Local Polynomial Order in Regression Discontinuity Designs[1]

David Card
UC Berkeley, NBER and IZA

Zhuan Pei
Brandeis University

David S. Lee
Princeton University and NBER

Andrea Weber
University of Mannheim and IZA

October 21, 2014

Abstract

The local linear estimator has become the standard in the regression discontinuity design literature, but we argue that it should not always dominate other local polynomial estimators in empirical studies. We show that the local linear estimator in the data generating processes (DGP's) based on two well-known empirical examples does not always have the lowest (asymptotic) mean squared error (MSE). Therefore, we advocate for a more flexible view towards the choice of the polynomial order, p, and suggest two complementary approaches for picking p: comparing the MSE of alternative estimators from Monte Carlo simulations based on an approximating DGP, and comparing the estimated asymptotic MSE using actual data.

Keywords: Regression Discontinuity Design; Regression Kink Design; Local Polynomial Estimation; Polynomial Order
JEL codes: C21, C13, C14

[1]We thank Pat Kline, Pauline Leung and seminar participants at Brandeis and George Washington University for helpful comments, and we thank Samsun Knight and Carl Lieberman for excellent research assistance.

1 Introduction

The seminal work of Hahn et al. (2001) has established local linear nonparametric regression as a standard approach for estimating the treatment effect in a regression discontinuity (RD) design. Recent influential studies on nonparametric estimation in RD designs have built upon the local linear framework. Imbens and Kalyanaraman (2012) propose bandwidth selectors optimal for the local linear RD estimator (henceforth the "IK bandwidth"), and Calonico et al. (Forthcoming) introduce a procedure to correct the bias in the local linear estimator and to construct robust confidence intervals.

The prevailing preference for the local linear estimator is based on the order of its asymptotic bias, but this alone cannot justify the universal dominance of the linear specification over other polynomial orders. Hahn et al. (2001) choose the local linear estimator over the local constant[1] for its smaller order of asymptotic bias – specifically, the bias of the local linear estimator is of order $O(h^2)$ and the local constant $O(h)$ (h here refers to the bandwidth that shrinks as the sample size n becomes large). However, the argument based on asymptotic order comparisons per se does not imply that local linear should always be preferred to alternative local polynomial estimators. Under standard regularity conditions, the asymptotic bias is of order $O(h^3)$ for the local quadratic RD estimator, $O(h^4)$ for local cubic, and $O(h^{p+1})$ for the p-th order local polynomial sharp RD estimator, $\hat{\tau}_p$ (Lemma A1 Calonico et al. (Forthcoming)). Therefore, if the goal is to maximize the shrinkage rate of the asymptotic bias, researchers should choose a p as large as possible. The fact that Hahn et al. (2001) do not recommend a very large p for RD designs implies that finite sample properties of the estimator must be taken into consideration. But if finite sample properties are important, the desired polynomial choice may depend on the sample size: the local constant estimator $\hat{\tau}_0$ may be preferred to $\hat{\tau}_1$ when the sample size is small, and higher-order local polynomial estimators may be preferred when the sample size is large.

In a given finite sample, the derivatives of the conditional expectation function of the outcome variable are also important for the local polynomial order choice, even though they are omitted under the $O(\cdot)$ notation of asymptotic rates. If the conditional expectation of the outcome variable Y is close to being a constant function of the assignment variable X, then the local constant specification will provide adequate approximation, and consequently $\hat{\tau}_0$ will perform well. On the other hand, if the said conditional expectation function has a large curvature, researchers may consider choosing a higher-order local polynomial estimator

[1] The local constant estimator is equivalent to a kernel regression estimator, which is the terminology used by Hahn et al. (2001).

instead.

Because the performance of a local polynomial estimator depends on the sample size and the properties of the data generating process (DGP), a single choice like $p = 1$, though convenient, may not be the best for all empirical RD applications. In this paper, we explore the best local polynomial order choice in the spirit of Fan and Gijbels (1996) by comparing the mean squared error (MSE) of $\hat{\tau}_p$ and its asymptotic approximation (AMSE) across p. Using the (A)MSE of the local estimator as a measuring stick is consistent with the optimal bandwidth literature and answers to the critique of Gelman and Imbens (2014) that the goodness of fit measure used in choosing a *global* polynomial order "is not closely related to the research objective of causal inference".

Similar to Imbens and Kalyanaraman (2012) and Calonico et al. (Forthcoming), we use the data generating processes (DGP's) based on Lee (2008) and Ludwig and Miller (2007) to illustrate the points above. We document that local regressions with orders other than $p = 1$ may perform better than with the local linear estimator $\hat{\tau}_1$. We provide details in the following section.

2 Mean Squared Error and the Local Polynomial Order

In this section, we rank the $\hat{\tau}_p$'s based on their (A)MSE for the approximating DGP's of Lee (2008) and Ludwig and Miller (2007). In subsection 2.1, we calculate the theoretical asymptotic mean squared error evaluated at the optimal bandwidth for the Lee and Ludwig-Miller DGP. Based on the calculation, we show that whether or not the local linear estimator $\hat{\tau}_1$ theoretically dominates an alternative $\hat{\tau}_p$ depends on the sample size as well as the DGP. In subsection 2.2, we examine the actual mean squared error of the local polynomial estimators via Monte Carlo simulation and confirm that $\hat{\tau}_1$ is not always the best-performing RD estimator. In subsection 2.3, we show that the *estimated* AMSE serves as a sensible basis for choosing the polynomial order. In subsection 2.4, we discuss the properties of the local polynomial estimators in light of the recent study, Gelman and Imbens (2014). Gelman and Imbens (2014) point out that an undesirable property of a high-order *global* polynomial estimator is that it may assign very large weights (henceforth "GI weights") to observations far away from the discontinuity threshold. We show in subsection 2.4 that this does not appear to be the case for high-order *local* polynomial estimators for the Lee and Ludwig-Miller data when using the corresponding optimal bandwidth selector. In subsection 2.5, we argue that the MSE-based methods for choosing the polynomial order can be easily applied to the fuzzy design and the regression kink design (RKD).

2.1 Theoretical AMSE

We specify the Lee and Ludwig-Miller DGP following Imbens and Kalyanaraman (2012) and Calonico et al. (Forthcoming). Let Y denote the outcome of interest, let X denote the normalized running variable, and let $D = 1_{[X \geqslant 0]}$ denote the treatment. For both DGP's, the running variable X follows the distribution $2\mathscr{B}(2,4) - 1$, where $\mathscr{B}(\alpha, \beta)$ denotes a beta distribution with shape parameters α and β. The outcome variable is given by $Y = E[Y|X = x] + \varepsilon$, where $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ with $\sigma_\varepsilon = 0.1295$ and the conditional expectation functions are specified as

$$\text{Lee: } E[Y|X = x] = \begin{cases} 0.48 + 1.27x + 7.18x^2 + 20.21x^3 + 21.54x^4 + 7.33x^5 & \text{if } x < 0 \\ 0.52 + 0.84x - 3.00x^2 + 7.99x^3 - 9.01x^4 + 3.56x^5 & \text{if } x \geqslant 0 \end{cases}$$

$$\text{Ludwig-Miller: } E[Y|X = x] = \begin{cases} 3.71 + 2.30x + 3.28x^2 + 1.45x^3 + 0.23x^4 + 0.03x^5 & \text{if } x < 0 \\ 0.26 + 18.49x - 54.81x^2 + 74.30x^3 - 45.02x^4 + 9.83x^5 & \text{if } x \geqslant 0 \end{cases}.$$

To obtain the conditional expectation functions, Imbens and Kalyanaraman (2012) and Calonico et al. (Forthcoming) first discard the outliers (i.e. observations for which the absolute value of the running variable is very large) and then fit a separate quintic function on each side of the threshold to the remaining observations.

Because the DGP is analytically specified, we can apply Lemma 1 of Calonico et al. (Forthcoming) to compute the theoretical AMSE-optimal bandwidth for the various local polynomial estimators and the corresponding AMSE's. Since the k-th order derivative of the conditional expectation functions is zero on both sides of the cutoff for $k > 5$, the highest-order estimator we allow is the local quartic in order to ensure the finiteness of the optimal bandwidth. Tables 1 and 2 summarize the results for two kernels and two sample sizes. The kernel choices are uniform and triangular, the most popular in the RD literature. The two sample sizes are $n = 500$ and $n = n_{actual}$. Imbens and Kalyanaraman (2012) and Calonico et al. (Forthcoming) use $n = 500$ in their simulations, while $n_{actual} = 6558$ is the actual sample size of the Lee data and $n_{actual} = 3138$ for the Ludwig-Miller data.

As summarized in Tables 1 and 2, $p = 4$ is the preferred choice based on theoretical AMSE. For the Lee

DGP, $\hat{\tau}_1$ dominates $\hat{\tau}_2$ when $n = 500$, but the AMSE of $\hat{\tau}_p$, denoted by $\text{AMSE}_{\hat{\tau}_p}$ monotonically decreases with p when $n = 6558$. For the Ludwig-Miller DGP, $\text{AMSE}_{\hat{\tau}_p}$ decreases with p for both $n = 500$ and $n = 6558$. In general, the $\text{AMSE}_{\hat{\tau}_p}$ is smaller under the triangular kernel than under the uniform kernel, confirming the boundary optimality of the triangular kernel per Cheng et al. (1997).

It can be easily shown that $\text{AMSE}_{\hat{\tau}_p}$ is proportional to $n^{-\frac{2p+2}{2p+3}}$, suggesting that a high-order estimator should have an asymptotically smaller AMSE. Therefore, when $q > p$, $\hat{\tau}_q$ either always has a lower AMSE than $\hat{\tau}_p$ or it does when the sample size exceeds a threshold. We compute the sample sizes for which $\text{AMSE}_{\hat{\tau}_p} < \text{AMSE}_{\hat{\tau}_1}$ for $p = 0, 2, 3, 4$, and the results are summarized in Table 3 and 4. Under the Lee DGP, $\text{AMSE}_{\hat{\tau}_0} < \text{AMSE}_{\hat{\tau}_1}$ when the sample size falls below 296 under the uniform kernel and 344 under the triangular kernel; similarly, a higher-order estimator ($p = 2, 3, 4$) has a smaller AMSE than $\hat{\tau}_1$ only when the sample size is large enough. In contrast, $\hat{\tau}_p$ has a smaller AMSE regardless of the sample size under the Ludwig-Miller DGP as a result of the large curvature therein.

We also compute the AMSE for the bias-corrected estimator from Calonico et al. (Forthcoming), $\hat{\tau}^{bc}_{p,p+1}$, henceforth "the CCT estimator". Calonico et al. (Forthcoming) propose to estimate the bias of the local RD estimator $\hat{\tau}_p$ by using a local regression of order $p+1$ and account for the variance in the bias estimation. The CCT estimator $\hat{\tau}^{bc}_{p,p+1}$ is equal to the sum of the conventional estimator $\hat{\tau}_p$ and the bias-correction term. We use results from Theorem A1 of Calonico et al. (Forthcoming) to compute the AMSE's of $\hat{\tau}^{bc}_{p,p+1}$ evaluated at optimal where $p = 0, 1, 2, 3$. We omit the $p = 4$ case to ensure that the optimal bandwidth used in bias estimation is finite.

The theoretical AMSE results for the CCT estimators are summarized in Table A.1 and A.2 in the Supplemental Appendix. Similar to the AMSE for the conventional estimators, higher-order estimators have a lower AMSE than local linear, and $\hat{\tau}^{bc}_{3,4}$ has the smallest AMSE for both the Lee and Ludwig-Miller DGP's. In fact, the relative ranking of $\text{AMSE}_{\hat{\tau}^{bc}_{p,p+1}}$ in Table A.1 and A.2 for each sample size and kernel choice is the same as that of $\text{AMSE}_{\hat{\tau}_{p+1}}$ in Table 1 and 2.[2] It is also worth noting that $\hat{\tau}^{bc}_{1,2}$, the local linear CCT estimator, has the largest AMSE among the four estimators when $n = 500$.

We summarize sample sizes for which $\text{AMSE}_{\hat{\tau}^{bc}_{p,p+1}} < \text{AMSE}_{\hat{\tau}^{bc}_{1,2}}$ for $p = 0, 2, 3$ in Table 3 and A.2. The results are similar to those in Table 3 and 4 for the conventional estimator. For the Lee DGP, $p = 0$ is preferred to $p = 1$ when n is small, and $p = 2, 3$ is preferred when n is large. For the Ludwig-Miller DGP,

[2] This is not surprising in light of Remark 7 in Calonico et al. (Forthcoming): when b, the pilot bandwidth for bias estimation, is equal to h, the main bandwidth for the conventional estimator, the estimator $\hat{\tau}^{bc}_{p,p+1}$ is the same as $\hat{\tau}_{p+1}$ and therefore has the same AMSE.

