A Permutation Test and Estimation Alternatives for the Regression Kink Design

Peter Ganong
Simon Jäger

I Z A

A Permutation Test and Estimation Alternatives for the Regression Kink Design

Peter Ganong
Harvard University

Simon Jäger
*Harvard University
and IZA*

ABSTRACT

A Permutation Test and Estimation Alternatives for the Regression Kink Design[*]

The Regression Kink (RK) design is an increasingly popular empirical method, with more than 20 studies circulated using RK in the last 5 years since the initial circulation of Card, Lee, Pei and Weber (2012). We document empirically that these estimates, which typically use local linear regression, are highly sensitive to curvature in the underlying relationship between the outcome and the assignment variable. As an alternative inference procedure, motivated by randomization inference, we propose that researchers construct a distribution of placebo estimates in regions *without* a policy kink. We apply our procedure to three empirical RK applications – two administrative UI datasets with true policy kinks and the 1980 Census, which has no policy kinks – and we find that statistical significance based on conventional p-values may be spurious. In contrast, our permutation test reinforces the asymptotic inference results of a recent Regression Discontinuity study and a Difference-in-Difference study. Finally, we propose estimating RK models with a modified cubic splines framework and test the performance of different estimators in a simulation exercise. Cubic specifications – in particular recently proposed robust estimators (Calonico, Cattaneo and Titiunik 2014) – yield short interval lengths with good coverage rates.

JEL Classification: C12, C13, C14, C31

Keywords: randomization inference, placebo test, cubic splines

Corresponding author:

Simon Jäger
Department of Economics
Harvard University
1805 Cambridge Street
Cambridge, MA 02138
USA
E-mail: jaeger@fas.harvard.edu

1 Introduction

The Regression Kink (RK) design has become an increasingly popular tool for causal inference in economics. Since the initial circulation of Card, Lee, Pei, and Weber (2012), abbreviated as CLPW in the following, more than 20 new studies have appeared using the RK design, as shown in Table 1. In the RK design, the effect of a continuous policy (e.g., the level of unemployment benefits, marginal tax rates) on an outcome of interest (e.g., unemployment duration, college enrollment) is evaluated by exploiting a change in the slope of the policy function that assigns the level of the policy based on a running variable (e.g., taxable income) at the kink point. The change in the slope of the relationship between the outcome variable and the running variable at the kink point divided by the change in the slope of the policy variable at the kink point identifies a treatment-on-the-treated parameter of interest under mild regularity conditions. In most applications of the RK design where the researchers use only data close to the kink point, local linear or quadratic regression is used.

Using a linear specification for estimating the change in slope at the kink point is problematic: any curvature in the conditional expectation function of the outcome variable results in a first-order bias in the RK estimate.[1] To illustrate how curvature in the relationship between the outcome and the running variable biases linear RK estimates, we have generated data with a piecewise linear data generating process (DGP) featuring a kink and a quadratic DGP with no kink. The top panel of Figure 1 shows the relationship between the outcome variable and the running variable for both the piecewise linear and the quadratic DGP. Visually, the relationships between the outcome and the running variable seem to be quite similar. In the second panel of Figure 1, we display the data for the piecewise linear DGP and add the predictions from a local linear model.

The point estimate for the slope change is statistically highly significant (t-statistic over 10) and close to the actual slope change in the underlying DGP.[2] The third panel of Figure

---

[1]See Calonico, Cattaneo, and Titiunik (forthcoming) who derive the asymptotic properties of regression discontinuity estimators. Applying their results to the case of linear RK estimators, one finds that the first order bias of the estimator is proportional to curvature, i.e., the sum of second derivatives at the kink point.

[2]In the RK literature, following the recommendation of CLPW, practitioners typically report heteroskedasticity-robust standard errors. We refer to these as "conventional" standard errors, so as to dis-

1 shows predictions from a linear RK model estimated on the quadratic DGP which features no kink. Even though the change in slope is in fact zero at the point where we allow for a slope change in the local linear regression model, the RK point estimate is similar to the one in the piecewise linear model and statistically highly significant (t-statistic over 10). This illustrates how curvature can bias linear RK estimates. In the regression discontinuity (RD) context (Hahn, Todd, and Van der Klaauw, 2001) – where the object of interest is a change in the level of a function at a point – it is widely accepted that local linear regression (or higher order polynomial regression) has superior properties compared to estimators which compute the difference in mean levels around the discontinuity (Porter, 2003; Imbens and Lemieux, 2008).

We explore ways in which data outside the estimation region typically used by local polynomial regressions can be used to improve inference and estimation in RK models. We propose a simple relabeling permutation test to assess how RK estimates are affected by curvature inspired by Fisher's randomization inference. Randomization inference has a long tradition in the statistics literature (Fisher, 1935; Lehmann and Stein, 1949; Welch and Gutierrez, 1988; Welch, 1990; Rosenbaum, 2001; Ho and Imai, 2006, see Rosenbaum, 2002, for an introduction) and has seen new interest in recent years from econometricians (see, for instance, Bertrand, Duflo, and Mullainathan, 2004; Imbens and Rosenbaum, 2005; Chetty, Looney, and Kroft, 2009; Abadie, Diamond, and Hainmueller, 2010; Abadie, Athey, Imbens, and Wooldridge, 2014; Cattaneo, Frandsen, and Titiunik, forthcoming). In the RD setting, for instance, Imbens and Lemieux (2008) suggest "testing for a zero effect in settings where it is known that the effect should be 0". Lee's (2008) seminal RD study of the impact of incumbency on reelection plots the data for a wide bandwidth and notes:

> As apparent from the figure, there is a striking discontinuous jump, right at the 0
> point. [...] The causal effect is enormous: about 0.45 in probability. Nowhere else
> is a jump apparent, as there is a well-behaved, smooth relationship between the

---

tinguish them from the Calonico et al. (forthcoming)'s inference procedure, which they call "robust" standard errors.

3

two variables, except at the threshold that determines victory or defeat.

In a sense, the Imbens and Lemieux (2008) suggestion can be thought of as a formalization of Lee's quoted argument. In this spirit, we relabel the location of the policy kink by estimating slope changes in regions where there is no change in the slope of the policy. Under a null hypothesis that treatment has no effect on the outcome and the assumption that the policy kink is placed at random, the placebo estimates provide a null distribution for the test statistic at the policy kink.

To assess the properties of RK estimators and the permutation test, we apply our methodology to three empirical settings. First, we evaluate its performance in a setting which has no policy kinks or discontinuities. We estimate changes in the slope of the relationship between earnings and age drawing on a 5% sample of the 1980 US Census (IPUMS).[3] Since there are no policy kinks, the change in slope of the age-earnings relationship should be zero at any given point under the assumptions of the RK design. The number of observations, 4.4 million, is similar in order of magnitude to that found in administrative data sets. We obtain a distribution of placebo RK estimates at other placebo kinks from age 22 to age 50. The p-values of a linear RK estimate at age 27 based on conventional standard errors are much lower than the p-values based on the distribution of placebo estimates.

Next, we apply the permutation test to two recent applications of the RK design that assess the effect of unemployment benefits on unemployment duration in Austria (CLPW) and the United States (Landais, 2012). In the Austrian setting, we find that the permutation test-based p-values are larger than the conventional ones based robust standard errors and, moreover, the placebo estimates are not centered at zero. This suggests that some of the estimated slope change at the policy kink in this setting may be unrelated to the impact of the policy itself. In the case of unemployment benefits in the US, we conduct permutation tests with bandwidths that are smaller than the one used in Landais (2012) due to data limitations. Here we find that the distribution of slope changes at 26 policy kinks we analyze is similar to the distribution of slope changes at placebo kinks which do not feature any policy slope

---

[3]See Ruggles et al. (2014).

4

changes. This suggest that some of the slope changes observed at the true policy kinks may be due to broader curvature in the underlying relationship between unemployment duration and earnings.

Finally, we apply the permutation test to the regression discontinuity study by Lee (2008) on the effect of incumbency on election outcomes and to work by Dube et al. (2010) which uses a difference-in-difference strategy to evaluate the impact of the minimum wage on employment. Here, we find that the intervals produced by the relabeling test are similar to those produced by asymptotic standard errors. These results affirm that our test produces sensible results in empirical settings where the behavior of asymptotic standard errors is well understood.

Next, we assess the performance of different estimation procedures in Monte Carlo simulations in settings where the outcome function has both curvature and a policy kink. Local linear RK specifications have poor empirical coverage because of the first-order bias discussed above. The local quadratic specification has either much longer confidence intervals under a cross-validation bandwidth selection procedure or empirical coverage below 75% using a bandwidth selection procedure for a quadratic bias-correction model proposed by Calonico, Cattaneo, and Titiunik (forthcoming) (abbreviated as CCT in the following). We also assess the properties of an estimator based on a modified cubic splines framework (see Green and Silverman, 1994, for an introduction) with equally spaced knots (chosen based on a generalized cross-validation criterion Wahba, 1990) and a potentially discontinuous slope and second derivative at the policy kink.

We find that CCT's bandwidth choice for a model with cubic bias-correction has excellent performance, with good empirical coverage and relatively small interval lengths. When there is little curvature in the data-generating process, local quadratic robust regressions also perform well. We also find that cubic splines estimators have good empirical coverage and interval lengths. The use of a flexible global model for RD/RK - such as cubic splines - has some precedent in the literature: Lee's (2008) seminal RD paper used a global fourth-order polynomial, rather than the local polynomial methods which are currently popular in the RD literature. Rau (2011) documents that a penalized spline approach for RD estimation yields

5

preferable results compared to local polynomial approaches in Monte Carlo experiments.

Our paper builds on important previous work critically evaluating RK and RD designs. CCT assess the role of bias for RD and RK estimators. Their procedure estimates higher-order derivatives around the discontinuity to correct for higher-order bias. Our placebo test, in contrast, uses data at non-kink points to assess the potential for higher-order bias (through randomization inference). These methods are complementary: CCT are careful to note that estimates of higher-order terms are usually imprecise, while using data away from the kink may be unattractive or infeasible in some empirical settings. Ando (2013) uses Monte Carlo simulations to argue that linear RK estimates are biased in the presence of plausible amounts of curvature. We share Ando's concerns about linear RK and make three additional contributions: we propose a specific placebo test for practitioners, we use our test to reassess existing empirical RK results and we suggest the use of cubic splines as an alternative estimation strategy and compare the empirical performance of different estimation methods for the RK design.

We recommend that practitioners: (1) avoid using linear and quadratic RK estimates, (2) use the distribution of placebo estimates to detect whether they will have power to detect economically meaningful results in their context, (3) report p-values constructed by comparing their point estimate to the distribution of placebo estimates, and (4) use cubic splines or CCT's robust procedure to estimate kinks.

