NOTES AND COMMENTS

INFERENCE ON CAUSAL EFFECTS IN A GENERALIZED REGRESSION KINK DESIGN

By David Card, David S. Lee, Zhuan Pei, and Andrea Weber[1]

We consider nonparametric identification and estimation in a nonseparable model where a continuous regressor of interest is a known, deterministic, but kinked function of an observed assignment variable. We characterize a broad class of models in which a sharp "Regression Kink Design" (RKD or RK Design) identifies a readily interpretable treatment-on-the-treated parameter (Florens, Heckman, Meghir, and Vytlacil (2008)). We also introduce a "fuzzy regression kink design" generalization that allows for omitted variables in the assignment rule, noncompliance, and certain types of measurement errors in the observed values of the assignment variable and the policy variable. Our identifying assumptions give rise to testable restrictions on the distributions of the assignment variable and predetermined covariates around the kink point, similar to the restrictions delivered by Lee (2008) for the regression discontinuity design. Using a kink in the unemployment benefit formula, we apply a fuzzy RKD to empirically estimate the effect of benefit rates on unemployment durations in Austria.

KEYWORDS: Regression discontinuity design, regression kink design, treatment effects, nonseparable models, nonparametric estimation.

1. INTRODUCTION

A GROWING BODY OF RESEARCH CONSIDERS the identification and estimation of nonseparable models with continuous endogenous regressors in semiparametric (e.g., Lewbel (1998, 2000)) and nonparametric settings (e.g., Blundell and Powell (2003), Chesher (2003), Florens et al. (2008), Imbens and Newey (2009)). The methods proposed in the literature so far rely on instrumental variables that are *independent* of the unobservable terms in the model. Unfortunately, independent instruments are often hard to find, particularly when the regressor of interest is a deterministic function of an endogenous assignment variable. Unemployment benefits, for example, are set as function of previous earnings in most countries. Any variable that is correlated with benefits is likely to be correlated with the unobserved determinants of previous wages and is therefore *unlikely* to satisfy the necessary independence assumptions for a valid instrument.

Nevertheless, many tax and benefit formulas are piecewise linear functions with kinks in the relationship between the assignment variable and the policy variable caused by minimums, maximums, and discrete shifts in the marginal tax or benefit rate. As noted by Classen (1977), Welch (1977), Guryan (2001), Dahlberg, Mork, Rattso, and Agren (2008), Nielsen, Sørensen, and Taber (2010), and Simonsen, Skipper, and Skipper (2015), a kinked assignment rule holds out the possibility for identification of the policy variable's effect even in the absence of traditional instruments. The idea is to look for an induced kink in the mapping between the assignment variable and the outcome variable that coincides with the kink in the policy rule, and compare the relative magnitudes of the two kinks.

This paper establishes conditions under which the behavioral response to a formulaic policy variable like unemployment benefits can be identified within a general class of nonparametric and nonseparable regression models. Specifically, we establish conditions for the regression kink design (RKD) to identify the "local average response" defined by Altonji and Matzkin (2005) or the "treatment-on-the-treated" parameter defined by Florens et al. (2008). The key assumptions are: (1) conditional on the unobservable determinants of the outcome variable, the density of the assignment variable is smooth (i.e., continuously differentiable) at the kink point in the policy rule, and (2) the treatment assignment rule is continuous at the kink point. We show that the smooth density condition rules out deterministic sorting while allowing less extreme forms of endogeneity—including, for example, situations where agents endogenously sort but make small optimization errors (e.g., Chetty (2012)). We also show that the smooth density condition generates testable predictions for the distribution of predetermined covariates among the population of agents located near the kink point. Thus, as in a regression discontinuity (RD) design (Lee and Lemieux (2010), DiNardo and Lee (2011)), the validity of the regression kink design can be evaluated empirically. The second key assumption—continuity of the treatment assignment rule—is important for RK identification in a general model with no restriction on treatment effect heterogeneity, but researchers can apply an RD design in its absence.

In many realistic settings, the policy rule of interest depends on unobserved individual characteristics or is implemented with error. In addition, both the assignment variable and the policy variable may be observed with error. We present a generalization of the RKD—which we call a "fuzzy regression kink design"—that allows for these features. The fuzzy RKD estimand replaces the known change in slope of the assignment rule at the kink with an estimate based on the observed data. Under a series of additional assumptions, including a monotonicity condition analogous to the one introduced by Imbens and Angrist (1994) (and implicit in latent index models (Vytlacil (2002))), we show

that the fuzzy RKD identifies a weighted average of marginal effects, where the weights are proportional to the magnitude of the individual-specific kinks.[2]

We then briefly review existing methods for the nonparametric estimation of RKD using local polynomial estimation, including Fan and Gijbels (1996)—hereafter, FG; Imbens and Kalyanaraman (2012)—hereafter, IK; and Calonico, Cattaneo, and Titiunik (2014)—hereafter, CCT. And finally, we use a fuzzy RKD approach to analyze the effect of unemployment insurance (UI) benefits on the duration of registered unemployment in Austria, focusing on the kink in the UI benefit formula at the maximum benefit level. Simple plots of the data show visual evidence of a kink in the relationship between base period earnings and unemployment durations around the earnings threshold associated with the maximum benefit. We present a range of alternative estimates of the behavioral effect of benefits on registered unemployment durations derived from local linear and local quadratic polynomial models obtained with various bandwidth selection algorithms (including FG, IK, and CCT, and extensions of IK and CCT for the fuzzy RKD case). For each of the alternative choices of polynomial order and bandwidth selector, we show the conventional kink estimates and the corresponding robust bias-corrected confidence intervals per Calonico, Cattaneo, and Titiunik (2014).

2. NONPARAMETRIC REGRESSION AND THE REGRESSION KINK DESIGN

2.1. *Background*

Consider the generalized nonseparable model

$$(2.1) \qquad Y = y(B, V, U),$$

where Y is an outcome, B is a continuous regressor of interest, V is another observed covariate, and U is a potentially multidimensional error term that enters the function y in a nonadditive way. This is a particular case of the model considered by Imbens and Newey (2009); there are two observable covariates and interest centers on the effect of B on Y. As noted by Imbens and Newey (2009), this setup is general enough to encompass a variety of treatment effect models. When B is binary, the treatment effect for a particular individual is given by $Y_1 - Y_0 = y(1, V, U) - y(0, V, U)$; when B is continuous, the treatment effect is $\frac{\partial}{\partial b} Y = \frac{\partial}{\partial b} y(b, V, U)$. In settings with discrete outcomes, Y could be defined as an individual-specific probability of a particular outcome (as in a binary response model) or as an individual-specific expected value (e.g., an expected duration) that depends on B, V, and U, where the

---

[2]The marginal effects of interest in this paper refer to derivatives of an outcome variable with respect to a continuous endogenous regressor, and should not be confused with the marginal treatment effects defined in Heckman and Vytlacil (2005), where the treatment is binary.

structural function of interest is the relation between B and the probability or expected value.[3]

For the continuous regressor case, Florens et al. (2008) defined the "treatment-on-the-treated" (TT) parameter as

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

where $F_{U|B=b,V=v}(u)$ is the c.d.f. of U conditional on $B = b, V = v$. As noted by Florens et al. (2008), this is equivalent to the "local average response" (LAR) parameter of Altonji and Matzkin (2005). The TT (or equivalently the LAR) gives the average effect of a marginal increase in b at some specific value of the pair (b, v), holding fixed the distribution of the unobservables, $F_{U|B=b,V=v}(\cdot)$.

Recent studies, including Florens et al. (2008) and Imbens and Newey (2009), have proposed methods that use an instrumental variable Z to identify causal parameters such as TT or LAR. An appropriate instrument Z is assumed to influence B but is also assumed to be *independent* of the nonadditive errors in the model. Chesher (2003) observed that such independence assumptions may be "strong and unpalatable," and hence proposed the use of local independence of Z to identify local effects.

As noted in the Introduction, there are some important contexts where no instruments can plausibly satisfy the independence assumption, either globally or locally. For example, consider the case where Y represents the expected duration of unemployment for a job loser, B represents the level of unemployment benefits, and V represents pre-job-loss earnings. Assume (as in many institutional settings) that unemployment benefits are a linear function of pre-job-loss earnings up to some maximum, that is, $B = b(V) = \rho \min(V, T)$. Conditional on V, there is no variation in the benefit level, so model (2.1) is not nonparametrically identified. One could try to get around this fundamental nonidentification by treating V as an error component correlated with B. But in this case, any variable that is independent of V will, by construction, be independent of the regressor of interest B, so it will not be possible to find instruments for B, holding constant the policy regime.

Nevertheless, it may be possible to exploit the kink in the benefit rule to identify the causal effect of B on Y. The idea is that if B exerts a causal effect on Y, and there is a kink in the deterministic relation between B and V at $v = T$, then we should expect to see an induced kink in the relationship between Y and V at $v = T$.[4] Using the kink for identification is in a similar spirit to the regression

---

[3]In these cases, one would use the observed outcome Y^0 (a discrete outcome, or an observed duration), and use the fact that the expectations of Y^0 and Y are equivalent given the same conditioning statement, in applying all of the identification results below.

[4]Without loss of generality, we normalize the kink threshold T to 0 in the remainder of our theoretical presentation.

discontinuity design of Thistlethwaite and Campbell (1960), but the RD approach cannot be directly applied when the benefit formula $b(\cdot)$ is continuous. This kink-based identification strategy has been employed in a few empirical studies. Guryan (2001), for example, used kinks in state education aid formulas as part of an instrumental variables strategy to study the effect of public school spending.[5] Dahlberg et al. (2008) used the same approach to estimate the impact of intergovernmental grants on local spending and taxes. More recently, Simonsen, Skipper, and Skipper (2015) used a kinked relationship between total expenditure on prescription drugs and their marginal price to study the price sensitivity of demand for prescription drugs. Nielsen, Sørensen, and Taber (2010), who introduced the term "Regression Kink Design" for this approach, used a kinked student aid scheme to identify the effect of direct costs on college enrollment.

