NBER WORKING PAPER SERIES

REGRESSION KINK DESIGN:
THEORY AND PRACTICE

David Card
David S. Lee
Zhuan Pei
Andrea Weber

Regression Kink Design: Theory and Practice
David Card, David S. Lee, Zhuan Pei, and Andrea Weber
NBER Working Paper No. 22781
October 2016
JEL No. C2,J65

ABSTRACT

A regression kink design (RKD or RK design) can be used to identify casual effects in settings where the regressor of interest is a kinked function of an assignment variable. In this paper, we apply an RKD approach to study the effect of unemployment benefits on the duration of joblessness in Austria, and discuss implementation issues that may arise in similar settings, including the use of bandwidth selection algorithms and bias-correction procedures. Although recent developments in nonparametric estimation (e.g. Imbens et al. (2012) and Calonico et al. (2014)) are sometimes interpreted by practitioners as pointing to a default estimation procedure, we show that in any given application different procedures may perform better or worse. In particular, Monte Carlo simulations based on data generating processes that closely resemble the data from our application show that some asymptotically dominant procedures may actually perform worse than "sub-optimal" alternatives in a given empirical application.

David Card
Department of Economics
549 Evans Hall, #3880
University of California, Berkeley
Berkeley, CA 94720-3880
and NBER
card@econ.berkeley.edu

David S. Lee
Princeton University
4 Nassau Hall
Princeton, NJ 08544J
and NBER
davidlee@princeton.edu

Zhuan Pei
Dept. of Policy Analysis and Management
134 Martha Van Rensselaer Hall
Cornell University
Ithaca, NY 14853-4401
peizhuan@gmail.com

Andrea Weber
Vienna University of Economics and Business
Economics Department
Welthandelsplatz 1
1020 Vienna
Austria
andrea.weber@wu.ac.at

1 Introduction

Regression discontinuity designs (RDD) have become part of the toolkit used by applied microeconomists to identify causal effects in observations settings. Among the recent literatures that have successfully adopted RDD techniques are studies of educational outcomes (e.g., Angrist and Lavy (1999); Chay et al. (2005)), health (e.g., Card et al. (2009a); McCrary and Royer (2011)), election outcomes (Lee (2008); Lee et al. (2004)), unions (DiNardo and Lee (2004); Lee and Mas (2012)), and unemployment (e.g., Card et al. (2007); Schmieder et al. (2012)). Concurrent with this growth in empirical applications is the widespread adoption of local polynomial methods for estimation and inference, following the recommendations of Porter (2003), Imbens and Lemieux (2008), and Lee and Lemieux (2010), and a series of recent papers offering guidance on how to select appropriate local polynomial specifications (e.g., Imbens and Kalyanaraman (2012), Calonico et al. (2014), Calonico et al. (2016a)).

In this paper, we discuss how the methods that are now widely used in RDD estimation can be extended to implement a *regression kink design* (RKD), a term first coined by Nielsen et al. (2010). Like an RDD, an RKD arises when a policy variable of interest is determined (at least in part) by a known assignment rule. While RDD posits a discontinuity in the assignment rule, in a regression kink design the policy rule is assumed to have a *kink* in the relationship between the policy variable and the underlying assignment variable. Such kinks arise, for example, when a benefit formula is subject to minimum or maximum values, or when marginal tax rates exhibit discrete jumps. The idea of the regression kink design is to examine the slope of the relationship between the outcome of interest and the assignment variable at the exact location of the kink in the policy formula. Provided that individuals on either side of the kink threshold are "similar", any kink in the outcome can be attributed to the treatment effect of the policy variable. As in an RDD, implementing an RKD involves the estimation of quantities "close" to a threshold using (local) polynomial regressions. Instead of estimating a shift in the intercept, however, we are interested in estimating a slope change.

We illustrate the approach by applying a "fuzzy" RKD to study the effects of unemployment insurance (UI) benefits on the duration of joblessness in Austria. The UI benefit schedule in Austria, as in other countries, is an increasing function of past earnings, subject to a minimum and a maximum benefit. The kinks in the benefit schedule generated by the minimum and maximum benefit levels allow us to estimate the elasticity of jobless duration with respect to benefit levels for a subpopulation of claimants with earnings

near the kink point.

The application is an extension of the example presented in Card et al. (2015b) – hereafter, CLPW. This paper presents a more extensive discussion of the potential issues that arise in applying local polynomial methods in the Austrian UI setting. We also use a different measure of joblessness as our key outcome variable, and compare results at the kinks created by the maximum and minimum benefit rules. As in CLPW, we present a range of alternative estimates of the elasticity of joblessness with respect to UI benefits using local linear and quadratic polynomial models, several bandwidth selection algorithms – including a rule of thumb procedure based on Fan and Gijbels (1996) (hereafter, FG) and Calonico et al. (2016b) (hereafter, CCFT) – and bias-corrected robust confidence intervals proposed by Calonico et al. (2014) (hereafter, CCT). These procedures have made a large impact on recent empirical work and are often interpreted by practitioners as providing a default estimation approach that yields "preferred" results. In our application, however, we find that alternative procedures give rise to widely disparate findings. Moreover, the default method suggested by CCT (and CCFT) – local quadratic estimator with regularized bandwidth selector and bias-correction – is typically uninformative, despite visual evidence of kinks in both the UI benefit assignment rule and in the outcomes of interest.

We then offer some suggestions for assessing the performance of alternative estimators by conducting Monte Carlo simulations of data generating processes (DGP's) that are closely based on the actual data in the specific application. In our setting, this approach suggests that local quadratic estimators have substantially larger (asymptotic) mean squared errors (MSE) than local linear estimators. Moreover, the bias correction procedure suggested by CCT leads to a large loss in precision with only modest offsetting reductions in bias. Of course, as we emphasize throughout this paper these conclusions are specific to our empirical context: The simulations and empirical RD examples of Calonico et al. (2014), for example, show limited precision loss associated with bias correction.

We close our discussion with a few brief comments on a related study by Card et al. (2015a). Card et al. (2015a) applies an RKD approach to study the affect of UI benefit levels on the duration of unemployment, using administrative data from the state of Missouri. In contrast to our Austrian example, the Missouri application has a much sharper "first stage" relationship between benefit levels and baseline earnings, leading to more stable estimates across alternative estimators.

The rest of the paper is structured as follows. We review the main identification results from CLPW in Section 2, and the theory of estimation and inference in Section 3. Our empirical analysis on the impact of

Austrian UI benefits follows in Section 4, where we discuss institutional background (Subsection 4.1), our data (Subsection 4.2), graphical and estimation results (Subsections 4.3 and 4.4), and related findings from Card et al. (2015a) (Subsection 4.5). Section 5 concludes.

2 Review of Identification

In this section, we review identification results in regression kink designs. Following the notation in CLPW, we use B to denote the treatment variable of interest (e.g., the unemployment benefit level), V the assignment variable (e.g., earnings in the period prior to the start of the benefit spell), U the error term and $Y \equiv y(B, V, U)$ the outcome of interest (e.g., the duration of time until the next job). We are interested in the causal effect of increasing B on Y, i.e. the partial derivative of y with respect to its first argument, which we denote by $y_1(B, V, U)$.

Identification of the treatment effect of B is formally investigated in Nielsen et al. (2010) and CLPW under the "sharp" regression kink design, where B is a *deterministic* function of V, i.e., $B = b(V)$ with a kink (slope change) at $V = 0$.[1] Nielsen et al. (2010) prove identification in the case where the function $y(B, V, U)$ is separably additive, while CLPW extend the result to the nonseparable case. In particular,

$$\frac{\lim_{v_0 \to 0^+} \frac{dE[Y|V=v]}{dv}\Big|_{v=v_0} - \lim_{v_0 \to 0^-} \frac{dE[Y|V=v]}{dv}\Big|_{v=v_0}}{\lim_{v_0 \to 0^+} \frac{db(v)}{dv}\Big|_{v=v_0} - \lim_{v_0 \to 0^-} \frac{db(v)}{dv}\Big|_{v=v_0}} = E[y_1(b_0, 0, U)|V=0], \text{ where } b_0 = b(0). \tag{1}$$

Equation (1) states that the RKD estimand – the slope change in the outcome variable (numerator) scaled by the change in the first stage (denominator) – identifies an average treatment effect. This treatment effect parameter is what Florens et al. (2008) call the treatment-on-the-treated (TOT) and what Altonji and Matzkin (2005) call the local average response (LAR). Analogous to Lee (2008), CLPW interpret the identified TOT/LAR parameter as a weighted average of the treatment effects across the population, where individuals receive higher weights for having a higher likelihood of being at the threshold ($V = 0$). Sufficient regularity conditions are proposed for identification. Intuitively speaking, these conditions require that everything evolves smoothly across the threshold *except* the derivative of the first-stage relationship, $b(\cdot)$. Similar to an RDD, these conditions can be tested by examining the smoothness (in the first derivative) of the observed density of V, f_V, and the conditional expectation of any predetermined covariate, $E[X|V=v]$.

---

[1] Unless otherwise mentioned, the kink threshold is normalized to zero throughout this article.

CLPW also prove identification in a "fuzzy" generalization of the sharp design, where the observed relationship between B and V deviates from the statutory rule, $b(\cdot)$. The deviation can arise due to measurement errors, noncompliance with the policy formula, or dependence of the formula for B on other unobserved variables. In this more general setup, the RK estimand still identifies a weighted average of the treatment effect subject to an expanded set of regularity conditions. As with the sharp case, CLPW show that the validity of these conditions can be tested by examining the smoothness of f_V and $E[X|V = v]$.[2]

In our experience (Card et al. (2009b), Card et al. (2012) and Card et al. (2015a)), potential applications of the regression kink idea using administrative data usually require a fuzzy design approach. In these datasets, it is almost inevitable that there are unexplained deviations from the prescribed benefit schedule. Nevertheless, the majority of the data points typically follow the statutory schedule exactly. As we will see below, the strength of this first stage relationship is important for obtaining precise and robust empirical estimates.

3 Review of Estimation and Inference

In this section, we review the theory of estimation and inference in a regression kink design, as developed in CCT and CLPW. The kinks in the outcome and treatment variable are measured by estimating local polynomial regressions of order p to the left and right of the kink point $(V = 0)$ with bandwidth h and kernel K. Denoting the resulting first-stage slope estimators above and below the threshold by $\hat{\kappa}_1^-$ and $\hat{\kappa}_1^+$ and outcome slope estimators by $\hat{\beta}_1^+$ and $\hat{\beta}_1^-$, the fuzzy RKD estimator is defined as

$$\hat{\tau} = \frac{\hat{\beta}_1^+ - \hat{\beta}_1^-}{\hat{\kappa}_1^+ - \hat{\kappa}_1^-}. \tag{2}$$

Estimation in a sharp design can be considered as a special case in which the estimators $\hat{\kappa}_1^+$ and $\hat{\kappa}_1^-$ are equated to the *known* slopes in the first stage: $\hat{\kappa}_1^+ = \lim_{v \to 0^+} b'(v)$ and $\hat{\kappa}_1^- = \lim_{v \to 0^-} b'(v)$.

When implementing the RKD estimator in practice, one must make choices for the kernel K, bandwidth h, and polynomial order p. Regarding K, Cheng et al. (1997) show that the triangular kernel is boundary optimal. In our experience based on actual applications and Monte Carlo examples, however, the efficiency losses from using a uniform kernel are small. Moreover, the ease of implementation of a uniform kernel –

---

[2]In this paper we focus on tests of the smoothness of the predetermined covariates at the kink. CLPW discuss the smoothness of the density of UI claimants around the kink point in the benefit formula in Austria.

which converts the RKD problem to a simple OLS/2SLS problem using data within a window of width h from the threshold – makes it an attractive choice, and we use it in our applications below.[3]

Following IK, CCT propose an MSE-optimal bandwidth selector for the sharp RKD. A feature of the IK procedure, also adopted by CCT, is the inclusion of a "regularization" term in the bandwidth formula that prevents the selection of large bandwidths. CLPW extend these selectors for the fuzzy design and experiment with removing the regularization term in the default CCT bandwidth. The presence or absence of the regularization term does not affect the optimality features of the bandwidth selector, since the term shrinks toward 0 in large samples. In the samples used in our application, however, the regularization term is often relatively large, leading to 30-70 percent reductions in the size of the bandwidth.

Very recent papers by Calonico et al. (2016a), Calonico et al. (2016c) and CCFT propose additional bandwidth choices. Building on Calonico et al. (2015a), Calonico et al. (2016a) derive a bandwidth selector in the RD/RK context that minimizes the asymptotic coverage error (CE) of the confidence interval; this CE-optimal bandwidth is obtained by shrinking the MSE-optimal bandwidth by a factor that depends on the sample size n and polynomial order p.[4] Calonico et al. (2016c) study the identification and local estimation of RDD and RKD after incorporating predetermined covariates in an additively separable way, and propose corresponding MSE-optimal and CE-optimal bandwidth selectors. CCFT implements Calonico et al. (2016a) and Calonico et al. (2016c) as well as additional bandwidth selectors in Stata and R.