Section 2 describes the permutation test. Section 3 implements the test in five empirical settings. Section 4 reports the results of RK estimation in Monte Carlo studies based on different estimators, including cubic splines. The last section concludes.

2 Relabeling: A Fisher-Style Permutation Test

We propose a simple permutation test to assess the null hypothesis that treatment has no effect on the outcome of interest. In Fisher's classic randomization test, he randomly relabeled the treatment status of his observations, while holding fixed their observed outcomes, and then re-calculated his test statistic. At the heart of our test is the assumption that the location of

the policy kink can be considered as randomly drawn from a known interval - an assumption that needs to be evaluated in the context of the specific research design under scrutiny. In this interval, we can reassign the location of the kink and calculate RK estimates at these placebo kinks. Our test procedure builds on a suggestion for RD by Imbens and Lemieux (2008) of "testing for a zero effect in settings where it is known that the effect should be 0."[4]

The thought experiment underlying randomization inference is different from the one underlying asymptotic inference. Whereas the idea underlying asymptotic inference is one of sampling observations from a large population, the thought experiment in randomization inference is based on a fixed population that the econometrician observes in the data and in which the assignment of treatment is sampled repeatedly. In the latter approach, treatment assignment is thought of as the random variable. Rosenbaum (2002) builds on one of the original settings of randomization inference - Fisher's (1935) lady tasting tea - to illuminate the differences between the thought experiment underlying randomization inference and the one underlying asymptotic inference and notes:

> The Lady is not a sample from a population of Ladies, and even if one could imagine that she was, there is but one Lady in the experiment and the hypothesis concerns her alone.

Analogously, our test does not treat the sample as being drawn from some (super) population but rather takes the observed sample as given and tests hypotheses regarding this particular sample, treating the location of the policy kink as a random variable.

By drawing on data away from the kink, this permutation test offers a finite sample inference procedure which researchers can use as alternative to inference based on conventional robust standard errors which will be misleading when curvature in the data biases RK estimates. This test assesses the extremeness of the estimated change in the slope at the kink point relative to estimated slope changes at non-kink points under the null hypothesis that the policy does not affect the outcome.

---

[4]Engström et al. (2011) consider a similar test for the RK design, holding bandwidth fixed and computing the regression kink estimator at other locations shown in Figure 11 in their paper. Unfortunately, they only report placebo tests of the linear RK estimator β_1 over a narrow region close to the policy kink.

Following CLPW, we assume a data generating process $Y = y(B, V, U)$ where V is a running variable, B is continuous treatment, U is unobservable, and y maps these variables into an outcome. There is a policy function $B = b(V)$ whose slope with respect to V changes from κ^- below the kink to κ^+ above the kink. Without loss of generality, the kink is located at $V = 0$. y_1 is the derivative of y with respect to B. CLPW develop conditions under which an RK estimator identifies the following average effect of treatment on outcome Y: $E(y_1(b(0), 0, U)|V = 0)$. Appendix A recaps the CLPW framework in more detail.

The data are a vector of n observations each with $(y_i, v_i, b(v_i))$. Let $\mathbf{y}$ denote the vector of y_i values, let $\mathbf{v}$ denote the vector of v_i realizations and let v_0^P be the policy kink. (In the next section, we label placebo kinks at locations v_c^P, $c \geq 1$.)

Define
$$
\tilde{\mathbf{v}}(v_c^P) \equiv \begin{pmatrix}
1 & (v_1 - v_c^P) & (v_1 - v_c^P)\mathbf{1}(v_1 \geq v_c^P) \\
\vdots & \vdots & \vdots \\
1 & (v_n - v_c^P) & (v_n - v_c^P)\mathbf{1}(v_n \geq v_c^P)
\end{pmatrix}
$$

and let $\mathbf{v}_c \equiv \tilde{\mathbf{v}}(v_c^P)$. CLPW propose a linear RK estimator for $E(y_1(b(v_0^P), v_0^P, U)|V = v_0^P)$ as $\frac{e_3'}{(\kappa^+ - \kappa^-)} (\mathbf{v}_0'\mathbf{v}_0)^{-1} \mathbf{v}_0\mathbf{y}_0$ where $e_3' = (\begin{array}{ccc} 0 & 0 & 1 \end{array})$, using observations within bandwidth h of the policy kink ($|v_i - v_0^P| \leq h$) and a uniform kernel.[5] Based on this estimator, we can define the following test statistic

$$
T(\mathbf{v}, \mathbf{y}, v_0^P) \equiv e_3' (\mathbf{v}_0'\mathbf{v}_0)^{-1} \mathbf{v}_0'\mathbf{y}, \quad |v_i - v_0^P| \leq h
$$

which corresponds to the reduced form of CLPW's estimator.

Suppose that we are interested in the distribution of the statistic $T(\mathbf{v}, \mathbf{y}, v_0^P)$. Consider the hypothesis:

Hypothesis: Treatment Irrelevance Outside Kink *Treatment does not affect outcomes outside the policy kink:* $y_1(b(v), v, U) = 0 \quad \forall v \in [\underline{v}, -h] \cup [h, \bar{v}]$.[6]

---

[5]In this section and in the next section, following CLPW and the majority of RK studies that we are aware of, we always use a uniform kernel.

[6]In Appendix B.1, we show that it is possible to calculate p-values under the assumption that treatment does not affect outcomes at any v: $y_1(b(v), v, U) = 0$. This assumption in the Appendix is closer in spirit to Fisher's randomization test. If the policy does have an impact ($y_1(b(0), 0, U) \neq 0$) the assumption above will

To test this hypothesis, the researcher must choose v_{min} and v_{max} as the domain over which she thinks that the slope relating control variable V and outcomes Y follows the same data-generating process. Given a bandwidth h that is used for estimation, permissible placebo kinks can be in the range $[\underline{v}, -h] \cup [h, \bar{v}]$ with $\underline{v} \equiv v_{min} + h$ and $\bar{v} \equiv v_{max} - h$.[7] Next, consider an alternative test statistic:

$$T(\mathbf{v}, \mathbf{y}, v_c^P) \equiv e_3' \left(\mathbf{v}_c' \mathbf{v}_c \right)^{-1} \mathbf{v}_c' \mathbf{y} \quad |v_i - v_c^P| \leq h$$

with $v_c^P \in [\underline{v}, \bar{v}]$. We can compute this test statistic with the same data used to calculate $T(\mathbf{v}, \mathbf{y}, v_0^P)$ by simply using different labels $v_c^P \in [\underline{v}, \bar{v}]$ to denote the location of the policy kink. We call this "relabeling" because we have held fixed the outcome data $\mathbf{y}$ and we have renamed the $\mathbf{v}$ values by using a different label for the location of the policy kink. This leads to the following assumption which is the basis for the permutation test of the hypothesis of treatment irrelevance:

Assumption: Random Kink Placement $(v_0^P \sim \text{Unif}[\underline{v}, \bar{v}])$.

Under the Treatment Irrelevance Hypothesis and the Assumption of Random Kink Placement, we can compute the distribution of possible estimates which could have arisen had the policy kink been at a different location in the same dataset. Because the kink locations v are continuous, there is a continuum of values for the test statistic $\{T(\mathbf{v}, \mathbf{y}, v)\}$. Under the assumption above, all kink locations v are equally likely, so the CDF of this distribution for test statistic T is

$$P(T > t) = \frac{\int_{\underline{v}}^{\bar{v}} \mathbf{1}(T(\mathbf{v}, \mathbf{y}, v) > t) dv}{\int_{\underline{v}}^{\bar{v}} dv}$$

<u>Interval:</u> Suppose the researcher is interested in determining the lowest and highest values of the test statistic which would achieve a p-value of α on each side. For a one-sided test, this value $\underline{\tau}^\alpha$ is identified as the $1 - \alpha$ quantile of the distribution of the test statistic. By finding

produce smaller p-values and is more likely to reject the hypothesis for a given set of data.

[7] A natural assumption that we make in the following is that $v_0^P \in [\underline{v}, \bar{v}]$.

$\bar{\tau}^\alpha$ such that $P(\bar{\tau}^\alpha > T) = 1 - \alpha$, and $\underline{\tau}^\alpha$ such that $P(T > \underline{\tau}^\alpha) = 1 - \alpha$ one can construct an interval $[\underline{\tau}^\alpha, \bar{\tau}^\alpha]$ with a coverage rate of $1 - 2\alpha$. Even before constructing an estimate at the policy kink, researchers can examine this interval to see whether they would be able to detect economically meaningful effects.

$$P(\bar{\tau}^\alpha > T > \underline{\tau}^\alpha) = \frac{\int_{\underline{v}}^{-h} \mathbf{1}(\bar{\tau}^\alpha > T(\mathbf{v}, \mathbf{y}, v) > \underline{\tau}^\alpha)dv + \int_{h}^{\bar{v}} \mathbf{1}(\bar{\tau}^\alpha > T(\mathbf{v}, \mathbf{y}, v) > \underline{\tau}^\alpha)dv}{\int_{\underline{v}}^{-h} dv + \int_{h}^{\bar{v}} dv} = 1 - 2\alpha$$

(1)

<u>P-value:</u> With a point estimate of $T(\mathbf{v}, \mathbf{y}, v_0^P)$ at the policy kink, the one-sided p-value for observing a test statistic of $T(\mathbf{v}, \mathbf{y}, v_0^P)$ or larger under the null hypothesis given this distribution is

$$P(T > T(\mathbf{v}, \mathbf{y}, v_0^P)) = \frac{\int_{\underline{v}}^{-h} \mathbf{1}(T(\mathbf{v}, \mathbf{y}, v) > T(\mathbf{v}, \mathbf{y}, v_0^P))dv + \int_{h}^{\bar{v}} \mathbf{1}(T(\mathbf{v}, \mathbf{y}, v) > T(\mathbf{v}, \mathbf{y}, v_0^P))dv}{\int_{\underline{v}}^{-h} dv + \int_{h}^{\bar{v}} dv}$$

(2)

We generalize this approach to incorporate alternative null hypotheses, alternative bandwidth choice procedures, discrete kink locations, and higher-order polynomials in Appendix B.1. We also show how a similar approach can be used to inference for Regression Discontinuity and Difference-in-Difference designs in Appendix B.2.

3 Randomization Inference: Empirical Examples

In this section, we demonstrate the usefulness of the empirical procedures from the previous section in three separate empirical RK settings: the age-earnings curve in the US (which features no policy kinks) to illustrate the permutation test, search duration and unemployment insurance in Austria (which features policy kinks) and search duration and unemployment insurance in the US (which also features policy kinks). For all three settings, we calculate the distribution for the RK estimator under the assumption that the kink location is chosen randomly. We find evidence of substantial curvature in regions with no policy kinks, casting doubt

on the accuracy of p-values when using the linear RK estimator with associated conventional standard errors.