$\hat{\tau}^{bc}_{1,2}$ is preferred to $\hat{\tau}^{bc}_{0,1}$ regardless of sample size, whereas $\hat{\tau}^{bc}_{2,3}$ and $\hat{\tau}^{bc}_{3,4}$ always have a smaller AMSE than $\hat{\tau}^{bc}_{1,2}$.

As its name suggests, the AMSE is an asymptotic approximation of the actual mean squared error, and the approximation may or may not be good for a given DGP and sample size. Therefore, the ranking of estimators by theoretical AMSE may not be the same as the ranking by MSE. We present the latter for the two DGP's in the following subsection.

2.2 MSE from Simulations

In this subsection, we present results from Monte Carlo simulations, which show that higher-order local estimators have lower MSE than their local linear counterpart for the actual sample sizes in the Lee and Ludwig-Miller application. Tables 5 and 6 report the MSE for $\hat{\tau}_p$ under the theoretical AMSE-optimal bandwidth for the Lee and Ludwig-Miller DGP respectively, where the MSE is computed over 10,000 repeated samples. For the Lee DGP, we report results for p between 0 and 4; for the Ludwidg-Miller DGP, we omit $\hat{\tau}_0$ because too few observations lie within its theoretical optimal bandwidth.

For the smaller sample size of $n = 500$, $\hat{\tau}_1$ appears to have the lowest MSE for the Lee DGP, which is in contrast with the AMSE ranking in Table 1. However, $\hat{\tau}_4$ does have the lowest MSE when $n = 6558$, suggesting that AMSE provides a better approximation under this larger sample size for the Lee DGP. As in Table 2, $\hat{\tau}_4$ has the lowest MSE for both $n = 500$ and $n = 3138$ for the Ludwig-Miller DGP as seen in Table 6. Again, the AMSE approximation is generally better for the larger sample size.

We also report the coverage rate of the 95% confidence interval constructed using $\hat{\tau}_p$, which is the focus of Calonico et al. (Forthcoming). The coverage rate is above 91% when $p \geqslant 1$ for both DGP's, and that of $\hat{\tau}_0$ for the Lee DGP is just under 90%. Therefore, all the estimators in Tables 5 and 6 appear to be sensible candidates in terms of coverage rates.

The theoretical AMSE optimal bandwidth is never known in any empirical application, and it has to be estimated. Consequently, we also evaluate the performance of alternative estimators with the *estimated* optimal bandwidth in Monte Carlo simulations. We adopt two alternative bandwidth choices, the CCT bandwidth with and without regularization,[3] and the corresponding results are reported in Tables 7-10.

For both the Lee and Ludwig Miller DGP, the MSE-preferred polynomial order is lower under the default

[3] The regularization Calonico et al. (Forthcoming) implement in their default bandwidth selector follows the spirit of Imbens and Kalyanaraman (2012). It decreases with the variance of the bias estimator and prevents the bandwidth from becoming large. We do not adopt the IK bandwidth here because it is only proposed for $\hat{\tau}_1$.

CCT bandwidth (i.e., with regularization), denoted by $\hat{h}_{CCT}$, than under the theoretical optimal bandwidth, h_{opt}. One explanation is that the average $\hat{h}_{CCT}$ is much smaller than h_{opt} for higher order p, and the corresponding variance of $\hat{\tau}_p$ is larger under $\hat{h}_{CCT}$ than under h_{opt}. In comparison, the average value of the CCT bandwidth without regularization, $\hat{h}_{CCT,noreg}$, is much closer to h_{opt} for $p = 3,4$. As a consequence, the MSE-preferred polynomial orders under $\hat{h}_{CCT,noreg}$ are closer to those under h_{opt}. For the Lee DGP, the MSE-preferred polynomial orders are the same as those for h_{opt}: $p = 1$ for $n = 500$ and $p = 4$ for $n = 6558$. For the Ludwig-Miller DGP, $\hat{\tau}_3$ is the MSE-preferred estimator when $n = 500$, and $\hat{\tau}_3$ and $\hat{\tau}_4$ have very similar MSE's when $n = 3138$. In this latter case, $\text{MSE}_{\hat{\tau}_3} > \text{MSE}_{\hat{\tau}_4}$ under the uniform kernel, $\text{MSE}_{\hat{\tau}_4} > \text{MSE}_{\hat{\tau}_3}$ under the triangular kernel. In summary, a higher order $\hat{\tau}_p$ has a lower MSE than $\hat{\tau}_1$ for the actual sample sizes in the two empirical applications.

We also examine the MSE of the CCT bias-corrected estimators, $\hat{\tau}^{bc}_{p,p+1}$, from Monte Carlo simulations, and the results are presented in Tables A.5-A.10. We find that the bias-corrected linear estimator, the default estimator in Calonico et al. (Forthcoming), never delivers the smallest MSE. In fact, $\hat{\tau}^{bc}_{0,1}$ or $\hat{\tau}^{bc}_{2,3}$ have the lowest MSE, depending on the sample size and kernel choice for the Lee DGP, whereas $\hat{\tau}^{bc}_{2,3}$ consistently have the lowest MSE for the Ludwig-Miller DGP. In the next subsection, we explore the use of the estimated AMSE for picking the polynomial order.

2.3 Estimated AMSE

When computing the optimal bandwidth for a local polynomial RD estimator, the asymptotic bias and variance are both estimated. It follows that the AMSE of the estimator, which is the sum of the squared bias and variance, can be easily estimated as well. As suggested by Fan and Gijbels (1996), comparing the estimated AMSE for alternative local polynomial estimators, $\widehat{\text{AMSE}}_{\hat{\tau}_p}$, can serve as a basis for choosing p. We adapt the suggestion by Fan and Gijbels (1996) to the RD design and investigate the choice of polynomial order p based on the estimated AMSE of $\hat{\tau}_p$ and $\hat{\tau}^{bc}_{p,p+1}$.

Tables 11-14 summarize the statistics of $\widehat{\text{AMSE}}_{\hat{\tau}_p}$ from Monte Carlo simulations. We report the average $\widehat{\text{AMSE}}_{\hat{\tau}_p}$ over 10,000 repeated samples, the fraction of times each $\hat{\tau}_p$ has the smallest $\widehat{\text{AMSE}}$ and the average computed optimal bandwidth. Comparing these four tables to the MSE Tables 7-10 reveals that the most likely choice of p based on $\widehat{\text{AMSE}}$ does not always have the smallest MSE, but it is nevertheless sensible in most cases. For the Lee DGP under $\hat{h}_{CCT}$, $\hat{\tau}_0$ and $\hat{\tau}_1$ are the most likely choice based $\widehat{\text{AMSE}}$ for $n = 500$ and $n = 6558$, respectively, and they have the second lowest MSE. For the Lee DGP under $\hat{h}_{CCT,noreg}$, $\hat{\tau}_1$ is the most likely choice based on $\widehat{\text{AMSE}}$ for both $n = 500$ and $n = 6558$; although it performs less well compared to the higher-order estimators for $n = 6558$, it does have the lowest MSE for $n = 500$. For the Ludwig-Miller DGP, $\widehat{\text{AMSE}}$-based order choice does well: in seven out of the eight cases, it has the smallest MSE; for the remaining case ($n = 3138$, uniform kernel and $\hat{h}_{CCT,noreg}$), its MSE comes as a close second.

We also estimate the AMSE of $\hat{\tau}^{bc}_{p,p+1}$ based on Theorem A1 of Calonico et al. (Forthcoming) and summarize the simulation results in Tables A.11-A.14. Again, we examine whether the $\widehat{\text{AMSE}}$-based choice of $\hat{\tau}^{bc}_{p,p+1}$ has the lowest MSE by way of comparison to Tables A.7-A.10. As with the conventional estimator $\hat{\tau}_p$, the most likely choice of p for the bias-corrected estimator based $\widehat{\text{AMSE}}$ does not always have the lowest MSE – it does so in ten out of the 16 cases. In the remaining six cases, whenever the most likely p is not 1, the associated MSE is lower than that of $\hat{\tau}^{bc}_{1,2}$. Therefore, for the Lee and Ludwig DGP, using $\widehat{\text{AMSE}}$ improves upon the fixed choice of the bias-corrected linear estimator $\hat{\tau}^{bc}_{1,2}$.

In summary, using $\widehat{\text{AMSE}}$ leads to a sensible polynomial order choice in many instances. In the vast majority of cases in our simulation, the most likely choice of p based on $\widehat{\text{AMSE}}$ does have the lowest or the second lowest MSE among alternative estimators. With only two exceptions out of 32 cases, the most likely choice of p has an MSE that is lower than or equal to that of the default $p = 1$. Therefore, estimating $\widehat{\text{AMSE}}$ can complement Monte Carlo simulations based on approximating DGP's for choosing the local polynomial order.

2.4 GI Weights for Local Regressions

As mentioned at the beginning of section 2, Gelman and Imbens (2014) recently raise concerns of using a global or local high-order polynomial (e.g. cubic or quartic) to estimate the RD treatment effect. One issue in particular is that estimators based on high-order global regressions sometimes assign too much weight to observations far away from the RD cutoff. Since we have demonstrated above that high-order *local* estimators may be desirable in certain cases for the Lee and Ludwig-Miller DGP's, we examine whether noisy weights are a problem for *local* regressions in the two applications.

Using the *actual* Lee and Ludwig-Miller data, Figures 1-8 plot the GI weights for the left and right intercept estimators that make up $\hat{\tau}_p$ for $p = 0, 1, ..., 5$. As with the previous subsections, we examine the weights for two kernel choices (uniform and triangular) and two bandwidth choices ($\hat{h}_{CCT}$ and $\hat{h}_{CCT,noreg}$). For high-order local estimators, observations far away from the threshold receive little weight as compared to those close to the threshold as desired. In fact, even the GI weights in the *global* estimators for the Lee and Ludwig-Miller data are reasonably well-behaved: as seen from Figures A.1 and A.2, observations far away from the RD cutoff never receive significantly larger weights than those close to the threshold.

The other two concerns regarding high-order global estimators voiced by Gelman and Imbens (2014) are 1) they are not chosen based on a criterion relevant for the causal parameter of interest and 2) the corresponding 95% confidence interval has incorrect coverage rates. As argued in section 1, the (A)MSE of the RD estimator is an important benchmark of the literature and therefore dispels the first concern. As demonstrated in the simulations, when a higher order p is preferred, the coverage rate of the corresponding 95% confidence interval is quite close to 95%, which helps to alleviate the second concern. Together with well-behaved GI weights, we believe that high-order local polynomial RD estimators in certain cases are good alternatives to the local linear.

2.5 Extensions: Fuzzy RD and Regression Kink Design

In this subsection, we briefly discuss how (A)MSE-based local polynomial order choice applies to two popular extensions of the sharp RD design. The first extension is the fuzzy RD design, where the treatment assignment rule is not strictly followed. In the existing RD literature, $p = 1$ is still the default choice in the fuzzy RD. But by the same argument as above, local linear is not necessarily the best estimator in all applications. In the same way that we can calculate the AMSE and simulate the MSE of sharp RD estimators, we can rely on Lemma A2 and Theorem A2 of Calonico et al. (Forthcoming) and do it for the fuzzy RD estimators. Similarly, the same principle can be applied to the regression kink design (RKD) proposed and explored by Nielsen et al. (2010) and Card et al. (2012). For RKD, Calonico et al. (Forthcoming) recommends using $p = 2$ as the default polynomial order following its RD analog, but again the best polynomial choice should depend on the particular data set. The ideas presented in this paper readily apply to the RKD case and may help researchers choose the best polynomial order for their study.

3 Conclusion

The local linear estimator has become the standard in the regression discontinuity literature. In this paper, we argue that $p = 1$ should not be the universally preferred polynomial order across all empirical applications. The mean squared error of the p-th order local estimator depends on the sample size and the intrinsic properties of the data generating process. In two well-known empirical examples, $p = 1$ is not necessarily the polynomial order that delivers the lowest (A)MSE.

We do not oppose the use of local linear estimator in RD studies; it is a convenient choice and performs quite well in many applications. However, we do oppose the notion that a single polynomial order can be optimal for all RD analyses. We advocate for a more flexible view that an empiricist should be able to adopt a different local polynomial estimator if it is better suited for the application. If the empiricist would like to explore the polynomial order choice, we suggest two complementary options to her: 1) estimate an approximating DGP to the data at hand and conduct Monte Carlo simulations to gauge the performance of alternative estimators; 2) estimate the AMSE and compare it across alternative estimators.

Each option has its own advantages and disadvantages. Option 1 reveals the exact MSE's (and gives the coverage rates of the 95% confidence interval) with an approximating DGP[4], whereas option 2 estimates an approximate MSE using the actual data. It is ideal if the two options point to the same p; when they do not, it is perhaps prudent to present the value of both estimators, as is typically done in empirical research.

