Why high-order polynomials should not be used in regression discontinuity designs*

Andrew Gelman† Guido Imbens‡

September 6, 2016

Abstract

It is common in regression discontinuity analysis to control for third, fourth, or higher-degree polynomials of the forcing variable. There appears to be a perception that such methods are theoretically justified, even though they can lead to evidently nonsensical results. We argue that controlling for global high-order polynomials in regression discontinuity analysis is a flawed approach with three major problems: it leads to noisy estimates, sensitivity to the degree of the polynomial, and poor coverage of confidence intervals. We recommend researchers instead use estimators based on local linear or quadratic polynomials or other smooth functions.

Keywords: causal 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 are a longstanding (going back to Thistlewaite and Campbell, 1960), and recently increasingly 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

*We thank Jennifer Hill, Joseph Cummins, and Karthik Rajkumar for helpful comments.
†Department of Statistics and Department of Political Science, Columbia University, New York.
‡Graduate School of Business, Stanford University, and NBER.

polynomials of the underlying continuous forcing variable. In this note we make the case that global 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 analyses:

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 minus a weighted average of the outcomes 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 can take on extreme, *prima facie* unattractive, values, relative to the weights based on local linear or quadratic regressions.

2. Results for the causal effects of interest based on global 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 (e.g., the methods suggested by Fan and Gijbels,1996), 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 the polynomial as an accurate approximation to the conditional expectation, are often misleading. Even if there is no discontinuity in the regression function, high-order polynomial regressions can lead to confidence intervals that fail to include zero with probability higher than the nominal Type 1 error rate.

Based on these arguments we recommend 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, as a credible way of estimating causal effects in settings where unconfoundedness is not plausible; Imbens and Lemieux (2008), Van Der Klaauw (2013), Lee and Lemieux (2010), DiNardo and Lee (2010), and Skovron and Titiunik (2015) provide recent reviews.

Regression discontinuity analyses are used to estimate the causal effect of a binary treatment on some outcome. Using the potential outcome approach (e.g., Imbens and Rubin, 2015), let $(y_i(0), y_i(1))$ denote 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 received is a monotone, deterministic, function of a pretreatment variable x_i, the forcing variable, with the threshold equal to zero:

$$w_i = \mathbf{1}_{x_i \geq 0}.$$

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:

$$\tau = \lim_{x \downarrow 0} \mathbb{E}(y_i^{\text{obs}} | x_i = x) - \lim_{x \uparrow 0} \mathbb{E}(y_i^{\text{obs}} | x_i = x).$$

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

$$\mu_+ = \lim_{x \downarrow 0} \mathbb{E}(y_i^{\text{obs}} | x_i = x), \qquad \text{and} \quad \mu_- = \lim_{x \uparrow 0} \mathbb{E}(y_i^{\text{obs}} | x_i = x).$$

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. Lee and Lemieux (2008) write: "From an applied perspective, a simple way of relaxing the linearity assumption is to include polynomial functions of x in the regression model." In this approach, researchers choose some integer K, possibly in a data-dependent way, and estimate the regression function,

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

on the N_+ units with values $x_i \geq 0$ and repeat the same procedure using the N_- units with values $x_i < 0$. 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 up to fifth or sixth order polynomials, often using statistical information criteria or cross-validation to 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 an algorithm for a data dependent bandwidth with a quadratic specification for the regression function.

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 (local linear or local quadratic) 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. We should note that these are not formal results. If a researcher is confident that the conditional means can be described with sufficient accuracy by a fifth order polynomial, than that would be a perfectly sensible method to use. However, in practice it is unlikely that a researcher is confident about this, and the approximation results available for polynomial approximations do not imply that in practical settings these methods will lead to reasonable results. We will attempt to make the case that in fact, they do not, and that local methods do better in empirically relevant settings.

2. Argument 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 weighted 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. This is true for both global and local polynomial methods, and we can therefore base comparisons of these methods on the form and values of these weights. We show that for global polynomial methods these weights can have unattractive properties. This is related to what is known in the approximation literature as Runge's phenomenon, that given a set of N pairs (x_i, y_i) on a compact interval $[a, b]$, the $N-1$th order polynomial that goes through all the pairs becomes increasingly erratic, as the number of points increases, close to the boundary of the interval, especially when there are relatively few points close to the boundary (Dahlquist and Bjork, 1974). See also Calonico, Cattaneo, and Titiunik, R. (2015).

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 $x_i \geq 0$:

$$\hat{\mu}_+ = \frac{1}{N_+} \sum_{i:x_i \geq 0} w_i \, y_i^{\text{obs}},$$

where the weights w_i have been normalized to have a mean of 1 over all N_+ units with a value of the forcing variable exceeding the threshold. The weights are an estimator-specific function of the full set of values $x_1, \ldots, x_N$ for the forcing variable that does not depend on the outcome values $y_1^{\text{obs}}, \ldots, y_N^{\text{obs}}$. Hence we can write the weights as

$$(w_1, \ldots, w_n) = w(x_1, \ldots, x_N).$$

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_+, $\hat{\mu}_+ = \sum_{i:x_i \geq 0} w_i \, y_i^{\text{obs}}/N_+$, is

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

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

There are two important features of these weights. First, the values of the weights have nothing to do with the actual shape of the conditional expectation 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 first 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. Its range is $[-199, 168]$. The outcome we look at here is the subsequent

Order of global polynomial	Normalized weight for individual with $x_i = 168$
1	-4.5
2	17.1
3	-16.3
4	8.3
5	-3.7
6	1.3

Table 1: Normalized weight for individuals with $x_i = 168$ for different orders of global polynomial, compared to average weight of 1.

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. First, we consider global polynomials up to sixth-degree. Second, 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.

Figures 1a–c and Table 1 present some of the results relevant for the discussion on the weights. Figure 1a gives the weights for the six global polynomial regressions, as a function of the forcing variable. Figure 1b gives the weights for the local linear regression with rectangular and triangular kernels, and the bandwidth equal to 27.6, again as a function of the forcing variable. 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 we present the weights for the individuals with the largest value for the forcing variable, $x_i = 168$, for the six polynomial regression specifications. Because this extreme value of 168 is outside the bandwidth, the weight for the local linear regression for individuals with such a value for x_i would be 0. Recall that the average value of the weights is 1 for individuals with a value of the forcing variable exceeding zero.