Data-dependent estimation approaches that build on MSE-optimal bandwidth selectors will in general introduce an asymptotic bias in the resulting estimator. CCT propose a procedure to estimate this bias term and construct robust confidence intervals that are re-centered *and* account for the estimation error in the bias correction term. Alternatively we can construct a bias-corrected estimator $\hat{\tau}^{bc}$ by subtracting the estimated bias term from $\hat{\tau}$, and an associated standard error that incorporates the uncertainty in the bias term.[5]

A final issue is the choice of the polynomial order, p. Standard arguments suggest that a local quadratic ($p = 2$) is preferred to local linear ($p = 1$) in estimating the boundary derivatives in the RK design, because

---

[3]As noted in Card et al. (2012) and CLPW, imposing continuity in $E[Y|V = v]$ and $E[B|V = v]$, which follows from the RKD identifying assumptions, does not affect the asymptotic properties of $\hat{\tau}$ when a uniform kernel and symmetric bandwidths are used in estimation. However, imposing continuity may change the asymptotic properties of $\hat{\tau}$ when *asymmetric* bandwidths are used. For our conventional estimates below, we use symmetric bandwidths and a uniform kernel and report results obtained by imposing continuity, and we find that removing this restriction has little impact on the estimates. In certain applications, there may be both a level and slope change in the first stage relationship. CLPW shows that in general the RK estimand does not identify a treatment effect, though it is possible to recover a causal parameter with additional model restrictions (e.g. Turner (2013)). In these applications, it is clearly sensible to not impose continuity.

[4]It is worth noting that the motivation of Calonico et al. (2015a) and Calonico et al. (2016a) echos the message of this paper that there is not a single procedure "optimal" for all empirical applications.

[5]For the bias-corrected estimates, we follow the CCFT procedure which does not impose continuity in $E[Y|V = v]$ or $E[B|V = v]$.

the former leads to a lower order of bias than the latter, allowing a sequence of bandwidths that shrinks at a slower rate and delivering a smaller order asymptotic mean squared error. As noted by Ruppert and Wand (1994) and Fan and Gijbels (1996), however, arguments based on asymptotic rates do not imply a universally preferred choice. Instead, as we argue in Card et al. (2014) the best choice for p depends on the sample size and the derivatives of the conditional expectation functions, $E[Y|V = v]$ and $E[B|V = v]$, in the particular data set of interest.[6]

In the following section, we illustrate the implementation of a fuzzy RKD approach with applications using unemployment insurance data from Austria over the 2001-2012 period. We exploit the kinks in the unemployment benefit schedule at the minimum and maximum benefit levels to estimate the effect of UI benefits on the duration of joblessness. We present a variety of estimates of this effect, which are obtained by using local linear and local quadratic regressions with several bandwidth selectors, with and without bias correction: the MSE-optimal bandwidth, calculated including the regularization term (which is the "default" option in the software package distributed by CCFT), the MSE-optimal bandwidth without regularization, the MSE-optimal bandwidth for a fuzzy design (without regularization) and the FG bandwidth as defined in Card et al. (2012). Using simulated data with DGP's closely resembling our actual data, we assess the performance of these alternative estimators.

4 Empirical Application: The Effect of UI Benefits on Unemployment Durations

4.1 The Unemployment Insurance Benefit Schedule in Austria

A UI claimant's benefit entitlement in Austria depends on his or her earnings in the "base period" prior to layoff. This period is either the previous calendar year or the second most recent year (depending on whether the claim starts early or late in the year). The UI replacement rate is 55%, subject to a maximum benefit level. Claimants with dependent family members are also eligible for a family allowance that is added to their basic UI benefit. For low-wage claimants, there is also a minimum benefit level that applies as long as benefits do not exceed 60% (for a single individual) or 80% (for a claimant with dependents) of base period

---

[6]In a related study, Ganong and Jäger (2014) raise concerns about the sensitivity of the RKD estimates when the relationship between the running variable and the outcome is highly nonlinear. They propose a permutation test to account for the estimation bias. We perform their test on our data and discuss the details in Subsection 4.4.4.

earnings. The end result of these rules is a schedule with multiple kink points.

To illustrate, Figure 1 plots actual daily UI benefits against annual base year earnings for a sample of UI claimants in 2004. The large fraction of claimants whose observed UI benefits are exactly equal to the amount predicted by the formula leads to a series of clearly discernible lines in the figure, though there are also many observations scattered above and below these lines.[7] Specifically, in the middle of the figure, there are 5 distinct upward-sloping linear segments, corresponding to claimants with 0, 1, 2, 3, or 4 dependents.

These schedules all reach an upper kink point at the maximum benefit threshold which we call T^{max}, shown in the graph by a solid vertical line on the right side of the figure. At the lower end, the situation is more complicated: each of the upward-sloping segments reaches the minimum daily benefit (22 euros per day in 2004) at a different level of earnings, reflecting the fact that the basic benefit includes family allowances, but the minimum does not. We define the minimum benefit T^{min} – shown by the solid vertical line on the left side of the figure – as the level of earnings at which a single claimant receives the basic minimum benefit amount.[8] Finally, very low-paid claimants receive a subminimum benefit, which is again an upward-sloping function of base period earnings. These subminimum claimants fall largely in two groups: single claimants (whose benefit is 60% of their base earnings, net of taxes) and those with dependents (whose benefits are 80% of their net base year earnings).[9]

4.2   Data and Analysis Sample

Our data are drawn from the Austrian Social Security Database and contain information on base period earnings, the level of benefits actually received, benefit duration, and the starting and ending dates of employment spells, as well as demographic information including age, gender, education, marital status, job tenure, and industry.[10] Our analysis sample includes qualifying claims filed between 2001 and 2012. Among other sample restrictions, we focus on claimants whose earnings are high enough to place them above the subminimum portion of the benefit schedule for a single claimant (i.e., above the level indicated by the

---

[7]The existence of these off-the-schedule observations is attributable to some combination of errors in the calculation of base year earnings (due to errors in the calculation of the claim start date, for example), errors in the Social Security earnings records that are overridden by benefit administrators, and mis-reported UI benefits. Similar errors have been found in many other settings – e.g., Kapteyn and Ypma (2007).

[8]Note that T^{min} is only a kink point for single claimants: the schedules for claimants with one or more dependents have kinks at earnings thresholds below T^{min}. Since we do not know the number of dependents a claimant reports, however, we use the kink point for the largest group of claimants, i.e. those with no dependents.

[9]The line for low-earning single claimants actually bends, reflecting the earnings threshold at which a single claimant begins paying income taxes.

[10]See Card et al. (2015c) for more details on the data and analysis samples.

vertical dashed line on the left side of Figure 1) and below the Social Security earnings cap (denoted by the vertical dashed line on the right side of Figure 1).

We study the effects of the kinks induced by the minimum and maximum benefit levels separately by defining two subsamples. Specifically, we divide the claimants in each year into two (roughly) equal groups based on their gross base year earnings: those below the 50th percentile are assigned to the "bottom kink" sample, while those above this threshold are assigned to the "top kink" sample. Since earnings have a right-skewed distribution, there is narrower support for our observed assignment variable (annual base year earnings) around the bottom kink relative to the top kink. We re-center base year earnings for observations in the bottom kink subsample around T^{min} (indicated by the solid vertical line just below €20,000 in Figure 1), and base year earnings for those in the top kink subsample around T^{max} (indicated by the solid vertical line just below €40,000 in Figure 1), so both kinks occur at $V = 0$. Each sample contains about 275,000 observations. As mentioned above, the existence of family allowances in Austria generates several distinct benefit schedules that all share the same upper kink location, but have different bottom kink thresholds. Therefore, in the top kink sample, the change in the *slope* of the benefit schedule is the same for all claimants at T^{max}. At the lower end of the benefit schedule, the kink location and magnitudes differ for families with or without dependents: To the left of T^{min}, benefits for claimants with no dependents are constant, whereas benefits for claimants with dependents continue to fall. Since we cannot distinguish between claimants who do or do not have dependents in our data, the measured kink in the average benefit at T^{min} will be proportional to the number of claimants without dependents.

Table 1 reports basic summary statistics for the bottom and top kink samples. Mean base year earnings for the bottom kink group are about €22,000, with a relatively narrow range of variation (standard deviation = €2,800), while mean earnings in the top kink group are higher (mean = €34,000) and more dispersed (standard deviation = €6,700). Mean daily UI benefits are €25.2 for the bottom kink group (implying an annualized benefit of €9,200, about 44% of T^{min}), while mean benefits for the top kink sample are €33.5 (implying an annualized benefit of €12,300, about 28% of T^{max}). Claimants in the bottom kink sample are more likely to be female, are a little younger, less likely to be married, more likely to have had a blue-collar occupation, and are less likely to have post-secondary education. Despite the differences in demographic characteristics and mean pay, the means of our main outcome variable, the time to next job, are quite similar in the two samples: the average duration of joblessness is around 150 days.[11] Only about 10 percent of

---

[11]In CLPW, the main outcome is registered unemployment duration, an alternative measure of unemployment.

claimants exhaust their regular UI benefits.

4.3 Graphical Overview of the Effect of Kinks in the UI Benefit Schedule

The first-stage relationships between (re-centered) base period earnings and UI benefits are presented in Figures 2a and 2b, which divide base period earnings into small bins and plot the average UI benefits in each bin.[12] These figures average away some of the noise seen in the raw scatter plots of Figures 1 and make it easier to visualize the conditional expectation function $E[B|V]$. Kinks are visible in both figures: we see an increase in slope as earnings go across T^{min} in Figure 2a and a decrease as they go across T^{max} in Figure 2b. As mentioned above, the UI schedule is only kinked for single claimants at T^{min}, whereas it is kinked for all claimants at T^{max}. As a result, the slope change at T^{max} appears sharper than that at T^{min}.[13] Figures 3a and 3b present analogous plots for our main outcome variable, the logarithm of time to the next job. These figures also show discernible kinks, though there is clearly more variability than in Figures 2a and 2b.

To check the validity of the kink design, we examine the patterns of the predetermined covariates around T^{min} and T^{max}. We do this by constructing a "covariate index" equal to the predicted value from a simple linear regression of the log of the time to next job on a total of 59 predetermined covariates, including gender, occupation, age, previous job tenure, quintile of the previous daily wage, industry, region, year of the claim, previous firm size, and the recall rates of the previous employer.[14] This estimated covariate index function can be interpreted as the best linear prediction of mean log time to next job given the vector of predetermined variables. Figures 4a and 4b plot the mean values of the estimated covariate indices around the top and bottom kinks, and they evolve relatively smoothly through both the top and bottom kinks.

---

[12]The bin sizes are 100 Euros and 300 Euros for the bottom (Figure 2a) and top kink (Figure 2b) samples, respectively. Figure 2b is reproduced from Figure 2 of CLPW. See Calonico et al. (2015b) for nonparametric procedures for picking the bin size in RD-type plots.

[13]As noted in CLPW, the slopes in the empirical benefit functions to the left of T^{min}, and to the right of T^{max} are mainly due to dependent allowances. Moving left from T^{min}, for example, the average number of dependent allowances is falling, as claimants with successively higher numbers of dependents hit the minimum benefit level (see Figure 1).

[14]We fit a single prediction model using the pooled bottom kink and top kink samples. The inclusion of the recall variable is motivated by the observation that many workers cycle between seasonal jobs (often with the same employer each year) and UI – see Del Bono and Weber (2008).

4.4 RKD Estimation Results

4.4.1 Reduced Form Kinks in Treatment and Outcome Variables

Table 2a presents the first stage and the reduced form estimates of the kinks in our endogenous policy variable (log daily benefits) and our main outcome variable (log of time to next job) around T^{min} and T^{max}. For each variable we show results using three different bandwidth selection procedures: the MSE-optimal choice with the regularization term included (i.e., the default procedure suggested by CCT procedure, updated using the procedures distributed by CCFT); the MSE-optimal bandwidth selection procedure without regularization; and the FG bandwidth. We show the estimated kink arising from each selected bandwidth, as well as the corresponding bias-corrected estimate and the associated robust 95% confidence interval. We present estimates from local linear models in columns 1 and 3, and from local quadratic models in columns 2 and 4.

Despite the visual evidence of kinks in the benefit formula in Figure 2, especially in the top kink sample, an examination of the estimated "first stage" kinks in Panel A of Table 2a suggests that not all the procedures yield statistically significant kink estimates. In particular, the default MSE-optimal bandwidth selector with regularization chooses relatively small bandwidths for the local linear model and yields an insignificant estimate of the bottom kink ($t = 1.6$) and a marginally significant estimate of the top kink ($t = -2.4$). The corresponding bias-corrected kink estimates are substantially less precise, with confidence intervals that are about 3 times wider. Although the default MSE-optimal procedure chooses somewhat larger bandwidths for the local quadratic models, this potential gain in efficiency is offset by the difficulty of precisely estimating the slopes on either side of the kink point once the quadratic terms are included, and neither the estimated bottom kink nor the estimated top kink in the quadratic models is close to being significant. As with the local linear models, the corresponding bias-corrected quadratic kink estimates are even less precise, with very wide confidence intervals.

