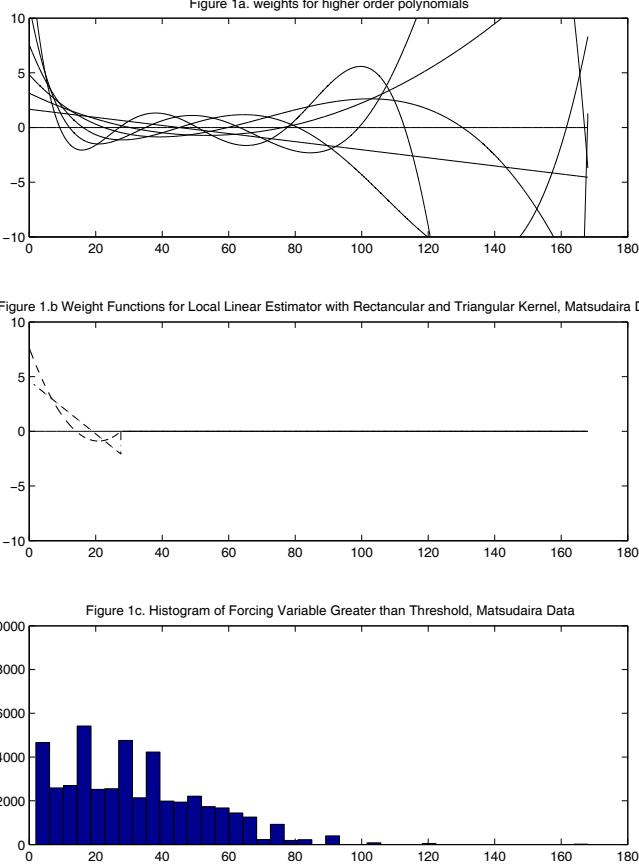


Figure 1: WEIGHTS, AS A FUNCTION OF THE FORCING VARIABLE X , FOR PERSONS IN THE MATSUDAIRA DATASET WITH POSITIVE VALUES OF X (A) WEIGHTS FOR FITTED POLYNOMIAL REGRESSIONS FROM FIRST TO SIXTH-ORDER: THE WEIGHTS OSCILLATE UNCOMFORTABLY AND ARE HIGHLY VARIABLE; (B) WEIGHTS FOR LOCAL LINEAR ESTIMATORS ARE MUCH BETTER BEHAVED; (C) HISTOGRAM OF THE POSITIVE VALUES OF X . HIGHER-ORDER REGRESSIONS CAN GIVE HUGE WEIGHTS TO POINTS THAT ARE FAR FROM THE DISCONTINUITY, THUS CREATING HIGHLY NOISY ESTIMATES OF THE CAUSAL ESTIMAND OF INTEREST.

gives the weights for the local linear regression with rectangular and triangular kernels, and the bandwidth equal to 27.6. Figure 1c presents a histogram of the distribution of the forcing variable for individuals with a value for the forcing variable greater than 0. In Table 1 2.2 we present the weight, under each of the six polynomial regression specifications, for the individuals with the largest value for the forcing variable, $X_i = 168$. Because this extreme value is outside the bandwidth, the weight for the local linear regression for individuals with such a value for X_i would be 0. The weights for all the points included in the regression have been normalized so that their average value is 1.

Figure 1a shows that the weight for the individuals with the largest values for the forcing variable are high in absolute value and highly sensitive to the order of the polynomial. It is difficult to imagine that, based on these figures, one would be comfortable with any of these six specifications. Figure 1b shows the weights for the local linear regression, which appear

Order of global polynomial	Normalized weight for the persons with the largest value of the forcing variable
1	-4.5
2	17.1
3	-16.3
4	8.3
5	-3.7
6	1.3

Table 1: FOR DIFFERENT ORDERS OF THE GLOBAL POLYNOMIAL: NORMALIZED WEIGHT FOR THE PERSONS WITH THE LARGEST VALUE OF THE FORCING VARIABLE IN THE MAT-SUDAIRA DATA. THE WEIGHTS ARE HIGHLY VARIABLE AND ALL ARE MUCH LARGER IN ABSOLUTE VALUE THAN THE AVERAGE WEIGHT OF 1. CONSIDERING THAT THIS POINT IS FAR AWAY FROM THE DISCONTINUITY, THIS SUGGESTS THAT THESE POLYNOMIAL REGRESSIONS ARE ASSIGNING MUCH TOO HIGH A WEIGHT TO POINTS IN THE DATA THAT ARE ESSENTIALLY IRRELEVANT TO THE RESEARCH QUESTION. IT WOULD BE MORE DESIRABLE FOR THE WEIGHT OF SUCH AN EXTREME POINT TO BE CLOSE TO ZERO.

more attractive: most of the weight goes to the individuals with values for X_i close to the threshold, and individuals with $X_i > 27.6$ have weights of 0.