Figure 1a shows that the weight for the individuals with large values for the forcing variable are quite sensitive to the order of the polynomial. Based on these figures, we would not be comfortable with any of these six specifications. Figure 1b shows the weights for the local linear regression, which appear 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 (by definition of the threshold).

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 often substantially larger than the average weight of one, whereas the local linear estimator attaches zero weight to these individuals.

2.3. Jacob-Lefgren data

In Figures 2a–c, we repeat these analyses for another dataset. Here the interest is also 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, with the range equal to $[-0.9, 6.8]$. Out of the 70,831 students, 29,900 score below the threshold of 2.75 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 weights for the individuals with large values for the forcing variable are quite sensitive to the order of the polynomial.

2.4. Lee data

In Figures 3a–c, we repeat these analysis for a third dataset, previously analyzed by Lee (2008). Lee analyzes the effect of one party winning an election on the voting shares in the next election, using data from congressional district elections. The Imbens-Kalyanaraman bandwidth here is 0.28. There are 3818 elections where the Democrats won, and 2740 where the Republicans won. Again we find that the weights far from the threshold can be quite sensitive to the order of the polynomial chosen to approximate the regression function. An additional feature of these data is that there are a fair number of elections that are uncontested, which clearly should have low weight in estimating the effect for close elections.

2.5. 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. Argument 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 three applications of regression discontinuity designs.

3.1. Matsudaira data

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 individuals in the sample.

	Order of polyn.	Matsudaira est.	(s.e.)	Jacob-Lefgren est.	(s.e.)	Lee est.	(s.e.)
global	1	−0.167	(0.008)	−0.024	(0.009)	0.118	(0.006)
global	2	0.079	(0.010)	−0.176	(0.012)	0.052	(0.007)
global	3	0.112	(0.011)	−0.209	(0.015)	0.112	(0.009)
global	4	0.077	(0.013)	−0.174	(0.018)	0.077	(0.011)
global	5	0.069	(0.016)	−0.164	(0.021)	0.043	(0.013)
global	6	0.104	(0.018)	−0.197	(0.025)	0.067	(0.015)
std		[0.025]		[0.024]		[0.017]	
local	1	0.080	(0.012)	−0.196	(0.018)	0.080	(0.008)
local	2	0.063	(0.017)	−0.176	(0.027)	0.067	(0.012)
std		[0.012]		[0.014]		[0.009]	

Table 2: Sensitivity of estimates to the order of the polynomial. The table reports estimates of the magnitude of the discontinuity in the conditional expectation of the outcome as a function of the forcing variable at the threshold.

Local linear or quadratic regression is based on a triangular kernel.

The first two numerical columns in Table 2 display 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, suggesting that the standard errors do not capture the full amount of uncertainty about the causal effects of interest. 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.

For the Matsudaira data we also present in Figures 4a and 4b the estimated regression functions based on the various specifications. From those figures there appears to be relatively little difference between the estimated regression functions over most of the range of values where the observations are. Nevertheless, these small differences matter for the estimated difference in the two regression functions.

3.2. Jacob-Lefgren and Lee data

We repeat these analyses for the Jacob-Lefgren and Lee data sets. The second pair of numerical columns in Table 2 reports the corresponding estimates for the Jacob-Lefgren dataset. Again the estimates based on the global polynomials have a wider range than the local linear and quadratic estimates. The third pair of numerical columns in Table 2 reports the corresponding estimates for the Lee congressional election dataset. Here the estimated effect based on a third order polynomial is 0.112, almost three times that based on a fifth order polynomial, 0.43. The local linear and local quadratic estimates are substantially closer, 0.080 and 0.067.

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

Our third point is that conventional inference for treatment effects in regression discontinuity settings can be misleading, in the sense that that conventional confidence intervals have lower than nominal coverage. We make that point by constructing confidence intervals for discontinuities in an artificial setting where we expect no discontinuities to be present.

We illustrate this point with two different datasets. The first contains 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 in thousands of dollars (the outcome y_i) given the average of earnings in 1974 and 1975 (the predictor x_i, in tens of thousands of dollars so that the coefficients of the higher powers are on a reasonable scale). Figure 5a gives a simple, histogram-based estimate of the conditional expectation, with a histogram of the marginal distribution of the conditioning variable in Figure 5b. Unsurprisingly, the conditional expectation looks fairly smooth and increasing. Overlaid with the histogram estimator are a first to sixth order polynomial approximations, with all approximations other than the sixth order one in dashes, and the sixth order one in a solid line. All approximations appear fairly accurate.

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 would expect to find an estimate close to zero. Doing so, for global and local polynomials of different degree, we find the estimates in Table 3. All estimates are in fact reasonably close to zero, with the nominal 95% confidence interval in most cases including zero. This exercise on its own is not particularly informative, because the estimates based on the different specifications are highly correlated. However, in the next step we assess whether the coverage found for this single case is typical. We do the following exercise. 20,000 times we randomly pick a single point from the empirical distribution of x_i between the 0.25 and 0.75 quantile that will serve as a pseudo threshold. We pretend this randomly drawn value

	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 3: Estimates of effect of pseudo treatment: Single replication on Lalonde data with pseudo threshold equal to 14.65.

of x_i is the threshold in a regression discontinuity design analysis. In each of the 20,000 replications we then draw $M = 1{,}000$ individuals randomly from the full sample of 15,992 individuals. Given this sample of size 1,000 and the randomly chosen threshold 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 these threshold, and so we should see that only 5% of the times we randomly pick a threshold the corresponding confidence interval should not include zero.

We do this exercise for the six global and the two local polynomial regressions. 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. For the local linear and local quadratic regressions 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). The results are in Table 4. The rejection rates for the global polynomials are substantially above the nominal rejection rate of 5%. In contrast the rejection rates for the local linear and local quadratic estimators are fairly close to the nominal rejection rate. Moreover, the median standard errors for the global estimators are substantially larger than the standard errors for the local estimators. Thus the global estimators combine large standard errors with under coverage for the confidence intervals, so that clearly the local polynomial estimators are superior in this setting.

We repeat this exercise for a second dataset. In this exercise we use a census data and consider the regression of years of education on earnings. This has the advantage that the data set is large, and the forcing variable is close to continuous. Substantively of course the regression is not of interest. However, qualitatively we find the same results: the global polynomial methods combine relatively poor coverage rates with substantially larger standard errors.