Nielsen, Sørensen, and Taber (2010) made precise the assumptions needed to identify the causal effects in the constant-effect, additive model

$$(2.2) \qquad Y = \tau B + g(V) + \varepsilon,$$

where $B = b(V)$ is assumed to be a deterministic (and continuous) function of V with a kink at $V = 0$. They showed that if $g(\cdot)$ and $E[\varepsilon|V = v]$ have derivatives that are continuous in v at $v = 0$, then

$$\tau = \frac{\lim\limits_{v_0 \to 0^+} \left. \dfrac{dE[Y|V=v]}{dv}\right|_{v=v_0} - \lim\limits_{v_0 \to 0^-} \left. \dfrac{dE[Y|V=v]}{dv}\right|_{v=v_0}}{\lim\limits_{v_0 \to 0^+} b'(v_0) - \lim\limits_{v_0 \to 0^-} b'(v_0)}.$$

The expression on the right-hand side of this equation—the RKD estimand—is simply the change in slope of the conditional expectation function $E[Y|V = v]$ at the kink point ($v = 0$), divided by the change in the slope of the deterministic assignment function $b(\cdot)$ at 0.[6]

Also related are papers by Dong and Lewbel (2014) and Dong (2013), which derive identification results using kinks in a regression discontinuity setting. Dong and Lewbel (2014) showed that the derivative of the RD treatment effect with respect to the running variable, which the authors called "TED,"

---

[5]Guryan (2001) described the identification strategy as follows: "In the case of the Overburden Aid formula, the regression includes controls for the valuation ratio, 1989 per-capita income, and the difference between the gross standard and 1993 education expenditures (the standard of effort gap). Because these are the only variables on which Overburden Aid is based, the exclusion restriction only requires that the functional form of the direct relationship between test scores and any of these variables is not the same as the functional form in the Overburden Aid formula."

[6]In an earlier working paper version, Nielsen, Sørensen, and Taber (2010) provided similar conditions for identification for a less restrictive, additive model, $Y = g(B, V) + \varepsilon$.

is nonparametrically identified. Under a local policy invariance assumption, TED can be interpreted as the change in the treatment effect that would result from a marginal change in the RD threshold. More closely related to our study is Dong (2013), which showed that identification in an RD design can be achieved in the absence of a first-stage discontinuity, provided there is a kink in the treatment probability at the RD cutoff. In Remark 6 below, we provide an example where such a kink could be expected. Dong (2013) also showed that a slope and level change in the treatment probability can both be used to identify the RD treatment effect with a local constant treatment effect restriction; we discuss an analogous point in the RK design in Remark 3.

Below, we provide the following new identification results. First, we establish identification conditions for the RK design in the context of the general nonseparable model (2.1). By allowing the error term to enter nonseparably, we permit unrestricted heterogeneity in the structural relation between the endogenous regressor and the outcome. As an example of the relevance of this generalization, consider the case of modeling the impact of UI benefits on unemployment durations with a proportional hazards model. Even if UI benefits enter the hazard function with a constant coefficient, the shape of the baseline hazard will, in general, cause the true model for expected durations to be incompatible with the constant-effects, additive specification in (2.2). The addition of multiplicative unobservable heterogeneity (as in Meyer (1990)) to the baseline hazard poses an even greater challenge to the justification of parametric specifications such as (2.2). The nonseparable model (2.1), however, contains the implied model for durations in Meyer (1990) as a special case, and goes further by allowing (among other things) the unobserved heterogeneity to be correlated with V and B. Having introduced unobserved heterogeneity in the structural relation, we show that the RKD estimand τ identifies an effect that can be viewed as the TT (or LAR) parameter. Given that the identified effect is an average of marginal effects across a heterogeneous population, we also make precise how the RKD estimand implicitly weights these heterogeneous marginal effects. The weights are intuitive and correspond to the weights that would determine the slope of the experimental response function in a randomized experiment.

Second, we generalize the RK design to allow for the presence of unobserved determinants of B and measurement errors in B and V. That is, while maintaining the model in (2.1), we allow for the possibility that the observed value for B deviates from the amount predicted by the formula using V, either because of unobserved inputs in the formula, noncompliance behavior, or measurement errors in V or B. This "fuzzy RKD" generalization may have broader applicability than the "sharp RKD."[7]

---

[7]The sharp/fuzzy distinction in the RKD is analogous to that for the RD Design (see Hahn, Todd, and der Klaauw (2001)).

Finally, we provide testable implications for a valid RK design. As we discuss below, a key condition for identification in the RKD is that the distribution of V for each individual is sufficiently smooth. This smooth density condition rules out the case where an individual can *precisely* manipulate V, but allows individuals to exert some influence over V.[8] We provide two tests that can be useful in assessing whether this key identifying assumption holds in practice.

2.2. *Identification of Regression Kink Designs*

2.2.1. *Sharp RKD*

We begin by stating the identifying assumptions for the RKD and making precise the interpretation of the resulting causal effect. In particular, we provide conditions under which the RKD identifies the $TT_{b|v}$ parameter defined above.

Sharp RK Design: Let (V, U) be a pair of random variables (with V observable and U unobservable). While the running variable V is one-dimensional, the error term U need not be, and this unrestricted dimensionality of heterogeneity makes the nonseparable model (2.1) equivalent to treatment effects models as mentioned in Section 2.1. Denote the c.d.f. and p.d.f. of V conditional on $U = u$ by $F_{V|U=u}(v)$ and $f_{V|U=u}(v)$. Define $B \equiv b(V)$, $Y \equiv y(B, V, U)$, $y_1(b, v, u) \equiv \frac{\partial y(b,v,u)}{\partial b}$, and $y_2(b, v, u) \equiv \frac{\partial y(b,v,u)}{\partial v}$. Let I_V be an arbitrarily small closed interval around the cutoff 0 and $I_{b(V)} \equiv \{b | b = b(v) \text{ for some } v \in I_V\}$ be the image of I_V under the mapping b. In the remainder of this section, we use the notation $I_{S_1,\ldots,S_k}$ to denote the product space $I_{S_1} \times \cdots \times I_{S_k}$, where the S_j's are random variables.

ASSUMPTION 1—Regularity: (i) *The support of U is bounded*: *it is a subset of the arbitrarily large compact set $I_U \subset \mathbb{R}^m$.* (ii) *$y(\cdot, \cdot, \cdot)$ is a continuous function and is partially differentiable w.r.t. its first and second arguments. In addition, $y_1(b, v, u)$ is continuous on $I_{b(V),V,U}$.*

ASSUMPTION 2—Smooth Effect of V: *$y_2(b, v, u)$ is continuous on $I_{b(V),V,U}$.*

ASSUMPTION 3—First Stage and Nonnegligible Population at the Kink: (i) *$b(\cdot)$ is a known function, everywhere continuous and continuously differentiable on $I_V \setminus \{0\}$, but $\lim_{v\to 0^+} b'(v) \neq \lim_{v\to 0^-} b'(v)$.* (ii) *The set $A_U = \{u : f_{V|U=u}(v) > 0 \, \forall v \in I_V\}$ has a positive measure under U: $\int_{A_U} dF_U(u) > 0$.*

---

[8]Lee (2008) required a similar identifying condition in a regression discontinuity design. Even though the smooth density condition is not necessary for an RD design, it leads to many intuitive testable implications, which the minimal continuity assumptions in Hahn, Todd, and der Klaauw (2001) do not.

ASSUMPTION 4—Smooth Density: *The conditional density $f_{V|U=u}(v)$ and its partial derivative w.r.t. v, $\frac{\partial f_{V|U=u}(v)}{\partial v}$, are continuous on $I_{V,U}$.*

Assumption 1(i) can be relaxed, but other regularity conditions, such as the dominance of y_1 by an integrable function with respect to F_U, will be needed instead to allow for the interchange of differentiation and integration in proving Proposition 1 below. Assumption 1(ii) states that the marginal effect of B must be a continuous function of the observables and the unobserved error U. Assumption 2 is considerably weaker than an exclusion restriction that dictates V not enter as an argument, because here V is allowed to affect Y, as long as its marginal effect is continuous.

Assumption 3(i) states that the researcher knows the function $b(v)$, and that there is a kink in the relationship between B and V at the threshold $V = 0$. The continuity of $b(v)$ is important and rules out the case where the level of $b(v)$ also changes at $v = 0$. Its necessity stems from the flexibility of our model, which we discuss in more detail in Remark 3. Assumption 3(ii) states that the density of V must be positive around the threshold for a nontrivial subpopulation.

Assumption 4 is another key identifying assumption for a valid RK design. But whereas continuity of $f_{V|U=u}(v)$ in v is sufficient for identification in the RD design, it is insufficient in the RK design. Instead, the sufficient condition is the continuity of the partial derivative of $f_{V|U=u}(v)$ with respect to v. In Card, Lee, Pei, and Weber (2012), we discussed a simple equilibrium search model where Assumption 4 may or may not hold. The importance of this assumption underscores the need to be able to empirically test its implications.

PROPOSITION 1: *In a valid Sharp RKD, that is, when Assumptions 1–4 hold:*
(a) $\Pr(U \leq u|V = v)$ *is continuously differentiable in v at $v = 0$ $\forall u \in I_U$.*
(b)

$$\frac{\lim_{v_0 \to 0^+} \frac{dE[Y|V=v]}{dv}\Big|_{v=v_0} - \lim_{v_0 \to 0^-} \frac{dE[Y|V=v]}{dv}\Big|_{v=v_0}}{\lim_{v_0 \to 0^+} \frac{db(v)}{dv}\Big|_{v=v_0} - \lim_{v_0 \to 0^-} \frac{db(v)}{dv}\Big|_{v=v_0}}$$

$$= E\big[y_1(b_0, 0, U)|V = 0\big] = \int_u y_1(b_0, 0, u)\frac{f_{V|U=u}(0)}{f_V(0)}\, dF_U(u)$$

$$= TT_{b_0|0},$$

where $b_0 = b(0)$.

PROOF: For part (a), we apply Bayes' rule and write

$$\Pr(U \leq u|V = v) = \int_A \frac{f_{V|U=u'}(v)}{f_V(v)}\, dF_U(u'),$$

where $A = \{u' : u' \le u\}$. The continuous differentiability of $\Pr(U \le u | V = v)$ in v follows from Lemma 1 and Lemma 2 in Section A.2 of the Supplemental Material (Card, Lee, Pei, and Weber (2015b)).