Relative to the default MSE-optimal bandwidth selector, the MSE-optimal selector *without regularization* yields much larger bandwidths – 3-5 times larger for the local linear models, and 1.3 to 2.7 times larger for the local quadratic models. These larger bandwidths lead to first stage kink estimates from the local linear models that are relatively precise (with t-ratios in excess of 15 in absolute value). The estimated kinks from the local quadratic estimates, however, are still relatively noisy ($t = 6.5$ for the bottom kink and $t = -2.1$ for the top kink). As is the case when the regularization term is included, the bias correction adds

substantial imprecision, leading to confidence intervals that are 8 times wider for the bottom kink estimates, and 3-5 times wider for the top kink estimates.

The third bandwidth selection procedure – the simple rule of thumb plug-in approach of FG – yields bandwidths for the lower kink that are very close to those selected by the MSE-optimal procedure without regularization, but 2-3 times bigger for the upper kink. Without accounting for bias correction, the implied kink estimates are relatively precise, even using local quadratic polynomials. However, the bias term is imprecisely estimated and the bias-corrected FG estimates have confidence intervals that are relatively wide.

Turning to the reduced form outcome results in Panel B, the estimated kinks in the duration of jobless-ness are less precise than the first-stage kinks in UI benefits. Again, the default MSE-optimal selector with regularization chooses relatively small bandwidths and yields imprecise estimates of the kink. The band-widths under the MSE-optimal procedure without regularization are substantially larger (3 to 4 times larger for the bottom kink, and 1.5 to 1.8 times larger for the top kink) and yield marginally significant estimated kinks in the outcome variable from the local linear models (with t-ratios in excess of 2.5 in absolute value). The FG bandwidths are even larger, and yield estimated kinks that are significant or marginally significant in both the linear and quadratic models. Nevertheless the bias-corrected confidence intervals for all three procedures are relatively wide and include 0 in all but two cases, reflecting the additional uncertainty asso-ciated with the bias correction term.[15] In finite samples, the use of higher order polynomial models and bias correction may come at the cost of an increase in variance relative to lower-order uncorrected models. In the remainder of this subsection we examine the tradeoff between bias and variance associated with the CCT bias correction procedure. We defer a discussion of the polynomial order choice to Subsection 4.4.3.

The intent of the bias correction is to eliminate the bias in the p-th order polynomial estimator $\hat{\tau}_p$ by subtracting the estimated asymptotic bias, $h^p \hat{\rho}_p$, where the constant ρ_p in part depends on the $(p+1)^{th}$ derivative of $E[Y|V=v]$ and $E[B|V=v]$. The cost of bias correction, however, is that the bias term is imprecisely estimated, leading to an increase in the overall variance of the corrected estimator $\hat{\tau}_p^{bc}$ relative to the uncorrected estimator $\hat{\tau}_p$.[16] The usual metric for trading off bias and variance is the (asymptotic) MSE, or AMSE, of the estimator, which is the sum of its squared bias and its variance. By Lemma A1 and Theorem A1 of CCT, the AMSE of $\hat{\tau}_p$ and $\hat{\tau}_{p+1}^{bc}$ are AMSE$(\hat{\tau}_p) = (h^p \rho_p)^2 + o_p(h^{2p}) + var(\hat{\tau}_p)$ and

---

[15]The exceptions are the bias-corrected local linear estimates for the top kink sample with the two larger bandwidths. In these cases, the robust confidence intervals are relatively wide but the bias-corrected point estimates are also large in magnitude.

[16]Remark 5 of CCT states that the variance of $\hat{\tau}_p^{bc}$ is smaller than that of $\hat{\tau}_p$ for a large n, but this asymptotic advantage may not materialize in a given finite sample and does not appear to hold in our data.

AMSE($\hat{\tau}_p^{bc}$) = $o_p(h^{2p}) + \text{var}_p^{bc}$. It follows that the *change* in the AMSE associated with bias correction is asymptotically $-(h^p \rho_p)^2 + \text{var}_p^{bc} - var(\hat{\tau}_p)$.

In Table 2b, we report the estimated bias $h^p \hat{\rho}_p$, its square, and the change in estimated variance, $\widehat{\text{var}}_p^{bc} - \widehat{var}(\hat{\tau}_p)$, for the first stage and reduced form estimators presented in Table 2a. The increase in variance is larger than the estimated squared bias for both the local linear and local quadratic estimates using either the default CCT bandwidth selector or the alternative version that ignores the regularization term. This is also the case for the FG bandwidth in the bottom kink sample. In the top kink sample, however, the estimated bias for the FG bandwidth is relatively large (-0.5), and bias correction appears to decrease the AMSE. This suggests that bias correction could be important for estimators based on the FG bandwidth in the top kink sample. At the same time, since the bias term is estimated (and the estimate is relatively imprecise), it may deviate from the actual bias. In Subsection 4.4.3 below, we evaluate the performance of the various estimators in Monte Carlo simulations using DGP's approximating our data, where we can directly obtain the mean squared errors without having to estimate the bias.

4.4.2 Fuzzy RKD Estimates

In this subsection, we present fuzzy RKD estimates of the elasticity of the duration of joblessness with respect to the level of UI benefits. Building on the discussion in Section 2 and Subsection 4.1, we interpret the fuzzy RK estimate in the bottom kink sample as an average behavioral response for claimants whose base period earnings are close to T^{min} and who follow the benefit schedule intended for single claimants. We interpret the fuzzy RK estimate in the top kink sample as the average elasticity for claimants whose base period earnings are close to T^{max} and who follow any of the benefit schedules seen in Figure 1. Therefore, applying a regression kink approach to the two samples allows us to estimate the elasticity of joblessness with respect to UI benefit generosity for two very different subpopulations.

Table 3 presents estimated elasticities for the local linear and quadratic estimators and the three bandwidth selection procedures discussed above. For reference, we also report the associated first stage estimates for each of the structural estimates. As in Tables 2a and 2b, we show both the conventional estimates and the bias-corrected estimates with robust confidence intervals that take account of the sampling variability in bias correction.

A first observation is that the MSE-optimal procedure with regularization yields highly variable and imprecise estimates of the elasticity of the duration of joblessness with respect to UI benefits. The MSE-

12

optimal procedure without regularization yields somewhat more precise estimates, particularly from the local linear specifications, though accounting for bias correction the confidence intervals include 0 for all specifications except the local linear estimate for the top kink. Again, as we saw in Tables 2a and 2b, the FG bandwidth selection procedure gives rise to estimates that are not to different from those obtained using the MSE-optimal selector without the regularization term, but the bias-corrected FG procedure yields relatively imprecise estimates.

Given the wide range of estimates in Table 3, which specification should we pick? The most conservative choice is the default CCFT procedure, which uses a richer model (local quadratic), bias correction, and small bandwidths in our application. For the bottom kink sample this choice suggests that the elasticity of joblessness with respect to UI benefits is 18.6, whereas for the top kink sample the estimate is wrong-signed and equal to -3.9. However, given the wide confidence intervals both estimates are essentially uninformative. At the opposite extreme, the uncorrected estimates from local linear specifications using the FG bandwidth selector are 1.4 for the bottom kink sample (conventional standard error = 0.4) and 2.4 for the top kink sample (conventional standard error = 1.2). In both cases the estimated bias correction term suggests that the uncorrected estimates are "too small" in magnitude, though the bias correction terms are imprecise.

4.4.3   Comparison of Alternative Estimators

To gain additional insight into the performance of the alternative estimators, we conduct a series of Monte Carlo simulations based on DGP's that closely resemble our actual samples, in the same spirit as the IK and CCT Monte Carlo simulations. Distinct from IK and CCT, however, we impose the first-stage kink parameter and the elasticity parameter in constructing the DGP's because we are interested in the power of the candidate estimators. For the bottom kink sample, we impose the first-stage kink as $\tau_B = 2.3 \times 10^{-5}$ and the elasticity $\tau_{FRKD} = 1.3$, which implies a reduced-form kink of $\tau_Y = \tau_B \cdot \tau_{FRKD} = 3.0 \times 10^{-5}$. For the top kink sample, we set $\tau_B = -1.4 \times 10^{-5}$ and $\tau_{FRKD} = 2.0$ with an implied $\tau_Y = -2.8 \times 10^{-5}$. As with the Monte Carlo simulations in IK and CCT, we specify the DGP's for $E[B|V]$ and $E[Y|V]$ as separate quintics on each side of the threshold. The parameters of the quintics are estimated by regressing, respectively, $B - \tau_B \cdot D \cdot V$ and $Y - \tau_Y \cdot D \cdot V$ on the polynomial terms V^j and $D \cdot V^k$ where $D = 1_{[V \geqslant 0]}$, $j = 0, 1..., 5$ and $k = 2, ..., 5$.[17] For our simulation, we sample V from its empirical distribution and the errors $(\varepsilon_B, \varepsilon_Y)$ jointly

---

[17]There is obviously room for discretion in specifying the approximating DGP, and we believe this is unavoidable. That said, we will also specify an alternative DGP below that imposes zero elasiticty, in order to examine the sensitivity of our conclusions with respect to the choice of DGP.

from the residuals from the quintic regressions, and construct $B = E[B|V] + \varepsilon_B$ and $Y = E[Y|V] + \varepsilon_Y$. We draw 1,000 repeated samples in our Monte Carlo exercise; each simulated sample contains the same number of observations as in the original dataset.

Panels A and B of Table 4 summarize the performance of the alternative estimators for the simulated bottom kink and top kink samples, respectively. The two polynomial orders (linear and quadratic), two bias-correction choices (uncorrected and bias-corrected) and four bandwidth selection procedures (MSE-optimal, MSE-optimal with no regularization, Fuzzy MSE-optimal with no regularization, and FG) give rise to 16 candidate estimators in total. For each estimator, we report the average value of the associated bandwidth(s) and its performance in estimating the first-stage kink (columns 3-5) and the structural elasticity parameter (columns 6-11).

With respect to the first stage kink estimates, we show 3 statistics: the fraction of replications in which the first stage would be declared "insignificant" (i.e., the confidence interval for the first stage kink estimate includes 0); the root mean squared error (RMSE) of the estimator (expressed as a fraction of the true value of the first stage kink); and the coverage rate of the confidence interval. With respect to the structural elasticity estimate we show the RMSE (with and without trimming the 5% of the simulation sample with the greatest deviation between $\hat{\tau}_{FRKD}$ and τ_{FRKD}), the confidence interval coverage rate, the bias and squared bias of the estimator, and the variance of the parameter estimate.

A comparison across the rows of Table 4 reveals several interesting patterns about the performance of the various estimators in simulated datasets designed to mimic our actual data. First, for a given bandwidth procedure and bias correction approach, local quadratic models have higher average RMSE's in both the first stage kink estimates and the structural elasticity estimates. Typically, the quadratic model has 2-3 times larger RMSE than the corresponding linear model. Moreover, the mean squared bias associated with the quadratic estimator is typically larger than the mean squared bias for the linear estimator with the same bandwidth selection and bias correction.[18] A second consistent pattern is that the introduction of bias correction leads to higher RMSE's for both linear and quadratic models and all 4 bandwidth selection procedures, and larger mean squared bias in every case but two (the exceptions are the quadratic model for the top kink sample using either the MSE-optimal bandwidth with regularization or the FG bandwidth). Typically bias correction doubles the RMSE of the kink estimates and leads to a 200-300 percent increase in the RMSE

---

[18]An exception is the bias-corrected estimators applied to the top kink sample – though both the linear and quadratic bias corrected estimators perform poorly in the top kink sample.

of the structural elasticity estimates. A third finding is that the inclusion of the regularization term in the calculation of the MSE-optimal bandwidth results in higher RMSE's for both the first stage kink estimates and (in all but one case) also leads to higher RMSE's for the structural elasticity estimate.[19] Finally, comparing across bandwidth procedures for a given polynomial choice and a given choice over bias correction or not, the FG bandwidth has uniformly the lowest RMSE for both the first stage and structural estimates. Indeed, only the FG procedure delivers structural estimates with an RMSE relative to the true value of the parameter that is less than 0.5. The other bandwidth procedures have proportional RMSE's in excess of 1 for the structural elasticity – a degree of variability which suggests that none of these procedures can provide useful information from our sample.[20]

Although simple comparisons of RMSE suggest that in our application a local linear specification using the FG procedure and no bias correction may be a plausible approach, in the top kink sample the confidence intervals for this estimator are too small, leading to an 11 percentage point under-coverage rate. Thus, we conclude that in these DGP's none of the procedures achieves both lowest RMSE and correct coverage rates.

In a parallel simulation study, we assess the performance of the estimators in DGP's where we impose the true elasticity τ_{FRKD} to be zero: the results are summarized in Table 5. In terms of RMSE, the results show that local linear specifications still dominate quadratic models even when there is no kink in the outcome variable; estimators without bias correction perform better than the corresponding bias-corrected estimators; the MSE-optimal bandwidth selector without regularization performs better than the selector with the regularization term included (with the exception of the quadratic model with bias correction for the top kink sample); and the FG bandwidth selection procedure yields the lowest RMSE among the alternative bandwidth selectors. In terms of coverage rates, however, local linear models using the FG selection procedure without bias correction perform quite poorly: the coverage rate is 2% for the bottom kink simulation and 19% for the top. The bias corrected local linear model with FG bandwidth selection has better coverage rates, but still below 90%.