Table 1 also shows the unattractiveness of the high order polynomial regressions. Whereas one would like to give little or zero weight to the individuals with extreme values for X_i , the global polynomial regressions attach large weights, sometimes positive, sometimes negative, to these individuals, and generally larger than the average weight of 1, whereas the local linear estimator attaches zero weight to these individuals.

2.3. General recommendation

Most, if not all, estimators for average treatment effects used in practice can be written as the difference between two weighted averages, one for the treated units and one for the control units. This includes estimators in regression discontinuity settings. In those cases it is useful to inspect the weights in the weighted average expression for the estimators to assess whether some units receive excessive weight in the estimators.

3. Issue 2: Estimates that are highly sensitive to the degree of the polynomial

The second argument against the high order global polynomial regressions is their sensitivity to the order of the polynomial. We illustrate that here using two applications of regression discontinuity designs.

	Order of polynomial	estimate (se)
global	1	-0.167 (0.008)
global	2	0.079 (0.010)
global	3	0.112 (0.011)
global	4	0.077 (0.013)
global	5	0.069 (0.016)
global	6	0.104 (0.018)
local	1	0.080 (0.012)
local	2	0.063 (0.017)

Table 2: FROM THE MATSUDAIRA DATA: ESTIMATES OF EFFECT OF SUMMER SCHOOL REQUIREMENT USING DIFFERENT REGRESSION DISCONTINUITY ESTIMATES.

3.1. Example: Matsudaira data (continued)

We return to the Matsudaira data. Here we use the outcome data and directly estimate the effect of the treatment on the outcome for units close to the threshold. To simplify the exposition, we look at the effect of being required to attend summer school, rather than actual attendance, analyzing the data as a sharp rather than a fuzzy regression discontinuity design. We consider global polynomials up to order six and local polynomials up to order two. The bandwidth is 27.6 for the local polynomial estimators, based on the Imbens-Kalyanaraman bandwidth selector, leaving 37,580 in the sample. Local regression is with a triangular kernel.

Table 2 displays the point estimates and standard errors. The variation in the global polynomial estimates over the six specifications is much bigger than the standard error for any of these six estimates. That the standard errors do not capture the full amount of uncertainty. The estimates based on third, fourth, fifth, and sixth order global polynomials range from 0.069 to 0.112, whereas the range for the local linear and quadratic estimates is 0.063 to 0.080, substantially narrower.

3.2. Example: Jacob-Lefgren data

Table 3 reports the corresponding estimates for a second dataset. Here the interest is again in the causal effect of a summer school program. The data were previously analyzed by Jacob and Lefgren (2004). There are observations on 70,831 students. The forcing variable is the minimum of a mathematics and reading test. Out of the 70,831 students, 29,900 score below the threshold on at least one of the tests, and so are required to participate in the summer program. The Imbens-Kalyanaraman bandwidth here is 0.57. As a result the local polynomial estimators are based on 31,747 individuals out of the full sample of 70,831, with 16,011 required and 15,736 not required to participate in summer school.

Again the estimates based on the global polynomials have a wider range than the local linear and quadratic estimates. In addition the range for the global polynomial estimates is again large compared to the standard errors.

	Order of polynomial	estimate (se)
global	1	0.024 (0.008)
global	2	0.176 (0.012)
global	3	0.209 (0.015)
global	4	0.174 (0.018)
global	5	0.164 (0.021)
global	6	0.197 (0.025)
local	1	0.196 (0.018)
local	2	0.176 (0.027)

Table 3: FROM THE JACOB-LEFGREN DATA: ESTIMATES OF THE EFFECT OF SUMMER SCHOOL REQUIREMENT USING DIFFERENT REGRESSION DISCONTINUITY ESTIMATES.

Predictor	Linear regression estimate (se)	6th order polynomial estimate (se)
intercept	5.0 (0.1)	3.6 (0.3)
X	7.1 (0.1)	27.6 (2.7)
X^2		-54 (11)
X^3		59 (18)
X^4		-32 (13)
X^5		8 (5)
X^6		0.9 (0.6)

Table 4: FROM THE LALONDE DATA: PARAMETER ESTIMATES FOR LINEAR AND SIXTH-ORDER POLYNOMIAL REGRESSION OF EARNINGS IN 1978 ON THE AVERAGE OF EARNINGS IN 1974 AND 1975.

4. Issue 3: Inferences that do not achieve nominal coverage

The third point we wish to make is that conventional inference for treatment effects in regression discontinuity settings can be misleading, in the sense that that confidence intervals are too narrow. We make that point by constructing confidence intervals for discontinuities in a setting where we expect none to be present.