For part (b), in the numerator

$$
\lim_{v_0 \to 0^+} \frac{dE[Y|V=v]}{dv}\bigg|_{v=v_0}
$$

$$
= \lim_{v_0 \to 0^+} \frac{d}{dv} \int y(b(v), v, u) \frac{f_{V|U=u}(v)}{f_V(v)} \, dF_U(u)\bigg|_{v=v_0}
$$

$$
= \lim_{v_0 \to 0^+} \int \frac{\partial}{\partial v} y(b(v), v, u) \frac{f_{V|U=u}(v)}{f_V(v)} \, dF_U(u)\bigg|_{v=v_0}
$$

$$
= \lim_{v_0 \to 0^+} b'(v_0) \int y_1(b(v_0), v_0, u) \frac{f_{V|U=u}(v_0)}{f_V(v_0)} \, dF_U(u)
$$

$$
+ \lim_{v_0 \to 0^+} \int \left\{ y_2(b(v_0), v_0, u) \frac{f_{V|U=u}(v_0)}{f_V(v_0)} \right.
$$
$$
\left. + y(b(v_0), v_0, u) \frac{\partial}{\partial v} \frac{f_{V|U=u}(v_0)}{f_V(v_0)} \right\} dF_U(u).
$$

A similar expression is obtained for $\lim_{v_0 \to 0^-} \frac{dE[Y|V=v]}{dv}|_{v=v_0}$. The bounded support and continuity in Assumptions 1–4 allow differentiating under the integral sign per Roussas (2004, p. 97). We also invoke the dominated convergence theorem allowed by the continuity conditions over a compact set in order to exchange the limit operator and the integral. It implies that the difference in slopes above and below the kink threshold can be simplified to

$$
\lim_{v_0 \to 0^+} \frac{dE[Y|V=v]}{dv}\bigg|_{v=v_0} - \lim_{v_0 \to 0^-} \frac{dE[Y|V=v]}{dv}\bigg|_{v=v_0}
$$

$$
= \left(\lim_{v \to 0^+} b'(v_0) - \lim_{v \to 0^-} b'(v_0) \right) \int y_1(b(0), 0, u) \frac{f_{V|U=u}(0)}{f_V(0)} \, dF_U(u).
$$

Assumption 3(i) states that the denominator $\lim_{v_0 \to 0^+} b'(v_0) - \lim_{v_0 \to 0^-} b'(v_0)$ is nonzero, and hence we have

$$
\frac{\lim_{v_0 \to 0^+} \frac{dE[Y|V=v]}{dv}\bigg|_{v=v_0} - \lim_{v_0 \to 0^-} \frac{dE[Y|V=v]}{dv}\bigg|_{v=v_0}}{\lim_{v_0 \to 0^+} b'(v_0) - \lim_{v_0 \to 0^-} b'(v_0)}
$$

$$
= E[y_1(b(0), 0, U)|V=0] = \int y_1(b(0), 0, u) \frac{f_{V|U=u}(0)}{f_V(0)} \, dF_U(u),
$$

which completes the proof. *Q.E.D.*

Part (a) states that the rate of change in the probability distribution of individual types with respect to the assignment variable V is continuous at $V = 0$.[9] This leads directly to part (b): as a consequence of the smoothness in the underlying distribution of types around the kink, the discontinuous change in the slope of $E[Y|V = v]$ at $v = 0$ divided by the discontinuous change in slope in $b(V)$ at the kink point identifies $TT_{b_0|0}$.[10]

REMARK 1: It is tempting to interpret $TT_{b_0|0}$ as the "average marginal effect of B for individuals with $V = 0$," which may seem very restrictive because the smooth density condition implies that $V = 0$ is a measure-zero event. However, part (b) implies that $TT_{b_0|0}$ is a weighted average of marginal effects across the entire population, where the weight assigned to an individual of type U reflects the *relative* likelihood that he or she has $V = 0$. In settings where U is highly correlated with V, $TT_{b_0|0}$ is only representative of the treatment effect for agents with realizations of U that are associated with values of V close to 0. In settings where V and U are independent, the weights for different individuals are equal, and RKD identifies the average marginal effect evaluated at $B = b_0$ and $V = 0$.

REMARK 2: The weights in Proposition 1 are the *same* ones that would be obtained from using a randomized experiment to identify the average marginal effect of B, evaluated at $B = b_0$, $V = 0$. That is, suppose that B was assigned randomly so that $f_{B|V,U}(b) = f(b)$. In such an experiment, the identification of an average marginal effect of b at $V = 0$ would involve taking the derivative of the experimental response surface $E[Y|B = b, V = v]$ with respect to b for

---

[9]Note also that Proposition 1(a) implies Proposition 2(a) in Lee (2008), that is, the continuity of $\Pr(U \leq u|V = v)$ at $v = 0$ for all u. This is a consequence of the stronger smoothness assumption we have imposed on the conditional distribution of V on U.

[10]Technically, the *TT* and *LAR* parameters do not condition on a second variable V. But in the case where there is a one-to-one relationship between B and V, the trivial integration over the (degenerate) distribution of V conditional on $B = b_0$ will imply that $TT_{b_0|0} = TT_{b_0} \equiv E[y_1(b_0, V, U)|B = b_0]$, which is literally the *TT* parameter discussed in Florens et al. (2008) and the *LAR* discussed in Altonji and Matzkin (2005). In our application to unemployment benefits, B and V are not one-to-one, since beyond $V = 0$, B is at the maximum benefit level. In this case, TT_b will, in general, be discontinuous with respect to b at b_0:

$$TT_b = \begin{cases} TT_{b|v}, & b < b_0, \\ \int TT_{b_0|v} f_{V|B}(v|b_0) \, dv, & b = b_0, \end{cases}$$

and the RKD estimand identifies $\lim_{b \uparrow b_0} TT_b$.

units with $V = 0$. This would yield

$$\frac{\partial E[Y|B = b, V = 0]}{\partial b}\bigg|_{b=b_0}$$

$$= \frac{\partial\left(\int y(b, 0, u)\, dF_{U|V=0, B=b}(u)\right)}{\partial b}\bigg|_{b=b_0}$$

$$= \frac{\partial\left(\int y(b, 0, u)\frac{f_{B|V=0, U=u}(b)}{f_{B|V=0}(b)}\frac{f_{V|U=u}(0)}{f_V(0)}\, dF_U(u)\right)}{\partial b}\bigg|_{b=b_0}$$

$$= \frac{\partial\left(\int y(b, 0, u)\frac{f_{V|U=u}(0)}{f_V(0)}\, dF_U(u)\right)}{\partial b}\bigg|_{b=b_0}$$

$$= \int y_1(b_0, 0, u)\frac{f_{V|U=u}(0)}{f_V(0)}\, dF_U(u).$$

Even though B is randomized in this hypothetical experiment, V is not. Intuitively, although randomization allows one to identify marginal effects of B, it cannot resolve the fact that units with $V = 0$ will, in general, have a particular distribution of U. Of course, the advantage of this hypothetical randomized experiment is that one could potentially identify the average marginal effect of B at all values of B and V, and not just at $B = b_0$ and $V = 0$.

REMARK 3: In the proof of Proposition 1, we need the continuity of $b(v)$ to ensure that the left and right limits of $y_1(b(v_0), v_0, u)$, $y_2(b(v_0), v_0, u)$, and $y(b(v_0), v_0, u)$ are the same as v_0 approaches 0. In the case where both the slope and the level of $b(v)$ change at $v = 0$, the RK estimand does not point identify an interpretable treatment effect in the nonseparable model (2.1). The RD estimand, however, still identifies an average treatment effect. In Section A.2 of the Supplemental Material, we show

$$\frac{\lim_{v_0 \to 0^+} E[Y|V = v_0] - \lim_{v_0 \to 0^-} E[Y|V = v_0]}{\lim_{v_0 \to 0^+} b(v_0) - \lim_{v_0 \to 0^-} b(v_0)} = E[y_1(\tilde{b}, 0, U)|V = 0],$$

where $\tilde{b}$ is a value between $\lim_{v_0 \to 0^-} b(v_0)$ and $\lim_{v_0 \to 0^+} b(v_0)$. In the special case of a constant treatment effect model like (2.2), the RD and RK design both identify the same causal effect parameter. In the absence of strong a priori

knowledge about treatment effect homogeneity, however, it seems advisable to use an RD design.[11]

2.2.2. *Fuzzy Regression Kink Design*

Although many important policy variables are set according to a deterministic formula, in practice there is often some slippage between the theoretical value of the variable as computed by the stated rule and its observed value. This can arise when the formula—while deterministic—depends on other (unknown) variables in addition to the primary assignment variable, when there is noncompliance with the policy formula, or when measurement errors are present. This motivates the extension to a fuzzy RKD.[12]

Specifically, assume now that $B = b(V, \varepsilon)$, where the presence of ε in the formula for B allows for unobserved determinants of the policy formula and noncompliant behavior. The vector ε is potentially correlated with U and therefore also with the outcome variable Y. As an illustration, consider the simple case where the UI benefit formula depends on whether or not a claimant has dependents. Let D be a claimant with dependents and let N be a claimant with no dependents, and let $b^1(v)$ and $b^0(v)$ be the benefit formulas for D and N, respectively. Suppose D and N both have base period earnings of v_0 and that the only noncompliant behavior allowed is for D to claim $b^0(v_0)$ or for N to claim $b^1(v_0)$. In this case, we have two potentially unobserved variables that determine treatment: (1) whether a claimant has dependents or not, and (2) whether a claimant "correctly" claims her benefits. We can represent these two variables with a two-dimensional vector $\varepsilon = (\varepsilon_1, \varepsilon_2)$. The binary indicator ε_1 is equal to 1 if a claimant truly has dependents, whereas ε_2 takes four values denoting whether a claimant with base period earnings v is an "always taker" (always claiming $b^1(v)$), a "never taker" (always claiming $b^0(v)$), a "complier" (claiming $b^{\varepsilon_1}(v)$), or a "defier" (claiming $b^{1-\varepsilon_1}(v)$). The representation $B = b(V, \varepsilon_1, \varepsilon_2)$ effectively captures the treatment assignment mechanism described in this simple example. With suitable definition of ε, it can also be used to allow for many other types of deviations from a deterministic rule. Except for a bounded support assumption similar to that for U, we do not need to impose any other restrictions on the distribution of ε. We will use $F_{U,\varepsilon}$ to denote the measure induced by the joint distribution of U and ε.