[4]In light of Gelman and Imbens (2014), if the approximating DGP is estimated using a high-order global polynomial, it may be advisable to trim the outliers as Imbens and Kalyanaraman (2012) and Calonico et al. (Forthcoming) have done.

References

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik, "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, Forthcoming.

_ , _ , **and** _ , "Robust Data-Driven Inference in the Regression-Discontinuity Design," *Stata Journal*, in press.

Card, David, David S. Lee, Zhuan Pei, and Andrea Weber, "Nonlinear Policy Rules and the Identification and Estimation of Causal Effects in a Generalized Regression Kink Design," NBER Working Paper 18564 November 2012.

Cheng, Ming-Yen, Jianqing Fan, and J. S. Marron, "On automatic boundary corrections," *The Annals of Statistics*, 08 1997, *25* (4), 1691–1708.

Fan, Jianqing and Irene Gijbels, *Local Polynomial Modelling and Its Applications*, Chapman and Hall, 1996.

Gelman, Andrew and Guido Imbens, "Why High-order Polynomials Should Not Be Used in Regression Discontinuity Designs," NBER Working Paper 20405 August 2014.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw, "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 2001, *69* (1), 201–209.

Imbens, Guido and Karthik Kalyanaraman, "Optimal Bandwidth Choice for the Regression Discontinuity Estimator.," *Review of Economic Studies*, 2012, *79* (3), 933 – 959.

Lee, David S., "Randomized Experiments from Non-random Selection in U.S. House Elections," *Journal of Econometrics*, February 2008, *142* (2), 675–697.

Ludwig, J. and D. Miller, "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design," *Quarterly Journal of Economics*, 2007, *122(1)*, 159–208.

Nielsen, Helena Skyt, Torben Sørensen, and Christopher R. Taber, "Estimating the Effect of Student Aid on College Enrollment: Evidence from a Government Grant Policy Reform," *American Economic Journal: Economic Policy*, 2010, *2* (2), 185–215.

Table 1: Theoretical AMSE of the Conventional Estimator: Lee DGP

AMSE and Optimal Bandwidth for the **Uniform Kernel**

	AMSE $\times$ 1000		h_{opt}	
poly. order p	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$
0	4.42	0.795	0.036	0.015
1	4.12	0.526	0.130	0.078
2	4.33	0.477	0.260	0.180
3	4.11	0.417	0.470	0.353
4	3.52	0.339	0.838	0.663
Preferred p	$p = 4$	$p = 4$		

AMSE and Optimal Bandwidth for the **Triangular Kernel**

	AMSE $\times$ 1000		h_{opt}	
poly. order p	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$
0	4.09	0.735	0.052	0.022
1	3.89	0.496	0.166	0.099
2	4.14	0.456	0.311	0.216
3	3.96	0.402	0.542	0.407
4	3.41	0.329	0.943	0.767
Preferred p	$p = 4$	$p = 4$		

Note: The function form of the Lee DGP is

$$E[Y|X=x] = \begin{cases} 0.48 + 1.27x + 7.18x^2 + 20.21x^3 + 21.54x^4 + 7.33x^5 & \text{if } x < 0 \\ 0.52 + 0.84x - 3.00x^2 + 7.99x^3 - 9.01x^4 + 3.56x^5 & \text{if } x \geqslant 0 \end{cases}$$

Table 2: Theoretical AMSE of the Conventional Estimator: Ludwig-Miller DGP

AMSE and Optimal Bandwidth for the **Uniform Kernel**

	AMSE $\times$ 1000		h_{opt}	
poly. order p	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$
0	20.3	5.98	0.008	0.004
1	8.28	1.90	0.065	0.045
2	5.74	1.19	0.196	0.151
3	4.48	0.88	0.431	0.351
4	3.46	0.65	0.854	0.722
Preferred p	$p = 4$	$p = 4$		

AMSE and Optimal Bandwidth for the **Triangular Kernel**

	AMSE $\times$ 1000		h_{opt}	
poly. order p	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$
0	18.8	5.52	0.011	0.006
1	7.81	1.80	0.082	0.057
2	5.49	1.14	0.235	0.181
3	4.32	0.84	0.497	0.405
4	3.35	0.63	0.961	0.813
Preferred p	$p = 4$	$p = 4$		

Note: The function form of the Ludwig-Miller DGP is

$$E[Y|X=x] = \begin{cases} 3.71 + 2.30x + 3.28x^2 + 1.45x^3 + 0.23x^4 + 0.03x^5 & \text{if } x < 0 \\ 0.26 + 18.49x - 54.81x^2 + 74.30x^3 - 45.02x^4 + 9.83x^5 & \text{if } x \geqslant 0 \end{cases}$$

Table 3: Comparison of Polynomial Orders by Theoretical AMSE for the Conventional Estimator: Lee DGP

Polynomial order p	is preferred to $p=1$ for the conventional RD estimator when ...	
	Uniform Kernel	Triangular Kernel
0	$n<296$	$n<344$
2	$n>1167$	$n>1466$
3	$n>476$	$n>607$
4	$n>177$	$n>150$

Note: The comparison is based on the theoretical AMSE evaluated at the optimal bandwidth.

Table 4: Comparison of Polynomial Orders by Theoretical AMSE for the Conventional Estimator: Ludwig-Miller DGP

Polynomial order p	is preferred to $p=1$ for the conventional RD estimator when ...	
	Uniform Kernel	Triangular Kernel
0	Never	Never
2	Always	Always
3	Always	Always
4	Always	Always

Note: The comparison is based on the theoretical AMSE evaluated at the optimal bandwidth(s).

Table 5: MSE of the Conventional Estimator and Coverage Rates of 95% CI from Simulations Using the Infeasible Optimal Bandwidth: Lee DGP

MSE, Coverage Rates and Infeasible Optimal Bandwidth for the **Uniform Kernel**

	MSE $\times$ 1000		Coverage Rates		h_{opt}	
poly. order p	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$
0	4.43	0.778	0.878	0.895	0.036	0.015
1	3.96	0.492	0.924	0.930	0.130	0.078
2	4.31	0.449	0.932	0.936	0.260	0.180
3	4.28	0.406	0.931	0.943	0.470	0.353
4	4.24	0.356	0.930	0.939	0.838	0.663
Preferred p	$p = 1$	$p = 4$				

MSE, Coverage Rates and Infeasible Optimal Bandwidth for the **Triangular Kernel**

	MSE $\times$ 1000		Coverage Rates		h_{opt}	
poly. order p	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$
0	4.09	0.720	0.871	0.891	0.036	0.015
1	3.67	0.463	0.920	0.929	0.130	0.078
2	4.08	0.431	0.926	0.938	0.260	0.180
3	4.13	0.391	0.929	0.942	0.470	0.353
4	4.18	0.349	0.927	0.938	0.838	0.663
Preferred p	$p = 1$	$p = 4$				

Note: The simulation is based 10,000 repetitions.

Table 6: MSE of the Conventional Estimator and Coverage Rates of 95% CI from Simulations Using the Infeasible Optimal Bandwidth: Ludwig-Miller DGP

MSE, Coverage Rates and Infeasible Optimal Bandwidth for the **Uniform Kernel**

	MSE $\times$ 1000		Coverage Rates		h_{opt}	
poly. order p	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$
1	9.28	1.862	0.921	0.923	0.065	0.045
2	5.98	1.119	0.935	0.939	0.196	0.151
3	4.65	0.842	0.943	0.942	0.431	0.351
4	3.97	0.643	0.950	0.949	0.854	0.722
Preferred p	$p = 4$	$p = 4$				

MSE, Coverage Rates and Infeasible Optimal Bandwidth for the **Triangular Kernel**

	MSE $\times$ 1000		Coverage Rates		h_{opt}	
poly. order p	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$
1	8.58	1.742	0.911	0.924	0.082	0.057
2	5.68	1.072	0.930	0.936	0.235	0.181
3	4.47	0.808	0.940	0.941	0.497	0.405
4	3.97	0.636	0.948	0.947	0.961	0.813
Preferred p	$p = 4$	$p = 4$				

Note: The simulation is based on 10000 repetitions. The program–available upon request–is a variant of the Stata package described in Calonico et al. (in press). The simulation for the Ludwig-Miller DGP does not include the local constant ($p = 0$) estimator, because too few observations lie within its corresponding bandwidth under when $n = 500$ or 3138.

Table 7: MSE of the Conventional Estimator and Coverage Rates of 95% CI from Simulations Using the CCT Bandwidth: Lee DGP

MSE, Coverage Rates and CCT Bandwidth for the **Uniform Kernel**

	MSE $\times$ 1000		Coverage Rates		$\hat{h}_{CCT}$	
poly. order p	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$
0	4.99	0.938	0.728	0.734	0.059	0.023
1	3.89	0.578	0.902	0.819	0.160	0.111
2	5.35	0.493	0.925	0.898	0.225	0.208
3	7.73	0.493	0.926	0.944	0.275	0.298
4	10.9	0.644	0.924	0.944	0.313	0.348
Preferred p	$p = 1$	$p = 2$				

MSE, Coverage Rates and CCT Bandwidth for the **Triangular Kernel**

	MSE $\times$ 1000		Coverage Rates		$\hat{h}_{CCT}$	
poly. order p	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$
0	4.54	0.853	0.722	0.743	0.084	0.032
1	3.77	0.533	0.892	0.822	0.205	0.139
2	5.24	0.472	0.916	0.898	0.271	0.248
3	7.66	0.489	0.920	0.939	0.316	0.344
4	10.8	0.644	0.916	0.940	0.350	0.390
Preferred p	$p = 1$	$p = 2$				

Note: The simulation is based on 10000 repetitions. The program–available upon request–is a variant of the Stata package described in Calonico et al. (in press).

The reported CCT bandwidth, $\hat{h}_{CCT}$, is the average over repeated simulation samples.

Table 8: MSE of the Conventional Estimator and Coverage Rates of 95% CI from Simulations Using the CCT Bandwidth: Ludwig-Miller DGP

MSE, Coverage Rates and CCT Bandwidth for the **Uniform Kernel**

	MSE $\times$ 1000		Coverage Rates		$\hat{h}_{CCT}$	
poly. order p	$n=500$	$n=3138$	$n=500$	$n=3138$	$n=500$	$n=3138$
1	10.0	1.99	0.878	0.889	0.076	0.050
2	6.45	1.22	0.920	0.908	0.206	0.166
3	7.58	1.04	0.940	0.940	0.280	0.293
4	10.7	1.34	0.936	0.946	0.319	0.343
Preferred p	$p=2$	$p=3$				

MSE, Coverage Rates and CCT Bandwidth for the **Triangular Kernel**

	MSE $\times$ 1000		Coverage Rates		$\hat{h}_{CCT}$	
poly. order p	$n=500$	$n=3138$	$n=500$	$n=3138$	$n=500$	$n=3138$
1	9.35	1.87	0.866	0.889	0.097	0.063
2	6.17	1.15	0.916	0.911	0.246	0.198
3	7.52	1.02	0.935	0.938	0.322	0.338
4	10.7	1.33	0.932	0.942	0.356	0.385
Preferred p	$p=2$	$p=3$				

Note: The simulation is based on 10000 repetitions. The program–available upon request–is a variant of the Stata package described in Calonico et al. (in press). The simulation for the Ludwig-Miller DGP does not include the local constant ($p=0$) estimator, because too few observations lie within its corresponding bandwidth under when $n=500$ or 3138.

The reported CCT bandwidth, $\hat{h}_{CCT}$, is the average over repeated simulation samples.

Table 9: MSE of the Conventional Estimator and Coverage Rates of 95% CI from Simulations Using the CCT Bandwidth without Regularization: Lee DGP

MSE, Coverage Rates and CCT Bandwidth for the **Uniform Kernel**

	MSE $\times$ 1000		Coverage Rates		$\hat{h}_{CCT,noreg}$	
poly. order p	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$
0	6.24	1.01	0.601	0.678	0.088	0.025
1	3.28	0.769	0.803	0.660	0.383	0.165
2	4.55	0.668	0.860	0.781	0.426	0.274
3	6.17	0.680	0.892	0.841	0.465	0.427
4	8.15	0.536	0.914	0.918	0.490	0.543
Preferred p	$p = 1$	$p = 4$				

MSE, Coverage Rates and CCT Bandwidth for the **Triangular Kernel**

	MSE $\times$ 1000		Coverage Rates		$\hat{h}_{CCT,noreg}$	
poly. order p	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$
0	5.48	0.898	0.620	0.706	0.116	0.035
1	3.19	0.639	0.809	0.715	0.472	0.188
2	4.55	0.566	0.861	0.817	0.482	0.309
3	6.37	0.592	0.886	0.860	0.519	0.475
4	8.47	0.548	0.905	0.917	0.539	0.594
Preferred p	$p = 1$	$p = 4$				

Note: The simulation is based on 10000 repetitions. The program–available upon request–is a variant of the Stata package described in Calonico et al. (in press).