In another exercise, we provide further evidence on the choice between linear and quadratic specifications in RKD applications by directly estimating the AMSE's for the local linear and quadratic estimators, per Card et al. (2014). In results available upon request, we find that using the MSE-optimal bandwidth

---

[19]The exception is the quadratic model with bias correction for the top kink sample.

[20]In results available upon request, we have also examined the performance of the coverage-error optimal bandwidth selector of Calonico et al. (2016a) and the asymmetric fuzzy MSE-optimal bandwidth selectors in Calonico et al. (2016b), and their corresponding RMSE's are also consistently dominated by that of FG in our simulations.

selection procedure with regularization, the AMSE for the local quadratic model is at least an order of magnitude larger than the AMSE for the local linear model. For the other bandwidth selectors, the linear AMSE is also much smaller (at least 68% smaller). These results lead us to conclude that in our dataset, local linear models are likely to be preferred to local quadratic models.

As a final robustness check, we investigate how sensitive the elasticity estimates are with respect to the choice of bandwidth. Figures 5a and 5b plot the local linear elasticity estimates for the bottom and top kink samples associated with a range of potential bandwidths. Ruppert (1997) argues that one can use the relationship between the point estimates and the bandwidth choice as an indicator of potential bias, with stability in the estimate indicating the absence of significant bias. Figure 5a shows that the estimated elasticity of time to next job with respect to UI benefits around the bottom kink is relatively stable at close to 1.4 for a very wide range of bandwidths. Figure 5b shows that the estimated elasticity around the top kink is a little more sensitive to bandwidth choice, with a larger estimate (between 2 and 3) for lower bandwidths, but an elasticity of 2 or less for bandwidths above €5,000.

4.4.4   Estimates with Covariates

As pointed out by Lee and Lemieux (2010) and Calonico et al. (2016c), incorporating covariates may improve the precision of the estimates. In Table A.1 of the online Supplemental Material (Card et al. (2016)), we present results with 79 covariates included by implementing the estimators in Calonico et al. (2016c) and Calonico et al. (2016b). Compared to their unadjusted counterparts in Table 3, all but one of the covariate-adjusted MSE-optimal main bandwidths in Table A.1 are narrower, by a factor ranging from 6% to 39%. The exception is the local quadratic estimator for the bottom kink sample, for which the covariate-adjusted main bandwidth is 32% wider.

In terms of the precision of the RK estimates, the results are mixed. Covariate adjustment increases the conventional first stage t-statistic in exactly half of the cases. In 2 of the 8 cases, the first stage robust confidence interval no longer covers 0 after covariate adjustment, but the adjustment also has the opposite effect in another case. For the structural elasticity estimator, covariate adjustment drastically reduces the standard error in the local-quadratic-bottom-kink case to 7% of its original level, as a result of the larger bandwidth. However, in all 7 other cases, the conventional standard error actually *increases* after covariate adjustment by as much as 150%. Similarly, the width of the robust confidence interval increases in 6 out of 8 cases after covariate adjustment.

Despite its inability to improve precision in the Austrian application, accounting for covariates may be important for other inference procedures, e.g. the permutation test suggested by Ganong and Jäger (2014). As a starting point for our discussion, note that if we plot the relationship between baseline earnings and the log time to next job not just locally around the kink points as in Figures 3a and 3b, but globally for the whole range of baseline earnings available in the data, the relationship is highly nonlinear. The upper- and lower-left panels in Figure 6 show the raw data plots, with observations centered at the annual cutoff values for the bottom and top-kink points in the unemployment benefit formula. Over the full range of baseline earnings the relationship appears to follow a u-shape, and this general pattern renders the kinks at the policy thresholds less salient.

Ganong and Jäger (2014) raise concerns about the sensitivity of the RKD estimates when the relationship between the running variable and the outcome is highly nonlinear. To assess this sensitivity, they suggest a permutation-type test which shifts the threshold value for the kink across the support of the running variable and estimates slope changes in these placebo samples. One example that they use to demonstrate their method is based on the raw data plot for the bottom kink sample in Figure 6. Ganong and Jäger (2014) argue that because of the curvature in the global relationships many of the placebo estimates result in large and significant slope changes.

We replicate this exercise based on the raw individual level data. Following Ganong and Jäger (2014), we randomly draw 200 cutoff values for both the bottom and top kink samples over the range of baseline earnings where there is no kink in the benefit formula. The empirical c.d.f.'s of the coefficient estimate and the corresponding t-statistic are shown by the solid lines in Figure 7, where the graphs in the top row correspond to the bottom kink sample and those in the bottom row correspond to the top kink sample.[21] The c.d.f. of the placebo coefficient estimate is similar to Figure 3 in Ganong and Jäger (2014), although the median value is somewhat lower in our sample. But the placebo estimates find a high share of coefficient estimates that are significantly different from zero. The same pattern appears for the top kink sample, although in this case the sign of the placebo estimates is the opposite of that of the reduced form RKD estimate in Table 2.

It is clear from visual inspection of Figure 6 that the global curvature of the outcome variable is driving the high number of statistically significant placebo estimates. In the right panels of Figure 6, we show that

---

[21]We estimate local linear models with a fixed bandwidth of 2500 for the bottom kink sample and a fixed bandwidth of 4000 for the top kink sample.

much of the curvature can be explained by the changing composition of workers over the distribution of baseline earnings. To do this, we regress log time to the next job on all observable characteristics on a 5% subsample over the full range of baseline earnings.[22] We then plot the residuals from predictions based on this regression for the remaining 95% of observations. The right panels of Figure 6 show that most of the curvature in the global relationships can be explained by changes in observable characteristics. The residual plots are mostly flat along the range of baseline earnings, and the nonlinearity at the kink points becomes even more salient. We then repeat the permutation test on the residuals. The corresponding c.d.f.'s of the placebo coefficient estimate and t-statistic are plotted using the dashed lines in the graphs in Figure 7. The placebo point estimates now have a high density around zero, and the absolute values of the t-statistics are below 2.

Our exercise highlights the importance of the role of curvature heterogeneity in the permutation test of Ganong and Jäger (2014). Specifically, the test may not be informative when the curvature of the conditional expectation function $\mu(v) \equiv E[Y|V = v]$ changes. Suppose $\mu(v)$ is locally piece-wise linear around the threshold $v = 0$ with a non-zero kink but has substantial curvature far away from the threshold, as appears to be true in our data. In this case, the permutation test may fail to detect the kink at the threshold, as the curvature away from the threshold can lead to many significant local linear estimates. Based on a similar argument, the permutation test may also have poor size control when the curvature of $\mu(v)$ changes. Hence, if researchers wish to conduct the permutation test, it will be important to control for confounding nonlinearities by taking the distribution of observables into account.[23] We conclude that after controlling for covariates our design passes the Ganong and Jäger (2014) permutation test.

4.5   Comparison with the Missouri RKD Application in Card et al. (2015a)

In this subsection, we compare our Austrian analysis with a related study by Card et al. (2015a) – hereafter CJLMP. CJLPM are also interested in the effects of the UI benefit level on unemployment duration, and implement an RKD using variation from a kinked benefit schedule in the state of Missouri. As in Austria, the Missouri UI benefit schedule is a piecewise linear function of past earnings, except past earnings in this case are defined as the highest quarterly earnings in the first four of the preceding five calendar quarters. It

---

[22]The list of covariates includes: indicator variables for gender, marital status, Austrian citizenship, education (6 categories), deciles for age, tenure, size, and recall rate at the firm of the last job, monthly wage last job (quintiles), days worked over the last 5 years (quartiles), 8 industry groups interacted with blue collar status, region, year, and month dummies.

[23]Ando (Forthcoming) makes a similar point regarding the role of covariates in RK estimation.

is simpler than Austria's, however, as it exhibits only one kink and does not change with dependents.

CJLMP examine how responsiveness to unemployment insurance benefit generosity varies before and after the Great Recession and conduct their RKD analysis separately using data from 2003 to 2007 and from 2008 to 2013. The data come from the Missouri Department of Labor and Industrial Relations, and the two analysis samples – "pre-" and "post-recession" contain 295,639 and 409,753 observations respectively. Figures 8a and 8b reproduce the two first stage plots from CJLPM over the two time periods, which show the average weekly UI benefit in each small bin of the normalized base period earnings. The sharpness of these figures, as compared to their Austrian counterparts in Figures 2a and 2b, are due to a much lower proportion of off-schedule observations in Missouri. In fact, Figure 8b appears deterministic to the naked eye even though a small fraction of observations still deviate from the prescribed UI rule. Figure 8a looks noisier to the right of the kink point, but this is due to the changing maximum UI benefit over the pre-recession period. As we will see below, the stronger first stage in Missouri yields both improved precision and stability in the empirical estimates.

Reproducing Figure 2 of CJLMP, Figures 9a and 9b depict the relationship between the outcome, log UI claim duration, and base period earnings in Missouri. Both plots show pronounced slope changes at the threshold where the benefit cap starts to bind, and the kink is especially salient in the 2008-2013 sample. Even more compelling, Figure A.1 in the Supplemental Material (reproducing Figure 5 of CJLMP) shows that when the benefit kink threshold increased over time, the UI duration kink had moved along with it.

Estimated benefit elasticities are shown in Table A.2 in the Supplemental Material (based on Appendix Table 1 of CJLMP) using a variety of estimators.[24] Overall, the results point to a larger benefit elasticity in the post-recession period relative to the pre- period. Looking at the local linear estimates, which yielded the smallest MSE's in the simulations, the estimated elasticities for the pre-period center around 0.35, whereas the post-period estimates indicate that the elasticity is at least 0.68. Compared to Austria, both conventional and robust confidence intervals from Missouri are much narrower, thanks in part to the sharp first stage. Here again, however, we see that the local quadratic estimates tend to be less precise than local linear, yielding robust confidence intervals that cover or nearly cover zero for all bandwidth choices.

---

[24]The Fuzzy IK bandwidth in Table A.2 refers to the fuzzy extension of the Imbens and Kalyanaraman (2012) bandwidth selector developed by Card et al. (2015b).

5   Conclusion

In this paper, we discuss and highlight potential issues in the implementation of a fuzzy RKD approach by studying the effect of unemployment benefits on the duration of joblessness in Austria, where the benefit schedule has kinks at the minimum and maximum benefit levels. We present graphical evidence that these benefit kinks are associated with kinks in the mean duration of total joblessness between the end of the previous job and the start of the next job, as well as a number of tests of the RKD identifying assumptions. We show that in our particular application, estimates are sensitive to alternative estimators, and discuss ways to evaluate the empirical performance of these estimators using Monte Carlo simulations. We also discuss another application in Card et al. (2015a) using data from Missouri, where the RKD produces more robust results across alternative specifications.

Through the exercises in this paper, we emphasize the following points. Although the burgeoning literature on nonparametric estimation in regression discontinuity designs (e.g. IK, CCT, and CCFT) may be interpreted by practitioners as pointing to a "default" estimation procedure, there is no single "optimal" approach. In practice, different estimators may have very different point estimates and sampling errors, requiring the researcher to try to "sift" the evidence from different estimators. We argue that some useful information is provided by the magnitude and sampling error of the asymptotic bias correction terms for different estimators. Calculation of these terms helps to interpret the differences between uncorrected and corrected estimators.

One way of choosing among alternative estimators is to conduct Monte Carlo simulations based on data generating processes that closely resemble the data. In our own simulations, we find that some of the asymptotically dominant procedures may actually perform worse than "suboptimal" alternatives in a given empirical application, even though the sample size is very large by conventional standard for parametric estimators. However, we also caution against generalizing conclusions based on specific Monte Carlo exercises to overarching guide for practitioners. While the recently proposed data-driven nonparametric procedures may provide a good starting point, we are skeptical of a one-size-fits-all solution and advocate for judiciously combining the art of sensitivity analysis and the science of econometric theory.

References

Altonji, Joseph G. and Rosa L. Matzkin, "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica*, 2005, *73* (4), 1053–1102.

Ando, Michihito, "How Much Should We Trust Regression-Kink-Design Estimates?," *Empirical Economics*, Forthcoming.

Angrist, Joshua D. and Victor Lavy, "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement," *Quarterly Journal of Economics*, May 1999, *114* (2), 533–575.

Calonico, Sebastian, Matias D. Cattaneo, and Max H. Farrell, "On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference," 2015.

_ , _ , **and** _ , "Coverage Error Optimal Confidence Intervals for Regression Discontinuity Designs," 2016.

_ , _ , **and Rocio Titiunik**, "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, November 2014, *82* (6), 2295–2326.

_ , _ , **and** _ , "Optimal Data-driven Regression Discontinuity Plots," *Journal of the American Statisitcal Association*, 2015, *110* (512), 1753–1769.

_ , _ , **Max H. Farrell, and Rocio Titiunik**, "rdrobust: Software for Regression Discontinuity Designs," 2016.

_ , _ , _ , **and** _ , "Regression Discontinuity Designs Using Covariates," 2016.

Card, David, Andrew Johnston, Pauline Leung, Alexandre Mas, and Zhuan Pei, "The Effect of Unemployment Benefits on the Duration of Unemployment Insurance Receipt: New Evidence from a Regression Kink Design in Missouri, 2003-2013," *American Economic Review: Papers and Proceedings*, May 2015, *105* (5), 126–130.