---

[11]Turner (2013) studied the effect of the Pell Grant program in the United States. The formula for these grants has both a discontinuity and a slope change at the Grant eligibility threshold. She argued that the status of being a Pell Grant recipient, D, may impact Y independently from the marginal financial effect of B on Y (i.e., $Y = y(B, D, V, U)$), and she studied the identification of the two treatment effects in a special case that restricts treatment effect heterogeneity.

[12]See Hahn, Todd, and der Klaauw (2001) for a definition of the fuzzy regression discontinuity design.

We also assume that the observed values of B and V, B^* and V^* respectively, differ from their true values as follows:

$$V^* \equiv V + U_V, \quad B^* \equiv B + U_B,$$
$$U_V \equiv G_V \cdot U_{V'}, \quad U_B \equiv G_B \cdot U_{B'},$$

where $U_{V'}$ and $U_{B'}$ are continuously distributed, and that their joint density conditional on U and ε is continuous and supported on an arbitrarily large compact rectangle $I_{U_{V'}, U_{B'}} \subset R^2$; G_V and G_B are binary indicators whose joint conditional distribution is given by the four probabilities $\pi^{ij}(V, U, \varepsilon, U_{V'}, U_{B'}) \equiv \Pr(G_V = i, G_B = j | V, U, \varepsilon, U_{V'}, U_{B'})$. Note that the errors in the observed values of V and B are assumed to be mixtures of conventional (continuously distributed) measurement error and a point mass at 0. The random variables $(V, U, \varepsilon, U_{V'}, U_{B'}, G_V, G_B)$ determine (B, B^*, V^*, Y), and we observe (B^*, V^*, Y).

ASSUMPTION 1a—Regularity: *In addition to the conditions in Assumption* 1, *the support of ε is bounded*: *it is a subset of the arbitrarily large compact set $I_\varepsilon \subset \mathbb{R}^k$.*

ASSUMPTION 3a—First Stage and Nonnegligible Population at the Kink: *$b(v, e)$ is continuous on $I_{V,\varepsilon}$ and $b_1(v, e)$ is continuous on $(I_V \setminus \{0\}) \times I_\varepsilon$. Let $b_1^+(e) \equiv \lim_{v \to 0^+} b_1(v, e)$, $b_1^-(e) \equiv \lim_{v \to 0^-} b_1(v, e)$, and $A_\varepsilon = \{e : f_{V|\varepsilon=e}(0) > 0\}$, then $\int_{A_\varepsilon} \Pr[U_V = 0 | V = 0, \varepsilon = e] |b_1^+(e) - b_1^-(e)| f_{V|\varepsilon=e}(0) \, dF_\varepsilon(e) > 0$.*

ASSUMPTION 4a—Smooth Density: *Let V, $U_{V'}$, $U_{B'}$ have a well-defined joint p.d.f. conditional on each $U = u$ and $\varepsilon = e$, $f_{V,U_{V'},U_{B'}|U=u,\varepsilon=e}(v, u_B, u_{V'})$. The density function $f_{V,U_{V'},U_{B'}|U=u,\varepsilon=e}(v, u_B, u_{V'})$ and its partial derivative w.r.t. v are continuous on $I_{V,U_{V'},U_{B'},U,\varepsilon}$.*

ASSUMPTION 5—Smooth Probability of No Measurement Error: *$\pi^{ij}(v, u, e, u_{V'}, u_{B'})$ and its partial derivative w.r.t. v are continuous on $I_{V,U,\varepsilon,U_{V'},U_{B'}}$ for all $i, j = 0, 1$.*

ASSUMPTION 6—Monotonicity: *Either $b_1^+(e) \geq b_1^-(e)$ for all e or $b_1^+(e) \leq b_1^-(e)$ for all e.*

Extending Assumption 1, Assumption 1a imposes the bounded support assumption for ε in order to allow the interchange of differentiation and integration. Assumption 3a modifies Assumption 3 and forbids a discontinuity in $b(\cdot, e)$ at the threshold. Analogously to the sharp case discussed in Remark 3, in the absence of continuity in $b(\cdot, e)$, the RK estimand does not identify a weighted average of the causal effect of interest, y_1, but the RD estimand does; see Section A.2 of the Supplemental Material for details. Assumption 3a also

requires a nonnegligible subset of individuals who simultaneously have a non-trivial first stage, have $U_V = 0$, and have positive probability that V is in a neighborhood of 0. It is critical that there is a mass point in the distribution of the measurement error U_V at 0. In the absence of such a mass point, we will not observe a kink in the first-stage relationship, and further assumptions must be made about the measurement error to achieve identification (as in the case with the RD design). In contrast, there is no need for a mass point in the distribution of U_B at 0, but we simply *allow* for the possibility here. In our empirical example based on UI benefits paid to job losers, we find that the majority of the data points appear to lie precisely on the benefit schedule (see Figure 1 of Card, Lee, Pei, and Weber (2015a)), a feature that we interpret as evidence of a mass point at zero in the joint distribution of (U_V, U_B). Assumption 3a can be formally tested by the existence of a first-stage kink in $E[B^*|V^* = v^*]$ as stated in Remark 4 below.

Assumption 4a modifies Assumption 4: for each $U = u$ and $\varepsilon = e$, there is a joint density of V and the measurement error components that is continuously differentiable in v. Note that this allows a relatively general measurement error structure in the sense that V, $U_{V'}$, $U_{B'}$ can be arbitrarily correlated. Assumption 5 states that the mass point probabilities, while potentially dependent on all other variables, are smooth with respect to V.

Assumption 6 states that the direction of the kink is either nonnegative or nonpositive for the entire population, and it is analogous to the monotonicity condition of Imbens and Angrist (1994). In particular, Assumption 6 rules out situations where some individuals experience a positive kink at $V = 0$, but others experience a negative kink at $V = 0$. In our application below, where actual UI benefits depend on the (unobserved) number of dependents, this condition is satisfied since the benefit schedules for different numbers of dependents are all parallel.

PROPOSITION 2: *In a valid Fuzzy RK Design, that is, when Assumptions* 1a, 2, 3a, 4a, 5, *and* 6 *hold*:

(a) $\Pr(U \leq u, \varepsilon \leq e|V^* = v^*)$ *is continuously differentiable in* v^* *at* $v^* = 0$ $\forall (u, e) \in I_{U,\varepsilon}$.

(b)

$$
\frac{\lim\limits_{v_0 \to 0^+} \left. \dfrac{dE[Y|V^* = v^*]}{dv^*} \right|_{v^* = v_0} - \lim\limits_{v_0 \to 0^-} \left. \dfrac{dE[Y|V^* = v^*]}{dv^*} \right|_{v^* = v_0}}{\lim\limits_{v_0 \to 0^+} \left. \dfrac{dE[B^*|V^* = v^*]}{dv^*} \right|_{v^* = v_0} - \lim\limits_{v_0 \to 0^-} \left. \dfrac{dE[B^*|V^* = v^*]}{dv^*} \right|_{v^* = v_0}}
$$

$$
= \int y_1\big(b(0, e), 0, u\big)\varphi(u, e)\, dF_{U,\varepsilon}(u, e),
$$

where

$$\varphi(u, e)$$

$$= \frac{\Pr[U_V = 0 | V = 0, U = u, \varepsilon = e](b_1^+(e) - b_1^-(e))\frac{f_{V|U=u,\varepsilon=e}(0)}{f_V(0)}}{\int \Pr[U_V = 0 | V = 0, \varepsilon = \omega](b_1^+(\omega) - b_1^-(\omega))\frac{f_{V|\varepsilon=\omega}(0)}{f_V(0)} \, dF_\varepsilon(\omega)}.$$

The proof is in Section A.1 of the Supplemental Material.

REMARK 4: The fuzzy RKD continues to estimate a weighted average of marginal effects of B on Y, but the weight is now given by $\varphi(u, e)$. Assumptions 3a and 6 ensure that the denominator of $\varphi(u, e)$ is nonzero. They also ensure a kink at $v^* = 0$ in the first-stage relationship between B^* and V^*, as seen from the proof of Proposition 2. It follows that the existence of a first-stage kink serves as a test of Assumptions 3a and 6.

REMARK 5: The weight $\varphi(u, e)$ has three components. The first component, $\frac{f_{V|U=u,\varepsilon=e}(0)}{f_V(0)}$, is analogous to the weight in a sharp RKD and reflects the relative likelihood that an individual of type $U = u$, $\varepsilon = e$ is situated at the kink (i.e., has $V = 0$). The second component, $b_1^+(e) - b_0^-(e)$, reflects the size of the kink in the benefit schedule at $V = 0$ for an individual of type e. Analogously to the LATE interpretation of a standard instrumental variables setting, the fuzzy RKD estimand upweights types with a larger kink at the threshold $V = 0$. Individuals whose benefit schedule is *not* kinked at $V = 0$ do not contribute to the estimand. An important potential difference from a standard LATE setting is that non-compliers may still receive positive weights if the schedule they follow as non-compliers has a kink at $V = 0$. Finally, the third component $\Pr[U_V = 0 | V = 0, U = u, \varepsilon = e]$ represents the probability that the assignment variable is correctly measured at $V = 0$. Again, this has the intuitive implication that observations with a mismeasured value of the assignment variable do not contribute to the fuzzy RKD estimand.