5. Discussion

Regression discontinuity designs have become increasingly popular in social sciences in the last twenty years as a credible method for obtaining causal estimates. One implementation relies on using global high-order polynomial

		Lalonde data $N = 15,992,\ M = 1,000$		Census data $N = 320,274,\ M = 1,000$	
	Order of polynomial	Rejection rate	Median s.e.	Rejection rate	Median s.e.
global	1	0.095	1.07	0.473	0.296
global	2	0.084	1.66	0.058	0.493
global	3	0.069	2.25	0.089	0.393
global	4	0.081	2.81	0.064	0.590
global	5	0.075	3.36	0.060	0.689
global	6	0.093	3.89	0.059	0.786
local	1	0.064	1.42	0.064	0.381
local	2	0.057	2.07	0.052	0.542

Table 4: Rejection rates for nominal 5% test under the null hypothesis of no true discontinuity. Contrary to naive intuition, rejection rates can be much higher than 5%, especially for the global fits.

approximations to the conditional expectation of the outcome given the forcing variable. Such models can give poor results in practice (see discussion from Gelman and Zelizer, 2015). This motivates the present paper in which we lay out the specific problems with the method and why we recommend against using high-order polynomials in regression discontinuity analyses. 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. We recommend that instead researchers use local low order polynomial methods (local linear or local quadratic) as discussed by Hahn, Todd, and VanderKlaauw (2001), Porter (2003), and Calonico, Cattaneo, and Titiunik (2014). In addition we recommend that researchers routinely present the implicit weights in the estimates of the causal estimands.

We present the arguments in the context of sharp regression discontinuity designs. The same arguments apply to fuzzy regression discontinuity designs, where we would recommend using local linear or quadratic methods for both the outcome and treatment received. In regression kink designs (e.g., Card, Lee, Pei, Z., and Weber, 2015, Dong, 2010), where the focus is on estimating a derivative of, rather than the level of the regression function at a point, one may wish to use local quadratic methods. The results in this paper suggest that such approaches would be superior to using global polynomial approximations.

Given all these problems, as well as the non-intuitive nature of high-degree polynomial fits, the natural question arises: what was the appeal of high-degree polynomials in the first place? We suspect this comes from three sources. First, the fact that high order polynomials can approximate any smooth function on a compact set arbitrarily well. While true, this does not address the issue that a high-order polynomial based on least square fit may not give a good approximation, especially close to the boundary. Second, in many of the paradigmatic examples of regression discontinuity analysis, the relation between the forcing variable and the outcome is very strong, so that even when a high-degree polynomial is fit to data, that fit is smooth and monotonic, in which case such polynomials can do less damage than they can in examples where the forcing variable is not a good predictor at all (as discussed in Gelman and Zelizer, 2015). The third implicit justification for high-degree polynomials, we suspect, is the recommendation given in many textbooks that, when performing causal inference, it is safest to include as many pre-treatment background variables as possible. The idea is that including relevant predictors should reduce bias, while including noise predictors will only induce slight increases in variance due to reduction in degrees of freedom. Thus when sample size is large, it can seem safe to include high-degree polynomial terms on the right hand side of the regression–especially if the coefficient on the discontinuity term is statistically significant, in which case the cost in increased variance would seem, in retrospect, minor in comparison to the gain in safety from including the more flexible predictor set. The flaw in this reasoning is that polynomials of the forcing term are not simply noise predictors and can induce bias as well as variance, thus damaging coverage, as discussed in section 4 of our paper.

6. References

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* **82** (6), 2295–2326.

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

Card, D., Lee, D. S., Pei, Z., & Weber, A. (2015). Inference on causal effects in a generalized regression kink design. *Econometrica*, **83**(6), 2453-2483.

Dahlquist, G.; Bjork, A. (1974), "Equidistant Interpolation and the Runge Phenomenon" *Numerical Methods*, pp. 101103, ISBN 0-13-627315-7.

Dehejia, R., and Wahba, S. (1999). Causal effects in non-experimental studies: Re-evaluating the evaluation of training programs. *Journal of the Ameri-*

can Statistical Association **94**, 1053–1062.

DiNardo, J., and Lee, D. (2010). Program evaluation and research designs. In Ashenfelter and Card (eds.), *Handbook of Labor Economics*, Vol. 4.

Dong, Y., 2010. Jumpy or kinky? regression discontinuity without the discontinuity.

Fan, J. and Gijbels, I., 1996. *Local polynomial modelling and its applications.* Monographs on statistics and applied probability 66 (Vol. 66). CRC Press.

Gelman, A., and Zelizer, A. (2015). Evidence on the deleterious impact of sustained use of polynomial regression on causal inference. *Research in Politics* **2** (1), 1–7.

Hahn, J., Todd, P., and Van Der Klaauw, W. (2001). Regression discontinuity. *Econometrica* **69** (1), 201–209.

Imbens, G., and Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies* **79** (3), 933–959.

Imbens, G., and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics* **142**, 615–635.

Imbens, G., and Rubin, D. (2015). *Causal Inference for Statistics, Social and Biomedical Sciences* **142**, Cambridge University Press.

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. (2008). Randomized experiments from non-random selection in U.S. House elections. *Journal of Econometrics* **142** (5), 675–697.

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, draft date September 25, 2003.

Skovron, C. and Titiunik, R. (2015) "A Practical Guide to Regression Discontinuity Designs in Political Science."

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** (2): 219–245.