_ , **Carlos Dobkin, and Nicole Maestas**, "Does Medicare Save Lives?," *Quarterly Journal of Economics*, 2009, *124(2)*, 597–636.

_ , **David S. Lee, and Zhuan Pei**, "Quasi-Experimental Identification and Estimation in the Regression Kink Design," Princeton University Industrial Relations Section Working Paper 553 November 2009.

_ , _ , _ , **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.

_ , _ , _ , **and** _ , "Local Polynomial Order in Regression Discontinuity Designs," October 2014.

_ , _ , _ , **and** _ , "Inference on Causal Effects in a Generalized Regression Kink Design," *Econometrica*, November 2015, *83* (6), 2453–2483.

_ , _ , _ , **and** _ , "Inference on Causal Effects in a Generalized Regression Kink Design," Technical Report 15-218, Upjohn Institute working paper 2015.

_ , _ , _ , **and** _ , "Supplemental Material to "Regression Kink Design: Theory and Practice"," 2016. https://sites.google.com/site/peizhuan/files/rkd-aie-supplemental-material.pdf.

_ , **Raj Chetty, and Andrea Weber**, "Cash-on-Hand and Competing Models of Intertemporal Behavior: New Evidence from the Labor Market," *Quarterly Journal of Economics*, 2007, *122* (4), 1511–1560.

Chay, Kenneth Y., Patrick J. McEwan, and Miguel Urquiola, "The Central Role of Noise in Evaluating Interventions that Use Test Scores to Rank Schools," *American Economic Review*, September 2005, *95* (4), 1237–1258.

Cheng, Ming-Yen, Jianqing Fan, and J. S. Marron, "On Automatic Boundary Corrections," *The Annals of Statistics*, August 1997, *25* (4), 1691–1708.

Del Bono, Emilia and Andrea Weber, "Do Wages Compensate for Anticipated Working Time Restrictions? Evidence from Seasonal Employment in Austria," *Journal of Labor Economics*, 2008, *26* (1), 181–221.

DiNardo, John and David S. Lee, "Economic Impacts of New Unionization on Private Sector Employers: 1984-2001," *Quarterly Journal of Economics*, November 2004, *119* (4), 1383–1441.

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

Florens, Jean-Pierre, James J. Heckman, Costas Meghir, and Edward J. Vytlacil, "Identification of Treatment Effects Using Control Functions in Models With Continuous, Endogenous Treatment and Heterogeneous Effects," *Econometrica*, 2008, *76* (5), 1191–1206.

Ganong, Peter and Simon Jäger, "A Permutation Test and Estimation Alternatives for the Regression Kink Design," June 2014.

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

— **and Thomas Lemieux**, "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, February 2008, *142* (2), 615–635.

Kapteyn, Arie and Jelmer Y. Ypma, "Measurement Error and Misclassification: A Comparison of Survey and Administrative Data," *Journal of Labor Economics*, 2007, *25* (3), 513–551.

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

— **and Alexandre Mas**, "Long-Run Impacts of Unions on Firms: New Evidence from Financial Markets, 1961-1999," *Quaterly Journal of Economics*, 2012, *127* (1), 333–378.

— **and Thomas Lemieux**, "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 2010, *48* (2), 281–355.

— **, Enrico Moretti, and Matthew Butler**, "Do Voters Affect or Elect Policies? Evidence from the U.S. House," *Quarterly Journal of Economics*, August 2004, *119* (3), 807–859.

McCrary, Justin and Heather Royer, "The Effect of Female Education on Fertility and Infant Health: Evidence from School Entry Policies Using Exact Date of Birth," *American Economic Review*, February 2011, *101* (1), 158–95.

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.

Porter, Jack, "Estimation in the Regression Discontinuity Model," *Department of Economics, University of Wisconsin*, 2003.

Ruppert, David, "Empirical-Bias Bandwidths for Local Polynomial Nonparametric Regression and Density Estimation," *Journal of the American Statistical Association*, 1997, *92* (439), 1049–1062.

— **and Matt P. Wand**, "Multivariate Locally Weighted Least Squares Regression," *The Annals of Statistics*, 1994, *22* (3), 1346–1370.

Schmieder, Johannes F., Till von Wachter, and Stefan Bender, "The Effects of Extended Unemployment Insurance Over the Business Cycle: Evidence from Regression Discontinuity Estimates Over 20 Years," *The Quarterly Journal of Economics*, 2012, *127* (2), 701–752.

Turner, Lesley J., "The Road to Pell is Paved with Good Intentions: The Economic Incidence of Federal Student Grant Aid," Technical Report, University of Maryland 2013.

Figure 1: UI Benefits in 2004 Austria

Figure 2a: Daily UI Benefits
Bottom Kink Sample

Figure 2b: Daily UI Benefits
Top Kink Sample

Figure 3a: Log Time to Next Job
Bottom Kink Sample

Figure 3b: Log Time to Next Job
Top Kink Sample

Figure 4a: Predicted Time to Next Job
Bottom Kink Sample

Figure 4b: Predicted Time to Next Job
Top Kink Sample
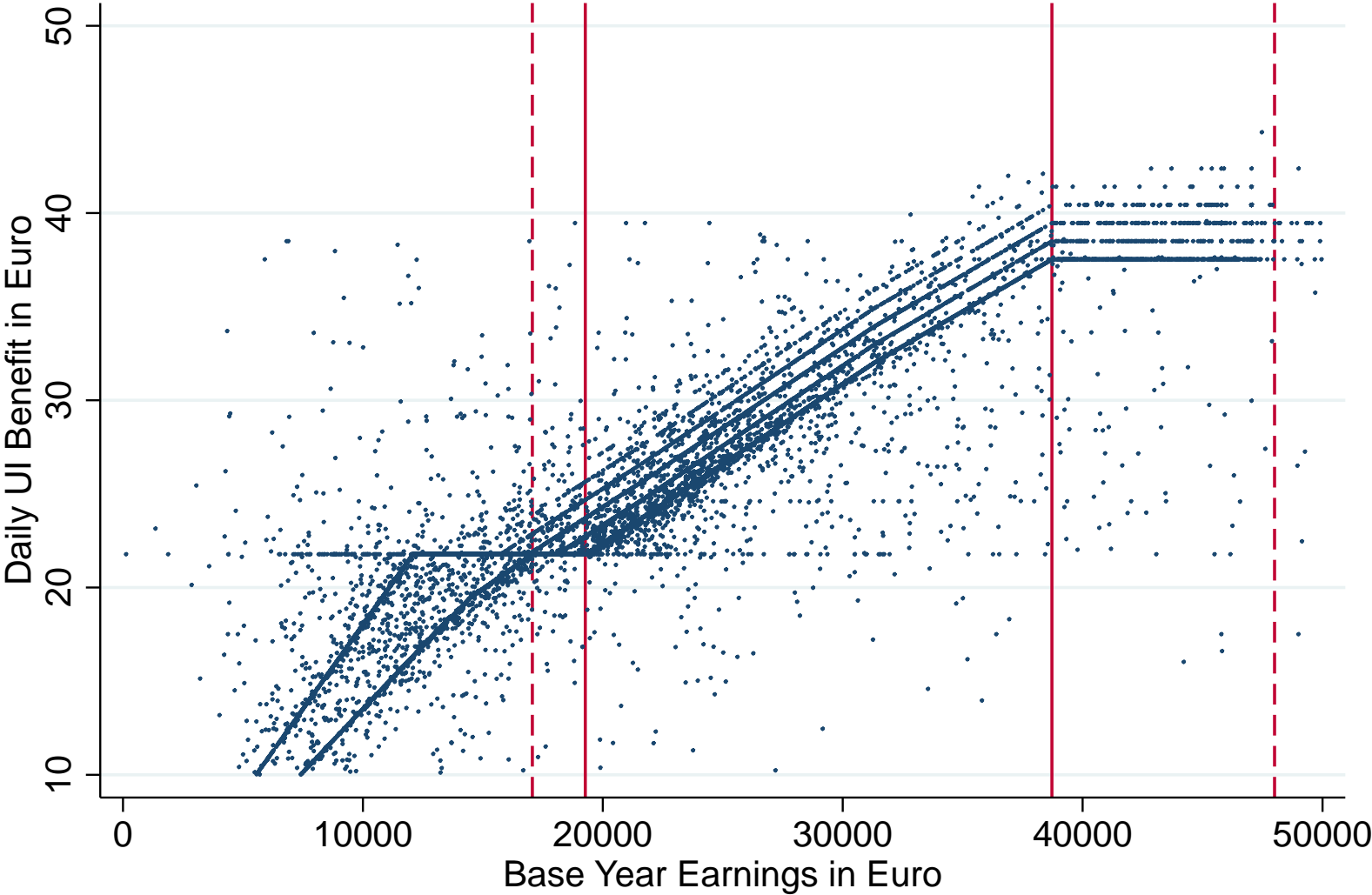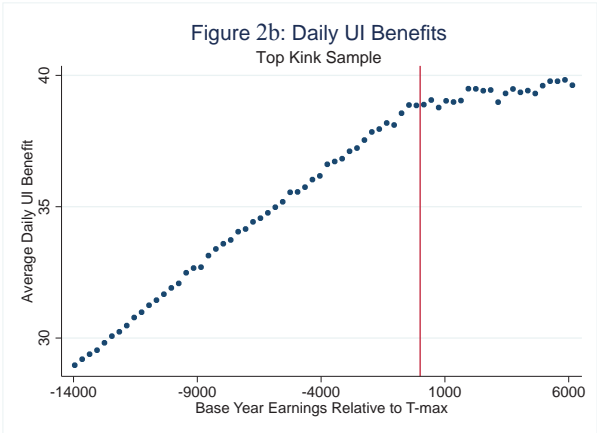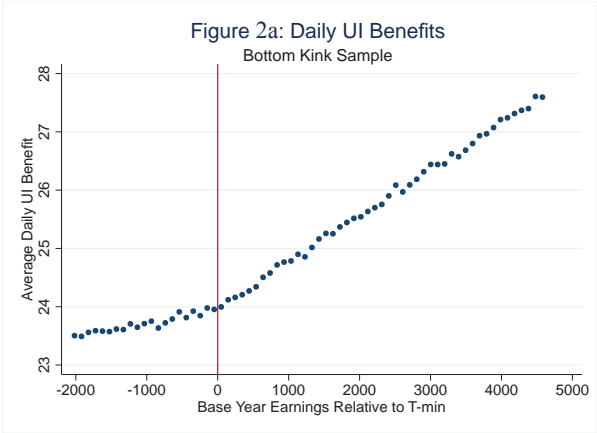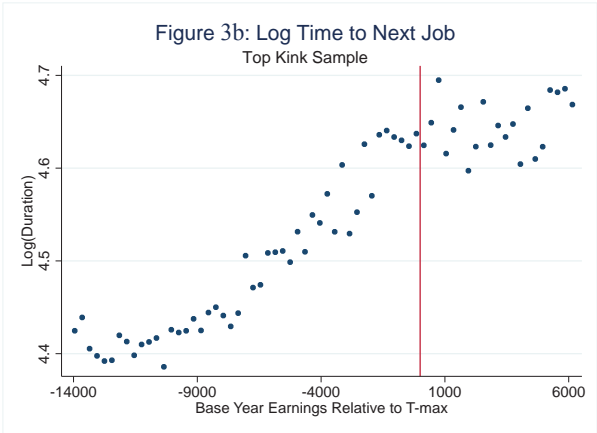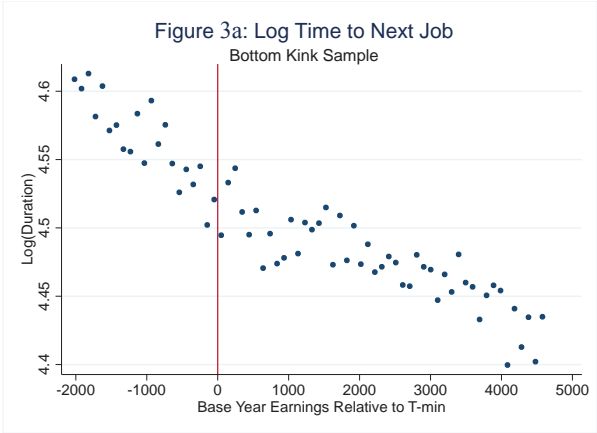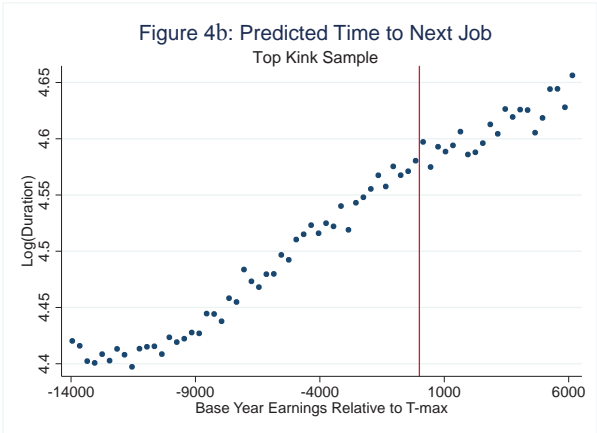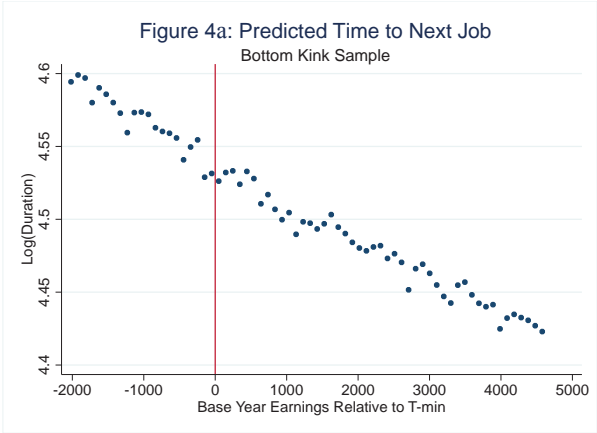
Figure 5a: Fuzzy RKD estimation with varying bandwidth