REMARK 6: So far, we have focused on a continuous treatment variable B, but the RKD framework may be applied to estimate the treatment effect of a binary variable as well. As mentioned above, Dong (2013) discussed the identification of the treatment effect within an RD framework where the treatment probability conditional on the running variable is continuous but kinked. Under certain regularity conditions, Dong (2013) showed that the RK estimand identifies the treatment effect at the RD cutoff for the group of compliers. In practice, it may be difficult to find policies where the probability of a binary treatment is statutorily mandated to have a kink in an observed running variable. One possibility, suggested by a referee, is that the kinked relationship

between two continuous variables B and V may *induce* a kinked relationship between T and V where T is a binary treatment variable of interest. In this case, we may apply the RK design to measure the treatment effect of T. To be more specific, let

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

$$T = 1_{[T^* \geq 0]} \quad \text{where} \quad T^* = t(B, V, \eta),$$

$$B = b(V) \text{ is continuous in } V \text{ with a kink at } V = 0.$$

As an example, B is the amount of financial aid available, which is a kinked function of parental income V. T^* is a latent index function of B, V, and a one-dimensional error term η. A student will choose to attend college ($T = 1$) if $T^* \geq 0$. We are interested in estimating the average returns to college education, an expectation of $y(1, V, U) - y(0, V, U)$. Assuming that t is monotonically increasing in its third argument and that, for every $(b, v) \in I_{b(V)} \times I_V$, there exists an n such that $t(b(v), v, n) = 0$, we can define a continuously differentiable function $\tilde{\eta} : I_{b(V)} \times I_V \to \mathbb{R}$ such that $t(b, v, \tilde{\eta}(b, v)) = 0$ by the implicit function theorem. We show in Section A.3 of the Supplemental Material that, under additional regularity conditions, we have the following identification result for the fuzzy RK estimand:

$$(2.3) \quad \frac{\lim_{v_0 \to 0^+} \frac{dE[Y|V=v]}{dv}\Big|_{v=v_0} - \lim_{v_0 \to 0^-} \frac{dE[Y|V=v]}{dv}\Big|_{v=v_0}}{\lim_{v_0 \to 0^+} \frac{dE[T|V=v]}{dv}\Big|_{v=v_0} - \lim_{v_0 \to 0^-} \frac{dE[T|V=v]}{dv}\Big|_{v=v_0}}$$

$$= \int_u \big[y(1, 0, u) - y(0, 0, u)\big] \frac{f_{V,\eta|U=u}(0, n^0)}{f_{V,\eta}(0, n^0)} \, dF_U(u),$$

where $n^0 \equiv \tilde{\eta}(b_0, 0)$ is the threshold value of η when $V = 0$ such that $n \geq n^0 \Leftrightarrow T(b_0, 0, n) = 1$. The right-hand side of equation (2.3) is similar to that in part (b) of Proposition 1, and the weights reflect the relative likelihood of $V = 0$ *and* $\eta = n^0$ for a student of type U.

Crucial to the point identification result above is the exclusion restriction that B does not enter the function y as an argument, that is, that the amount of financial aid does not have an independent effect on future earnings conditional on parental income and college attendance. When this restriction is not met, the RK estimand can be used to bound the effect of T on Y if theory can shed light on the sign of the independent effect of B on Y. The details are in Section A.3 of the Supplemental Material.

We can also allow the relationship between B and V to be fuzzy by writing $B = b(V, \varepsilon)$ and introducing measurement error in V as above. Similarly to Proposition 2, we show that the fuzzy RK estimand still identifies a weighted

average of treatment effects under certain regularity assumptions. The weights are similar to those in Proposition 2, and the exact expression is in the Supplemental Material.

2.3. *Testable Implications of the RKD*

In this section, we formalize the testable implications of a valid RK design. Specifically, we show that the key smoothness conditions given by Assumptions 4 and 4a lead to two strong testable predictions. The first prediction is given by the following corollary of Propositions 1 and 2:

COROLLARY 1: *In a valid Sharp RKD*, $f_V(v)$ *is continuously differentiable in* v. *In a valid Fuzzy RKD*, $f_{V^*}(v^*)$ *is continuously differentiable in* v^*.

The key identifying assumption of the sharp RKD is that the density of V is sufficiently smooth for every individual. This smoothness condition cannot be true if we observe either a kink or a discontinuity in the density of V. That is, evidence that there is "deterministic sorting" in V at the kink point implies a violation of the key identifying sharp RKD assumption. This is analogous to the test of manipulation of the assignment variable for RD designs, discussed in McCrary (2008). In a fuzzy RKD, both Assumption 4a, the smooth-density condition, and Assumption 5, the smooth-probability-of-no-measurement-error condition, are needed to ensure the smoothness of f_{V^*} (see the proof of Lemma 5), and a kink or a discontinuity in f_{V^*} indicates that either or both of the assumptions are violated.

The second prediction presumes the existence of data on "baseline characteristics"—analogous to characteristics measured prior to treatment assignment in an idealized randomized controlled trial—that are determined prior to V.

ASSUMPTION 7: *There exists an observable random vector*, $X = x(U)$ *in the sharp design and* $X = x(U, \varepsilon)$ *in the fuzzy design, that is determined prior to* V. X *does not include* V *or* B, *since it is determined prior to those variables*.

In conjunction with our basic identifying assumptions, this leads to the following prediction:

COROLLARY 2: *In a valid Sharp RKD, if Assumption 7 holds, then* $\frac{d\Pr[X \leq x|V=v]}{dv}$ *is continuous in* v *at* $v = 0$ *for all* x. *In a valid Fuzzy RKD, if Assumption 7 holds, then* $\frac{d\Pr[X \leq x|V^*=v^*]}{dv^*}$ *is continuous in* v^* *at* $v^* = 0$ *for all* x.

The smoothness conditions required for a valid RKD imply that the conditional distribution function of any predetermined covariates X (given V or V^*) cannot exhibit a kink at $V = 0$ or $V^* = 0$. Therefore, Corollary 2 can be used to

test Assumption 4 in a sharp design and Assumption 4a and 5 jointly in a fuzzy design. This test is analogous to the simple "test for random assignment" that is often conducted in a randomized trial, based on comparisons of the baseline covariates in the treatment and control groups. It also parallels the test for continuity of $\Pr[X \leq x | V = v]$ emphasized by Lee (2008) for a regression discontinuity design. Importantly, however, the assumptions for a valid RKD imply that the *derivatives* of the conditional expectation functions (or the conditional quantiles) of X with respect to V (or V^*) are continuous at the kink point—a stronger implication than the continuity implied by the sufficient conditions for a valid RDD.

3. NONPARAMETRIC ESTIMATION AND INFERENCE IN A REGRESSION KINK DESIGN

In this section, we review the theory of estimation and inference in a regression kink design. We assume that estimation is carried out via local polynomial regressions. For a sharp RK design, the first-stage relationship $b(\cdot)$ is a known function, and we only need to solve the following least squares problems:

$$(3.1) \qquad \min_{\{\hat{\beta}_j^-\}} \sum_{i=1}^{n^-} \left\{ Y_i^- - \sum_{j=0}^{p} \tilde{\beta}_j^- (V_i^-)^j \right\}^2 K\left(\frac{V_i^-}{h}\right),$$

$$(3.2) \qquad \min_{\{\hat{\beta}_j^+\}} \sum_{i=1}^{n^+} \left\{ Y_i^+ - \sum_{j=0}^{p} \tilde{\beta}_j^+ (V_i^+)^j \right\}^2 K\left(\frac{V_i^+}{h}\right),$$

where the $-$ and $+$ superscripts denote quantities in the regression on the left and right side of the kink point, respectively, p is the order of the polynomial, K the kernel, and h the bandwidth. Since $\kappa_1^+ = \lim_{v \to 0^+} b'(v)$ and $\kappa_1^- = \lim_{v \to 0^-} b'(v)$ are known quantities in a sharp design, the sharp RKD estimator is defined as

$$\hat{\tau}_{\text{SRKD}} = \frac{\hat{\beta}_1^+ - \hat{\beta}_1^-}{\kappa_1^+ - \kappa_1^-}.$$

In a fuzzy RKD, the first-stage relationship is no longer deterministic. We need to estimate the first-stage slopes on two sides of the threshold by solving[13]

$$(3.3) \qquad \min_{\tilde{\kappa}_j^-} \sum_{i=1}^{n^-} \left\{ B_i^- - \sum_{j=0}^{p} \tilde{\kappa}_j^- (V_i^-)^j \right\}^2 K\left(\frac{V_i^-}{h}\right),$$

---

[13]We omit the asterisk in B^* and V^* notations in the fuzzy design to ease exposition.

$$(3.4) \qquad \min_{\tilde{\kappa}_j^+} \sum_{i=1}^{n^+} \left\{ B_i^+ - \sum_{j=0}^{p} \tilde{\kappa}_j^+ (V_i^+)^j \right\}^2 K\left(\frac{V_i^+}{h} \right).$$

The fuzzy RKD estimator $\hat{\tau}_{\text{FRKD}}$ can then be defined as

$$(3.5) \qquad \hat{\tau}_{\text{FRKD}} = \frac{\hat{\beta}_1^+ - \hat{\beta}_1^-}{\hat{\kappa}_1^+ - \hat{\kappa}_1^-}.$$

Lemmas A1 and A2 of Calonico, Cattaneo, and Titiunik (2014) establish the asymptotic distributions of the sharp and fuzzy RKD estimators, respectively. It is shown that, under certain regularity conditions, the estimators obtained from local polynomial regressions of order p are asymptotically normal:

$$\sqrt{nh^3}\big(\hat{\tau}_{\text{SRKD},p} - \tau_{\text{SRKD}} - h^p \varrho_{\text{SRKD},p}\big) \Rightarrow N(0, \Omega_{\text{SRKD},p}),$$
$$\sqrt{nh^3}\big(\hat{\tau}_{\text{FRKD},p} - \tau_{\text{FRKD}} - h^p \varrho_{\text{FRKD},p}\big) \Rightarrow N(0, \Omega_{\text{FRKD},p}),$$

where ϱ and Ω denote the asymptotic bias and variance, respectively.[14] Given the identification assumptions above, one expects the conditional expectation of Y given V to be continuous at the threshold. A natural question is whether imposing continuity in estimation (as opposed to estimating separate local polynomials on either side of the threshold) may affect the asymptotic bias and variance of the kink estimator. Card et al. (2012) showed that when K is uniform, the asymptotic variances are not affected by imposing continuity. A similar calculation reveals that the asymptotic biases are not affected either.