Then, we show the results of our permutation test in two other empirical settings in order to assess whether our plabeo test delivers p-values that are quantitatively comparable to ones based on asymptotic inference in settings where the properties of asymptotic standard errors are well understood. To this end, we apply our test to a regression discontinuity model based on work by Lee (2008) for the effect of incumbency on elections and a difference-in-difference model for the effect of minimum wage on employment by Dube et al. (2010).

3.1 RK: Earnings and Age

We study the age-earnings curve in the US as an example of how curvature may affect an RK estimator. This is an attractive setting to study the performance of RK estimators: because there are no policy kinks with age among working-age adults, the RK estimator relating earnings and age should be zero. The top-left panel of Figure 2 plots the age-earnings curve from the 5% 1980 Census sample. This sample has about 4.4 million people with positive earnings, comparable to sample sizes one might find when working with administrative data.

Consider a sample linear RK estimate at age 27 drawing on data from age 24 to 30 (or 12 quarters of birth on each side). When using robust standard errors, as is standard practice in the RK literature, we find a highly statistically significant slope change at higher bandwidths.

We construct point estimates and one-sided p-values for the "kink" at age 27 at a variety of bandwidths. We compute the cumulative distribution function of the RK estimator under the assumption that the "kink" was placed randomly between age 22 and 50. At a bandwidth of four quarters of birth, we calculate a p-value of 0.73 using equation 2, implying that 73% of placebo estimates have more extreme values than the estimate at age 27. This p-value is larger than that which emerges from the robust standard errors (0.45). With a bandwidth of eight quarters, the p-value based on conventional standard errors is much lower ($p < 0.0001$) than the one based on the placebo exercise ($p = 0.30$). This illustrates the idea that in a region with substantial curvature, the linear RK estimator as conventionally implemented can yield misleading results. With a bandwidth of 12 quarters, the point estimate is even larger because

bias grows when larger bandwidths are used, and we find an even lower p-value. Even though the RK estimator is larger in absolute terms with this higher bandwidth and the asymptotic p-value indicates more precision, the Fisher p-values indicates no statistically significant change in slope. The remaining three panels of Figure 2 illustrate the results visually. A linear specification with an overly large bandwidth yields false precision.

3.2 RK: Search Duration and Unemployment Insurance in Austria

Next, we apply our permutation test to a setting with a true policy kink – the unemployment insurance system in Austria analyzed in CLPW. The top panels of Figure 3 show that benefits rise more sharply with prior earnings above a policy kink and that the slope of durations with respect to prior earnings changes as well. The bottom-left panel of Figure 3 shows the global distribution of search duration with respect to income. Much as we used the age-earnings distribution from ages 22 to 50 to compute a null distribution for the test statistic, here we use other parts of the income distribution to compute a null distribution, with the CDF shown in the bottom-right panel of Figure 3. Our analysis is not based on the administrative data directly but rather uses the data outside the CLPW estimation region provided in the the bottom-left panel of Figure 3 for estimation. We scale up the reduced form placebo coefficients by the coefficient of the first stage relationship reported in CLPW so that the estimates are interpretable in elasticity units.

We use equation 1 to compute the lowest value of the test statistic under which we would reject the null hypothesis with 95% confidence using a two-sided test. We compute that the minimum detectable elasticity is 1.7.[8] CLPW review the literature on benefit elasticities in Appendix Table 6 of their paper. In that table, the median estimate is 0.7.[9] If our assumption about the randomly selected kink location holds, then the linear RK estimator cannot detect economically meaningful parameter values given these data.

Next, the solid vertical line in the top panel of Figure 3 shows the point estimate reported in CLPW Figure 10. This slope change at the true policy kink is larger than many of the

---

[8] The minimum detectable elasticity for a two-sided test with 90% coverage is 1.58.

[9] When a study reports a range of estimates, we take the mean of the min and the max and count it as one observation.

placebo estimates and we compute a p-value of 0.038 using equation 2. Although this p-value is larger than CLPW's asymptotic p-value (t-statistic around 4 implying a p-value of less than 0.001), it is consistent with rejecting the null hypothesis that the level of unemployment benefits has no impact on search duration at a 5% level (one-sided test).

3.3   RK: Search Duration and Unemployment Insurance in the US

As a third empirical example, we analyze the US UI system studied by Landais (2012) using the Continuous Wage and Benefit History (CWBH) sample.[10] In US states, benefits are rising as a function of prior income, up to a maximum benefit. This induces a kinked relationship between prior income and UI benefits. Landais uses linear RK models to estimate the impact of benefits on search duration. His bandwidth choice of \$2500 uses 89% of the population data. Unlike in the Austrian setting, there is very little data far from the kink which can be used to generate placebo estimates. As a substitute, we compare estimates from a two-part linear spline (one potential kink) and from a four-part linear spline (three potential kinks).

The left panel of Figure 4 shows estimates from Louisiana. Here, a two-part linear spline finds a highly statistically significant kink at zero. Our estimates with a four-part linear spline find a highly statistically significant slope change at -\$1000 and a wrong-signed slope change at zero. This suggests that the apparent kink at zero in the two-part linear spline may arise from global curvature.

Next, we construct linear RK estimates for 26 different state-period cells in the relevant CWBH sample. The right panel of Figure 4 displays the distribution of point estimates based on four-part linear splines. The blue line denotes the distribution of point estimates for the actual policy kink. In the spirit of the permutation test, the green and maroon lines denote the distribution of point estimates at placebo kinks at \$1000 and -\$1000, respectively. The distribution of slope changes are the true policy kink and at the -\$1000 kink are fairly similar. Two lessons arise from this exercise. The results from the -\$1000 kink suggest that the slope changes observed at the true policy kink may be consistent with broader curvature of expected

---

[10]Our empirical estimates reported below with a single kink are broadly consistent with those in Landais (2012)

duration with respect to earnings.

3.4 Comparison to RD and DD and Summary

We also implement our permutation test for two well-known empirical examples: Lee's (2008) RD study of the effect of winning an election on a party's subsequent victory in a congressional house election two years later and Dube et al. (2010)'s (henceforth, DLR) Difference-in-Difference study of the effect of the minimum wage on employment using contiguous county pairs. Appendix C describes construction of placebo estimates in more detail.

To summarize our results from this section, we compare intervals from our permutation test to confidence intervals based on asymptotic standard errors in Figure 5. For the age-earnings curve and for CLPW with linear RK models, the permutation intervals are much longer than the asymptotic intervals. Evidently, first-order bias is empirically important here. In contrast, for Lee (2008) and DLR, the permutation intervals are, if anything, shorter than the asymptotic intervals. This suggests that our relabeling procedure delivers sensible results in empirical settings where the properties of asymptotic standard errors are well understood.

4 Simulation Study

To understand the coverage properties of different estimators for the RK model, we conduct a Monte Carlo simulation study to assess the performance of local polynomial specifications and recently proposed robust bias-corrected estimators (CCT). In addition, we assess the performance of modified cubic spline models (see Green and Silverman, 1994, for an introduction to cubic splines that we follow in this paragraph). Given a set of knots $\{t_j\}_{j=1}^n$ on an interval $[a, b]$, a cubic spline g is a cubic polynomial on each of the intervals (t_j, t_{j+1}) for $j \in \{1, .., n\}$ and on (a, t_1) and (t_n, b) such that g and its first and second derivatives are continuous at each t_j. Cubic splines are an attractive framework for estimation because they are shown to be the solution to an optimal interpolation problem.[11] We adjust the cubic splines framework so that it can be used for estimation of an RK model by placing a knot at 0 and additionally

---

[11]More precisely, the solution to the problem of finding the smoothest function that interpolates point (t_i, y_i) is a natural cubic spline. A cubic spline is a natural cubic spline if it is linear on the extreme intervals $[a, t_1]$ and $[t_n, b]$.

allowing for a change in the first and second derivative at this special knot. We choose the total number of equally-spaced knots based on a generalized cross-validation criterion (Wahba, 1990).

We assume a data-generating process with substantial curvature as well as a true policy kink at zero:[12]

$$E(y|x) = 10x\mathbf{1}_{x>0} + sin(15(x-0.1)) + x^2$$

This function is shown in Figure 6. x is distributed uniformly on [-1,1]. A single draw consists of 10,000 observations with $y = E(y|x) + \varepsilon$ with $\varepsilon \sim N(0, 0.5)$. We consider two different methods for bandwidth choice: generalized cross-validation (GCV) which seeks to minimize the MSE of the estimated conditional mean function (minimizes $E(\hat{y} - y)^2$) and methods proposed by CCT which seek to minimize the MSE of the point estimate (minimizes $E(\hat{\tau}_{RK} - \tau_{RK})^2$).

For each polynomial order from linear to cubic, we evaluate models with (1) bandwidth chosen by the GCV, (2) estimation bandwidths chosen by CCT, and (3) CCT's bias-corrected specification with bandwidths chosen by CCT.[13] Because CCT's bias-correction procedure adds an additional polynomial term, we label their procedure by the highest polynomial used in estimation.[14] At the optimal bandwidth(s), we compute an RK estimate and a 95% confidence interval using asymptotic heteroskedasticity-robust standard errors or CCT's standard errors for their bias-corrected estimators. Coverage denotes the fraction of confidence intervals - for a given specification - that cover the true slope change of 10. |Error| denotes the mean absolute error of the estimate of the slope change at zero. Interval length is the average length of the confidence interval for a given estimation method.

Both linear models have quite low empirical coverage of the true estimate. The quadratic model, with bandwidth chosen using the GCV, has an empirical coverage rate of 74% but the average length of the nominal confidence intervals is very large (61.77). While interval

---

[12]This data-generating process - combining a sine function and polynomials - is similar to processes evaluated in Sun's (2005) analysis of RD estimators.

[13]CCT use two different bandwidths in estimation: one for point estimation ("$CCT\ h$") and a second for bias-correction ("$CCT\ b$").

[14]For example, rdrobust, p(1) q(2) is labeled "Local Quad (CCT Bias-Correct)".

lengths are shorter for linear and quadratic specifications with CCT bandwidth choice or CCT bias correction, the empirical coverage rates are at most 72% in all specifications (with the exception of model 8, discussed below).