log time to next job bottom kink sample

Notes: local linear estimation, estimated coefficients (blue) with confidence bounds (dash)
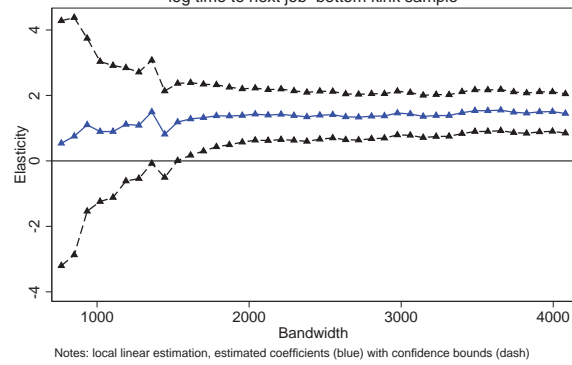
Figure 5b: Fuzzy RKD estimation with varying bandwidth

log time to next job top kink sample

Notes: local linear estimation, estimated coefficients (blue) with confidence bounds (dash)
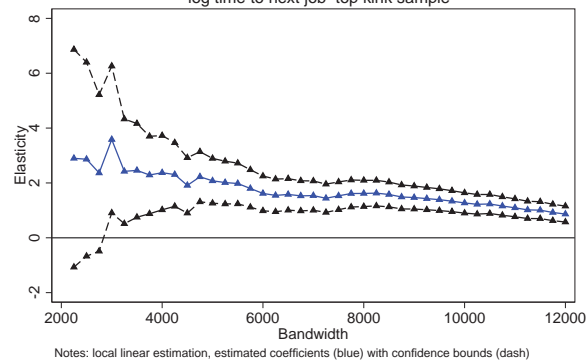
Figure 6: Global Relationship between Log Time to Next Job and Baseline Earnings

Figure 7: Distribution of the Coefficient Estimate and t-Statistic in the Permutation Test
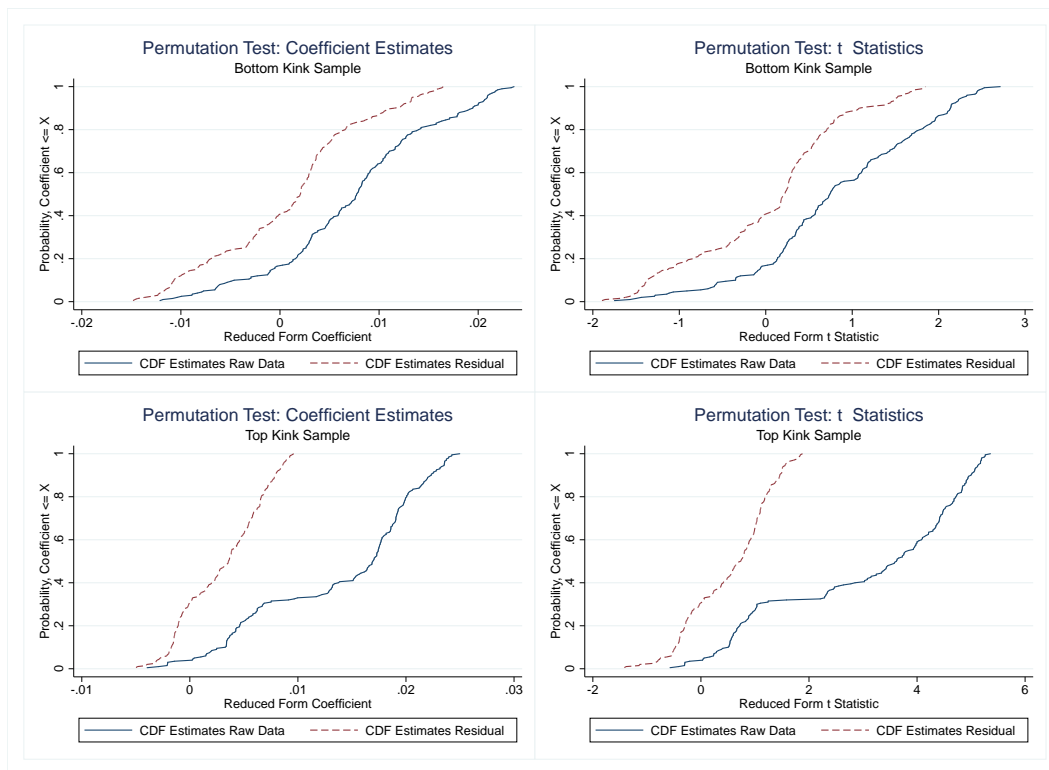
Figure 8a: Missouri First Stage, 2003-2007
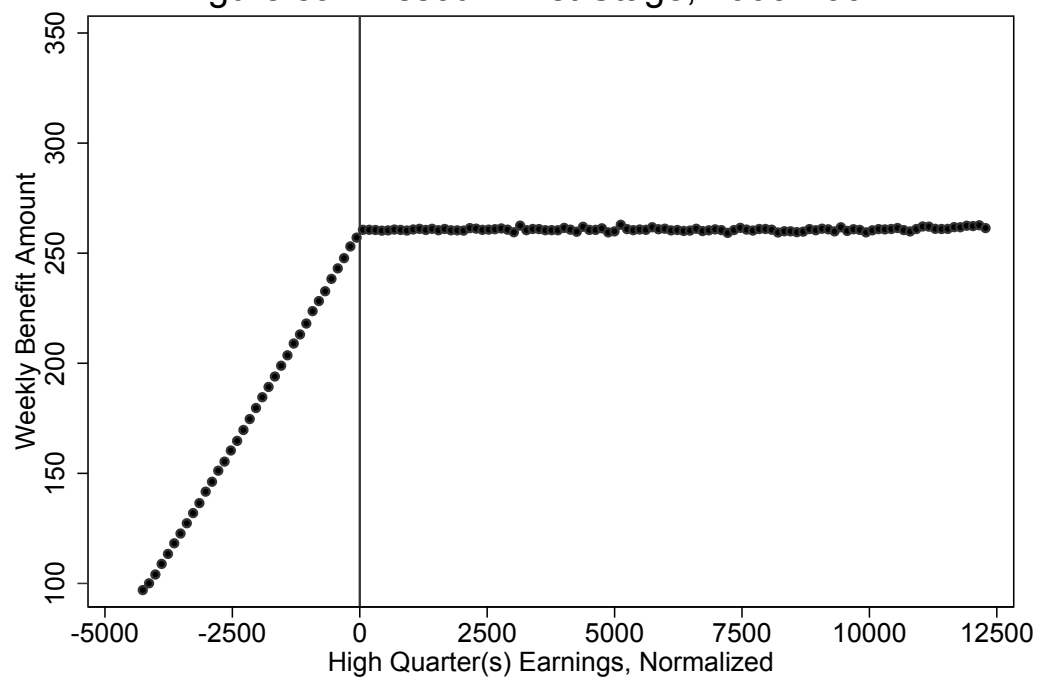
Figure 8b: Missouri First Stage, 2008-2013
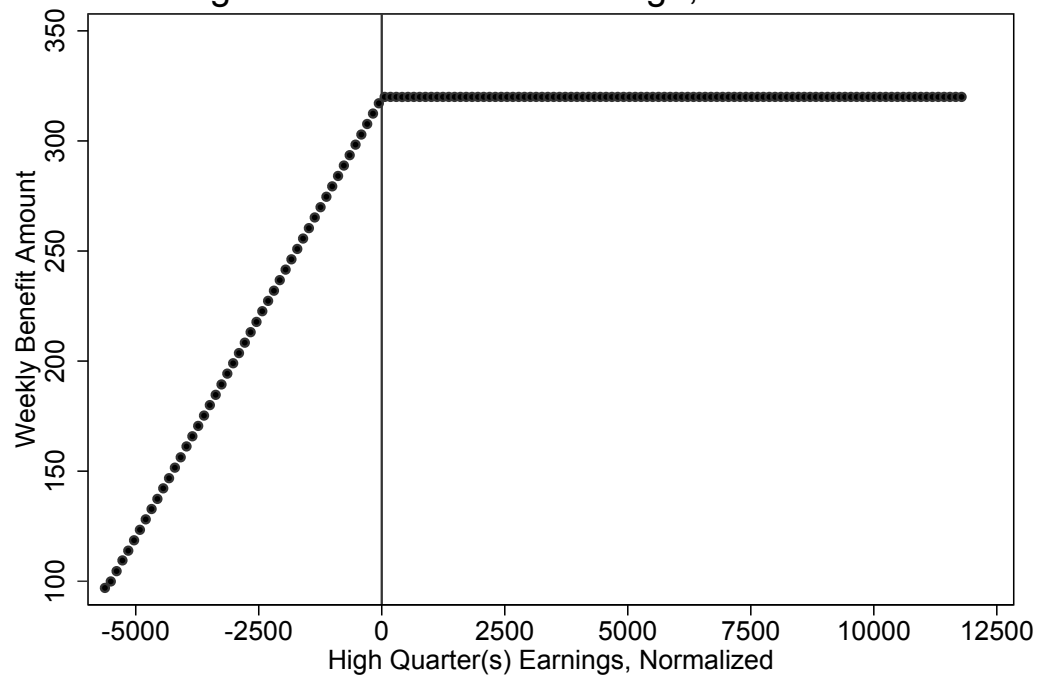
Notes: Reproduced from Card et al (2015a).

Figure 9a: Log Initial UI Spell Duration, Missouri, 2003-2007
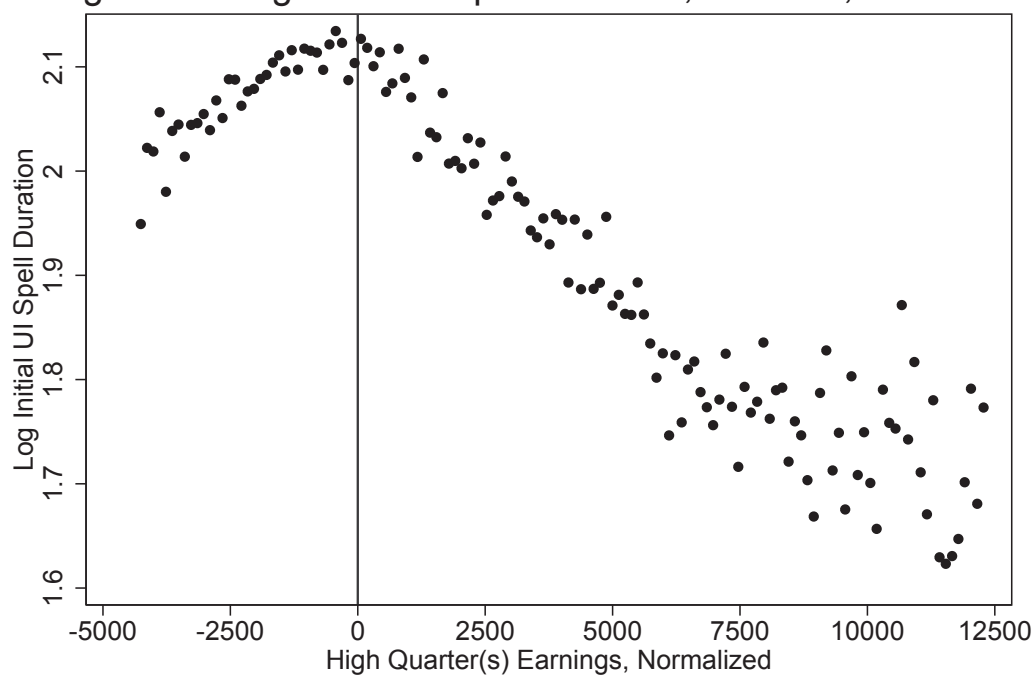
Figure 9b: Log Initial UI Spell Duration, Missouri, 2008-2013
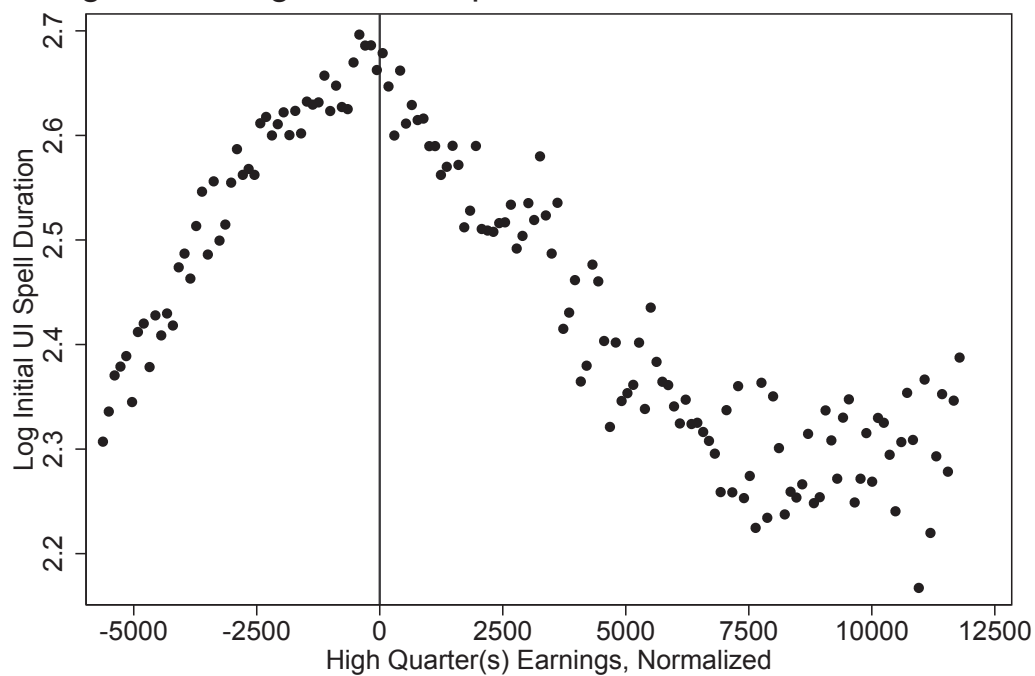
Notes: Reproduced from Card et al (2015a).