The reported CCT bandwidth, $\hat{h}_{CCT,noreg}$, is the average over repeated simulation samples.

Table 10: MSE of the Conventional Estimator and Coverage Rates of 95% CI from Simulations Using the CCT Bandwidth without Regularization: Ludwig-Miller DGP

MSE, Coverage Rates and CCT Bandwidth for the **Uniform Kernel**

	MSE × 1000		Coverage Rates		$\hat{h}_{CCT,noreg}$	
poly. order p	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$
1	10.2	2.00	0.866	0.886	0.079	0.051
2	7.16	1.24	0.877	0.902	0.240	0.173
3	6.17	1.07	0.908	0.897	0.446	0.388
4	7.87	1.04	0.931	0.934	0.504	0.539
Preferred p	$p = 3$	$p = 4$				

MSE, Coverage Rates and CCT Bandwidth for the **Triangular Kernel**

	MSE × 1000		Coverage Rates		$\hat{h}_{CCT,noreg}$	
poly. order p	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$
1	9.51	1.88	0.858	0.885	0.100	0.064
2	6.74	1.16	0.882	0.905	0.283	0.204
3	6.18	1.02	0.908	0.899	0.502	0.440
4	8.15	1.07	0.922	0.930	0.555	0.595
Preferred p	$p = 3$	$p = 3$				

Note: The simulation is based on 10000 repetitions. The program–available upon request–is a variant of the Stata package described in Calonico et al. (in press). The simulation for the Ludwig-Miller DGP does not include the local constant ($p = 0$) estimator, because too few observations lie within its corresponding bandwidth under when $n = 500$ or 3138.

The reported CCT bandwidth, $\hat{h}_{CCT,noreg}$, is the average over repeated simulation samples.

Table 11: Estimated AMSE from Simulations Using the CCT Bandwidths: Lee DGP

Estimated AMSE, Preferred p and CCT Bandwidths for the **Uniform Kernel**

	$\widehat{\text{AMSE}} \times 1000$		Fraction of Times Preferred		$\hat{h}_{CCT}$	
poly. order p	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$
0	2.59	0.527	0.719	0.001	0.059	0.023
1	3.14	0.346	0.277	0.777	0.160	0.111
2	4.92	0.391	0.004	0.183	0.225	0.208
3	7.28	0.466	0.000	0.039	0.275	0.298
4	10.4	0.619	0.000	0.001	0.313	0.348
Most Likely Choice of p			$p = 0$	$p = 1$		

Estimated AMSE, preferred p and CCT Bandwidths for the **Triangular Kernel**

	$\widehat{\text{AMSE}} \times 1000$		Fraction of Times Preferred		$\hat{h}_{CCT}$	
poly. order p	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$
0	2.41	0.505	0.729	0.000	0.084	0.032
1	2.91	0.334	0.270	0.861	0.205	0.139
2	4.64	0.377	0.001	0.109	0.271	0.248
3	6.97	0.450	0.000	0.030	0.316	0.344
4	10.0	0.602	0.000	0.000	0.350	0.390
Most Likely Choice of p			$p = 0$	$p = 1$		

Note: The simulation is based 10,000 repetitions. The program–available upon request–is a variant of the Stata package described in Calonico et al. (in press). The reported AMSE measure and CCT bandwidth, $\widehat{\text{AMSE}} \times 1000$ and $\hat{h}_{CCT}$ are the averages over repeated simulation samples.

Table 12: Estimated AMSE from Simulations Using the CCT Bandwidths: Ludwig-Miller DGP

Estimated AMSE, Preferred p and CCT Bandwidths for the **Uniform Kernel**

	$\widehat{\text{AMSE}} \times 1000$		Fraction of Times Preferred		$\hat{h}_{CCT}$	
poly. order p	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$
1	8.97	1.73	0.022	0.000	0.076	0.050
2	6.33	1.06	0.874	0.284	0.206	0.166
3	8.43	1.00	0.103	0.705	0.280	0.293
4	12.6	1.33	0.001	0.011	0.319	0.343
Most Likely Choice of p			$p = 2$	$p = 3$		

Estimated AMSE, Preferred p and CCT Bandwidths for the **Triangular Kernel**

	$\widehat{\text{AMSE}} \times 1000$		Fraction of Times Preferred		$\hat{h}_{CCT}$	
poly. order p	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$
1	8.60	1.63	0.015	0.000	0.097	0.063
2	6.17	1.02	0.915	0.271	0.246	0.198
3	8.30	0.96	0.069	0.725	0.322	0.338
4	12.5	1.29	0.001	0.005	0.356	0.385
Most Likely Choice of p			$p = 2$	$p = 3$		

Note: The simulation is based on 10000 repetitions. The program–available upon request–is a variant of the Stata package described in Calonico et al. (in press). Unlike in Tables 6, 8 and 10, results are reported for the local constant estimator. This is because only the optimal bandwidth for the local estimator is calculated and not the value of local estimator itself.

The reported AMSE measure and CCT bandwidths, $\widehat{\text{AMSE}} \times 1000$ and $\hat{h}_{CCT}$, are the averages over repeated simulation samples.

Table 13: Estimated AMSE from Simulations Using the CCT Bandwidths without Regularization: Lee DGP

Estimated AMSE, Preferred p and CCT Bandwidths w/o Regularization for the **Uniform Kernel**

	$\widehat{\text{AMSE}} \times 1000$		Fraction of Times Preferred		$\hat{h}_{CCT,noreg}$	
poly. order p	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$
0	2.31	0.500	0.383	0.004	0.087	0.025
1	2.38	0.320	0.575	0.644	0.383	0.165
2	4.97	0.408	0.041	0.175	0.426	0.274
3	10.5	0.585	0.001	0.121	0.465	0.427
4	13.4	0.776	0.000	0.056	0.490	0.543
Most Likely Choice of p			$p = 1$	$p = 1$		

Estimated AMSE, Preferred p and CCT Bandwidths w/o Regularization for the **Triangular Kernel**

	$\widehat{\text{AMSE}} \times 1000$		Fraction of Times Preferred		$\hat{h}_{CCT,noreg}$	
poly. order p	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$
0	2.19	0.488	0.405	0.001	0.116	0.035
1	2.19	0.310	0.567	0.753	0.472	0.188
2	3.84	0.363	0.027	0.086	0.482	0.309
3	6.27	0.472	0.001	0.097	0.518	0.475
4	10.6	0.556	0.000	0.064	0.539	0.594
Most Likely Choice of p			$p = 1$	$p = 1$		

Note: The simulation is based 10,000 repetitions. The program–available upon request–is a variant of the Stata package described in Calonico et al. (in press). The reported AMSE measure and CCT bandwidths, $\widehat{\text{AMSE}} \times 1000$ and $\hat{h}_{CCT,noreg}$, are the averages over repeated simulation samples.

Table 14: Estimated AMSE from Simulations Using the CCT Bandwidths: Ludwig-Miller DGP

Estimated AMSE, Preferred p and CCT Bandwidths w/o Regularization for the **Uniform Kernel**

	$\widehat{\text{AMSE}} \times 1000$		Fraction of Times Preferred		$\hat{h}_{CCT,noreg}$	
poly. order p	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$
1	8.77	1.72	0.006	0.000	0.079	0.051
2	5.89	1.08	0.448	0.058	0.240	0.173
3	8.11	1.13	0.495	0.680	0.446	0.388
4	18.8	1.76	0.051	0.263	0.504	0.539
Most Likely Choice of p			$p = 3$	$p = 3$		

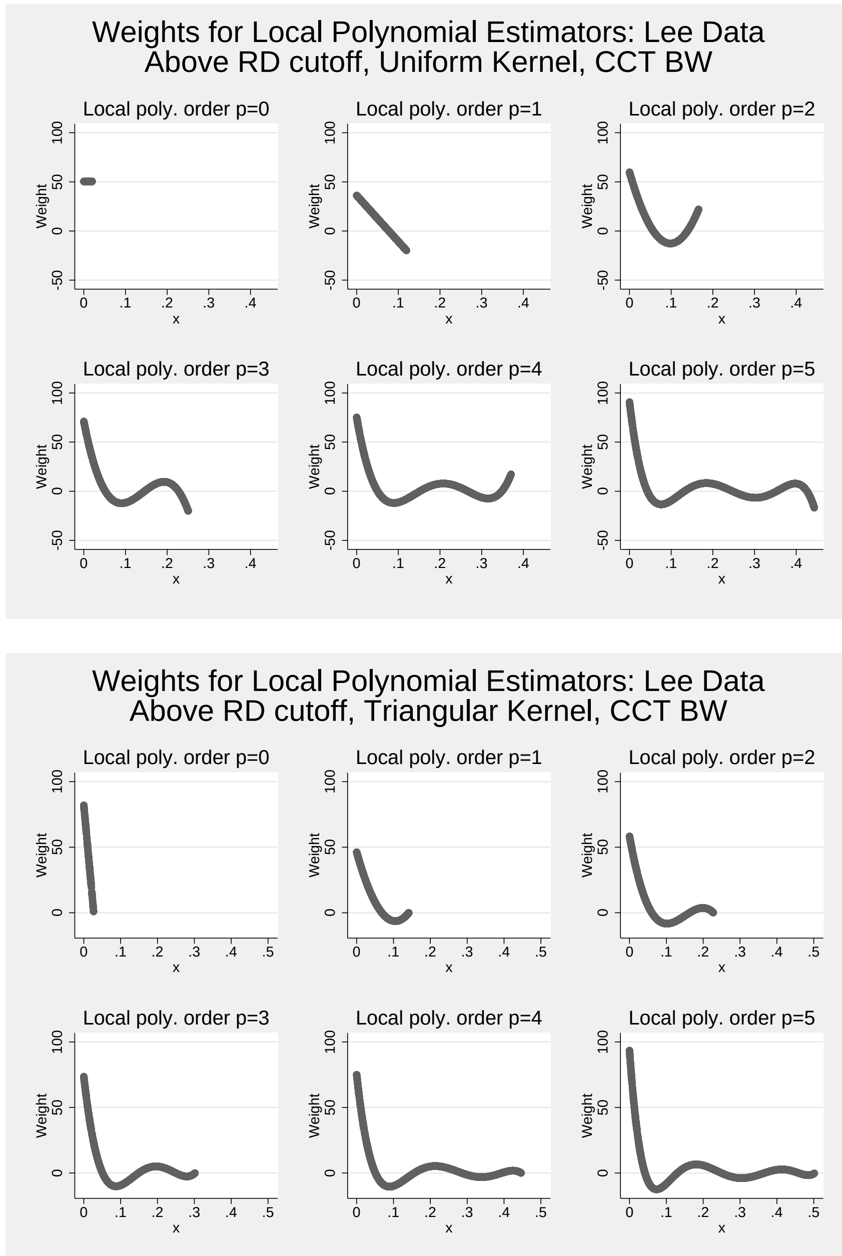
Estimated AMSE, Preferred p and CCT Bandwidths w/o Regularization for the **Triangular Kernel**

	$\widehat{\text{AMSE}} \times 1000$		Fraction of Times Preferred		$\hat{h}_{CCT,noreg}$	
poly. order p	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$
1	8.44	1.63	0.007	0.000	0.100	0.064
2	5.96	1.01	0.438	0.069	0.283	0.204
3	7.17	0.921	0.511	0.665	0.502	0.440
4	14.0	1.19	0.044	0.266	0.555	0.595
Most Likely Choice of p			$p = 3$	$p = 3$		

Note: The simulation is based on 10000 repetitions. The program–available upon request–is a variant of the Stata package described in Calonico et al. (in press). Unlike in Tables 6, 8 and 10, results are reported for the local constant estimator. This is because only the optimal bandwidth for the local estimator is calculated and not the value of local estimator itself.