Figure 1a. weights for higher order polynomials, Matsudaira data

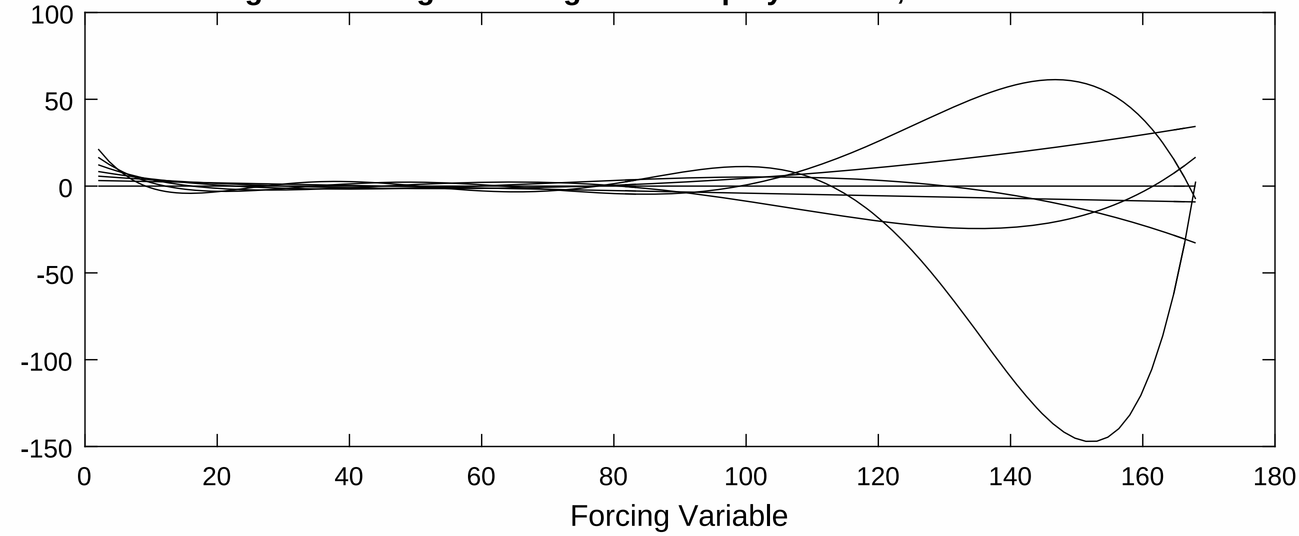

Figure 1.b Weight Functions for Local Linear Estimator with Rectancular and Triangular Kernel, Matsudaira Data