When implementing the RKD estimator in practice, one must make choices for the polynomial order p, kernel K, and bandwidth h. In the RD context where the quantities of interest are the intercept terms on two sides of the threshold, Hahn, Todd, and der Klaauw (2001) proposed local linear ($p = 1$) over local constant ($p = 0$) regression because the former leads to a smaller order of bias ($O_p(h^2)$) than the latter ($O_p(h)$). Consequently, the local linear model affords the econometrician a sequence of bandwidths that shrinks at a slower rate, which in turn delivers a smaller order of the asymptotic mean squared error (MSE). The same logic would imply that a local quadratic ($p = 2$) should be preferred to local linear ($p = 1$) in estimating boundary derivatives in the RK design. As noted by Ruppert and Wand (1994) and Fan and Gijbels (1996) and as we discussed in detail in Card, Lee, Pei, and Weber (2014), however, arguments based solely on asymptotic rates cannot justify

---

[14]In categorizing the asymptotic behavior of fuzzy estimators, both Card et al. (2012) and Calonico, Cattaneo, and Titiunik (2014) assumed that the researcher observes the joint distribution (Y, B, V). In practice, there may be applications where (B, V) is observed in one data source whereas (Y, V) is observed in another, and the three variables do not appear in the same data set. We investigate the two-sample estimation problem in Section B.1 of the Supplemental Material.

$p = 1$ as the universally preferred choice for RDD or $p = 2$ as the universally preferred choice for RKD. Rather, the best choice of p in the mean squared error sense 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. In Card et al. (2014), we proposed two methods for picking the polynomial order for interested empiricists: (1) evaluate the empirical performance of the alternative estimators using simulation studies of DGPs closely based on the actual data;[15] (2) estimate the asymptotic mean squared error (AMSE) and compare it across alternative estimators.

For the choice of K, we adopt a uniform kernel following Imbens and Lemieux (2008) and the common practice in the RD literature. The results are similar when the boundary optimal triangular kernel (cf. Cheng, Fan, and Marron (1997)) is used.

For the bandwidth choice h, we use and extend existing selectors in the literature. Imbens and Kalyanaraman (2012) proposed an algorithm to compute the MSE-optimal RD bandwidth. Building on Imbens and Kalyanaraman (2012), Calonico, Cattaneo, and Titiunik (2014) developed an optimal bandwidth algorithm for the estimation of the discontinuity in the νth derivative, which contains RKD ($\nu = 1$) as a special case.[16]

We examine alternatives to the direct analogs of the default IK and the CCT bandwidths for RKD, addressing two specific issues that are relevant for our setting. First, both bandwidth selectors involve a regularization term, which reflects the variance in the bias estimation and guards against large bandwidths. While IK and CCT argued that the regularized bandwidth selector performs well for several well-known regression discontinuity designs, we find that the RK counterparts of these regularized selectors yield bandwidths that tend to be too small in our empirical setting. Since omitting the regularization term does not affect the asymptotic properties of the bandwidth selector, we also investigate the performance of bandwidth selectors without the regularization term. Second, the CCT bandwidth is asymptotically MSE-optimal for the reduced-form kink in a fuzzy design, even though the fuzzy estimator $\hat{\tau}_{\text{FRKD}}$ defined in (3.5) is the main object of interest. Based on the asymptotic theory in Calonico, Cattaneo, and Titiunik (2014), we propose fuzzy analogs of the IK and CCT bandwidths that are optimal for $\hat{\tau}_{\text{FRKD}}$ and state their asymptotic properties—see Section B.2 of the Supplemental Material for details.

A complication of using the optimal bandwidth is that the asymptotic bias is, in general, nonzero. As a result, conventional confidence intervals that ignore the bias may not have correct coverage rates. Calonico, Cattaneo, and

---

[15]Clearly, this method depends on how a researcher specifies the approximating DGP, but whether the "right" DGP is specified is to some degree untestable.

[16]The optimal bandwidth in Calonico, Cattaneo, and Titiunik (2014) is developed for the unconstrained RKD estimator, that is, without imposing continuity in the conditional expectation of Y, but the bandwidth is also optimal for the constrained RKD estimator because it has the same asymptotic distribution as stated above.

Titiunik (2014) offered a solution by deriving robust confidence intervals for the RD and RK estimands that account for this asymptotic bias. For an RK design, they first estimated the asymptotic bias ϱ_p of a pth order local polynomial estimator $\hat{\tau}_p$ by using a qth order local polynomial regression ($q \geq p + 1$) with pilot bandwidth h_q, then estimated the variance var_p^{bc} of the bias-corrected estimator $\hat{\tau}_p^{bc} \equiv \hat{\tau}_p - h^p \hat{\varrho}_p$, by accounting for the sampling variation in both $\hat{\tau}_p$ and $h^p \hat{\varrho}_p$.[17] Finally, they constructed a robust 95% confidence interval as $\hat{\tau}_p^{bc} \pm 1.96\sqrt{\text{var}_p^{bc}}$. Using Monte Carlo simulations, Calonico, Cattaneo, and Titiunik (2014) demonstrated that the confidence intervals constructed using their bias-corrected procedure perform well in RDDs, and that the associated coverage rates are robust to different choices of h.[18]

In the following section, we present a variety of alternative estimates of the behavioral effect of higher benefits on unemployment durations. We employ the local linear and local quadratic estimators with several bandwidth selectors—default CCT, CCT without regularization, Fuzzy CCT, Fuzzy IK, and the rule of thumb FG bandwidth.[19] We report uncorrected RKD estimates and the associated (conventional) sampling errors, as well as the robust bias-corrected confidence intervals per Calonico, Cattaneo, and Titiunik (2014).

4. THE EFFECT OF UI BENEFITS ON UNEMPLOYMENT DURATIONS

In this section, we illustrate the estimation procedures by using a fuzzy RKD approach to estimate the effect of higher unemployment benefits on the duration of registered unemployment among UI claimants in Austria. The precise magnitude of the disincentive effect of UI benefits is of substantial policy interest. As shown by Baily (1978), for example, an optimal unemployment insurance system trades off the moral hazard costs of reduced search effort against the risk-sharing benefits of more generous payments to the unemployed. Obtaining credible estimates of this effect is difficult, however, because UI benefits are determined by previous earnings, and are likely to be correlated with unobserved characteristics of workers that affect both wages and the expected duration of unemployment. Since the Austrian UI benefit formula—similar to

---

[17]A crucial assumption in estimating var_p^{bc} is that the pilot bandwidth h_q and the optimal bandwidth h have the same shrinkage rate, that is, $\frac{h_q}{h} \to \rho \in (0, \infty)$ as $n \to \infty$.

[18]In a related study, Ganong and Jäger (2014) raised concerns about the sensitivity of the RKD estimates when the relationship between the running variable and the outcome is highly nonlinear. They proposed a permutation test to account for the estimation bias. We perform the test on our data and discuss the details in Card et al. (2015a).

[19]See Card et al. (2012) for the definition of the FG bandwidth. We apply the same logic to derive the pilot FG bandwidth for bias estimation.

those of many other countries—has a maximum, a regression kink approach can provide new evidence on the impact of higher UI benefits.[20]

4.1. *Institutional Setting and Data*

Job losers in Austria who have worked at least 52 weeks in the past 24 months are eligible for UI benefits, with a rate that depends on their average daily earnings in the "base year" for their benefit claim. The daily UI benefit is calculated as 55% of net daily earnings, subject to a maximum benefit level that is adjusted each year. Claimants with dependent family members are eligible for supplemental benefits based on the number of dependents.

This rule creates a piecewise linear relationship between base year earnings and UI benefits that depends on the Social Security and income tax rates as well as the replacement rate and the maximum benefit amount. Since we do not observe the number of dependents claimed by a job loser, we adopt a fuzzy RKD approach in which the number of dependents is treated as an unobserved determinant of benefits.

Our data are drawn from the Austrian Social Security Database (ASSD), which records employment and unemployment spells on a daily basis for all individuals employed in the Austrian private sector (see Zweimüller, Winter-Ebmer, Lalive, Kuhn, Wuellrich, Ruf, and Büchi (2009)). The ASSD contains information on starting and ending dates of employment spells and earnings (up to the Social Security contribution cap) received by each individual from each employer in a calendar year. We merge the ASSD with UI claims records that include the claim date, the daily UI benefit received by the claimant, and the duration of the registered unemployment spell. We use the UI claim dates to assign the base year for each claim and then calculate base year earnings for the claim, which is the observed assignment variable for our RKD analysis (i.e., V^* in the notation of Section 2). The outcome variable we analyze here is the duration of registered unemployment (which we censor at one year).

Our analysis sample includes claimants from 2001 to 2012 with at least one year of tenure on their previous job who initiated their claim within four weeks of the job ending date (eliminating job quitters, who face a four-week waiting period). We drop individuals with zero earnings in the base year, claimants older than 50, and those whose earnings are above the Social Security earnings cap or whose earnings are so low that they are affected by other nonlinearities in the benefit schedule (see Card et al. (2015a) for analysis of further kinks). We pool observations from different years by centering around the respective income thresholds for the maximum benefit level. This yields a sample of about 275,000 observations.

---

[20]Due to space constraints, we include only a short empirical illustration. For a detailed empirical analysis of the effects of UI benefits on labor market outcomes in Austria, see Card et al. (2015a).
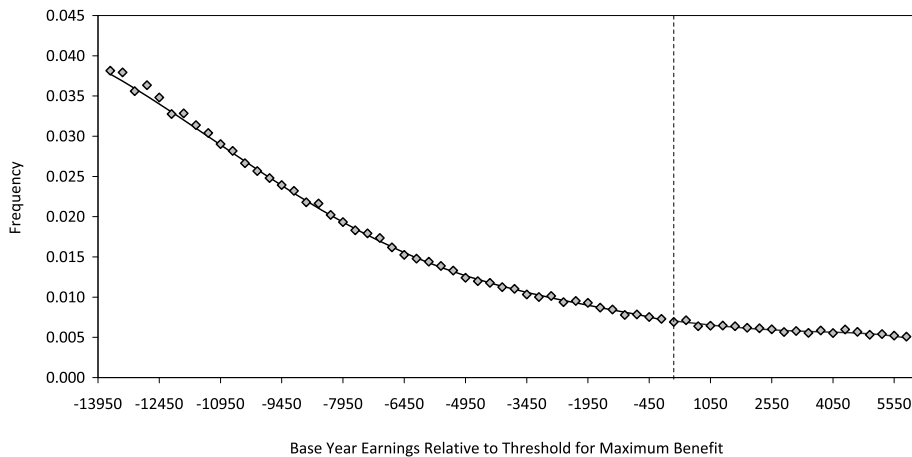
FIGURE 1.—Frequency distribution of base year earnings. *Note*: Figure shows estimated and predicted frequency distributions, using 300 Euro bins. Predicted frequencies are from a fourth order polynomial model with unrestricted first and higher-order derivatives on each side of the threshold. T-test statistic for change in derivative at the threshold is 1.61.