Table 1: Summary Statistics for Bottom and Top Kink Samples of UI Claimants

	Bottom Kink Sample		Top Kink Sample	
	Mean	Std. Dev.	Mean	Std. Dev.
	(1)	(2)	(3)	(4)
Baseline earnings (euros)	22,142	2,805	33,847	6,730
Daily UI benefit (euros)	25.2	3.0	33.5	5.7
Time to next job (days)*	148.2	129.1	147.4	131.7
Duration of initial UI spell (days)**	77.6	67.0	79.8	70.3
Total days of UI received	1944.4	1707.0	2679.7	2480.1
Fraction exhausted benefits***	0.09	0.29	0.09	0.29
Fraction with time to next job censored	0.19	0.39	0.20	0.40
Fraction eligible for extended benefits	0.78	0.41	0.90	0.29
Fraction female	0.38	0.48	0.21	0.41
Mean Age	33.06	8.48	36.04	7.45
Fraction Austrian nationals	0.80	0.40	0.86	0.35
Fraction married	0.34	0.47	0.40	0.49
Fraction bluecollar occupation	0.65	0.48	0.56	0.50
Fraction with higher education	0.11	0.31	0.18	0.38
Fraction in Vienna	0.19	0.39	0.21	0.41
Tenure in most recent job (Years)	3.39	3.27	4.23	4.20
Recalled to last job	0.20	0.40	0.25	0.44
Industry:				
Construction	0.13	0.34	0.23	0.42
Manufacturing	0.20	0.40	0.23	0.42
Trade	0.20	0.40	0.15	0.35
Services	0.26	0.44	0.23	0.42
Number observations	275,293		275,665	

Notes: Sample contains UI claimants under the age of 50 with claims in 2001-2012, who had at least 1 year of tenure on their previous job, began their claim within 4 weeks of losing their past job, and had a valid UI claim record and non-missing earnings in the base period prior to the claim. Observations in the bottom kink sample have base period earnings in a range around the bottom kink in the UI benefit schedule; observations in the top kink sample have base period earnings in a range around top kink. See text.
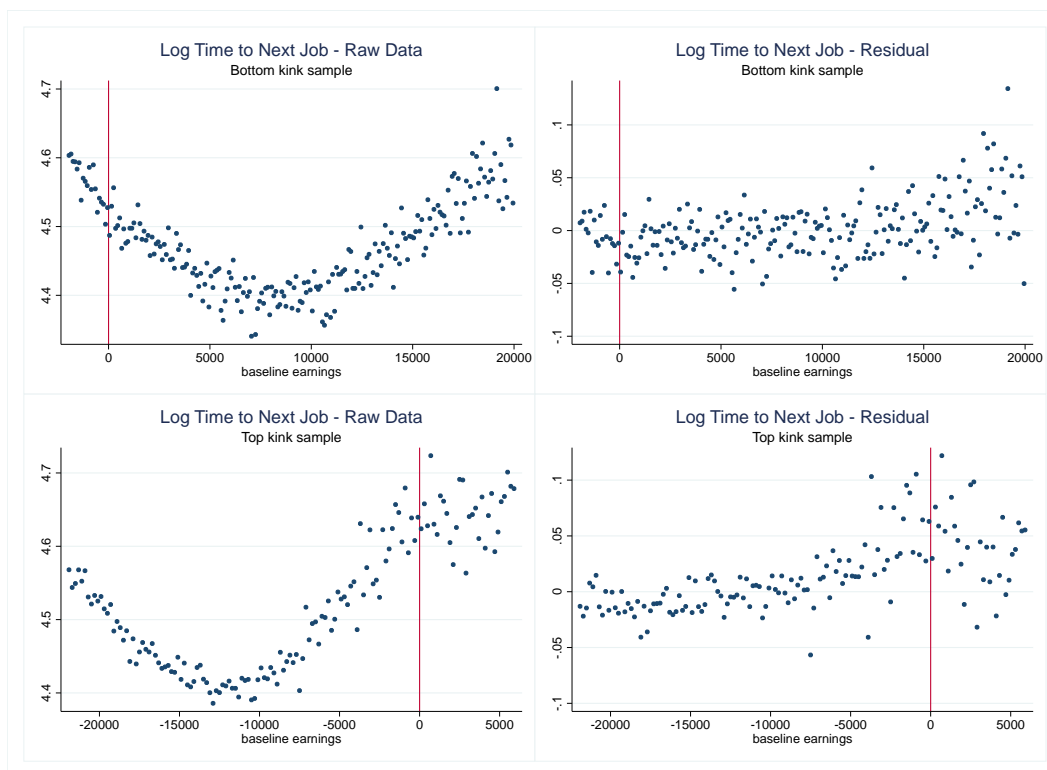
* Time to next job is censored at 365 days.
** Claim duration is censored at 39 weeks (maximum entitlement).
*** Indicator equals 1 if claim duration = maximum entitlement.

Table 2a: First Stage and Reduced Form Estimated Kinks

	Bottom Kink		Top Kink	
	Local Linear	Local Quad.	Local Linear	Local Quad.
	(1)	(2)	(3)	(4)
A. First Stage Model (Dependent variable = log daily benefit)				
MSE optimal (with regularization)				
Main (Pilot) Bandwidth	461 (940)	818 (1,346)	1,400 (2,808)	1,571 (2,619)
Estimated Kink	1.5	1.9	-1.6	-1.1
(conventional std error)	(0.9)	(1.5)	(0.7)	(2.2)
Estimated Kink - Bias Corrected	1.1	1.1	-1.8	-2.2
[robust confidence interval]	[-1.6,3.9]	[-2.8,5.0]	[-3.8,0.2]	[-7.6,3.1]
MSE optimal with no regularization				
Main (Pilot) Bandwidth	2,388 (2,022)	2,243(1,770)	5,189 (6,746)	2,034 (3,037)
Estimated Kink	2.2	2.4	-1.5	-3.1
(conventional std error)	(0.1)	(0.4)	(0.1)	(1.5)
Estimated Kink - Bias Corrected	2.2	2.0	-1.4	-4.4
[robust confidence interval]	[1.4,3.0]	[-1.1,5.2]	[-1.9,-0.9]	[-8.5,-0.3]
FG				
Main (Pilot) Bandwidth	2,466 (2,380)	2,628 (4,146)	11,603 (6,343)	7,003 (13,081)
Estimated Kink	2.2	2.2	-1.9	-1.4
(conventional std error)	(0.1)	(0.3)	(0.4)	(0.2)
Estimated Kink - Bias Corrected	2.4	1.9	-1.4	-0.8
[robust confidence interval]	[1.7,3.1]	[0.4,3.4]	[-2.1,-0.7]	[-1.9,0.3]
B. Reduced Form Model (Dependent variable = log time to next jobt)				
MSE optimal (with regularization)				
Main (Pilot) Bandwidth	497 (983)	599 (987)	1,725 (3,574)	2,571 (4,041)
Estimated Kink	6.4	30.7	1.4	6.9
(conventional std error)	(7.5)	(22.8)	(3.0)	(6.6)
Estimated Kink - Bias Corrected	6.1	42.5	3.2	8.7
[robust confidence interval]	[-17.1,29.3]	[-15.9,101]	[-5.6,12.0]	[-9.0,26.5]
MSE optimal with no regularization				
Main (Pilot) Bandwidth	1,866 (1,097)	1,418 (1,647)	3,208 (9,972)	3,717 (5,191)
Estimated Kink	3.3	6.1	-3.0	-1.8
(conventional std error)	(1.0)	(6.3)	(1.2)	(3.8)
Estimated Kink - Bias Corrected	4.5	11.6	-4.5	1.6
[robust confidence interval]	[-25.2,34.2]	[-13.3,36.5]	[-7.9,-1.1]	[-10.1,13.2]
FG				
Main (Pilot) Bandwidth	2,501 (3,858)	4,259 (4,918)	4,465 (7,255)	8,001 (12,383)
Estimated Kink	3.0	4.8	-3.0	-4.1
(conventional std error)	(0.8)	(2.4)	(0.7)	(1.5)
Estimated Kink - Bias Corrected	4.2	9.9	-5.2	0.0
[robust confidence interval]	[-1.4,9.8]	[-3.6,23.4]	[-8.4,-1.9]	[-7.0,6.9]

Notes: Estimated kinks, conventional standard errors and robust confidence intervals are all multiplied by 10^5. Point estimates and standard errors are obtained from regressions described in Card et al (2012). The MSE optimal bandwidths and the Robust CI's are obtained by the Stata package "rdrobust" (August, 2016) by Calonico, Cattaneo, Farrell and Titiunik.

Table 2b: Bias and Variance Tradeoff -- Conventional versus Bias-corrected Estimators

	Bottom Kink		Top Kink	
	Local Linear	Local Quad.	Local Linear	Local Quad.
	(1)	(2)	(3)	(4)
A. First Stage Model (Dependent variable = log daily benefit)				
Default CCT (with regularization)				
Estimated Bias	0.4	0.8	0.2	1.1
Estimated Bias Squared	0.2	0.7	0.0	1.3
Change in Estimated Variance	1.1	1.6	0.6	2.8
CCT with no regularization				
Estimated Bias	0.0	0.4	-0.1	1.2
Estimated Bias Squared	0.0	0.1	0.0	1.5
Change in Estimated Variance	0.2	2.5	0.1	2.2
FG				
Estimated Bias	-0.1	0.3	-0.2	-0.5
Estimated Bias Squared	0.0	0.1	0.0	0.3
Change in Estimated Variance	0.1	0.5	0.1	0.2
B. Reduced Form Model (Dependent variable = log time to next jobt)				
Default CCT (with regularization)				
Estimated Bias	0.2	-11.8	-1.8	-1.8
Estimated Bias Squared	0.0	139.5	3.2	3.4
Change in Estimated Variance	83.5	367.5	11.1	38.0
CCT with no regularization				
Estimated Bias	-1.2	-5.4	1.5	-3.4
Estimated Bias Squared	1.4	29.7	2.1	11.6
Change in Estimated Variance	228.4	122.1	1.5	21.0
FG				
Estimated Bias	-1.3	-5.5	2.2	-4.4
Estimated Bias Squared	1.7	30.3	4.6	18.9
Change in Estimated Variance	7.5	39.6	2.3	10.0

Notes: Estimated bias is the difference between the value of the conventional linear estimator and the its bias-corrected counterpart, which is the center of the robust confidence interval. The change in estimated variance is the difference between the variance of the bias corrected estimator and its conventional counterpart. Estimated bias is multiplied by 10^5, and estimated bias squared and change in estimated variance are both multiplied by 10^{10}.