Cubic specifications come close to attaining 95% nominal coverage. The local cubic model with bandwidth chosen by GCV is not attractive, however, as it is associated with very large interval lengths (141.40). Local cubic and local cubic robust regressions, both with bandwidths chosen by CCT's procedure, have relatively similar coverage rates (95% and 96%) and similar mean interval lengths (22.26 and 26.80). Model #8 (local quadratic model with bandwidth chosen as part of the cubic bias-correction) delivers similar coverage (91%) and interval length (21.25). Evidently, the effectiveness of CCT's procedure comes primarily from good bandwidth choice in standard local polynomial models, rather than bias-correction in point estimation. The cubic spline model attains coverage of 88% but has short confidence intervals and mean absolute error comparable to local cubic and local cubic robust regressions with CCT bandwidth choice. We conclude that in this setting, "nothing beats cubic", in the sense that cubic models always deliver good coverage, while this is not true for lower order polynomials.

We conduct an additional Monte Carlo where we repeat the same exercises as described above but choose a function with less curvature ($E(y|x) = 10x\mathbf{1}_{x>0} + sin(5(x-0.1)) + x^2$, see Figure 7). Again, we find that the cubic models that performed well in the previous exercise have good coverage properties and relatively short intervals. In addition, we find that local quadratic regression with CCT bandwidth choice as well as local quadratic robust regression have good coverage (96%) and feature short intervals comparable to those of cubic specifications.

This suggests that quadratic specifications are a viable complement to cubic specifications when there is little curvature in the data-generating process. As our simulations document that cubic specifications have good coverage with reasonably short intervals in both settings we consider, researchers are well advised to use such cubic specifications as preferred estimation method.

5 Conclusion

We conclude with some recommendations for practitioners:

1. Avoid using the linear and quadratic RK specification when possible.

2. Assess power by comparing the distribution of placebo RK estimates to economically meaningful magnitudes.

3. Compute p-values for the test statistic using the distribution of placebo RK estimates.

4. Use cubic specifications, in particular, CCT's rdrobust command or, in addition, when there is sufficient data, global cubic splines, with generalized cross-validation for knot placement.

In this paper, we examined the properties of RK estimators. The RK estimator has become increasingly popular in economics research in the last five years. When there is curvature in the global function, linear RK estimates are first-order biased. Linear RK specifications fail to follow the advice of Fan and Gijbels (1996) and Porter (2003) in using a model with higher-order polynomial terms than the object of interest. In three empirical examples, we showed that such bias is quantitatively important and that standard asymptotic inference may be misleading.

References

Abadie, A., Athey, S., Imbens, G., and Wooldridge, J. "Finite Population Causal Standard Errors." *mimeo* (2014).

Abadie, A., Diamond, A., and Hainmueller, J. "Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program." *Journal of the American Statistical Association*, 105(490) (2010).

Ando, M. "How Much Should We Trust Regression-Kink-Design Estimates?" *Department of Economics Working Paper 2013:22 Uppsala University* (2013).

Bertrand, M., Duflo, E., and Mullainathan, S. "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics*, 119:249–75 (2004).

Böckerman, P., Kanninen, O., and Suoniemi, I. "A Kink that Makes You Sick: The Incentive Effect of Sick Pay on Absence." *IZA DP No. 8205* (2014).

Bravo, J. "The Effects of Intergovernmental Grants on Local Revenue: Evidence from Chile." *Documentos de Trabajo (Instituto de Economía PUC)*, (393):1 (2011).

Calonico, S., Cattaneo, M. D., and Titiunik, R. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* (forthcoming).

Card, D., Lee, D., Pei, Z., and Weber, A. "Nonlinear Policy Rules and the Identification and Estimation of Causal Effects in a Generalized Regression Kink Design." Technical report, National Bureau of Economic Research (2012).

Cattaneo, M., Frandsen, B., and Titiunik, R. "Randomization Inference in the Regression Discontinuity Design: An Application to Party Advantages in the U.S. Senate." *Journal of Causal Inference* (forthcoming).

Caughey, D. and Sekhon, J. S. "Elections and the regression discontinuity design: Lessons from close us house races, 1942–2008." *Political Analysis*, 19(4):385–408 (2011).

Chetty, R., Looney, A., and Kroft, K. "Salience and Taxation: Theory and Evidence." *The American Economic Review*, 99(4):1145–1177 (2009).

Dahlberg, M., Mörk, E., Rattsø, J., and Ågren, H. "Using a discontinuous grant rule to identify the effect of grants on local taxes and spending." *Journal of Public Economics*, 92(12):2320–2335 (2008).

Dobbie, W. and Skiba, P. M. "Information Asymmetries in Consumer Credit Markets: Evidence from Payday Lending." *American Economic Journal: Applied Economics*, 5(4):256–282 (2013).

Dong, Y. "Jumpy or Kinky? Regression Discontinuity without the Discontinuity." *Working Paper* (2010).

Dube, A., Lester, W., and Reich, M. "Minimum Wage Effects Across State Borders: Estimates Using Contiguous Counties." *Review of Economics and Statistics,*, 92(4):945–964 (2010).

Engström, P., Nordblom, K., Ohlsson, H., and Persson, A. "Loss evasion and tax aversion." *Working Paper* (2011).

Fan, J. and Gijbels, I. *Local Polynomial Modelling and its Applications.*, volume 66. Chapman and Hall (1996).

Fe, E. and Hollingsworth, B. "Estimating the effect of retirement on mental health via panel discontinuity designs." *Working Paper* (2012).

Fidrmuc, J. and Tena, J. d. D. "National minimum wage and employment of young workers in the UK." *CESifo Working Paper, No. 4286* (2013).

Fisher, R. *The Design of Experiments*. Oliver and Boyd, Oxford, England (1935).

Garmann, S. "The Causal Effect of Coalition Governments on Fiscal Policies: Evidence from a Regression Kink Design." *Available at SSRN 2271082* (2013).

Green, P. and Silverman, B. *Nonparametric regression and generalized linear models: a roughness penalty approach*. Chapman and Hall, London (1994).

Hahn, J., Todd, P., and Van der Klaauw, W. "Identification and estimation of treatment effects with a regression-discontinuity design." *Econometrica*, 69(1):201–209 (2001).

Hanson, A. "The Incidence of the Mortgage Interest Deduction: Evidence from the Market for Home Purchase Loans." *Public Finance Review*, 40(3):339–359 (2012).

Ho, D. E. and Imai, K. "Randomization inference with natural experiments: An analysis of ballot effects in the 2003 California recall election." *Journal of the American Statistical Association*, 101(475):888–900 (2006).

Imbens, G. W. and Lemieux, T. "Regression discontinuity designs: A guide to practice." *Journal of Econometrics*, 142(2):615–635 (2008).

Imbens, G. W. and Rosenbaum, P. R. "Robust, accurate confidence intervals with a weak instrument: quarter of birth and education." *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(1):109–126 (2005).

Jones, M. R. "The EITC and Labor Supply: Evidence from a Regression Kink Design." *mimeo Cornell University* (2011).

Kolsrud, J. "Precaution versus Risk Aversion: Decomposing the effect of Unemployment Benefits on Saving." *Working Paper* (2012).

Kristensen, S. R., Fe, E., Bech, M., and Mainz, J. "Is the quality of hospital care price sensitive? Regression kink estimates from a volume dependent price setting scheme." *COHERE-Centre of Health Economics Research, University of Southern Denmark* (2013).

Landais, C. "Assessing the Welfare Effects of Unemployment Benefits Using the Regression Kink Design." *Working Paper* (2012).

Lee, D. S. "Randomized experiments from non-random selection in US House elections." *Journal of Econometrics*, 142(2):675–697 (2008).

Lehmann, E. and Stein, C. "On the theory of some non-parametric hypotheses." *The Annals of Mathematical Statistics*, 20(1):28–45 (1949).

Lundqvist, H., Dahlberg, M., and Mörk, E. "Stimulating local public employment: Do general grants work?" *American Economic Journal: Economic Policy*, 6(1):167–92 (2014).

Manoli, D. and Turner, N. "Cash-on-Hand and College Enrollment: Evidence from Population Tax Data and Policy Nonlinearities." *NBER Working Paper 19836* (2014).

Marx, B. and Turner, L. "Borrowing Trouble? Student Loans, the Cost of Borrowing, and Implications for the Effectiveness of Need-Based Grant Aid." *Working Paper* (2014).

Nielsen, H. S., Sørensen, T., and Taber, C. R. "Estimating the effect of student aid on college enrollment: Evidence from a government grant policy reform." *American Economic Journal: Economic Policy*, 2:185–215 (2010).

Peck, J. R. "Can Hiring Quotas Work? The Effect of the Nitaqat Program on the Saudi Private Sector." *Working Paper, MIT* (2014).

Porter, J. "Estimation in the regression discontinuity model." *Unpublished Manuscript, Department of Economics, University of Wisconsin at Madison* (2003).

Rau, T. "Bayesian inference in the regression discontinuity model." *Working Paper, Pontificia Universidad Católica de Chile* (2011).

Rosenbaum, P. R. "Stability in the absence of treatment." *Journal of the American Statistical Association*, 96(453):210–219 (2001).

—. *Observational studies*. New York: Springer, 2 edition (2002).

Ruggles, S., Alexander, J., Genadek, K., Goeken, R., Schroeder, M., and Sobek, M. "Integrated public use microdata series: Version 5.0 [machine-readable database]." *University of Minnesota, Minneapolis* (2014).

Scharlemann, T. C. and Shore, S. H. "Does Reducing Underwaterness Prevent Mortgage Default? Evidence from HAMP PRA." *Working Paper* (2013).

Seim, D. "Behavioral Responses to an Annual Wealth Tax: Evidence from Sweden." *Working Paper, University of Toronto* (2014).

Simonsen, M., Skipper, L., and Skipper, N. "Price sensitivity of demand for prescription drugs: Exploiting a regression kink design." *Univ. of Aarhus Dept. of Economics Working Paper* (2010).

Sukhatme, N. U. and Cramer, J. N. "Optimal Patent Term and Cross-Industry Measures of Patent Term Sensitivity." *Working Paper, Princeton University* (2014).

Sun, Y. "Adaptive estimation of the regression discontinuity model." *Available at SSRN 739151* (2005).

Turner, L. J. "The incidence of student financial aid: Evidence from the Pell grant program." *mimeo Columbia University* (2012).

Wahba, G. *Spline models for observational data*, volume 59. Siam (1990).

Welch, W. J. "Construction of permutation tests." *Journal of the American Statistical Association*, 85(411):693–698 (1990).

Welch, W. J. and Gutierrez, L. G. "Robust permutation tests for matched-pairs designs." *Journal of the American Statistical Association*, 83(402):450–455 (1988).

White, H. "A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity." *Econometrica: Journal of the Econometric Society*, 817–838 (1980).