We look at a dataset containing information on yearly earnings in 1974, 1975, and 1978 for 15,992 individuals for whom there is information from the Current Population Survey. These data were previously used for different purposes in work by Lalonde (1986) and Dehejia and Wahba (1999). We look at the conditional expectation of earnings in 1978 (the outcome Y_i) given the average of earnings in 1974 and 1975 (with the predictor X_i divided by 10 for convenience so that the coefficients of the higher powers are on a reasonable scale). Figure 2 gives a simple, histogram-based estimate of the conditional expectation. Unsurprisingly, the conditional expectation looks fairly smooth and increasing. Overlaid with the histogram estimator are a linear approximation and a sixth order polynomial approximation. The linear approximation is fairly accurate, and so is the sixth order polynomial approximation. We

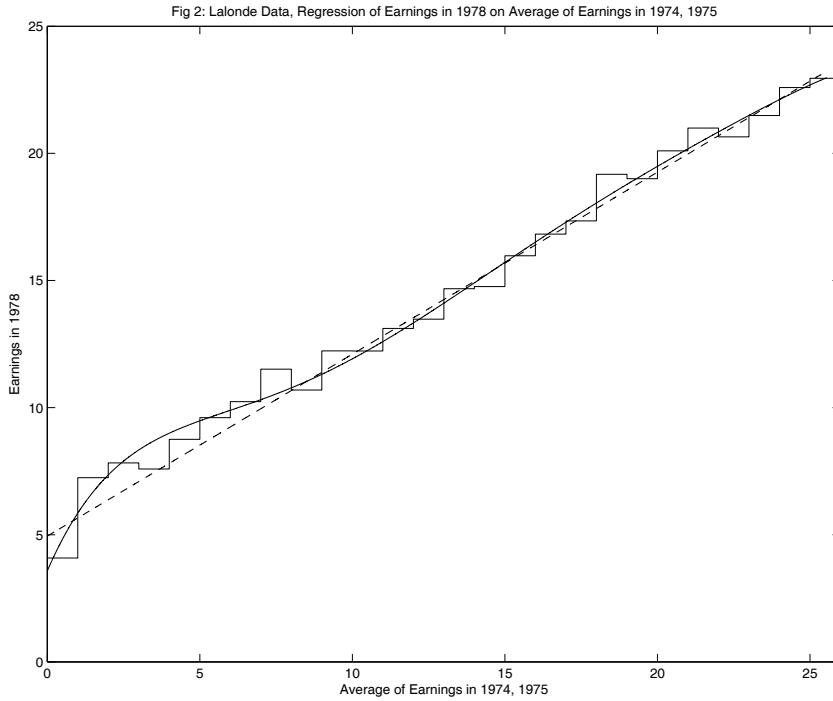


Figure 2: FROM THE LALONDE DATA: EARNINGS IN 1978 VS. AVERAGE EARNINGS IN 1974–1975. WE WILL USE THESE DATA TO FIT A REGRESSION DISCONTINUITY ESTIMATE IN A CASE WHERE THERE IS NO ACTUAL DISCONTINUITY.

present the parameter estimates for those two regression functions in Table 4.

Now suppose we pretend the median of the average of earnings in 1974 and 1975 (equal to 14.65) was the threshold, and we estimate the discontinuity in the conditional expectation of earnings in 1978. We should expect to find an estimate close to zero. The estimates and standard errors for polynomials of different degree appear in Table 5.

All estimates are in fact reasonably close to zero, with the nominal 95% confidence interval in all cases including zero, though this does not hold for the nominal 90% confidence interval.

The question arises whether this is typical. To assess this we do the following exercise. 5,000 times we randomly pick a single point from the empirical distribution of X_i that will serve as a pseudo threshold. We exclude values for the threshold less than 1 and greater than 25, to ensure a sufficient number of observations on both sides of the threshold.¹ We pretend this randomly drawn value of X_i is the threshold in a regression discontinuity design analysis. In each of the 5,000 replications we then estimate the average effect of the pseudo treatment, its standard error, and check whether the implied 95% confidence interval excludes zero. There is no reason to expect a discontinuity in this conditional expectation at this threshold, and so we would like to see that only 5% of the times we randomly pick a threshold the corresponding confidence interval should not include zero.

We compare the eight different estimates of the pseudo treatment effect we used in the previous section. The first six are global polynomial regressions of order ranging from 1 to

¹Out of the sample of 15,992 individuals there are 2044 individuals with values for X less than 1 and 2747 individuals with values for X greater than 25.