A key assumption for valid inference in an RK design is that the density of the assignment variable is smooth at the kink point. Figure 1 shows the frequency distribution of base year earnings around the threshold for maximum benefits using 300-Euro bins with an average of 4200 observations per bin. While the histogram looks quite smooth, we test this more formally by fitting a series of polynomial models that allow the first- and higher-order derivatives of the binned density function to jump at the kink point. This test confirms the smoothness of the density.[21]

4.2. *Graphical Overview and Estimation Results*

Figure 2 shows the relationships between base year earnings and actual UI benefits around the kink. We plot the data using the same bin sizes as in Figure 1.[22] The figure shows a clear kink in the empirical relationship between average benefits and base year earnings, with a sharp decrease in the slope

---

[21]We fit a series of polynomial models of different orders by minimum chi-squared, imposing continuity but allowing the first derivative and all higher-order derivatives to vary at the threshold. An Akaike criterion selects a fourth-order polynomial model, which has an overall goodness of fit statistic of 75.6 (p-value = 0.05). The estimated change in the first derivative of the density function at the threshold is 2.00×10^{-4}, with a standard error of 1.24×10^{-4}.

[22]See Calonico, Cattaneo, and Titiunik (2015) for nonparametric procedures for picking the bin size in RD-type plots.
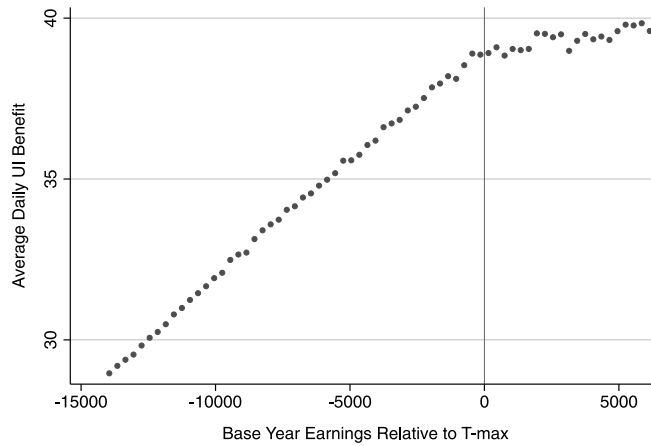
FIGURE 2.—Daily UI benefits.

as they pass through the threshold $T^{\max}$.[23] Figure 3 presents the parallel picture for mean unemployment durations, which also shows a discernible kink, though there is clearly more variability in the relationship with base year earnings.[24]
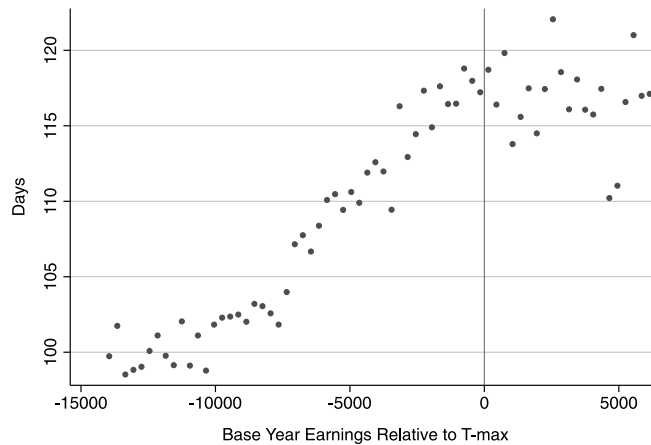

FIGURE 3.—Unemployment duration.

[23]The slope in the mean benefit function to the right of the threshold for the maximum benefit is attributable to family allowances, which are added to the base benefit amount (and are not capped). Moving right from the threshold, the average number of allowances is rising, reflecting larger family sizes for higher-earning claimants.

[24]For additional graphical analyses and robustness checks, see Card et al. (2015a).

4.2.1. *Fuzzy RKD Estimates and Comparison of Alternative Estimators*

In Table I, we present fuzzy RKD estimates of the elasticity of the unemployment duration with respect to the level of UI benefits along with first-stage estimates. We present estimates from local linear models in columns 1 and 2, and estimates from local quadratic models in columns 3 and 4. We present results under five alternative bandwidth selection procedures: default CCT, CCT

TABLE I

IV (Fuzzy Kink) Estimates of Benefit Elasticity,
Alternative Estimators and Bandwidths[a]

	Local Linear		Local Quadratic	
	First Stage (Coeff. $\times 10^5$) (1)	Struct. Model (2)	First Stage (Coeff. $\times 10^5$) (3)	Struct. Model (4)
A. Default CCT (*with regularization*)				
Main bandwidth (pilot)	1460 (2976)		2033 (3160)	
Estimated kink	−1.5	0.1	−3.1	−2.2
(conventional std error)	(0.6)	(2.4)	(1.4)	(3.0)
CCT robust confidence interval	[−3.4, 0.2]	[−8.4, 6.0]	[−8.5, −0.7]	[−10.1, 6.5]
B. CCT with no regularization				
Main bandwidth (pilot)	2947 (4898)		5058 (3870)	
Estimated kink	−1.1	2.7	−0.9	0.1
(conventional std error)	(0.2)	(1.2)	(0.4)	(2.4)
CCT robust confidence interval	[−1.4, 0.2]	[−3.9, 5.6]	[−5.0, 3.3]	[−28.7, 31.2]
C. Fuzzy CCT (*no regularization*)				
Main bandwidth (pilot)	4091 (7162)		3095 (3112)	
Estimated kink	−1.4	1.9	−0.6	−2.4
(conventional std error)	(0.1)	(0.6)	(0.8)	(7.7)
CCT robust confidence interval	[−1.8, −0.8]	[0.1, 5.1]	[−7.9, −0.5]	[−34.6, 47.4]
D. Fuzzy IK (*no regularization*)				
Main bandwidth (pilot)	5411 (5210)		4990 (5070)	
Estimated kink	−1.4	1.7	−0.8	−0.8
(conventional std error)	(0.1)	(0.4)	(0.4)	(2.8)
CCT robust confidence interval	[−1.7, −0.3]	[−2.0, 4.6]	[−3.2, 1.1]	[−17.3, 11.8]
E. FG (*no regularization*)				
Main bandwidth (pilot)	8959 (8482)		9365 (13,276)	
Estimated kink	−1.5	1.4	−1.3	2.1
(conventional std error)	(0.1)	(0.2)	(0.2)	(1.0)
CCT robust confidence interval	[−1.9, −0.9]	[−0.6, 3.4]	[−1.9, 0.3]	[−6.1, 5.0]

[a]See text for description of estimation methods.

with no regularization, Fuzzy CCT, Fuzzy IK, and FG. (Expressions for the Fuzzy CCT and IK bandwidths are given in Section B.2 of the Supplemental Material.) For each bandwidth selector, we present the value of the main and pilot bandwidth (the pilot bandwidth is for bias estimation in constructing the CCT robust confidence interval), the uncorrected first stage coefficient and structural elasticity (with their associated standard errors), and the CCT robust 95% confidence intervals for the bias corrected first stage coefficient and structural elasticity.[25]

Despite the strong visual evidence of the kink in the benefit formula in Figure 2, an examination of the estimated "first-stage" kinks in Table I suggests that not all the procedures yield statistically significant kink estimates. In particular, the default CCT bandwidth selector (panel A) chooses relatively small bandwidths for the local linear model and yields only a marginally significant estimate ($t = 2.5$). The corresponding bias-corrected kink estimate is even less precise, and its confidence interval, shown in square brackets, includes 0. The default CCT procedure chooses a somewhat larger bandwidth for the local quadratic model, but this is offset by the difficulty of precisely estimating the slopes on either side of the kink point once the quadratic terms are included, resulting in a first-stage kink estimate that is only marginally significant.

As shown in panel B, use of the CCT selector without regularization yields substantially larger bandwidths than when the regularization term is included. These larger bandwidths lead to some gain in precision, but allowance for the bias-correction term again yields CCT robust confidence intervals for the first-stage model that include 0. By comparison, the bandwidths selected by the Fuzzy CCT, Fuzzy IK, and the FG procedures (panels C, D, and E) are relatively large and deliver relatively stable and significant first-stage estimates in the local linear case even under the bias correction. The local quadratic estimates are still much less precise, however. The bias-corrected estimates are, in all but one case, insignificant.

Turning to the elasticity estimates, we observe that they are generally less precisely estimated than the first-stage coefficients. In the local linear case, estimates from procedures selecting small bandwidths, such as the default CCT estimator and the CCT estimator without regularization, are very unstable and have large standard errors. Estimates that are based on larger selected bandwidths yield more stable point estimates of the elasticity, ranging from 1.4 to 1.9. The CCT robust confidence intervals indicate a large degree of uncertainty, however, with only one of three estimates being statistically significant.