Wong, M. "Estimating Ethnic Preferences Using Ethnic Housing Quotas in Singapore." *The Review of Economic Studies*, 80(3):1178–1214 (2013).

TABLE 1: OVERVIEW OF EXISTING RK PAPERS

Paper:	Policy Variable	Outcome Variable	Preferred Polynomial/Estimation
Ando (2013)	Federal Subsidy	Gov't Expenditure	Linear/Quadratic
Böckerman, Kanninen, and Suoniemi (2014)	Sickness Insurance	Duration of Sickness Absence	Linear
Bravo (2011)	Federal Subsidy	Local Revenue	Quad
Dahlberg, Mörk, Rattsø, and Ågren (2008)	Federal Grant	Gov't Expenditure	Quad
Dobbie and Skiba (2013)	Paycheck	Loan Default	Linear
Dong (2010)	Retirement	Food Expenditure	Quad
Engström, Nordblom, Ohlsson, and Persson (2011)	Tax Liability	Tax Behavior	Linear, Quadratic, Cubic
Fe and Hollingsworth (2012)	Retirement	Health Outcomes	5th order polyn.
Fidrmuc and Tena (2013)	Age relative to cutoff	Employment Status	Quadratic
Garmann (2013)	Vote Share	Municipal Expenditure/Taxes	Cubic to 6th order polyn.
Hanson (2012)	Loan Amount	Interest Rates	Linear
Jones (2011)	EITC	Hours Worked	Linear/Quadratic
Kolsrud (2012)	UI Benefits	Asset Accumulation	Linear/Quadratic
Kristensen, Fe, Bech, and Mainz (2013)	Reimbursement Rate	Hospital Process Quality	qMLE
Landais (2012)	UI Benefits	Search Duration	Linear
Lundqvist, Dahlberg, and Mörk (2014)	Federal Subsidy	Local Employment	Quadratic
Manoli and Turner (2014)	EITC	College Enrollment	Linear
Marx and Turner (2014)	College Subsidy	Student Loans	Linear
Nielsen, Sørensen, and Taber (2010)	College Subsidy	Enrollment	Linear
Peck (2014)	% Employees	Hiring	Linear/Quadratic
Scharlemann and Shore (2013)	Mortgage Balance Reduction	Mortgage Delinquency	Linear
Seim (2014)	Tax Liability	Wealth Accumulation	Linear
Simonsen, Skipper, and Skipper (2010)	Drug Subsidy	Drug Expenditure	Linear
Sukhatme and Cramer (2014)	Application Pendency of Patent	Stock Price	Linear
Turner (2012)	College Subsidy	Enrollment	Linear
Wong (2013)	%Ethnic groups	Housing Prices	Quartic
CLPW (2012)	UI Benefits	Search Duration	Linear
Example used in current paper	None/Age	Earnings	Linear, Quadratic, Cubic, Cubic Splines

21

FIGURE 1 – PIECEWISE LINEAR AND QUADRATIC SIMULATED DGPS
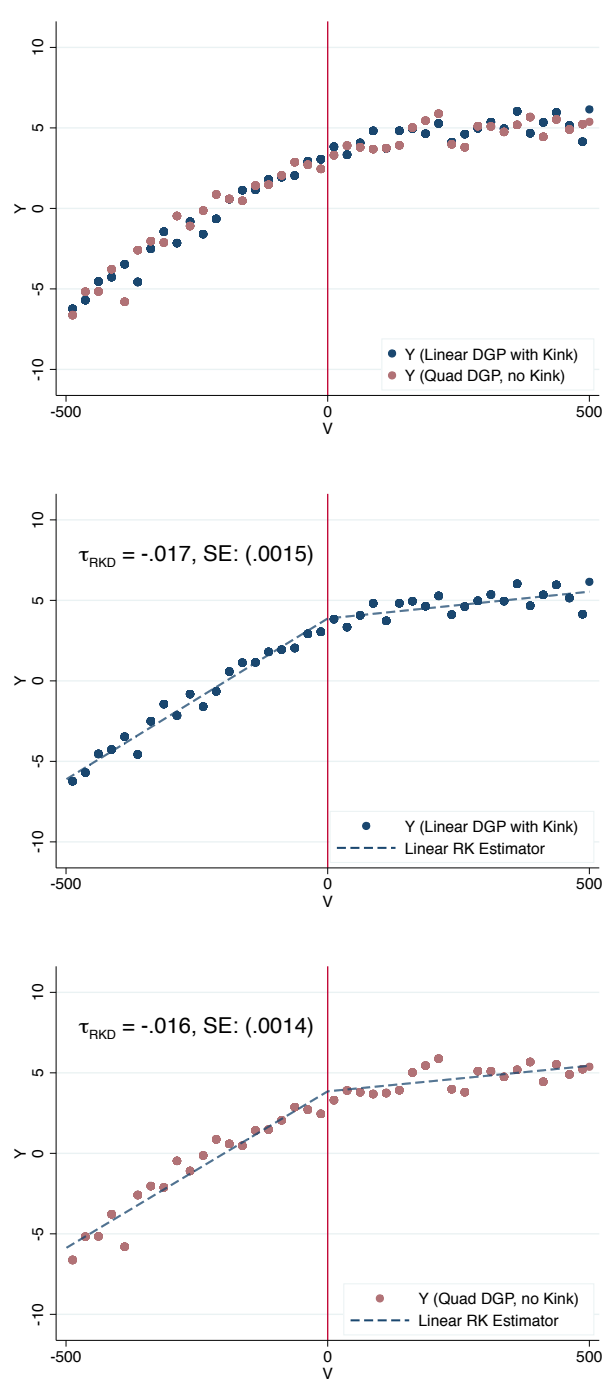
Notes: The data generating process (DGP) is either linear with a kink (blue dots) or quadratic (red dots) without a kink. We generate 1000 observations with a variance of 12 and plot the data in 40 bins. We estimate a linear Regression Kink model with heteroskedasticity-robust standard errors and a uniform kernel. Both the true linear DGP with a kink and the quadratic no-kink DGP yield highly significant regression kink estimates.

FIGURE 2 – RK INFERENCE EXAMPLE WITH NO POLICY KINK: AGE-EARNINGS CURVE IN US
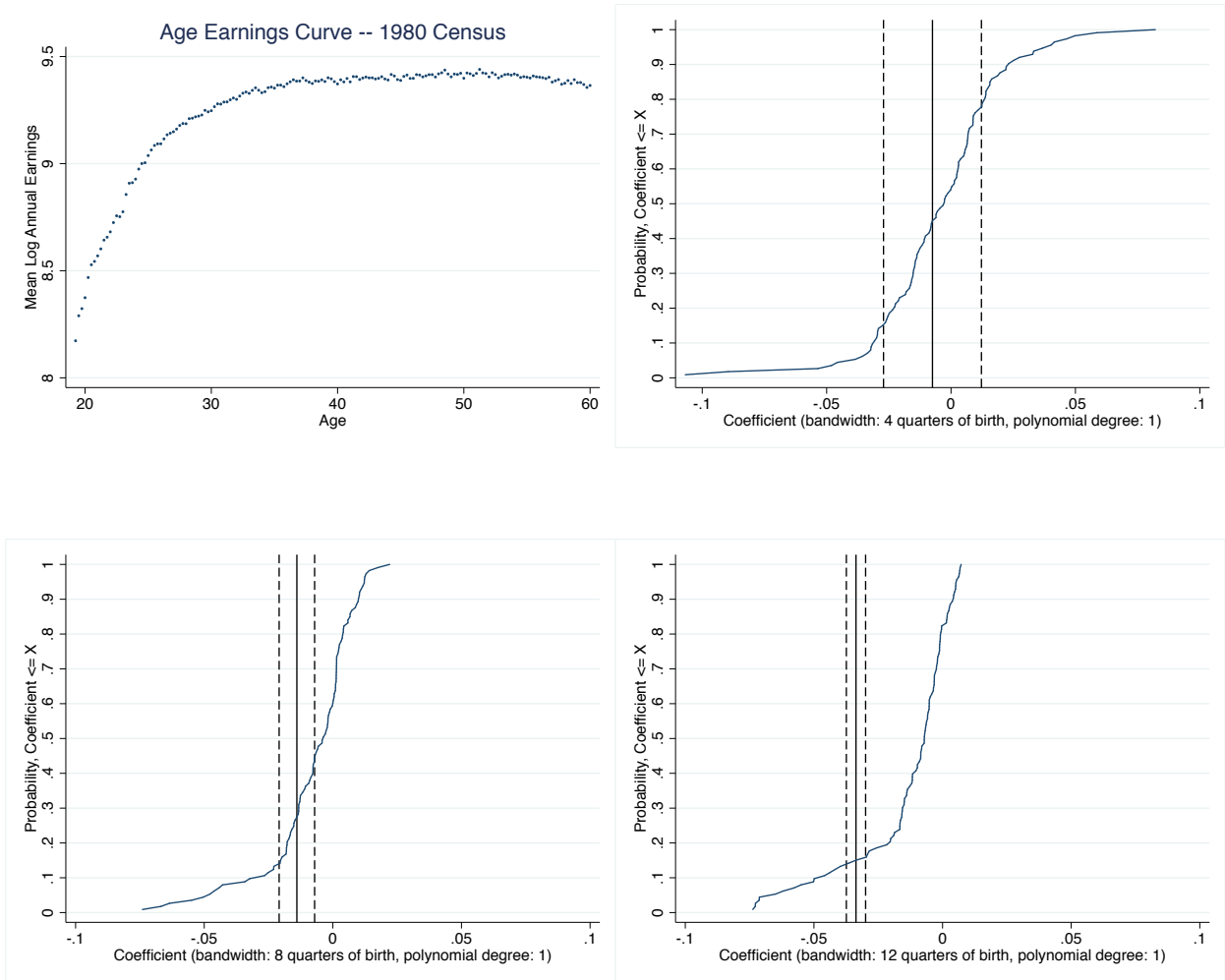
Notes: The first panel shows age-earnings relationship in the 1980 US Census. The next three panels assess the behavior of asymptotic and permutation-based intervals. The solid vertical lines denote the linear RK point estimates for the slope change at age 27; the dashed lines mark the corresponding 95% asymptotic confidence intervals based on heteroskedasticity-robust standard errors. Using linear RK specifications, we estimate the slope change in the age earnings relationship for each quarter of birth between ages 22 and 50 for bandwidths of 4, 8, and 12 quarters of birth and report the distribution of point estimates based on these specifications in the three figures above. Although the asymptotic confidence intervals suggest that there is a highly statistically significant slope change at age 27, the placebo distributions show that the age 27 slope change is in fact not unusual.