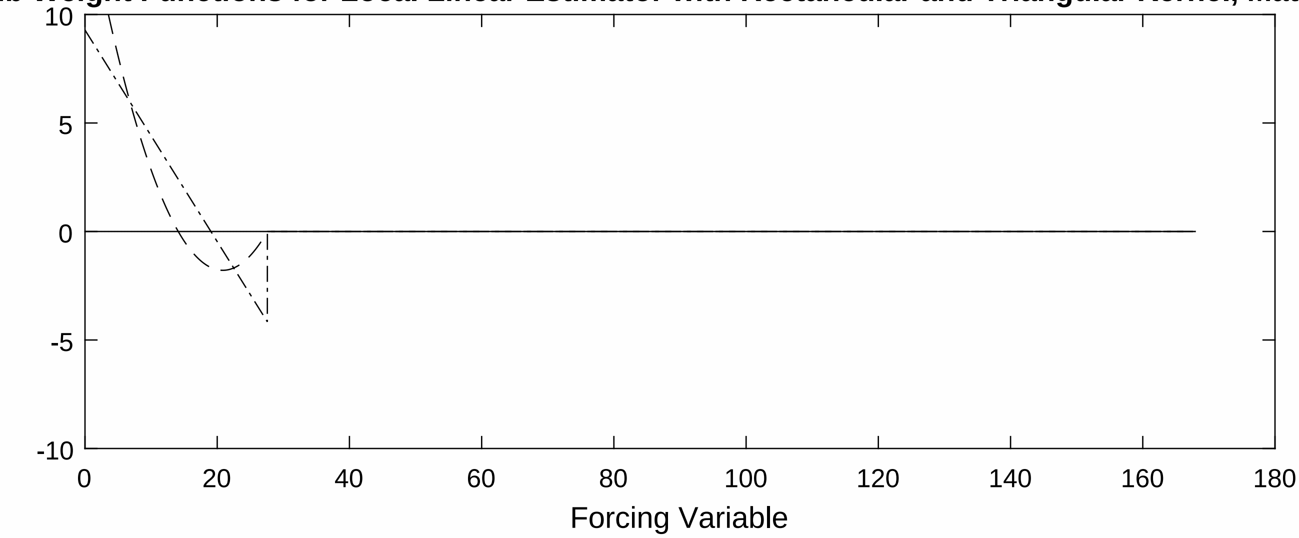

Figure 1c. Histogram of Forcing Variable Greater than Threshold, Matsudaira Data

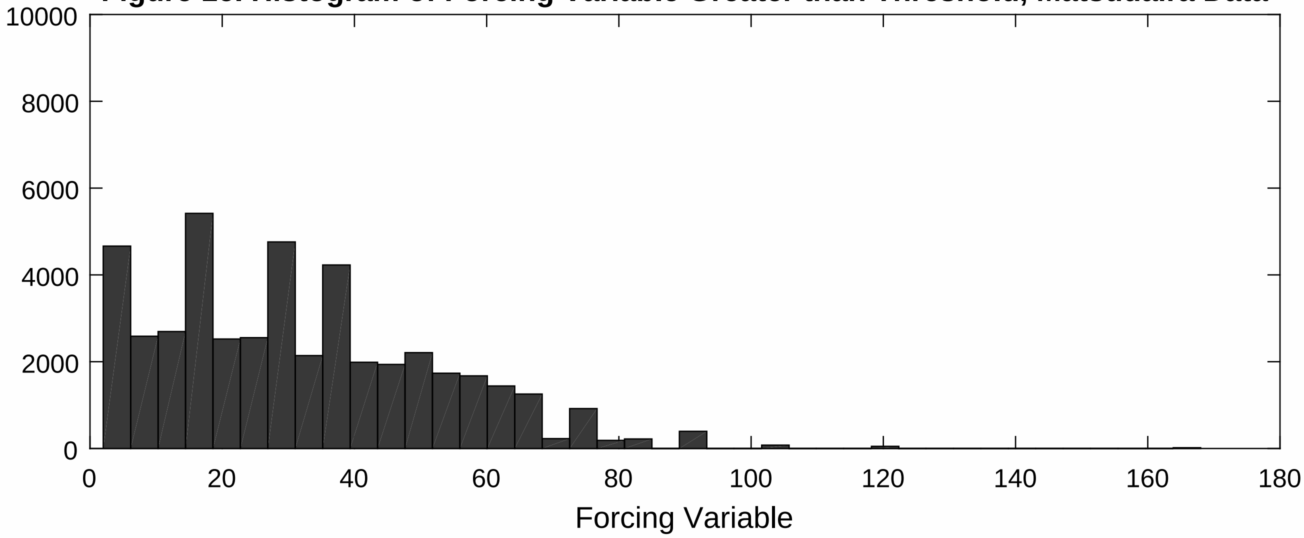

Figure 2a. weights for higher order polynomials, Jacob-Lefgren data

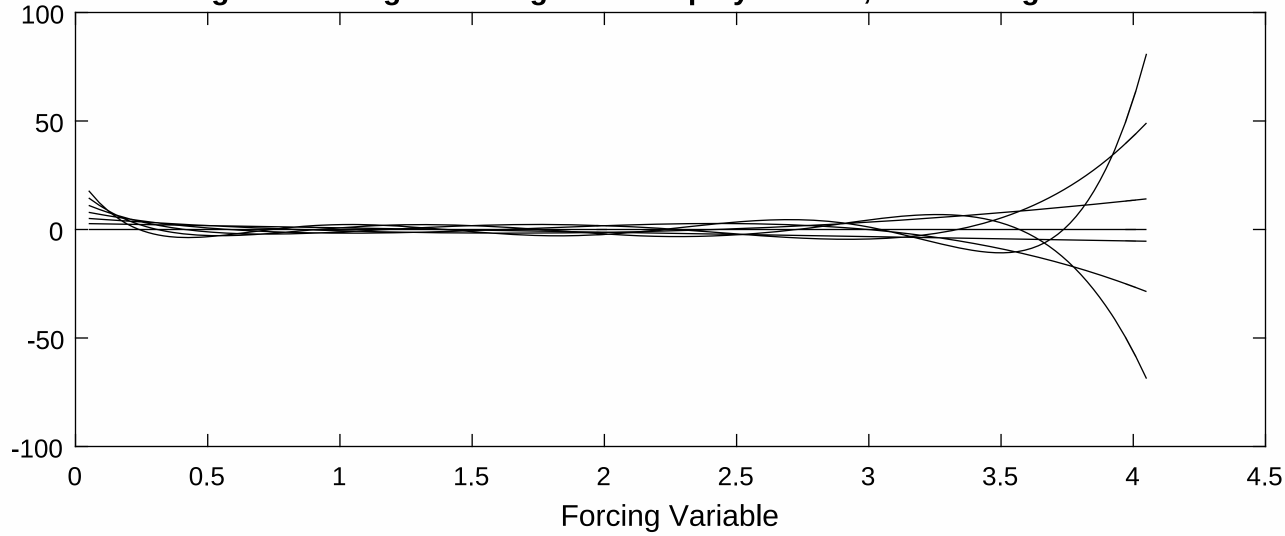

Figure 2.b Weight Functions for Local Linear Estimator with Rectancular and Triangular Kernel, Jacob-Lefgren Da

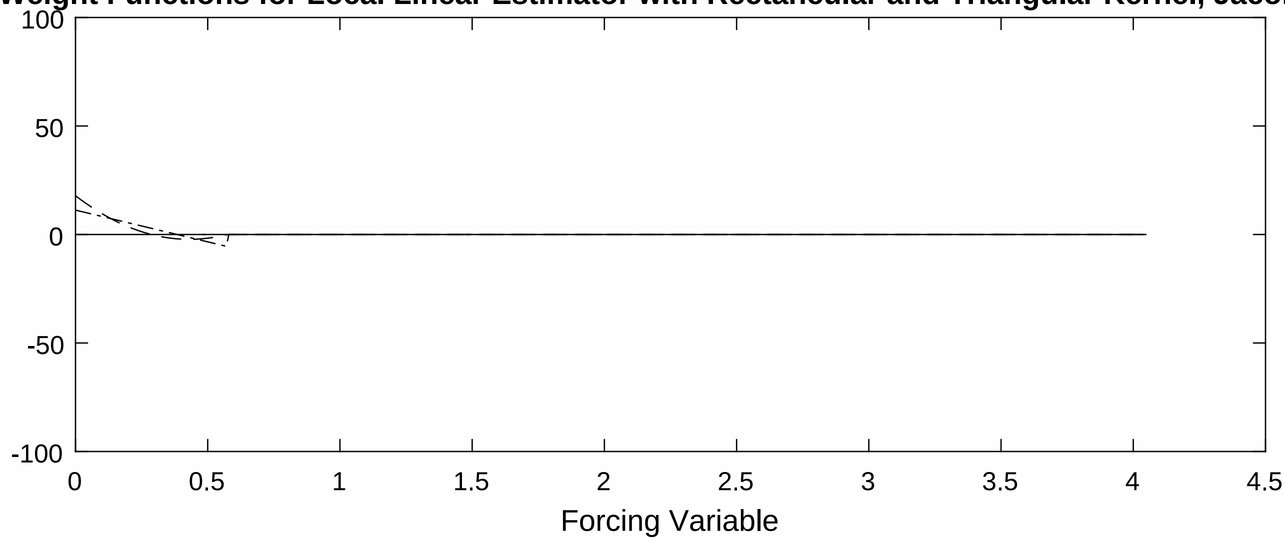

Figure 2c. Histogram of Forcing Variable Greater than Threshold, Jacob-Lefgren Data