	Order of polynomial	estimate (se)
global	1	-0.02 (0.25)
global	2	0.71 (0.25)
global	3	-0.98 (0.53)
global	4	0.24 (0.66)
global	5	-1.22 (0.80)
global	6	-0.13 (0.93)
local	1	0.02 (0.37)
local	2	-0.39 (0.55)

Table 5: ESTIMATES OF EFFECT OF A PSEUDO TREATMENT: SINGLE REPLICATION ON LALONDE DATA WITH PSEUDO THRESHOLD EQUAL TO 14.65.

	Order of polynomial	Rejection rate	Median standard error
global	1	0.49	0.29
global	2	0.34	0.45
global	3	0.21	0.61
global	4	0.14	0.77
global	5	0.11	0.93
global	6	0.11	1.09
local	1	0.07	0.46
local	2	0.07	0.67

Table 6: REJECTION RATES FOR NOMINAL 5% TEST IN A SIMULATION EXPERIMENT IN WHICH THE TRUE DISCONTINUITY EFFECT IS ZERO. THE OVER-REJECTION OF THE GLOBAL POLYNOMIAL REGRESSIONS SUGGEST THAT THEY ARE NOT ACCURATE ENOUGH TO ALLOW THE RESEARCHER TO IGNORE THE BIAS IN THE ESTIMATES OF THE TREATMENT EFFECTS.

6. If, say, the regression functions on both sides of the threshold are truly linear, than the estimator based on linearity should be approximately unbiased for the average treatment effect (which is zero here), and the corresponding 95% confidence interval should include zero 95% of the time. If, on the other hand, the regression function is not truly linear, the confidence intervals based on linearity are likely to include the true value of zero less than 95% of the time. The last two estimators are based on local linear and local quadratic regressions, where we drop observations with values of X more than h away from the threshold (where the distance h is chosen using the Imbens-Kalyanaram bandwidth procedure).²

We present the results for the rejection rates and median standard errors in Table 6. The estimates based on the global polynomial regressions perform poorly, making our point that inference for estimates based on global polynomial regressions is not reliable. For low order

²In 50,000 replications the median bandwidth was 5.7.

polynomials we reject the null of no effect for a large fraction of the replications. For high order global polynomials we still over-reject, and in addition the standard errors tend to be large compared to those for the local polynomial estimators. The over-rejection suggests that the global polynomial approximations are not accurate enough to allow the researcher to ignore the bias in the estimates of the treatment effects. The local linear and quadratic regressions work substantially better in that the rejection rates are close to nominal levels, and the standard errors are substantially smaller than those based on the high order global polynomial approximations.

5. Discussion

Regression discontinuity designs have become very popular in social sciences in the last twenty years. One implementation relies on using high-order polynomial approximations to the conditional expectation of the outcome given the forcing variable. In this paper we recommend against using this method. We present three arguments for this position: the implicit weights for high order polynomial approximations are not attractive, the results are sensitive to the order of the polynomial approximation, and conventional inference has poor properties in these settings. These issues are complementary, in that the noisiness of the implicit weights explains how the global polynomial regressions can have poor coverage and wide confidence intervals at the same time. In addition we recommend that researchers routine present the implicit weights arising from regression estimates of causal estimands.

6. References

- Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. Technical report, Department of Economics, University of Michigan.
- Dehejia, R., and Wahba, S. (1999). Causal effects in non-experimental studies: re-evaluating the evaluation of training programs. *Journal of the American Statistical Association* **94**, 1053–1062.
- DiNardo, J., and D. Lee. (2010). Program evaluation and research designs. In Ashenfelter and Card (eds.), *Handbook of Labor Economics*, Vol. 4.
- Hahn, J., Todd, P., and Van Der Klaauw, W. (2001). Regression discontinuity, *Econometrica* **69**, 201–209.
- Imbens, G., and K. Kalyanaraman. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies* **79**, 933–959.
- Imbens, G., and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics* **142**, 615–635.
- Jacob, B., and Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity design. *Review of Economics and Statistics* **86**, 226–244.
- LaLonde, R. J. (1986). Evaluating the econometric evaluations of training programs using experimental data. *American Economic Review* **76**, 604–620.

- Lee, D. S., and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature* **48**, 281-355.
- Matsudaira, J, (2008). Mandatory summer school and student achievement. *Journal of Econometrics* **142**, 829–850.
- Porter, J. (2003). Estimation in the regression discontinuity model. Working Paper, Harvard University Department of Economics.
- Thistlewaite, D., and Campbell, D. (1960). Regression-discontinuity analysis: An alternative to the ex-post facto experiment. *Journal of Educational Psychology* **51**, 309–317.
- Van Der Klaauw, W. (2008). Regression-discontinuity analysis: A survey of recent developments in economics. *Labour* **22**, 219–245.