FIGURE 3 – RK INFERENCE EXAMPLE: UI BENEFITS IN AUSTRIA
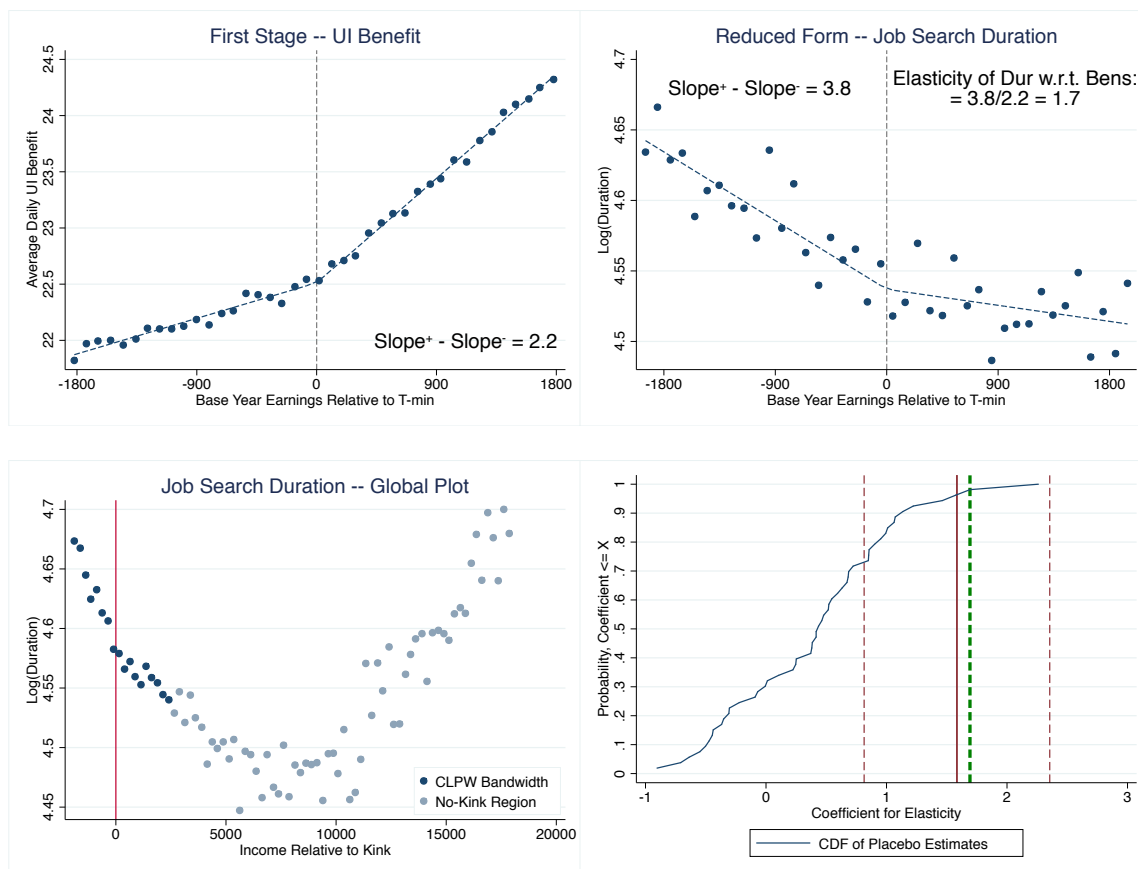
Notes: Top panel plots are from Figure 3 and Figure 5 of Card et al. (2012). T-min refers to the earnings threshold at which benefits start to rise. Coefficients are from Table 2. The bottom-left panel is supplemental data shared with the authors by Andrea Weber.

The bottom-right figure shows the cumulative distribution functions of placebo RK estimates of the relationship between unemployment duration and previous earnings. Using linear RK specifications, we estimate slope changes at placebo kinks outside the CLPW estimation region. This analysis is not based on the administrative data directly but rather uses the data provided in the the bottom-left panel for estimation. The coefficients are scaled up by the coefficient of the first stage relationship at the true bottom policy kink to be interpretable in elasticity units. The dashed green line denotes the 97.5th percentile of the distribution of placebo RK estimates. The solid maroon line denotes the reduced form estimate reported in CLPW. The dashed vertical lines in maroon denote the 95% confidence interval based on the standard errors reported in CLPW Table 2.

Figure 4 – RK Inference Example: UI Benefits in US
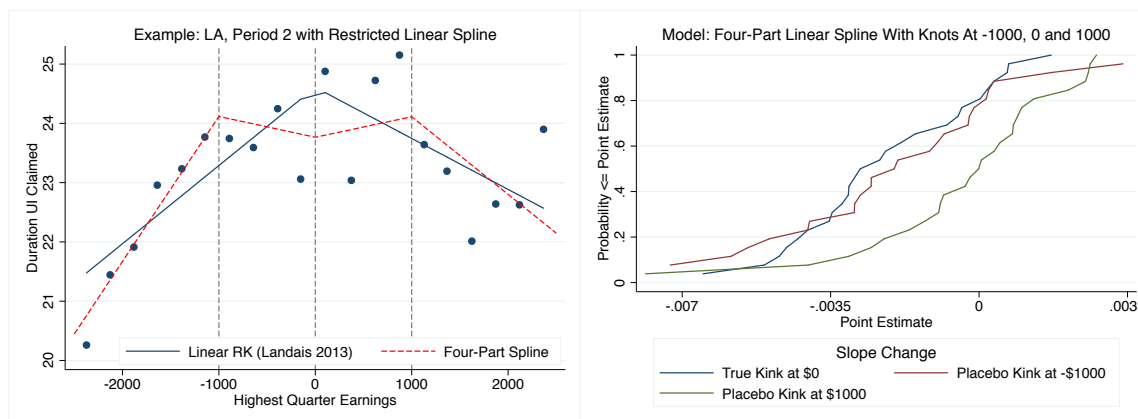
Notes: Landais (2012) implements linear RK estimates for the US using data from the Continuous Wage and Benefit History data. Following Landais, we estimate a linear RK specification (two-part linear spline in blue) and find a highly statistically significant kink at zero, as shown in the top panel. For comparison, we estimate four part linear spline shown in red and find negative slope changes at the placebo kinks.

In the bottom panel, we estimate four-part linear splines for each of the 26 state-period pairs of the CWBH data, constrained to be continuous at each knot, so that we have estimates for two placebo kinks and one true policy kink. (Because the estimation bandwidth (from -$2500 to $2500) is typically about 90% of the population support, there is not enough data to estimate placebo kinks using data outside the kink region.) The distribution of these estimates is plotted in the bottom panel. The distribution of slope changes are the true policy kink and at the -$1000 kink are fairly similar, suggesting that the slope changes observed at the true policy kink may be consistent with broader curvature of expected duration with respect to earnings. Two point estimates at placebo kink of -$1000 (0.007 and -0.052) are omitted for readability.

FIGURE 5 – RANDOMIZATION INFERENCE FOR RK, RD, AND DD
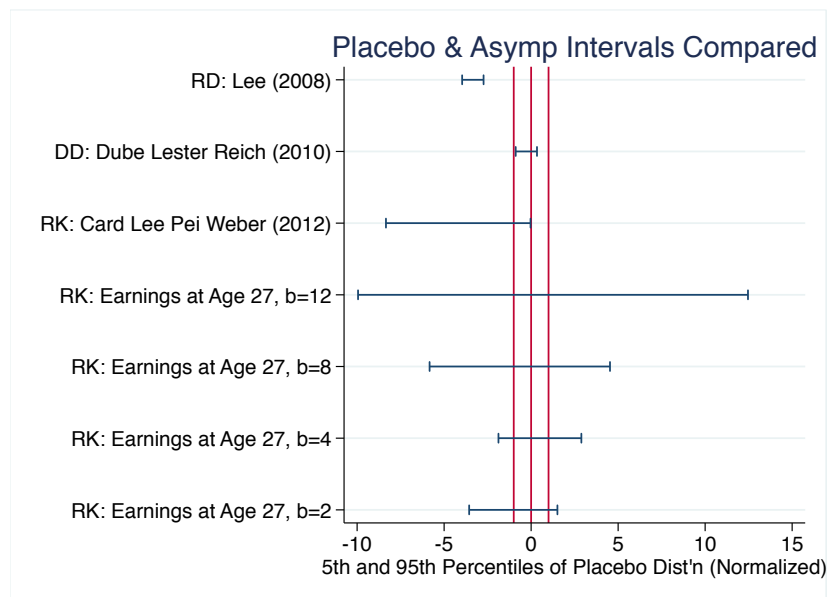
Notes: This figure plots an interval of the 5th and 95th percentile of placebo estimates for a variety of specifications. The 5th and 95th percentiles of the distribution were normalized by subtracting the actual point estimate and dividing by 1.645 times the asymptotic standard errors. This implies that the 5th and 95th percentile of the distributions of placebo estimates will be plotted at -1 and 1, respectively, if they correspond to the lower and upper limit of the asymptotic confidence interval, respectively.

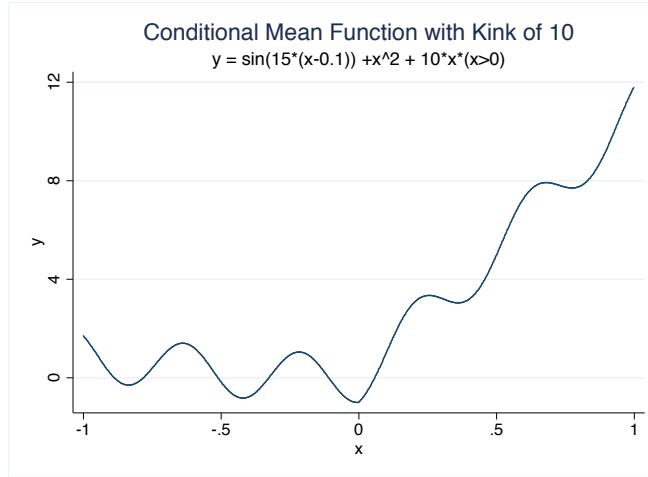# FIGURE 6 – LOCAL POLYNOMIALS AND CUBIC SPLINES SIMULATION I

Conditional Mean Function with Kink of 10
y = sin(15*(x-0.1)) +x^2 + 10*x*(x>0)