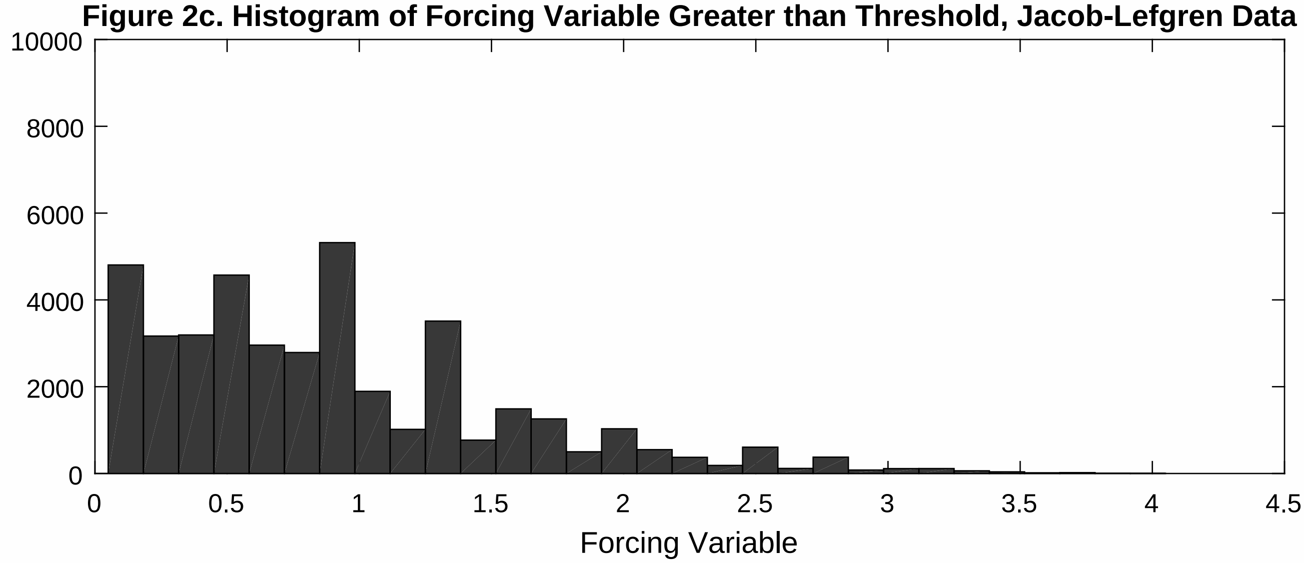

Figure 3a. weights for higher order polynomials, Lee data

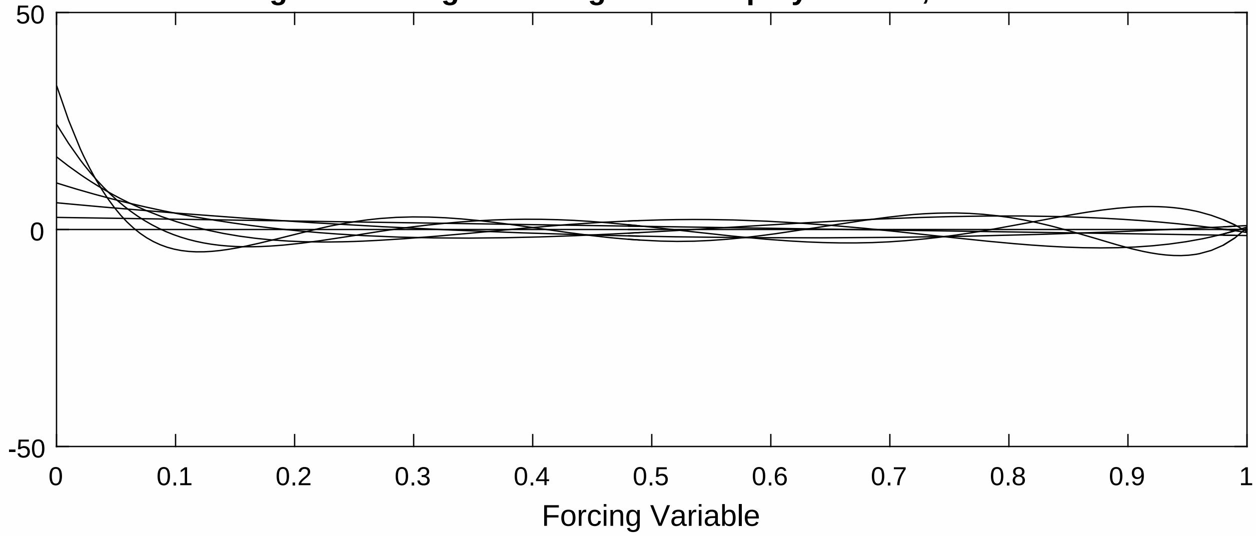

Figure 3.b Weight Functions for Local Linear Estimator with Rectancular and Triangular Kernel, Lee Data