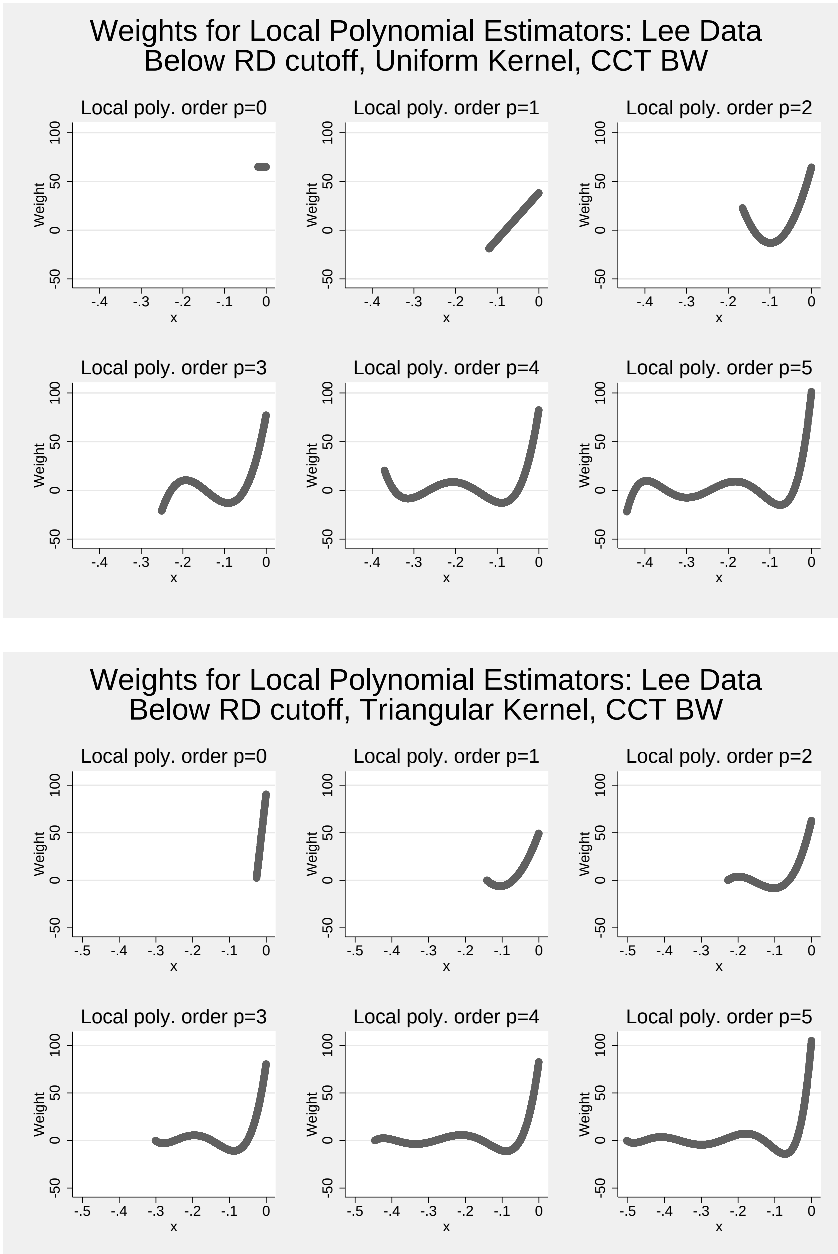
The reported AMSE measure and CCT bandwidths, $\widehat{\text{AMSE}} \times 1000$ and $\hat{b}_{CCT,noreg}$, are the averages over repeated simulation samples.

Figure 1: Weights for Local Polynomial Estimators: Lee Data, above the RD Cutoff, CCT Bandwidth



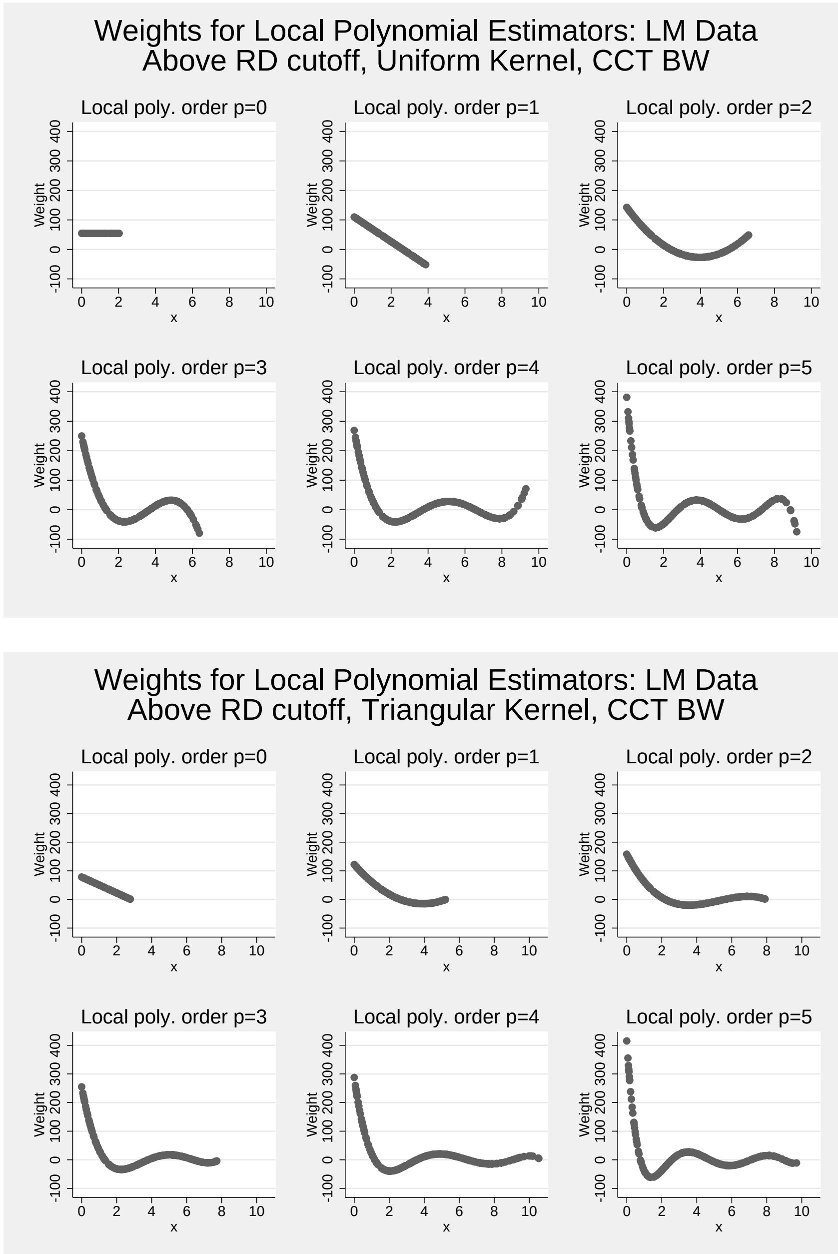
Note: The graphs plot the GI weights for the local estimator $\hat{\tau}_p$ with the default CCT bandwidth, for $p = 0, 1, ..., 5$.

Figure 2: Weights for Local Polynomial Estimators: Lee Data, below the RD Cutoff, CCT Bandwidth



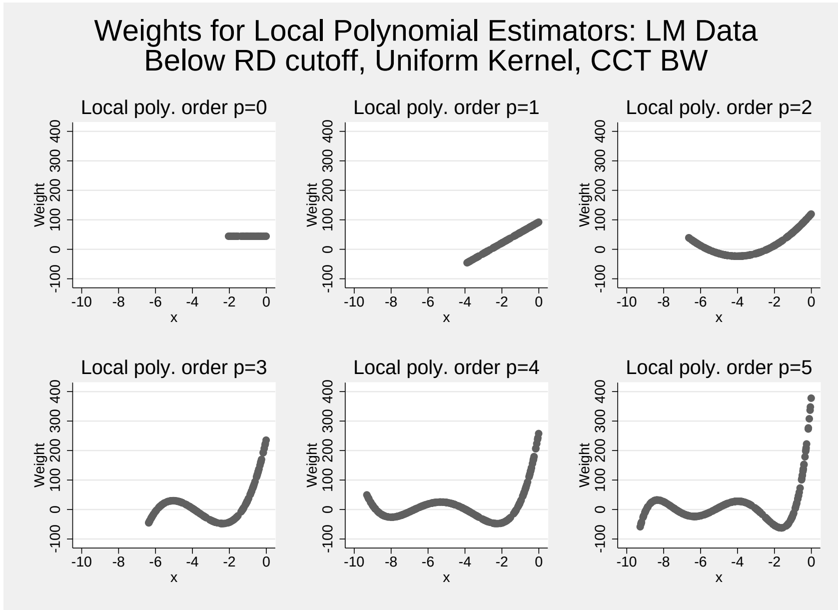
Note: The graphs plot the GI weights for the local estimator $\hat{\tau}_p$ with the default CCT bandwidth, for $p = 0, 1, \ldots, 5$.

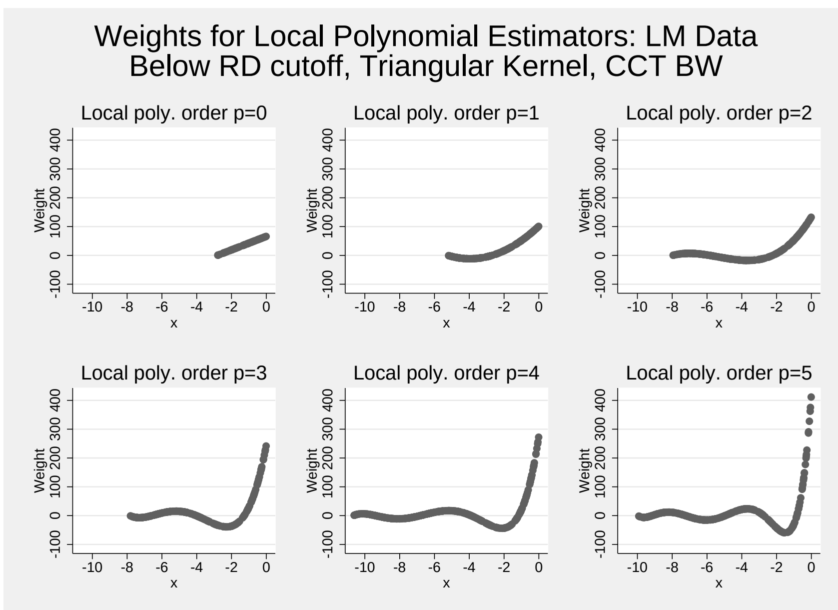
Figure 3: Weights for Local Polynomial Estimators: Ludwig-Miller Data, above the RD Cutoff, CCT Bandwidth



Note:The graphs plot the GI weights for the local estimator $\hat{\tau}_p$ with the default CCT bandwidth, for $p = 0, 1, ..., 5$.

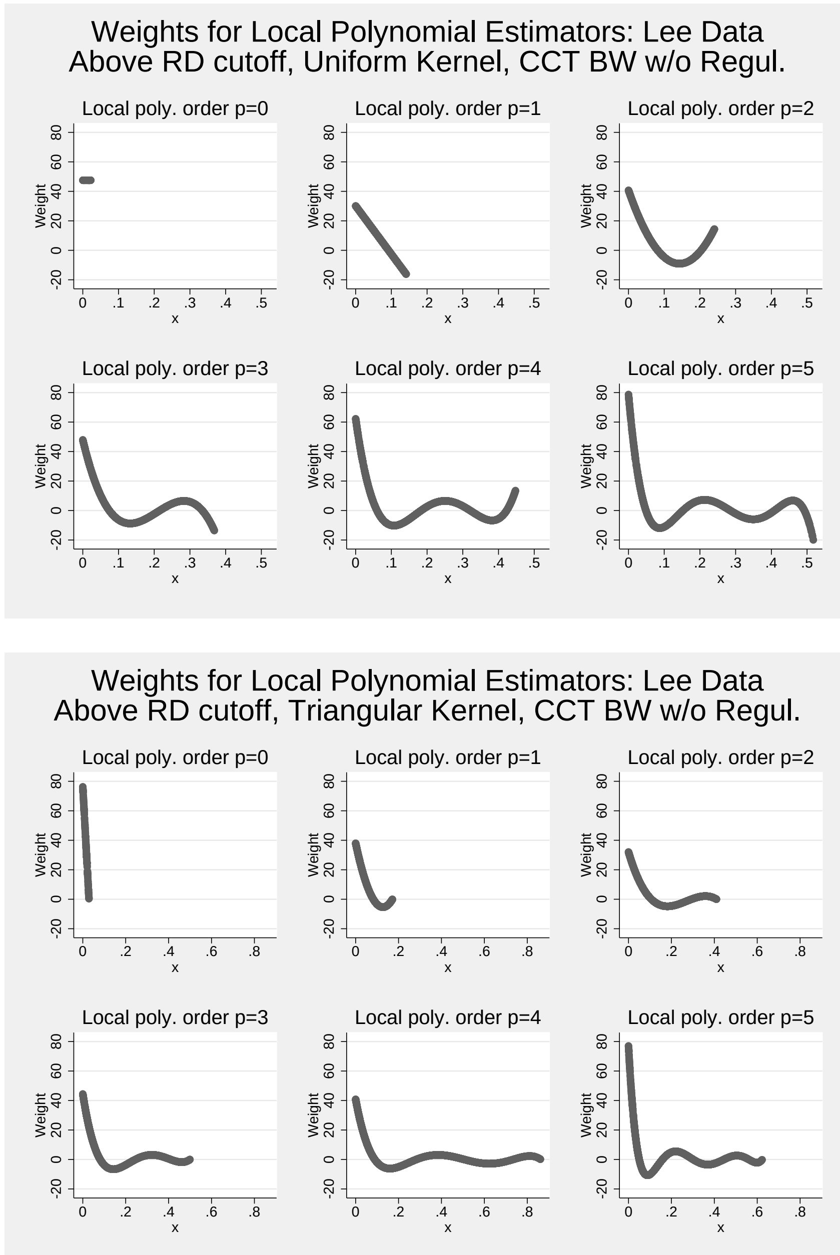
Figure 4: Weights for Local Polynomial Estimators: Ludwig-Miller Data, below the RD Cutoff, CCT Bandwidth