Table 2 - Simulation Study of RK Estimation Models

			Mean from 250 Replications		
Model	h Criteria	Coverage	h	|Error|	Interval Len
1. Local Linear	GCV	0.40	0.06	12.74	19.11
2. Local Quadratic	GCV	0.74	0.08	15.76	61.77
3. Local Cubic	GCV	0.86	0.11	28.67	130.90
4. Cubic Spline	GCV	0.88	0.10	4.84	20.58
5. Local Linear	CCT h	0.23	0.04	9.87	13.97
6. Local Quadratic	CCT b	0.11	0.14	9.01	10.05
7. Local Quad (Bias-Correct)	CCT	0.72	0.14	5.96	16.13
8. Local Quadratic	CCT h	0.93	0.08	4.65	21.25
9. Local Cubic	CCT b	0.95	0.15	4.42	22.26
10. Local Cubic (Bias-Correct)	CCT	0.96	0.15	5.18	26.80

Note: To understand the coverage properties of different models used for RK estimation, we conduct a Monte Carlo exercise based on 250 iterations. We assume a data-generating process with substantial curvature as well as a true policy kink at zero: $E(y|x) = 10x\mathbf{1}_{x>0} + sin(15(x-0.1)) + x^2$. x is distributed uniformly on [-1,1]. We construct a single Monte Carlo draw by drawing 10,000 observations wit $y = E(y|x) + \varepsilon$ where $\varepsilon \sim N(0, 0.5)$. In rows 1-3, we evaluate the performance of local linear, local quadratic and local cubic models. We also analyze a cubic spline model in row 4 which analyzes all the data using splines between equally-spaced knots, allowing for a discontinuous slope change at the policy kink. In rows 1-4, we choose the bandwidth - or the knot spacing - based on a generalized cross-validation criterion (Wahba, 1990). At the optimal bandwidth, we compute an RK estimate and a 95% confidence interval using robust standard errors. We also analyze the performance of different versions of the "robust" estimator proposed by CCT. CCT propose using two bandwidths: one for point estimation ("h") and a second for bias-correction ("b"). In rows 5-7, we evaluate the linear model with quadratic bias-correction and in rows 8-10, we evaluate the quadratic model with cubic bias-correction. Coverage denotes the fraction of confidence intervals - for a given specification - that cover the true slope change of 10. |Error| denotes the mean absolute error of the estimate of the slope change at zero. Interval length is the average length of the confidence interval for a given estimation method.

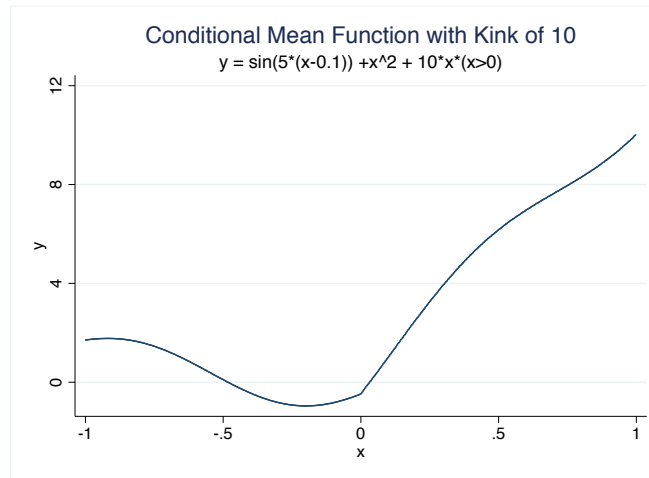FIGURE 7 – LOCAL POLYNOMIALS AND CUBIC SPLINES SIMULATION
II

Conditional Mean Function with Kink of 10
y = sin(5*(x-0.1)) +x^2 + 10*x*(x>0)

Table 3 - Simulation Study of RK Estimation Models

Model	h Criteria	Coverage	Mean from 250 Replications		
			h	|Error|	Interval Len
1. Local Linear	GCV	0.60	0.09	4.13	16.12
2. Local Quadratic	GCV	0.96	0.12	12.00	54.51
3. Local Cubic	GCV	0.93	0.12	27.86	129.39
4. Cubic Spline	GCV	0.95	0.25	1.37	5.83
5. Local Linear	CCT h	0.70	0.10	1.57	4.41
6. Local Quadratic	CCT b	0.96	0.24	0.93	4.57
7. Local Quad (Bias-Correct)	CCT	0.95	0.24	1.15	5.88
8. Local Quadratic	CCT h	0.96	0.25	0.91	4.24
9. Local Cubic	CCT b	0.96	0.43	0.96	4.72
10. Local Cubic (Bias-Correct)	CCT	0.99	0.43	1.03	5.55

Note: See notes to Figure 6. The only difference is that $E(y|x) = 10x\mathbf{1}_{x>0} + sin(5(x-0.1)) + x^2$.

A   A Review of the Regression Kink Design

In this section, we recap some of the key results derived in the canonical CLPW framework to formally introduce the RK design. The framework builds on the following non-separable model:

$$Y = y(B, V, U)$$

with Y as outcome, B a continuous regressor of interest or policy variable, V an observed covariate, and an error term U. To fix ideas, it may be useful to think of Y as duration of unemployment of a job-loser, B as level of unemployment benefits, and V as pre-job-loss earnings. The policy function $B = b(V)$ maps the covariate V into B. For instance, in the setting of unemployment insurance (UI), many UI systems – such as the one in Austria – feature a linear relationship between earnings V and benefits B up until some maximum level of the benefit after which the level of the benefit is constant. The RK design exploits such kinks in the policy function $b(V)$ by estimating whether the slope change in the relationship between B and V at the kink point can also be found in the relationship between the outcome variable Y and V around the kink point. To illustrate, Figure 2 shows plots of unemployment duration and benefits plotted against earnings in the previous year based on Austrian UI data (CLPW).[15]

The goal of the RK framework is to identify a causal effect of B on Y. CLPW define a "treatment on the treated" parameter of interest:

$$TT_{b|v}(b, v) = \int \frac{\partial y(b, v, u)}{\partial b} dF_{U|B=b, V=v}(u)$$

where $F_{U|B=b, V=v}(u)$ denotes the c.d.f. of U conditional on B, V equal to b and v, respectively. This parameter captures the average effect of a marginal increase in b at a specific value of (b, v) holding fixed the distribution of the unobservables, $F_{U|B=b, V=v}(.)$.

CLPW prove that under mild regularity conditions this "treatment on the treated" parameter is identified by the change in the slope of the outcome variable w.r.t. V at the kink point (normalized to be at $V = 0$ here) divided by the change in the slope of the policy function B at the kink point:[16]

$$\frac{\lim_{v \to 0^+} dE(Y|V=v)/dv - \lim_{v \to 0^-} dE(Y|V=v)/dv}{\lim_{v \to 0^+} db(v)/dv - \lim_{v \to 0^-} db(v)/dv} = TT_{b|v} = \tau_{RK}$$

Analogous to the RD setting, there are "sharp" and "fuzzy" versions of the RK design. In the

---

[15]The policy kink that is featured in these plots is the earnings threshold for minimum benefits: unemployed individuals below this threshold only qualify for minimum benefits; benefits increase linearly with earnings above the threshold.

[16]These assumptions are:

(1) (Regularity) $y(.,.,.)$ is a continuous function with $\frac{\partial y(b,v,u)}{\partial b}$ continuous in b for all b, v, and u.

(2) (Smooth effect of V) $\frac{\partial y(b,v,u)}{\partial v}$ for all b, v, and u.

(3) (First stage) $b(\cdot)$ is a known function, everywhere continuous and continuously differentiable on $(-\infty, 0)$ and $(0, \infty)$, but $\lim_{v \to 0^+} b'(v) \neq \lim_{v \to 0^-} b'(v)$. In addition, $f_{V|U=u}(0)$ is strictly positive for all $u \in A$, where $\sum_{u \in A} \Pr(U = u) > 0$.

(4) (Smooth density) $F_{V|U=u}(v)$ is twice continuously differentiable in v for all v, u. That is, the derivative of the conditional probability density function $f_{V|U=u}(v)$, $\frac{\partial f_{V|U=u}(v)}{\partial v}$, is continuous in v for all u.

fuzzy setting, the econometrician estimates the change in the slop of the policy function at the kink point – the denominator in the expression above – as in the top panel of Figure 2. In the sharp RK design, in contrast, the change in the slope of the policy function is treated as known. To keep our exposition of the RK design as concise as possible, we focus on the sharp RK design in this paper or, put alternatively, the reduced form relationship between Y and V (numerator in expression above).

Local polynomial regression techniques (Fan and Gijbels, 1996) are used for estimation of the change in slope of the outcome variable Y at the kink point. The data is split into two subsamples to the left and right of the kink point (denoted by $+$ and $-$, respectively) and a local polynomial regression is estimated separately for each subsample. This amounts to solving the following least squares problem:

$$\min_{\{\beta_j^-\}} \quad \sum_{i=1}^{N^-}\{Y_i^- - \sum_{j=0}^{p}\tilde{\beta}_j^-(V_i^-)^j\}^2 K\left(\frac{V_i^-}{h}\right)$$

$$\min_{\{\beta_j^+\}} \quad \sum_{i=1}^{N^+}\{Y_i^+ - \sum_{j=0}^{p}\tilde{\beta}_j^+(V_i^+)^j\}^2 K\left(\frac{V_i^+}{h}\right)$$

$$\text{subject to } \tilde{\beta}_0^- = \tilde{\beta}_0^+$$

$$\hat{\tau}_{RK}^P \equiv \qquad\qquad \tilde{\beta}_1^+ - \tilde{\beta}_1^-$$

Here, p denotes the order of the polynomial, K the kernel function, and h the bandwidth used for estimation. The denominator of the left-hand side of equation X is identified as $\hat{\beta}_1^+ - \hat{\beta}_1^-$. All papers in the RK literature that we are aware of have adopted a uniform kernel as choice of K. CLPW impose the restriction that $\tilde{\beta}_0^+ = \tilde{\beta}_0^-$ so that the regression function is continuous at the kink.

CLPW establish the rate of shrinkage for the bandwidth necessary for valid inference and prove the following asymptotic property of local linear and quadratic sharp RK estimators:[17]

$$\sqrt{nh^3}(\hat{\tau}_{RK}^P - \tau_{RK}) \xrightarrow{d} N(0, C_P \cdot \Omega_{RK})$$

where P denotes the order of the polynomial used for estimation and $\Omega_{RK} = \frac{\sigma_Y^2(0^+)+\sigma_Y^2(0^-)}{(\kappa_1^+-\kappa_1^-)^2 f(0)}$.[18] C_P is a constant that depends on the order of the polynomial. For the uniform kernel case, CLPW establish that $C_1 = 12$ and $C_2 = 192$. Building on these results, inference for the RK design can be conducted using robust standard errors (White, 1980).