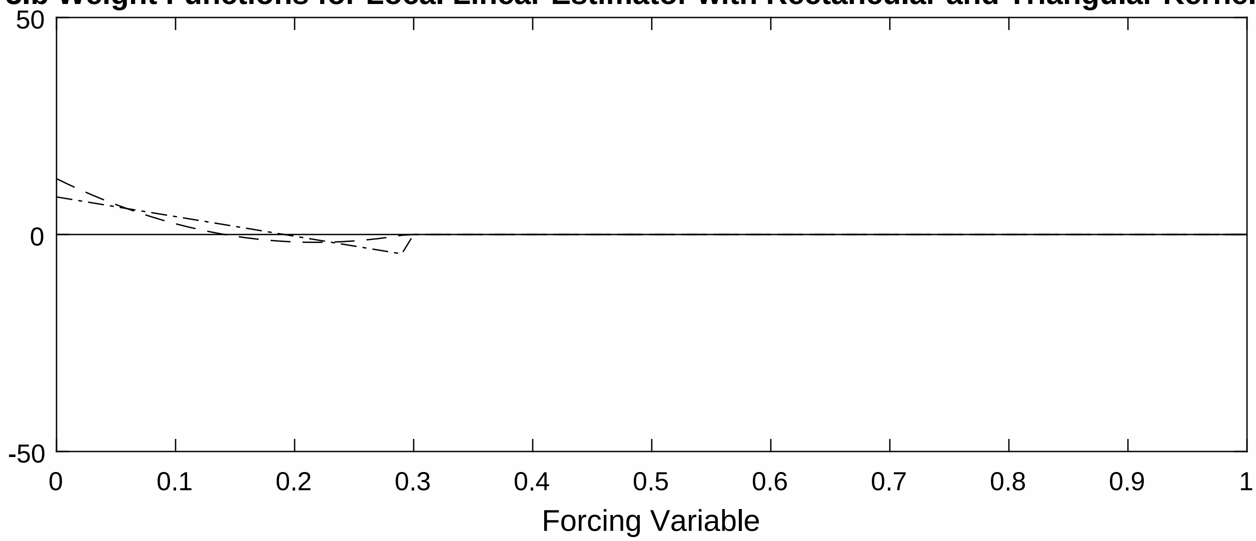

Figure 3c. Histogram of Forcing Variable Greater than Threshold, Lee Data

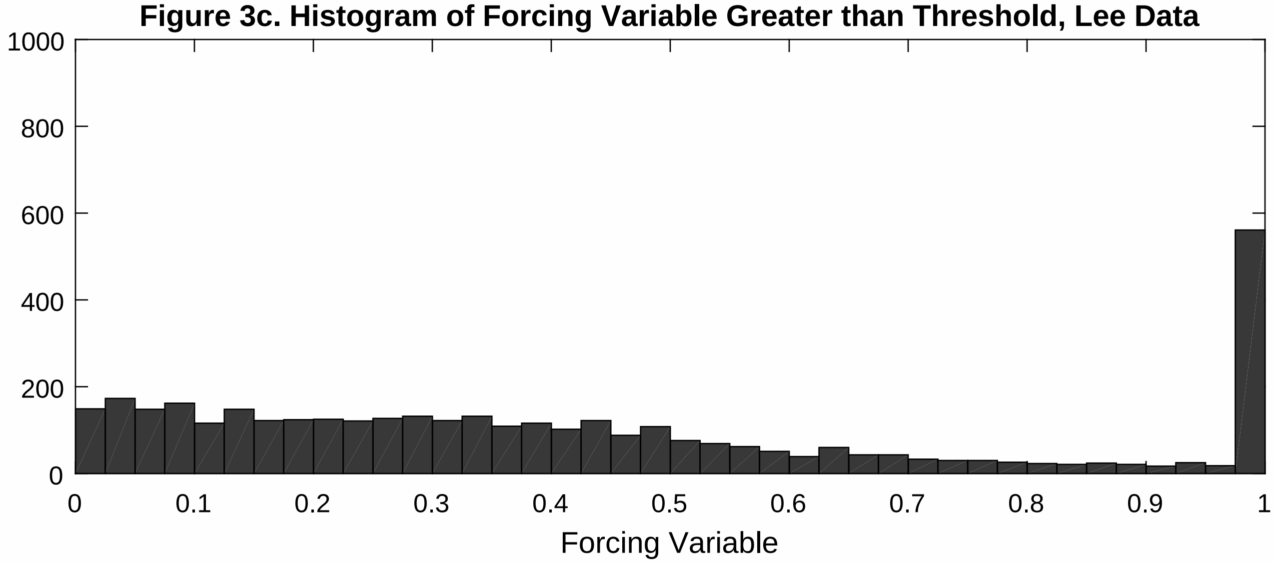

Fig 4a: Matsudaira Data, Regression of Test Score on Forcing Variable, Forcing Variable Positive

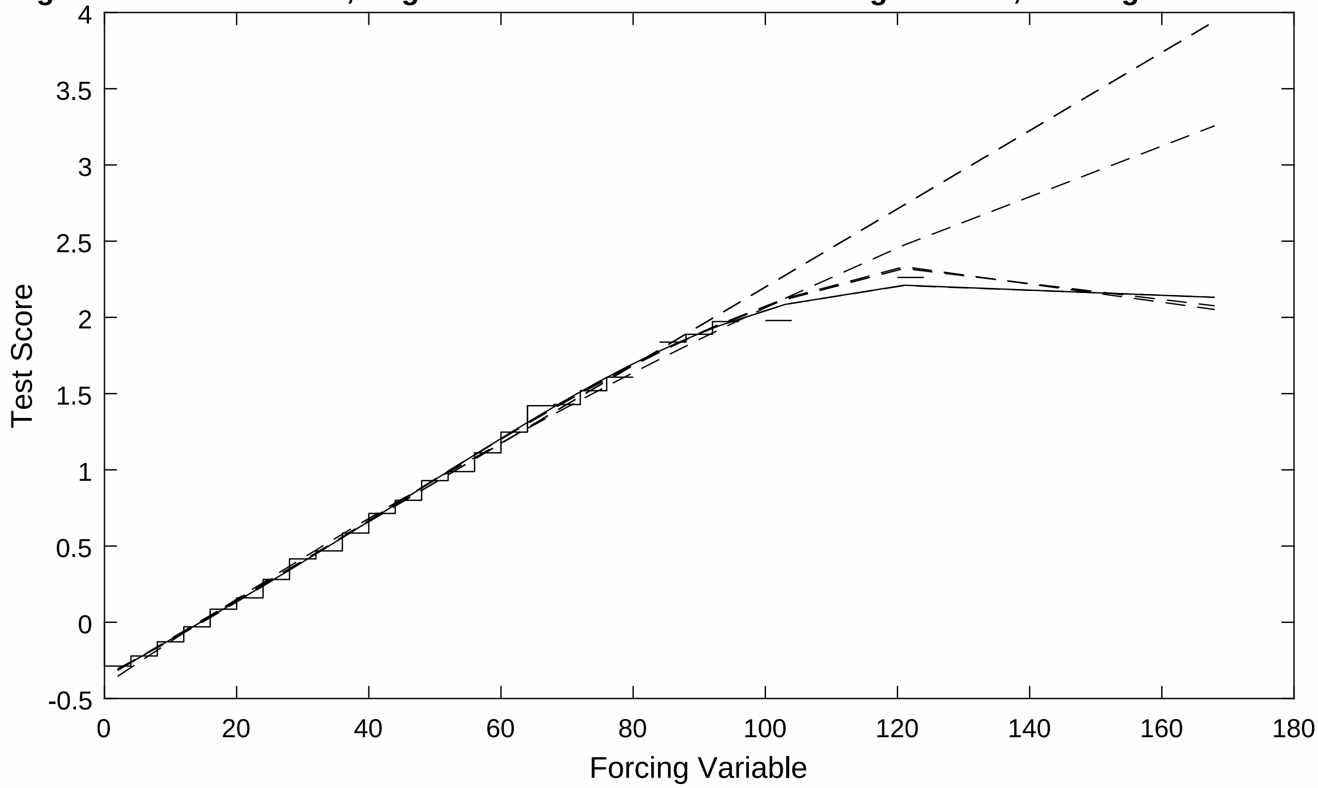

Fig 4b: Matsudaira Data, Regression of Test Score on Forcing Variable, Forcing Variable Negative