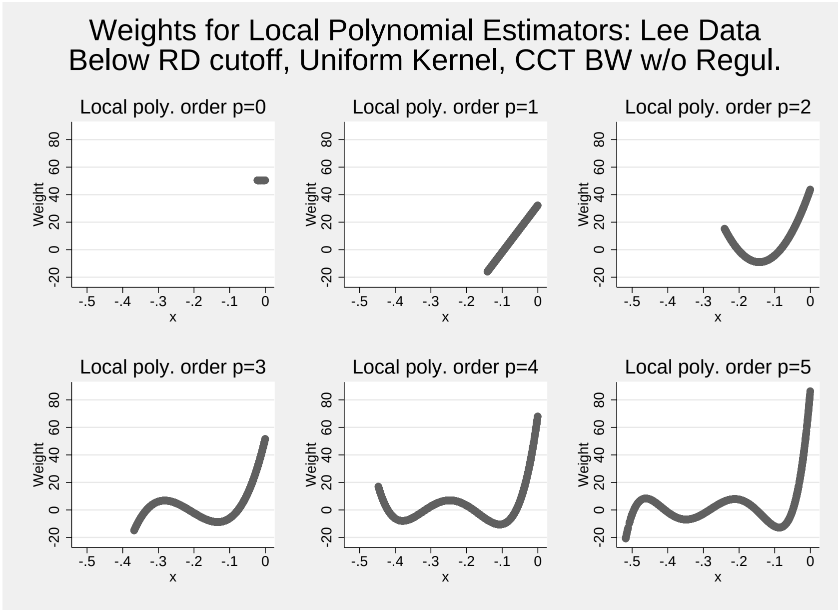
Note: The graphs plot the GI weights for the local estimator $\hat{\tau}_p$ with the default CCT bandwidth, for $p = 0, 1, ..., 5$.

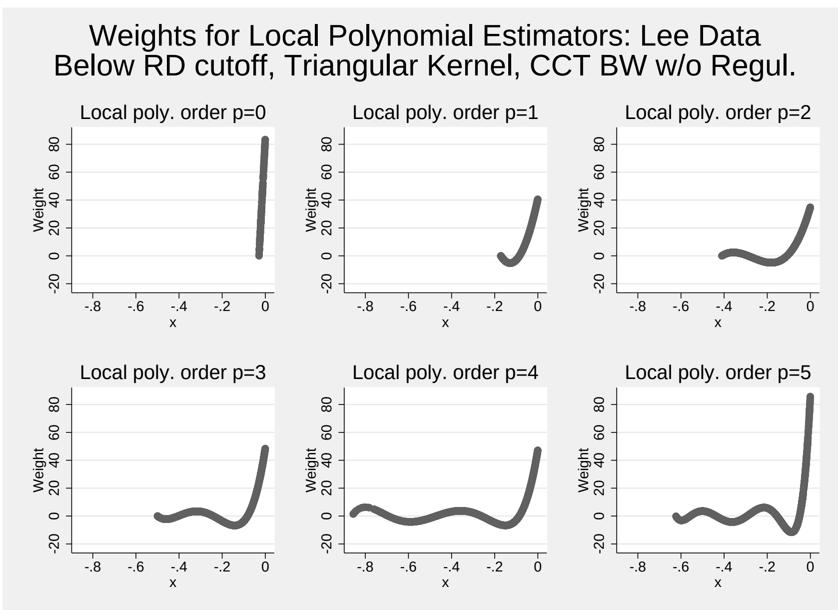
Figure 5: Weights for Local Polynomial Estimators: Lee Data, above the RD Cutoff, CCT Bandwidth without Regularization



Note: The graphs plot the GI weights for the local estimator $\hat{\tau}_p$ with the no-regularization CCT bandwidth, for $p = 0, 1, \ldots, 5$.

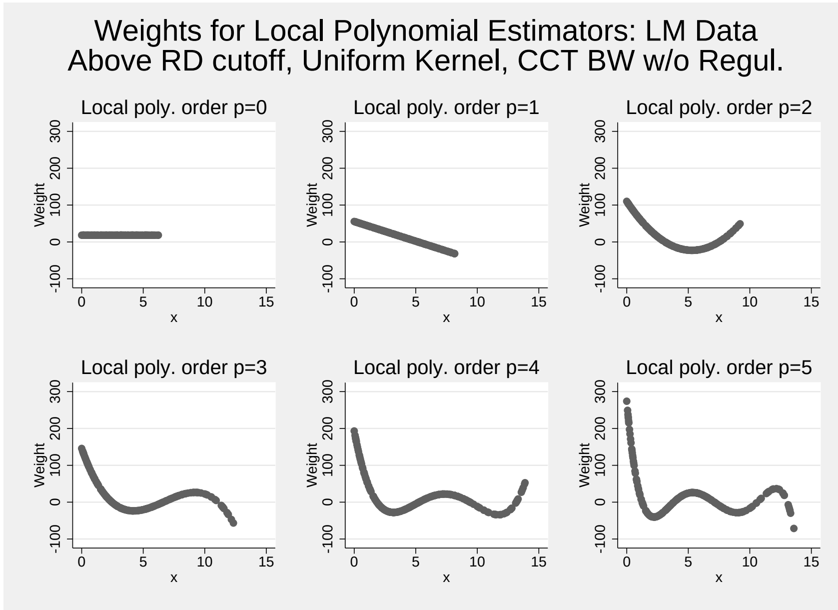
Figure 6: Weights for Local Polynomial Estimators: Lee Data, below the RD Cutoff, CCT Bandwidth without Regularization

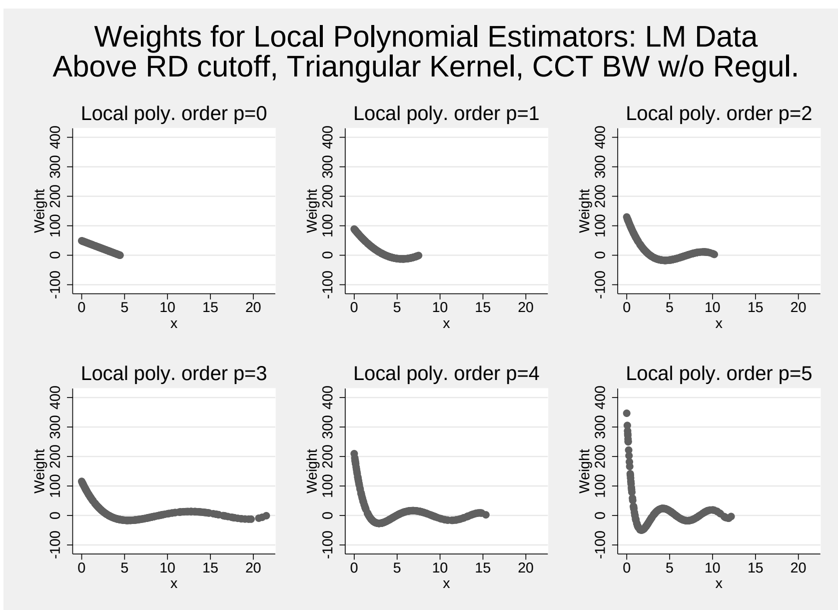




Note: The graphs plot the GI weights for the local estimator $\hat{\tau}_p$ with the no-regularization CCT bandwidth, for $p = 0, 1, \ldots, 5$.

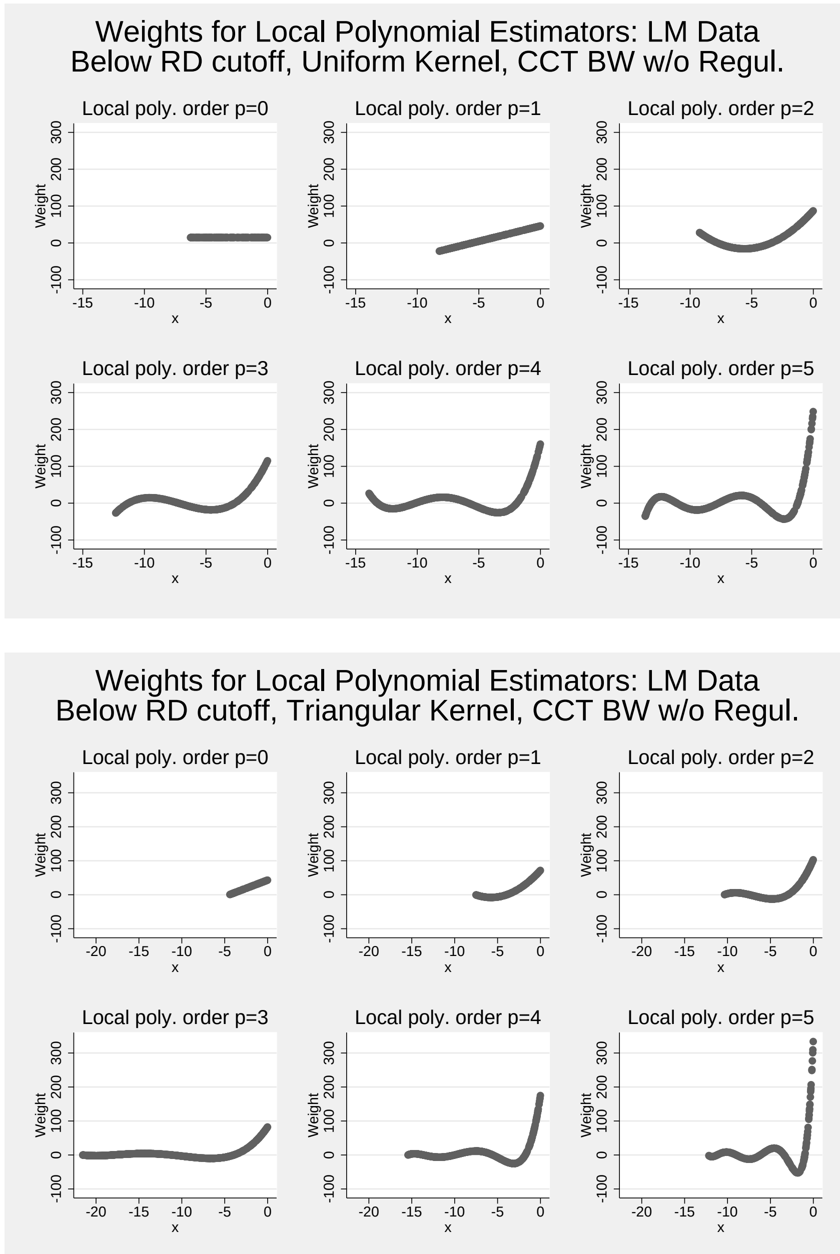
Figure 7: Weights for Local Polynomial Estimators: Ludwig-Miller Data, above the RD Cutoff, CCT Bandwidth without Regularization





Note:The graphs plot the GI weights for the local estimator $\hat{\tau}_p$ with the no-regularization CCT bandwidth, for $p = 0, 1, \ldots, 5$.

Figure 8: Weights for Local Polynomial Estimators: Ludwig-Miller Data, below the RD Cutoff, CCT Bandwidth without Regularization



Note: The graphs plot the GI weights for the local estimator $\hat{\tau}_p$ with the no-regularization CCT bandwidth, for $p = 0, 1, ..., 5$.

Supplemental Appendix

Table A.1: Theoretical AMSE of the Bias-corrected Estimator: Lee DGP

AMSE and Optimal Bandwidths for the **Uniform Kernel**

	AMSE $\times$ 1000		h_{opt}		b_{opt}	
poly. order p	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$
0	4.52	0.671	0.036	0.015	0.130	0.078
1	4.55	0.523	0.130	0.078	0.260	0.180
2	4.38	0.461	0.260	0.180	0.470	0.353
3	3.87	0.388	0.470	0.353	0.838	0.663
Preferred p	$p = 3$	$p = 3$				

AMSE and Optimal Bandwidths for the **Triangular Kernel**

	AMSE $\times$ 1000		h_{opt}		b_{opt}	
poly. order p	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$
0	4.26	0.682	0.052	0.022	0.166	0.099
1	4.28	0.525	0.166	0.099	0.311	0.216
2	4.18	0.460	0.311	0.216	0.542	0.407
3	3.41	0.329	0.542	0.407	0.943	0.767
Preferred p	$p = 3$	$p = 3$				

Note: The bias-corrected estimator is $\hat{\tau}^{bc}_{p,p+1}$. It is equal to the sum of the conventional RD estimator from a p-th order local regression and the bias correction term estimated from a $(p+1)$-th order local regression.