---

[17]See 3.1.2 of CLPW for the assumptions necessary for the following statement. In particular, they assume that the bandwidth sequence satisfies $h \propto N^{-\rho}$ with $\rho \in (\frac{1}{5}, \frac{1}{3})$ for the linear and and $\rho \in (\frac{1}{7}, \frac{1}{3})$ for the quadratic case. For this choice of h, the asymptotic approximation for the bias converges to zero faster than the asymptotic approximation for the variance does. In a finite sample with positive bandwidth the bias will not be zero, as we detail below.

[18]Here, $\sigma_Y^2(.)$ denotes the conditional variance of Y and $(\kappa_1^+ - \kappa_1^-)$ the change in the derivative of the policy function at the kink point.

B Extensions to the Permutation Test

B.1 Extensions of the RK Permutation Test

This framework can be modified easily to account for alternative null hypotheses, discrete distributions, bandwidth selection procedures and quadratic specifications.

Alternative Null Hypothesis: The null hypothesis can be that treatment has no impact anywhere, including at the policy kink.

Hypothesis: Treatment Irrelevance RK *Treatment does not affect outcomes at any* v: $y_1(b(v), v, U) = 0$.

An interval with coverage rate 1-α is

$$P(T > \underline{\tau}^\alpha) = \frac{\int_{\underline{v}}^{\overline{v}} \mathbf{1}(T(\mathbf{v}, \mathbf{y}, v) > \underline{\tau}^\alpha) dv}{\int_{\underline{v}}^{\overline{v}} dv} = 1 - \alpha$$

Under this hypothesis, the one-sided "leave-out" p-value is computed as

$$P(T > T(\mathbf{v}, \mathbf{y}, v_0^P)) = \frac{\int_{\underline{v}}^{\overline{v}} \mathbf{1}(T(\mathbf{v}, \mathbf{y}, v) > T(\mathbf{v}, \mathbf{y}, v_0^P)) dv}{\int_{\underline{v}}^{\overline{v}} dv}$$

Discrete distributions with J potential locations for a policy kink can be accounted for by computing the CDF as $P(T > t) = \frac{1}{J} \sum_j \mathbf{1}(T(\mathbf{v}, \mathbf{y}, j) > t)$.

Bandwidth Choice: So far, we have assumed that a fixed bandwidth h is used at each placebo kink point. However, it is also possible to choose a bandwidth as a function of the data around each kink, so long as the same bandwidth selection procedure is used at both the actual policy kink and the placebo kinks.

Alternative Estimators: Finally, the framework can be modified to incorporate other estimators, e.g., higher order polynomials or recently proposed robust estimators (CCT). For instance, in the case of a local quadratic specification, the following design matrix $\tilde{\mathbf{v}}^Q(v_c^P)$ is used instead of $\tilde{\mathbf{v}}(v_c^P)$:

$$\mathbf{v}_c^Q \equiv \tilde{\mathbf{v}}^Q(v_c^P) \equiv \begin{pmatrix} 1 & (v_1 - v_c^P) & (v_1 - v_c^P)^2 & (v_1 - v_c^P)\mathbf{1}(v_1 \geq v_c^P) & (v_1 - v_c^P)^2\mathbf{1}(v_1 \geq v_c^P) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & (v_n - v_c^P) & (v_n - v_c^P)^2 & (v_n - v_c^P)\mathbf{1}(v_n \geq v_c^P) & (v_n - v_c^P)^2\mathbf{1}(v_n \geq v_c^P) \end{pmatrix}$$

The test statistic corresponding to a quadratic RK estimator can then be calculated as follows:
$$T^Q(\mathbf{v}, \mathbf{y}, v_0^P) \equiv e_4'(\mathbf{v}_c^{Q'}\mathbf{v}_c^Q)\mathbf{v}_c^{Q'}\mathbf{y}, \quad |v_i - v_0^P| \leq h$$

B.2   Relabeling for Regression Discontinuity and Difference-in-Difference

For Regression Discontinuity (RD), consider a data-generating process $Y = y(B, V, U)$. where V is a continuous running variable, B is a binary indicator for treatment, and U is a continuous unobservable. The treatment policy is zero for all observations with $V \leq v_0^P$ and one for all observations with $V > v_0^P$. Normalize $v_0^P = 0$. The object of interest is the causal impact of the policy at $V = 0$: $E(Y(1, 0, U) - Y(0, 0, U))$. We have n observations (y_i, b_i, v_i) Suppose that a researcher decides to use a uniform kernel and a local linear regression to estimate an RD model. Define

$$
\mathbf{v}_c^{RD} \equiv \begin{pmatrix} 1 & v_1 & \mathbf{1}(v_1 \geq v_c^P) \\ \vdots & \vdots & \vdots \\ 1 & v_n & \mathbf{1}(v_n \geq v_c^P) \end{pmatrix}
$$

The regression discontinuity estimator, which can be formulated as a test statistic, is:
$$
T^{RD}(\mathbf{v}, \mathbf{y}, v_0^P) \equiv e_3' \left((\mathbf{v}_0^{RD})' \mathbf{v}_0^{RD} \right)^{-1} (\mathbf{v}_0^{RD})' \mathbf{y}, \quad |v_i - v_0^P| \leq h.
$$

Hypothesis: Treatment Irrelevance RD $y(1, v, U) = y(0, v, U) \forall v.$

Under this null hypothesis, and the same assumption of random kink placement from the RK discussion, we can define a placebo test statistic as $T^{RD}(\mathbf{v}, \mathbf{y}, v_c^P) \equiv e_3' \left(\mathbf{v}_c' \mathbf{v}_c \right)^{-1} \mathbf{v}_c' \mathbf{y} \quad |v_i - v_c^P| \leq h.$

For Difference-in-Difference (DD) models, consider a data-generating process $Y = y(X, U)$ where X is a continuous treatment and U is a continuous unobservable. We are interested in a linear approximation to the regression function of Y on X: $\frac{Cov(X,Y)}{Var(X)}$. There are n sample units. Each panel unit i is a vector of y values $\mathbf{y}_i = \{y_{i1} \ldots y_{iJ}\}'$ and a vector of x values $\mathbf{x}_i = \{x_{i1} \ldots x_{iJ}\}'$. Define

$$
\mathbf{x} \equiv \begin{pmatrix} \mathbf{1} & & & \mathbf{x}_1 \\ & \mathbf{1} & & \mathbf{x}_2 \\ & & \ddots & \vdots \\ & & \mathbf{1} & \mathbf{x}_n \end{pmatrix}
$$

where $\mathbf{1}$ is a vector of ones with J elements and $\mathbf{y} = \{\mathbf{y}_1, \mathbf{y}_2 \ldots \mathbf{y}_n\}'$. A researcher interested in $\frac{Cov(X,Y)}{Var(X)}$ would compute test statistic: $T^{DD}(\mathbf{y}, \mathbf{x}) \equiv e_{n+1}' (\mathbf{x}'\mathbf{x})^{-1} \mathbf{x}'\mathbf{y}$. Again, we are interested in testing the hypothesis of treatment irrelevance:

Hypothesis: Treatment Irrelevance DD $Y(x, U) = Y(x', u) \forall x, x', u$

In the RK and RD settings, we had exactly one location for the policy kink or discontinuity and we assumed that location was chosen at random. Here, the treatment structure is more complex: each sample unit i has multiple outcome realizations and multiple x realizations. We pursue a relabeling strategy which preserves the serial correlation within the x realizations and similarly within the y realizations. Consider an alternative design matrix where $\mathbf{k} = \{k_1, k_2 \ldots k_n\}$ is a random re-ordering of integers from 1 to n.

$$
\tilde{\mathbf{x}}(\mathbf{k}) \equiv \begin{pmatrix} \mathbf{1} & & & \mathbf{x}_{k_1} \\ & \mathbf{1} & & \mathbf{x}_{k_2} \\ & & \ddots & \vdots \\ & & \mathbf{1} & \mathbf{x}_{k_n} \end{pmatrix}
$$

Using matrix, we can compute a placebo test statistic: $T(\mathbf{y}, \tilde{\mathbf{x}}(\mathbf{k})) \equiv e'_{n+1} (\tilde{\mathbf{x}}'\tilde{\mathbf{x}})^{-1} \tilde{\mathbf{x}}'\mathbf{y}$. Note that we can construct this placebo statistic under many different $\mathbf{k}$ vectors.

Assumption: Random Assignment of Treatment Sequences *Treatment sequences* $\mathbf{x}_i$ *are assigned randomly to each panel unit* i

Under this assumption, the vector $\{T(\mathbf{y}, \tilde{\mathbf{x}}(\mathbf{k}))\}$ with all possible orderings of $\mathbf{k}$ is the distribution of the test statistic under Treatment Irrelevance DD.

C   Data Appendix on Lee (2008) and Dube, Lester, and Reich (2010)

Lee analyzes a sample which uses all of the data and an estimating equation with a fourth-order polynomial for Democratic vote share and a dummy variable for vote share greater than fifty percent. Our test formalizes Lee's remark that there are no jumps in the estimated conditional mean function except at the policy discontinuity. We use a dataset posted by Caughey and Sekhon (2011) which analyzes elections for the US House of Representatives from 1942 to 2008. Following what has become the recommended practice in the RD literature, we use a local linear regression. At a bandwidth of 2 percent for relative Democratic vote share (so from -2 percent to 2 percent), we estimate a treatment effect of incumbency of about 0.5, which is very similar to Lee's estimate. Next, we construct placebo estimates by relabeling other vote share points from -48 percent through -2 percent and 2 percent through 48 percent (there is little sample mass below -48 percent and above 48 percent).

DLR's preferred specification, in their notation, is

$$\ln y_{ipt} = \alpha + \eta \ln(w_{it}^M) + \gamma \ln(pop_{it}) + \phi_i + \tau_{pt} + \varepsilon_{ipt}$$

where i indexes counties, t indexes quarters, and p indexes pairs of contiguous counties. y_{ipt} is the employment of restaurant workers, w_{it}^M is the local minimum wage, and pop_{it} is county population. The results are reported in Table 2, column 6 of their paper.

DLR have made their code and analysis sample publicly available.[19] Although DLR's analysis uses county pairs, the variation in their data comes from state-level changes to the minimum wage over time. We apply the relabeling procedure outlined in Section B.2 to their data, by taking each state's minimum wage sequence and applying it to a different, randomly chosen state's county-level data on employment and population. Note that we preserve the joint distribution of county employment and county population because county population, in the language of Abadie, Athey, Imbens, and Wooldridge (2014), is a fixed "attribute" rather than a "potential cause".

---

[19]

http://dvn.iq.harvard.edu/dvn/dv/restat/faces/study/StudyPage.xhtml?studyId=70784&versionNumber=2