Table 3: Estimates of Benefit Elasticity (Fuzzy RK), Alternative Estimators and Bandwidths

	Bottom Kink				Top Kink			
	Local Linear		Local Quadratic		Local Linear		Local Quadratic	
	First Stage (Coeff×10^5)	Struct. Model	First Stage (Coeff×10^5)	Struct. Model	First Stage (Coeff×10^5)	Struct. Model	First Stage (Coeff×10^5)	Struct. Model
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
MSE optimal (with regularization)								
Main Bandwidth (Pilot)	548 (1094)		619 (992)		1678 (992)		2591 (4060)	
Estimated Kink	1.7	3.5	0.9	31.0	-1.9	-0.7	-1.6	-3.1
(conventional std error)	(0.7)	(4.1)	(2.3)	(81.7)	(0.5)	(1.7)	(1.0)	(4.6)
Bias-corrected Estimate	1.5	3.9	1.7	18.6	-1.5	-1.8	-1.9	-3.9
[robust conf. interval]	[-0.6,3.6]	[-8.7,16.43]	[-4.4,7.8]	[-196.3,233.4]	[-2.9,-0.1]	[-6.6,3.0]	[-4.6,0.8]	[-16.2,8.4]
MSE optimal with no regularization								
Main Bandwidth (Pilot)	1414 (1614)		1294 (1062)		4893 (6921)		7174 (6233)	
Estimated Kink	2.2	0.7	2.3	-0.1	-1.5	2.1	-1.4	4.0
(conventional std error)	(0.2)	(0.7)	(0.8)	(3.2)	(0.1)	(0.4)	(0.2)	(1.4)
Bias-corrected Estimate	1.9	0.0	0.7	17.4	-1.5	3.6	-0.8	-0.6
[robust conf. interval]	[0.8,3.0]	[-4.6,4.6]	[-6.4,7.8]	[-11.6,46.4]	[-2.0,-1.0]	[1.2,5.9]	[-2.3,0.6]	[-9.1,7.9]
FG								
Main Bandwidth (Pilot)	2,501 (3,858)		4,259 (4,918)		4,465 (7,255)		8,011 (12,383)	
Estimated Kink	2.2	1.4	2.0	2.0	-1.4	2.4	-1.4	2.7
(conventional std error)	(0.1)	(0.4)	(0.3)	(0.5)	(0.1)	(1.2)	(0.2)	(1.1)
Bias-corrected Estimate	2.1	2.0	1.9	5.0	-1.0	3.7	-1.1	1.2
[robust conf. interval]	[1.5, 2.7]	[-0.6,4.5]	[0.4, 3.4]	[-1.9,11.9]	[-1.9, -0.1]	[1.2,6.1]	[-2.0, 0.2]	[-4.5,6.9]

Notes: Point estimates and standard errors are obtained from 2SLS regressions described in Card et al (2012). The MSE optimal bandwidths and the Robust CI's are obtained by the Stata package "rdrobust" (August, 2016) by Calonico, Cattaneo, Farrell and Titiunik. The fuzzy FG bandwidths are authors' calculations.

Table 4: Summary of Monte Carlo Studies: DGP with Kink

	Median Main b.w. (1)	Median Pilot b.w. (2)	First Stage Model Estimation Summary			Elasticity Estimation Summary					
			Fraction of Replications: C.I. includes 0 (3)	RMSE/true value (trimmed) (4)	C.I. Coverage Rate (5)	RMSE/true value (6)	RMSE/true value (trimmed) (7)	C.I. Coverage Rate (8)	Bias (trimmed) (9)	Bias^2 (trimmed) (10)	Variance (trimmed) (11)
Panel A. DGP Design Based on Bottom Kink Sample											
A.1. Local Linear, No Bias Correction											
MSE-optimal	456	-	0.34	0.42	0.94	35.9	4.13	1.00	1.90	3.61	25.3
MSE-opt, no reg	878	-	0.05	0.20	0.90	5.8	1.82	0.95	1.01	1.03	4.55
Fuzzy MSE-opt, no reg	943	-	0.04	0.21	0.90	7.45	1.47	0.95	0.74	0.55	3.10
FG	2,437	-	0.00	0.07	0.66	0.28	0.24	0.96	0.11	0.01	0.08
A.2. Local Linear, Bias-Corrected											
MSE-optimal	456	947	0.63	0.63	0.94	3778	7.28	1.00	3.31	11.0	78.6
MSE-opt, no reg	878	1,223	0.38	0.47	0.94	56.7	3.71	0.96	2.01	4.04	19.3
Fuzzy MSE-opt, no reg	943	1,317	0.34	0.48	0.92	496.0	3.53	0.95	2.03	4.11	17.0
FG	2,437	3,795	0.00	0.16	0.89	1.22	1.07	0.90	0.39	0.16	1.77
A.3. Local Quadratic, No Bias Correction											
MSE-optimal	676	-	0.79	0.94	0.95	99	10.4	1.00	1.24	1.54	181
MSE-opt, no reg	1,186	-	0.38	0.53	0.93	183.9	6.15	0.99	2.53	6.40	57.5
Fuzzy MSE-opt, no reg	1,261	-	0.32	0.50	0.91	23.7	4.58	0.98	2.46	6.04	29.5
FG	4,190	-	0.00	0.15	0.84	1.48	1.32	0.79	0.99	0.98	1.98
A.4. Local Quadratic, Bias-Corrected											
MSE-optimal	676	1,076	0.85	1.26	0.94	2.05E+08	17.7	1.00	2.26	5.11	527
MSE-opt, no reg	1,186	1,373	0.77	1.14	0.93	9188	11.7	0.99	3.21	10.28	222
Fuzzy MSE-opt, no reg	1,261	1,450	0.75	1.05	0.93	2.07E+06	10.9	0.97	3.65	13.3	188
FG	4,190	4,914	0.25	0.36	0.91	4.46	3.93	0.77	3.88	15.0	11.1
Panel B. DGP Design Based on Top Kink Sample											
B.1. Local Linear, No Bias Correction											
MSE-optimal	1,402	-	0.60	0.59	0.79	57.6	3.88	0.99	0.36	0.13	60.1
MSE-opt, no reg	2,757	-	0.11	0.32	0.72	22.84	1.08	0.92	0.23	0.05	4.66
Fuzzy MSE-opt, no reg	2,708	-	0.15	0.37	0.72	7.9	1.11	0.93	0.31	0.10	4.9
FG	4,464	-	0.00	0.15	0.79	0.49	0.36	0.84	0.11	0.01	0.51
B.2. Local Linear, Bias-Corrected											
MSE-optimal	1,402	2,871	0.87	0.90	0.77	2777	10.6	0.99	2.16	4.68	446
MSE-opt, no reg	2,757	3,779	0.65	0.77	0.65	44280	2.73	0.93	0.62	0.39	29.3
Fuzzy MSE-opt, no reg	2,708	3,856	0.62	0.76	0.69	1092	3.01	0.92	0.91	0.82	36
FG	4,464	7,256	0.01	0.19	0.89	1.21	0.96	0.81	1.52	2.32	1.40
B.3. Local Quadratic, No Bias Correction											
MSE-optimal	2,024	-	0.89	1.14	0.87	96.3	8.05	0.99	-1.37	1.89	258
MSE-opt, no reg	3,481	-	0.65	0.76	0.74	209.9	7.50	0.97	-0.09	0.01	225
Fuzzy MSE-opt, no reg	3,551	-	0.63	0.74	0.71	66.3	5.28	0.96	0.09	0.01	112
FG	8,012	-	0.00	0.15	0.92	1.09	0.94	0.82	1.61	2.60	0.93
B.4. Local Quadratic, Bias-Corrected											
MSE-optimal	2,024	3,188	0.91	1.43	0.90	3776	13.6	1.00	-0.27	0.07	743
MSE-opt, no reg	3,481	4,064	0.93	1.41	0.79	1.12E+05	30.4	0.98	4.74	22.44	3686
Fuzzy MSE-opt, no reg	3,551	4,302	0.92	1.99	0.76	1.40E+05	23.6	0.97	0.26	0.1	2228
FG	8,012	12,409	0.70	0.64	0.61	2.05	1.55	0.93	0.54	0.30	9.3

Notes: based on 1,000 simulations. In Panel A, DGP is based on 5th order polynomial approximation of bottom kink sample. True kink in first stage is: 2.3 ×10^{-5}; true elasticity is: 1.3. In Panel B, DGP is based on 5th order polynomial approximation of top kink sample. True kink in first stage is: -1.4 ×10^{-5}; true elasticity is: 2.0. The trimmed statistic are obtained by first trimming the 5% sample in which the estimates deviate the most from the true parameter value.

Table 5: Summary of Monte Carlo Studies: DGP without Kink

	Median Main b.w. (1)	Median Pilot b.w. (2)	First Stage Model Estimation Summary			Elasticity Estimation Summary					
			Fraction of Replications: C.I. includes 0 (3)	RMSE/true value (trimmed) (4)	C.I. Coverage Rate (5)	RMSE (6)	RMSE (trimmed) (7)	C.I. Coverage Rate (8)	Bias (trimmed) (9)	Bias^2 (trimmed) (10)	Variance (trimmed) (11)
Panel A. DGP Design Based on Bottom Kink Sample											
A.1. Local Linear, No Bias Correction											
MSE-optimal	464	-	0.33	0.41	0.94	24.3	4.46	1.00	0.35	0.12	19.8
MSE-opt, no reg	1,015	-	0.03	0.18	0.89	5.3	1.83	0.75	0.67	0.45	2.88
Fuzzy MSE-opt, no reg	976	-	0.03	0.20	0.88	135.9	1.71	0.78	0.68	0.46	2.48
FG	2,602	-	0.00	0.08	0.59	1.40	1.36	0.02	1.33	1.77	0.09
A.2. Local Linear, Bias-Corrected											
MSE-optimal	464	959	0.63	0.62	0.94	1050.03	7.5	1.00	0.35	0.12	56.6
MSE-opt, no reg	1,015	1,323	0.34	0.46	0.93	15.11	3.60	0.97	0.29	0.09	12.9
Fuzzy MSE-opt, no reg	976	1,371	0.32	0.47	0.93	16242.67	3.72	0.94	0.25	0.06	13.8
FG	2,602	3,364	0.00	0.16	0.89	1.91	1.70	0.83	0.80	0.64	2.24
A.3. Local Quadratic, No Bias Correction											
MSE-optimal	683	-	0.79	0.94	0.94	168	14.6	1.00	-0.02	0.00	213
MSE-opt, no reg	1,223	-	0.36	0.49	0.92	305	6.4	0.98	0.27	0.07	41.4
Fuzzy MSE-opt, no reg	1,261	-	0.33	0.50	0.92	71	5.0	0.98	0.26	0.07	25.2
FG	3,714	-	0.00	0.14	0.88	2.13	1.91	0.71	1.14	1.29	2.34
A.4. Local Quadratic, Bias-Corrected											
MSE-optimal	683	1,087	0.85	1.26	0.94	51580.95	25.0	1.00	0.61	0.37	625
MSE-opt, no reg	1,223	1,401	0.76	1.17	0.94	4.33E+07	13.0	1.00	0.22	0.05	168
Fuzzy MSE-opt, no reg	1,261	1,444	0.76	1.05	0.94	26966.60	12.0	0.99	-0.16	0.03	145
FG	3,714	4,947	0.24	0.36	0.90	3.60	3.07	0.95	0.06	0.00	9.45
Panel B. DGP Design Based on Top Kink Sample											
B.1. Local Linear, No Bias Correction											
MSE-optimal	1,405	-	0.60	0.60	0.77	38.0	7.34	1.00	0.79	0.62	53.3
MSE-opt, no reg	2,998	-	0.10	0.31	0.71	25.3	2.18	0.66	0.98	0.97	3.79
Fuzzy MSE-opt, no reg	2,798	-	0.13	0.35	0.74	14.8	2.21	0.71	1.03	1.07	3.8
FG	5,621	-	0.00	0.18	0.66	1.31	1.23	0.19	1.12	1.25	0.25
B.2. Local Linear, Bias-Corrected											
MSE-optimal	1,405	2,877	0.86	0.91	0.76	5.17E+06	20.9	0.99	1.23	1.52	435
MSE-opt, no reg	2,998	3,951	0.61	0.73	0.70	2883.03	4.87	0.93	0.86	0.73	23.0
Fuzzy MSE-opt, no reg	2,798	3,931	0.62	0.75	0.68	243.42	5.34	0.92	0.98	0.96	27.6
FG	5,621	9,306	0.00	0.17	0.90	2.02	1.84	0.65	1.54	2.37	1.02
B.3. Local Quadratic, No Bias Correction											
MSE-optimal	2,037	-	0.88	1.10	0.87	102	17.6	1.00	0.06	0.00	309
MSE-opt, no reg	3,605	-	0.60	0.74	0.70	342	13.9	0.95	0.90	0.80	192
Fuzzy MSE-opt, no reg	3,634	-	0.61	0.74	0.70	257	9.1	0.94	0.67	0.45	82
FG	10,274	-	0.00	0.16	0.81	3.18	3.00	0.26	2.77	7.69	1.32
B.4. Local Quadratic, Bias-Corrected											
MSE-optimal	2,037	3,219	0.91	1.40	0.89	2.8E+07	32.9	1.00	0.02	0.00	1085
MSE-opt, no reg	3,605	4,134	0.93	1.53	0.77	3.68E+05	44.2	0.99	1.15	1.33	1951
Fuzzy MSE-opt, no reg	3,634	4,335	0.92	1.66	0.76	20482.07	32.8	0.98	-0.07	0.01	1078
FG	10,274	13,142	0.70	0.66	0.57	3.65	3.02	0.92	0.82	0.67	8.44

Notes: based on 1,000 simulations. In Panel A, DGP is based on 5th order polynomial approximation of bottom kink sample. True kink in first stage is: 2.3 ×10^{-5}; true elasticity is: 0. In Panel B, DGP is based on 5th order polynomial approximation of top kink sample. True kink in first stage is: -1.4 × 10^{-5}; true elasticity is: 0. The trimmed statistic are obtained by first trimming the 5% sample in which the estimates deviate the most from the true parameter value.