Table A.2: Theoretical AMSE of the Bias-corrected Estimator: Ludwig-Miller DGP

AMSE and Optimal Bandwidths for the **Uniform Kernel**

	AMSE × 1000		h_{opt}		b_{opt}	
poly. order p	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$
0	14.9	4.22	0.008	0.004	0.065	0.045
1	7.32	1.64	0.065	0.045	0.196	0.151
2	5.32	1.08	0.196	0.151	0.431	0.351
3	4.16	0.806	0.431	0.351	0.854	0.722
Preferred p	$p = 3$	$p = 3$				

AMSE and Optimal Bandwidths for the **Triangular Kernel**

	AMSE × 1000		h_{opt}		b_{opt}	
poly. order p	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$
0	13.9	4.06	0.011	0.006	0.082	0.057
1	6.95	1.61	0.082	0.057	0.235	0.181
2	5.10	1.07	0.235	0.181	0.497	0.405
3	4.01	0.794	0.497	0.405	0.961	0.813
Preferred p	$p = 3$	$p = 3$				

Note: The bias-corrected estimator is $\hat{\tau}^{bc}_{p,p+1}$. It is equal to the sum of the conventional RD estimator from a p-th order local regression and the bias correction term estimated from a $(p+1)$-th order local regression.

Table A.3: Comparison of Polynomial Orders by Theoretical AMSE for the Bias-Corrected Estimator: Lee DGP

Polynomial order p	is preferred to $p=1$ for the bias-corrected RD estimator when ...	
	Uniform Kernel	Triangular Kernel
0	$n < 531$	$n < 530$
2	$n > 137$	$n > 253$
3	$n > 10$	$n > 30$

Note: The comparison is based on the theoretical AMSE evaluated at the optimal bandwidths.

Table A.4: Comparison of Polynomial Orders by Theoretical AMSE for the Bias-Corrected Estimator: Ludwig-Miller DGP

Polynomial order p	is preferred to $p=1$ for the bias-corrected RD estimator when ...	
	Uniform Kernel	Triangular Kernel
0	Never	Never
2	Always	Always
3	Always	Always

Note: The comparison is based on the theoretical AMSE evaluated at the optimal bandwidths.

Table A.5: MSE of the Bias-Corrected Estimator and Coverage Rates of 95% Robust CI from Simulations Using the Infeasible Optimal Bandwidths: Lee DGP

MSE, Coverage Rates and Infeasible Optimal Bandwidths for the **Uniform Kernel**

	MSE × 1000		Coverage Rates		h_{opt}		b_{opt}	
poly. order p	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$
0	4.73	0.668	0.925	0.943	0.036	0.015	0.130	0.078
1	4.80	0.523	0.933	0.943	0.130	0.078	0.260	0.180
2	4.77	0.465	0.934	0.946	0.260	0.180	0.470	0.353
3	4.59	0.405	0.937	0.950	0.470	0.353	0.838	0.663
Preferred p	$p = 3$	$p = 3$						

MSE, Coverage Rates and Infeasible Optimal Bandwidths for the **Triangular Kernel**

	MSE × 1000		Coverage Rates		h_{opt}		b_{opt}	
poly. order p	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$
0	4.41	0.633	0.923	0.941	0.052	0.022	0.166	0.099
1	4.44	0.495	0.929	0.945	0.166	0.099	0.311	0.216
2	4.53	0.447	0.932	0.944	0.311	0.216	0.542	0.407
3	4.46	0.392	0.934	0.948	0.542	0.407	0.943	0.767
Preferred p	$p = 0$	$p = 3$						

Note: The bias-corrected estimator is $\hat{\tau}^{bc}_{p,p+1}$. It is equal to the sum of the conventional RD estimator from a p-th order local regression and the bias correction term estimated from a $(p+1)$-th order local regression.

The simulation is based 10,000 repetitions. The program–available upon request–is a variant of the Stata package described in Calonico et al. (in press).

Table A.6: MSE of the Bias-Corrected Estimator and Coverage Rates of 95% Robust CI from Simulations Using the CCT Bandwidth: Ludwig-Miller DGP

MSE, Coverage Rates and Infeasible Optimal Bandwidths for the **Uniform Kernel**

	$\text{MSE} \times 1000$		Coverage Rates		h_{opt}		b_{opt}	
poly. order p	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$
1	8.57	1.65	0.942	0.947	0.065	0.045	0.196	0.151
2	5.90	1.07	0.946	0.950	0.196	0.151	0.431	0.351
3	4.77	0.828	0.948	0.950	0.431	0.351	0.854	0.722
Preferred p	$p = 3$	$p = 3$						

MSE, Coverage Rates and Infeasible Optimal Bandwidths for the **Triangular Kernel**

	$\text{MSE} \times 1000$		Coverage Rates		h_{opt}		b_{opt}	
poly. order p	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$
1	7.91	1.56	0.933	0.944	0.082	0.057	0.235	0.181
2	5.62	1.04	0.941	0.946	0.235	0.181	0.497	0.405
3	4.64	0.804	0.945	0.948	0.497	0.405	0.961	0.813
Preferred p	$p = 3$	$p = 3$						

Note: The bias-corrected estimator is $\hat{\tau}^{bc}_{p,p+1}$. It is equal to the sum of the conventional RD estimator from a p-th order local regression and the bias correction term estimated from a $(p+1)$-th order local regression.

The simulation is based on 10000 repetitions. The program–available upon request–is a variant of the Stata package described in Calonico et al. (in press). The simulation for the Ludwig-Miller DGP does not include the local constant ($p = 0$) estimator, because too few observations lie within its corresponding bandwidth under when $n = 500$ or 3138.

Table A.7: MSE of the Bias-Corrected Estimator and Coverage Rates of 95% Robust CI from Simulations Using the CCT Bandwidths: Lee DGP

MSE, Coverage Rates and CCT Bandwidths for the **Uniform Kernel**

	MSE $\times$ 1000		Coverage Rates		$\hat{h}_{CCT}$		$\hat{b}_{CCT}$	
poly. order p	$n=500$	$n=6558$	$n=500$	$n=6558$	$n=500$	$n=6558$	$n=500$	$n=6558$
0	3.74	0.550	0.896	0.899	0.059	0.023	0.225	0.138
1	4.70	0.529	0.920	0.893	0.160	0.111	0.296	0.219
2	6.36	0.492	0.933	0.934	0.225	0.208	0.350	0.333
3	9.13	0.552	0.931	0.951	0.275	0.298	0.386	0.424
Preferred p	$p=0$	$p=2$						

MSE, Coverage Rates and CCT Bandwidths for the **Triangular Kernel**

	MSE $\times$ 1000		Coverage Rates		$\hat{h}_{CCT}$		$\hat{b}_{CCT}$	
poly. order p	$n=500$	$n=6558$	$n=500$	$n=6558$	$n=500$	$n=6558$	$n=500$	$n=6558$
0	3.46	0.521	0.900	0.905	0.084	0.032	0.271	0.161
1	4.63	0.495	0.914	0.905	0.205	0.139	0.333	0.245
2	6.33	0.479	0.926	0.933	0.271	0.248	0.382	0.364
3	9.15	0.555	0.925	0.945	0.316	0.344	0.413	0.453
Preferred p	$p=0$	$p=2$						

Note: The bias-corrected estimator is $\hat{\tau}^{bc}_{p,p+1}$. It is equal to the sum of the conventional RD estimator from a p-th order local regression and the bias correction term estimated from a $(p+1)$-th order local regression.
The simulation is based 10,000 repetitions. The program–available upon request–is a variant of the Stata package described in Calonico et al. (in press). The reported CCT bandwidths, $\hat{h}_{CCT}$ and $\hat{b}_{CCT}$, are the averages over repeated simulation samples.

Table A.8: MSE of the Bias-Corrected Estimator and Coverage Rates of 95% Robust CI from Simulations Using the CCT Bandwidth: Ludwig-Miller DGP

MSE, Coverage Rates and CCT Bandwidths for the **Uniform Kernel**

	MSE $\times$ 1000		Coverage Rates		$\hat{h}_{CCT}$		$\hat{b}_{CCT}$	
poly. order p	$n=500$	$n=3138$	$n=500$	$n=3138$	$n=500$	$n=3138$	$n=500$	$n=3138$
1	7.87	1.57	0.935	0.945	0.076	0.050	0.197	0.145
2	6.56	1.08	0.944	0.942	0.206	0.166	0.343	0.316
3	9.06	1.17	0.945	0.948	0.280	0.293	0.389	0.418
Preferred p	$p=2$	$p=2$						

MSE, Coverage Rates and CCT Bandwidths for the **Triangular Kernel**

	MSE $\times$ 1000		Coverage Rates		$\hat{h}_{CCT}$		$\hat{b}_{CCT}$	
poly. order p	$n=500$	$n=3138$	$n=500$	$n=3138$	$n=500$	$n=3138$	$n=500$	$n=3138$
1	7.41	1.51	0.931	0.941	0.097	0.063	0.223	0.164
2	6.44	1.05	0.939	0.943	0.246	0.198	0.376	0.346
3	9.03	1.17	0.939	0.944	0.322	0.338	0.418	0.448
Preferred p	$p=2$	$p=2$						

Note: The bias-corrected estimator is $\hat{\tau}^{bc}_{p,p+1}$. It is equal to the sum of the conventional RD estimator from a p-th order local regression and the bias correction term estimated from a $(p+1)$-th order local regression.

The simulation is based on 10000 repetitions. The program–available upon request–is a variant of the Stata package described in Calonico et al. (in press). The simulation for the Ludwig-Miller DGP does not include the local constant ($p=0$) estimator, because too few observations lie within its corresponding bandwidth under when $n=500$ or 3138.

The reported CCT bandwidths, $\hat{h}_{CCT}$ and $\hat{b}_{CCT}$, are the averages over repeated simulation samples.

Table A.9: MSE of the Bias-Corrected Estimator and Coverage Rates of 95% Robust CI from Simulations Using the CCT Bandwidths without Regularization: Lee DGP

MSE, Coverage Rates and CCT Bandwidths w/o Regularization for the **Uniform Kernel**

	MSE $\times$ 1000		Coverage Rates		$\hat{h}_{CCT,noreg}$		$\hat{b}_{CCT,noreg}$	
poly. order p	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$
0	4.00	0.573	0.842	0.866	0.088	0.025	0.440	0.177
1	4.84	0.626	0.869	0.843	0.383	0.165	0.433	0.248
2	7.08	0.595	0.915	0.903	0.426	0.274	0.473	0.380
3	12.9	0.640	0.934	0.940	0.465	0.427	0.502	0.537
Preferred p	$p = 0$	$p = 0$						

MSE, Coverage Rates and CCT Bandwidths for the **Triangular Kernel**

	MSE $\times$ 1000		Coverage Rates		$\hat{h}_{CCT,noreg}$		$\hat{b}_{CCT,noreg}$	
poly. order p	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$
0	3.52	0.530	0.848	0.880	0.116	0.035	0.548	0.191
1	4.57	0.517	0.867	0.873	0.472	0.188	0.487	0.271
2	6.05	0.489	0.906	0.910	0.482	0.309	0.518	0.414
3	9.04	0.545	0.914	0.936	0.519	0.475	0.539	0.581
Preferred p	$p = 0$	$p = 2$						

Note: The bias-corrected estimator is $\hat{\tau}^{bc}_{p,p+1}$. It is equal to the sum of the conventional RD estimator from a p-th order local regression and the bias correction term estimated from a $(p+1)$-th order local regression.
The simulation is based 10,000 repetitions. The program–available upon request–is a variant of the Stata package described in Calonico et al. (in press). The reported CCT bandwidths, $\hat{h}_{CCT,noreg}$ and $\hat{b}_{CCT,noreg}$, are the averages over repeated simulation samples.