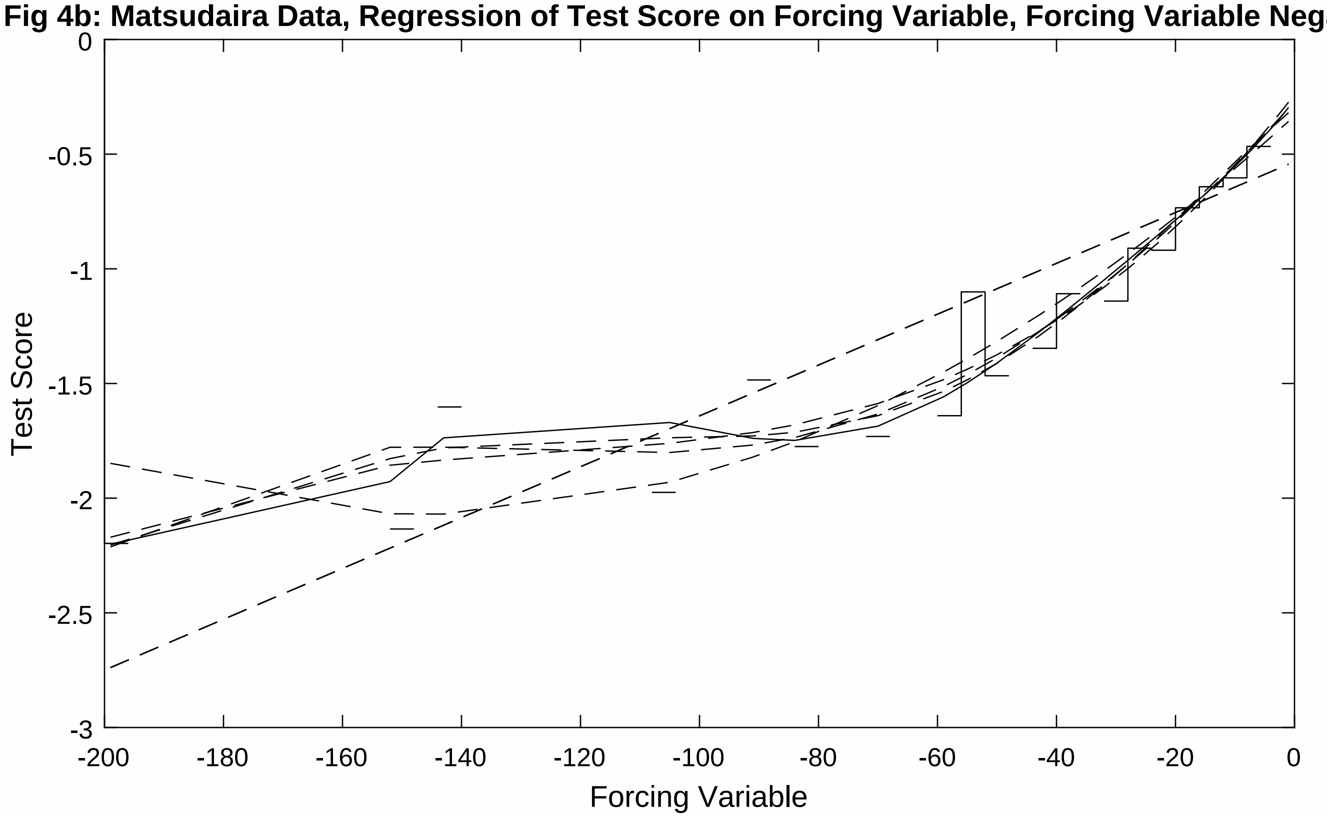

Fig 5a: Lalonde Data, Regression of Earnings in 1978 on Average of Earnings in 1974, 1975

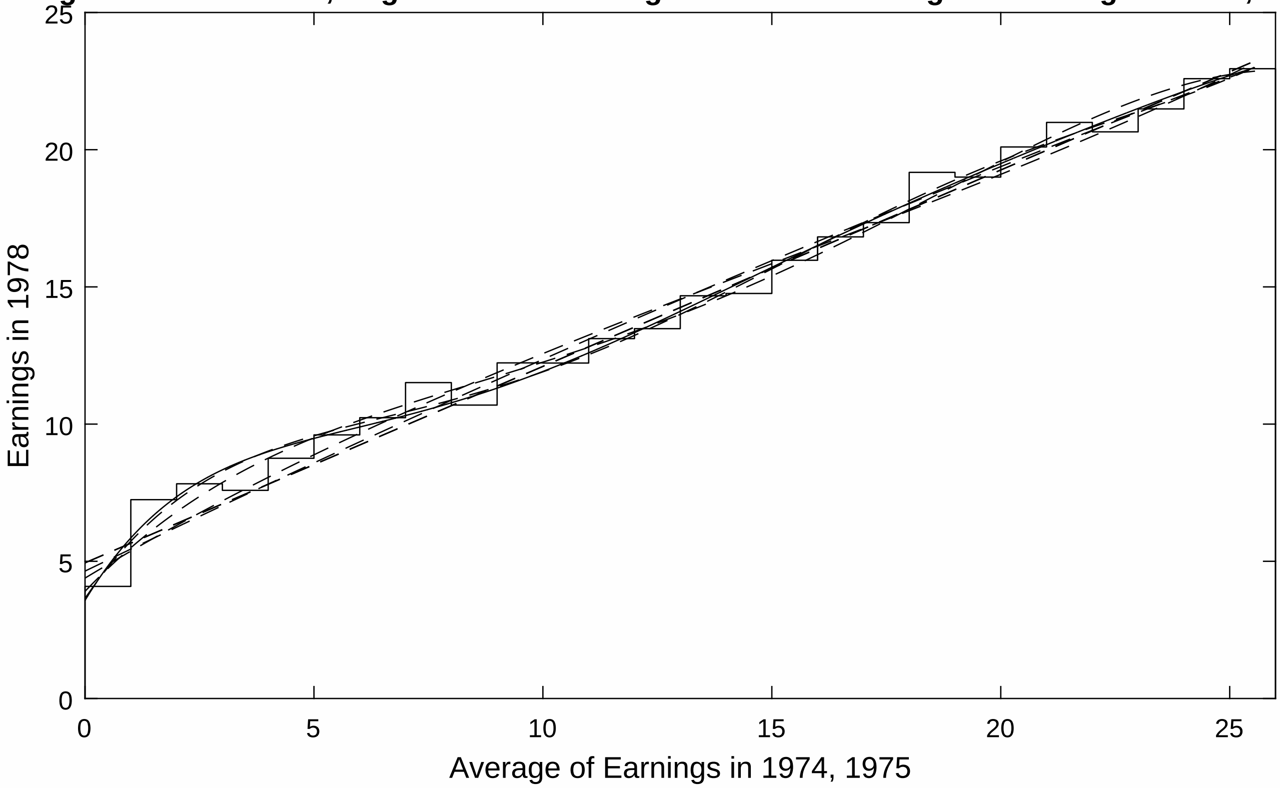

Figure 5b. Histogram of Average Earnings, Lalonde Data