Table A.10: MSE of the Bias-Corrected Estimator and Coverage Rates of 95% Robust CI from Simulations Using the CCT Bandwidth without regularization: Ludwig-Miller DGP

MSE, Coverage Rates and CCT Bandwidths for the **Uniform Kernel**

	MSE $\times$ 1000		Coverage Rates		$\hat{h}_{CCT,noreg}$		$\hat{b}_{CCT,noreg}$	
poly. order p	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$
1	7.87	1.57	0.935	0.945	0.076	0.050	0.197	0.145
2	6.56	1.08	0.944	0.942	0.206	0.166	0.343	0.316
3	9.06	1.17	0.945	0.948	0.280	0.293	0.389	0.418
Preferred p	$p = 2$	$p = 2$						

MSE, Coverage Rates and CCT Bandwidths for the **Triangular Kernel**

	MSE $\times$ 1000		Coverage Rates		$\hat{h}_{CCT,noreg}$		$\hat{b}_{CCT,noreg}$	
poly. order p	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$
1	7.41	1.51	0.931	0.941	0.097	0.063	0.223	0.164
2	6.44	1.05	0.939	0.943	0.246	0.198	0.376	0.346
3	9.03	1.17	0.939	0.944	0.322	0.338	0.418	0.448
Preferred p	$p = 2$	$p = 2$						

Note: The bias-corrected estimator is $\hat{\tau}^{bc}_{p,p+1}$. It is equal to the sum of the conventional RD estimator from a p-th order local regression and the bias correction term estimated from a $(p+1)$-th order local regression.
The simulation is based on 10000 repetitions. The program–available upon request–is a variant of the Stata package described in Calonico et al. (in press). The simulation for the Ludwig-Miller DGP does not include the local constant ($p = 0$) estimator, because too few observations lie within its corresponding bandwidth under when $n = 500$ or 3138.
The reported CCT bandwidths, $\hat{h}_{CCT,noreg}$ and $\hat{b}_{CCT,noreg}$, are the averages over repeated simulation samples.

Table A.11: Estimated AMSE of the Bias-corrected Estimator from Simulations Using the CCT Bandwidths: Lee DGP

Estimated AMSE, Preferred p and CCT Bandwidths for the **Uniform Kernel**

	$\widehat{AMSE} \times 1000$		Fraction of Times Preferred		$\hat{h}_{CCT}$		$\hat{b}_{CCT}$	
poly. order p	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$
0	2.77	0.422	0.929	0.299	0.059	0.023	0.225	0.138
1	4.06	0.400	0.070	0.566	0.160	0.111	0.296	0.219
2	6.05	0.448	0.000	0.120	0.225	0.208	0.350	0.333
3	8.73	0.537	0.001	0.015	0.275	0.298	0.386	0.424
Most Likely Choice of p			$p = 0$	$p = 1$				

Estimated AMSE, Preferred p and CCT Bandwidths for the **Triangular Kernel**

	$\widehat{\text{AMSE}} \times 1000$		Fraction of Times Preferred		$\hat{h}_{CCT}$		$\hat{b}_{CCT}$	
poly. order p	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$
0	2.62	0.415	0.977	0.212	0.084	0.032	0.271	0.161
1	3.87	0.391	0.023	0.674	0.205	0.139	0.333	0.245
2	5.82	0.436	0.000	0.102	0.271	0.248	0.382	0.364
3	8.49	0.525	0.000	0.012	0.316	0.344	0.413	0.453
Most Likely Choice of p			$p = 0$	$p = 1$				

Note: The bias-corrected estimator is $\hat{\tau}^{bc}_{p,p+1}$. It is equal to the sum of the conventional RD estimator from a p-th order local regression and the bias correction term estimated from a $(p+1)$-th order local regression.
The simulation is based 10,000 repetitions. The program–available upon request–is a variant of the Stata package described in Calonico et al. (in press). The reported CCT bandwidths, $\hat{h}_{CCT}$ and $\hat{b}_{CCT}$, are the averages over repeated simulation samples.

Table A.12: Estimated AMSE of the Bias-corrected Estimator from Simulations Using the CCT Bandwidths: Ludwig-Miller DGP

Estimated AMSE, Preferred p and CCT Bandwidths for the **Uniform Kernel**

	$\widehat{AMSE} \times 1000$		Fraction of Times Preferred		$\hat{h}_{CCT}$		$\hat{b}_{CCT}$	
poly. order p	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$
1	8.70	1.56	0.178	0.000	0.076	0.050	0.197	0.145
2	7.23	1.07	0.789	0.683	0.206	0.166	0.343	0.316
3	10.4	1.15	0.034	0.317	0.280	0.293	0.389	0.418
Most Likely Choice of p			$p = 2$	$p = 2$				

Estimated AMSE, Preferred p and CCT Bandwidths for the **Triangular Kernel**

	$\widehat{AMSE} \times 1000$		Fraction of Times Preferred		$\hat{h}_{CCT}$		$\hat{b}_{CCT}$	
poly. order p	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$
1	8.48	1.50	0.193	0.000	0.097	0.063	0.223	0.164
2	7.20	1.04	0.788	0.692	0.246	0.198	0.376	0.346
3	10.4	1.12	0.019	0.308	0.322	0.338	0.418	0.448
Most Likely Choice of p			$p = 2$	$p = 2$				

Note: The bias-corrected estimator is $\hat{\tau}^{bc}_{p,p+1}$. It is equal to the sum of the conventional RD estimator from a p-th order local regression and the bias correction term estimated from a $(p+1)$-th order local regression.
The simulation is based on 10000 repetitions. The program–available upon request–is a variant of the Stata package described in Calonico et al. (in press). The simulation for the Ludwig-Miller DGP does not include the local constant ($p = 0$) estimator, because too few observations lie within its corresponding bandwidth under when $n = 500$ or 3138.
The reported CCT bandwidths, $\hat{h}_{CCT}$ and $\hat{b}_{CCT}$, are the averages over repeated simulation samples.

Table A.13: Estimated AMSE of the Bias-corrected Estimator from Simulations Using the CCT Bandwidths without Regularization: Lee DGP

Estimated AMSE, Preferred p and CCT Bandwidths w/o Regularization for the **Uniform Kernel**

	$\widehat{\text{AMSE}} \times 1000$		Fraction of Times Preferred		$\hat{h}_{CCT,noreg}$		$\hat{b}_{CCT,noreg}$	
poly. order p	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$
0	2.44	0.383	0.884	0.324	0.087	0.025	0.439	0.177
1	8.78	0.946	0.111	0.423	0.383	0.165	0.433	0.248
2	23.5	1.156	0.006	0.135	0.426	0.274	0.472	0.380
3	73.0	1.700	0.000	0.118	0.465	0.427	0.502	0.537
Most Likely Choice of p			$p = 0$	$p = 1$				

Estimated AMSE, Preferred p and CCT Bandwidths for the **Triangular Kernel**

	$\widehat{\text{AMSE}} \times 1000$		Fraction of Times Preferred		$\hat{h}_{CCT,noreg}$		$\hat{b}_{CCT,noreg}$	
poly. order p	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$	$n = 500$	$n = 6558$
0	2.10	0.39	0.914	0.196	0.116	0.035	0.548	0.191
1	4.70	0.426	0.083	0.559	0.472	0.188	0.487	0.271
2	8.41	0.425	0.003	0.102	0.482	0.309	0.517	0.414
3	20.1	0.699	0.000	0.144	0.518	0.475	0.539	0.581
Most Likely Choice of p			$p = 0$	$p = 1$				

Note: The bias-corrected estimator is $\hat{\tau}^{bc}_{p,p+1}$. It is equal to the sum of the conventional RD estimator from a p-th order local regression and the bias correction term estimated from a $(p+1)$-th order local regression.
The simulation is based 10,000 repetitions. The program–available upon request–is a variant of the Stata package described in Calonico et al. (in press). The reported CCT bandwidths, $\hat{h}_{CCT,noreg}$ and $\hat{b}_{CCT,noreg}$, are the averages over repeated simulation samples.

Table A.14: Estimated AMSE of the Bias-corrected Estimator from Simulations Using the CCT Bandwidth without regularization: Ludwig-Miller DGP

Estimated AMSE, Preferred p and CCT Bandwidths for the **Uniform Kernel**

	$\widehat{AMSE} \times 1000$		Fraction of Times Preferred		$\hat{h}_{CCT,noreg}$		$\hat{b}_{CCT,noreg}$	
poly. order p	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$
1	8.28	1.53	0.072	0.000	0.079	0.051	0.206	0.148
2	8.71	1.38	0.748	0.323	0.240	0.173	0.424	0.335
3	68.2	3.96	0.180	0.677	0.446	0.388	0.507	0.538
Most Likely Choice of p			$p = 2$	$p = 3$				

Estimated AMSE, Preferred p and CCT Bandwidths for the **Triangular Kernel**

	$\widehat{AMSE} \times 1000$		Fraction of Times Preferred		$\hat{h}_{CCT,noreg}$		$\hat{b}_{CCT,noreg}$	
poly. order p	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$	$n = 500$	$n = 3138$
1	8.09	1.48	0.080	0.000	0.100	0.064	0.232	0.167
2	9.42	1.00	0.727	0.302	0.283	0.204	0.464	0.366
3	25.4	1.91	0.194	0.698	0.502	0.440	0.550	0.583
Most Likely Choice of p			$p = 2$	$p = 3$				

Note: The bias-corrected estimator is $\hat{\tau}^{bc}_{p,p+1}$. It is equal to the sum of the conventional RD estimator from a p-th order local regression and the bias correction term estimated from a $(p+1)$-th order local regression.
The simulation is based on 10000 repetitions. The program–available upon request–is a variant of the Stata package described in Calonico et al. (in press). The simulation for the Ludwig-Miller DGP does not include the local constant ($p = 0$) estimator, because too few observations lie within its corresponding bandwidth under when $n = 500$ or 3138.

The reported CCT bandwidths, $\hat{h}_{CCT,noreg}$ and $\hat{b}_{CCT,noreg}$, are the averages over repeated simulation samples.

Figure A.1: Weights for Global Polynomial Estimators: Lee Data

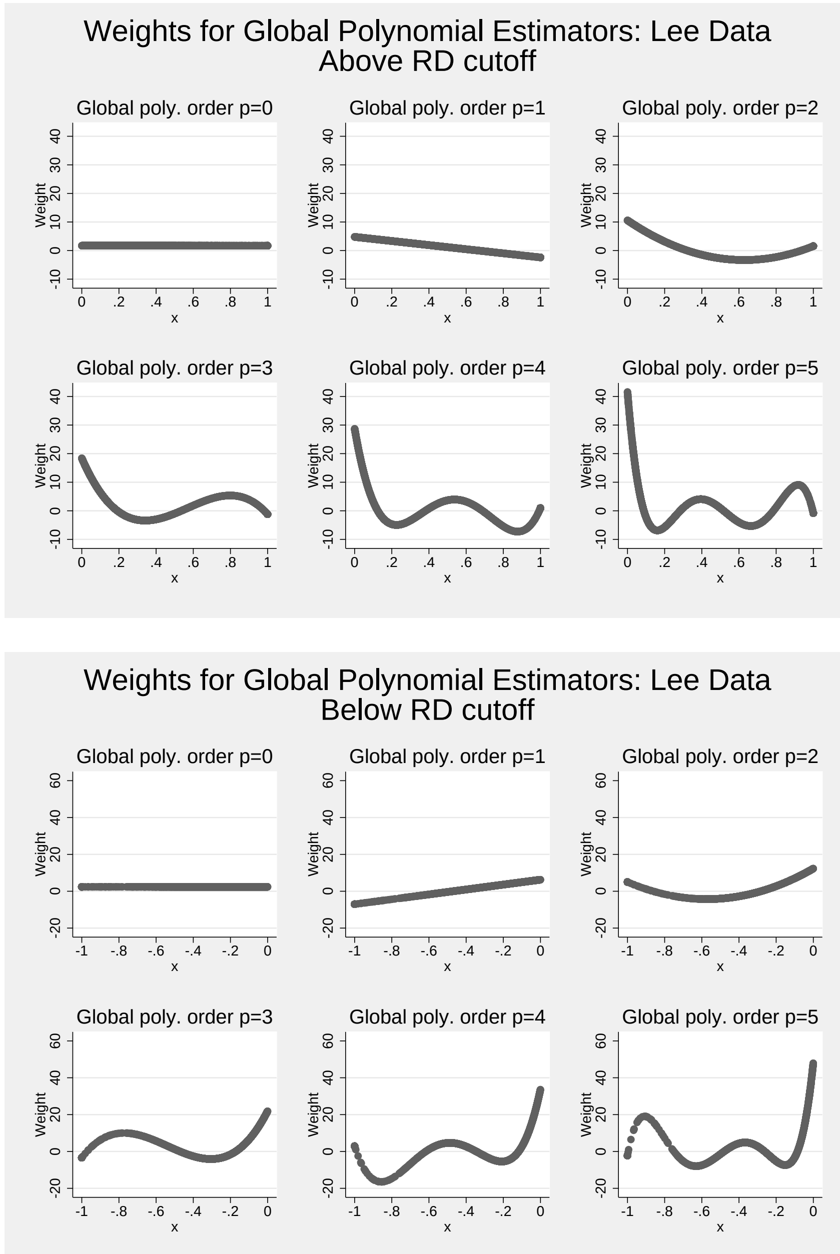


Figure A.2: Weights for Global Polynomial Estimators: Ludwig-Miller Data