Figure 4 visualizes the structural estimates in the local linear case for a range of different bandwidths along with estimated confidence intervals based on the conventional standard errors. This graph highlights that point estimates stabilize for bandwidths in the range from 4000 to 9000, which are selected by

---

[25]CCT robust confidence intervals and the CCT bandwidths are obtained based on a variant of the Stata package described in Calonico, Cattaneo, and Titiunik (2014) with the nearest neighbor variance estimator. Using the CCT Stata package generates very similar empirical results.
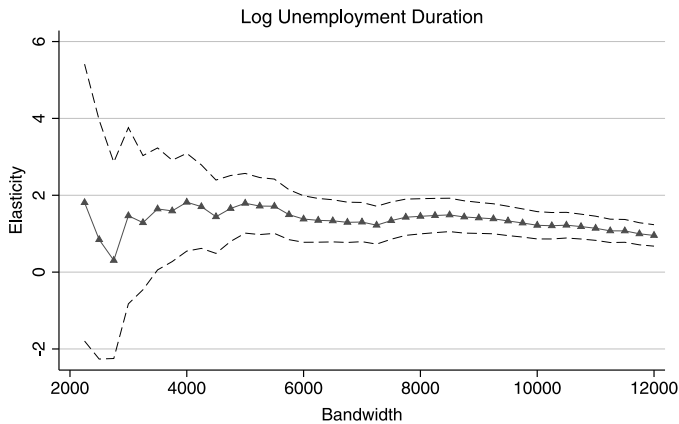
FIGURE 4.—Fuzzy estimation with varying bandwidth. *Notes*: local linear estimation, estimated elasticity (solid line with triangles) with conventional confidence bounds (dash).

the Fuzzy CCT, Fuzzy IK, and the FG procedures. But they are highly volatile at smaller bandwidth levels. The local quadratic elasticity estimates, although based on larger bandwidths, are much less precise. We get negative point estimates for the elasticity in three out of five cases, which renders the local quadratic procedures uninformative.

Overall, the pattern of estimates in Table I points to three main conclusions. First, many of the bandwidth selectors choose relatively small bandwidths that lead to relatively imprecise first-stage and structural coefficient estimates. A second observation is that the local quadratic estimators are generally quite noisy. Third, the bias-corrected estimates from the local linear models are typically not too different from the uncorrected estimates, but the added imprecision associated with uncertainty about the magnitude of the bias-correction factor is large, leading to relatively wide confidence intervals for the bias-corrected estimates.

5. CONCLUSION

In many institutional settings, a key policy variable (like unemployment benefits or public pensions) is set by a deterministic formula that depends on an endogenous assignment variable (like previous earnings). Conventional approaches to causal inference, which rely on the existence of an instrumental variable that is correlated with the covariate of interest but independent of underlying errors in the outcome, will not work in these settings. When the policy function is continuous but kinked (i.e., non-differentiable) at a known threshold, a regression kink design provides a potential way forward (Guryan (2001), Nielsen, Sørensen, and Taber (2010), Simonsen, Skipper, and Skipper (2015)). The sharp RKD estimand is simply the ratio of the estimated kink

in the relationship between the assignment variable and the outcome of interest at the threshold point, divided by the corresponding kink in the policy function. In settings where there is incomplete compliance with the policy rule (or measurement error in the actual assignment variable), a fuzzy RKD replaces the denominator of the RKD estimand with the estimated kink in the relationship between the assignment variable and the policy variable.

In this paper, we provide sufficient conditions for a sharp and fuzzy RKD to identify interpretable causal effects in a general nonseparable model (e.g., Blundell and Powell (2003)). A key assumption is that the conditional density of the assignment variable, given the unobserved error in the outcome, is continuously differentiable at the kink point. This smooth density condition rules out situations where the value of the assignment variable can be precisely manipulated, while allowing the assignment variable to be correlated with the latent errors in the outcome. Thus, extreme forms of "bunching" predicted by certain behavioral models (e.g., Saez (2010)) violate the smooth density condition, whereas similar models with errors in optimization (e.g., Chetty (2010)) are potentially consistent with an RKD approach. In addition to yielding a testable smoothness prediction for the observed distribution of the assignment variable, we show that the smooth density condition also implies that the conditional distributions of any predetermined covariates will be smooth functions of the assignment variable at the kink point. These two predictions are very similar in spirit to the predictions for the density of the assignment variable and the distribution of predetermined covariates in a regression discontinuity design (Lee (2008)).

We also provide a precise characterization of the treatment effects identified by a sharp or fuzzy RKD. The sharp RKD identifies a weighted average of marginal effects, where the weight for a given unit reflects the relative probability of having a value of the assignment variable close to the kink point. Under an additional monotonicity assumption, we show that the fuzzy RKD identifies a slightly more complex weighted average of marginal effects, where the weight also incorporates the relative size of the kink induced in the actual value of the policy variable for that unit.

We illustrate the use of a fuzzy RKD approach by studying the effect of unemployment benefits on the duration of registered unemployment in Austria, where the benefit schedule has a kink at the maximum benefit level. We present simple graphical evidence showing that this kink induces a kink in the duration of unemployment. We also present a test of the smooth density assumption around the maximum benefit threshold. Finally, we report a range of estimates of the behavioral effect of higher benefits on unemployment durations by using alternative local nonparametric estimators.

REFERENCES

ALTONJI, J. G., AND R. L. MATZKIN (2005): "Cross Section and Panel Data Estimators for Non-separable Models With Endogenous Regressors," *Econometrica*, 73, 1053–1102. [2454,2456, 2462]

BAILY, M. N. (1978): "Some Aspects of Optimal Unemployment Insurance," *Journal of Public Economics*, 10, 379–402. [2473]

BLUNDELL, R., AND J. L. POWELL (2003): "Endogeneity in Nonparametric and Semiparametric Regression Models," in *Advances in Economics and Econometrics Theory and Applications: Eighth World Congress*, Vol. II, ed. by M. Dewatripont, L. P. Hansen and S. J. Turnovsky. Econometric Society Monographs, Vol. 36. Cambridge: Cambridge University Press, 312–357. [2453,2480]

CALONICO, S., M. D. CATTANEO, AND R. TITIUNIK (2014): "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82, 2295–2326. [2455, 2471-2473,2478]

——— (2015): "Optimal Data-Driven Regression Discontinuity Plots," *Journal of the American Statisitcal Association* (forthcoming). [2475]

CARD, D., D. S. LEE, Z. PEI, AND A. WEBER (2012): "Nonlinear Policy Rules and the Identification and Estimation of Causal Effects in a Generalized Regression Kink Design," Working Paper 18564, NBER. [2460,2471,2473]

——— (2014): "Local Polynomial Order in Regression Discontinuity Designs." Unpublished Manuscript. [2471,2472]

——— (2015a): "Inference on Causal Effects in a Generalized Regression Kink Design," Technical Report 15-218, Upjohn Institute Working Paper. [2466,2473,2474,2476]

——— (2015b): "Supplement to 'Inference on Causal Effects in a Generalized Regression Kink Design'," *Econometrica Supplemental Material*, 83, http://dx.doi.org/10.3982/ECTA11224. [2461]

CHENG, M.-Y., J. FAN, AND J. S. MARRON (1997): "On Automatic Boundary Corrections," *The Annals of Statistics*, 25, 1691–1708. [2472]

CHESHER, A. (2003): "Identification in Nonseparable Models," *Econometrica*, 71, 1405–1441. [2453,2456]

CHETTY, R. (2010): "Moral Hazard versus Liquidity and Optimal Unemployment Insurance," *Journal of Political Economy*, 116, 173–234. [2480]

——— (2012): "Bounds on Elasticities With Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," *Econometrica*, 80, 969–1018. [2454]

CLASSEN, K. P. (1977): "The Effect of Unemployment Insurance on the Duration of Unemployment and Subsequent Earnings," *Industrial and Labor Relations Review*, 30 (8), 438–444. [2454]

DAHLBERG, M., E. MORK, J. RATTSO, AND H. AGREN (2008): "Using a Discontinuous Grant Rule to Identify the Effect of Grants on Local Taxes and Spending," *Journal of Public Economics*, 92, 2320–2335. [2454,2457]

DINARDO, J. E., AND D. S. LEE (2011): "Program Evaluation and Research Designs," in *Handbook of Labor Economics*, Vol. 4A, ed. by O. Ashenfelter and D. Card. Amsterdam: Elsevier. [2454]

DONG, Y. (2013): "Regression Discontinuity Without the Discontinuity," Technical Report, University of California Irvine. [2457,2458,2467]

DONG, Y., AND A. LEWBEL (2014): "Identifying the Effect of Changing the Policy Threshold in Regression Discontinuity Models." Unpublished Manuscript. [2457]

FAN, J., AND I. GIJBELS (1996): *Local Polynomial Modelling and Its Applications*. London: Chapman & Hall. [2455,2471]

FLORENS, J.-P., J. J. HECKMAN, C. MEGHIR, AND E. J. VYTLACIL (2008): "Identification of Treatment Effects Using Control Functions in Models With Continuous, Endogenous Treatment and Heterogeneous Effects," *Econometrica*, 76, 1191–1206. [2453,2454,2456,2462]

GANONG, P., AND S. JÄGER (2014): "A Permutation Test and Estimation Alternatives for the Regression Kink Design." Unpublished Manuscript. [2473]

GURYAN, J. (2001): "Does Money Matter? Regression-Discontinuity Estimates From Education Finance Reform in Massachusetts," Working Paper 8269, National Bureau of Economic Research. [2454,2457,2479]

HAHN, J., P. TODD, AND W. V. DER KLAAUW (2001): "Identification and Estimation of Treatment Effects With a Regression-Discontinuity Design," *Econometrica*, 69, 201–209. [2458,2459,2464, 2471]

HECKMAN, J. J., AND E. J. VYTLACIL (2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73, 669–738. [2455]

IMBENS, G. W., AND J. D. ANGRIST (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467–475. [2454,2466]

IMBENS, G. W., AND K. KALYANARAMAN (2012): "Optimal Bandwidth Choice for the Regression Discontinuity Estimator," *Review of Economic Studies*, 79, 933–959. [2455,2472]

IMBENS, G. W., AND T. LEMIEUX (2008): "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 142, 615–635. [2472]

IMBENS, G. W., AND W. K. NEWEY (2009): "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 77, 1481–1512. [2453,2455,2456]

LEE, D. S. (2008): "Randomized Experiments From Non-Random Selection in U.S. House Elections," *Journal of Econometrics*, 142, 675–697. [2453,2459,2462,2470,2480]

LEE, D. S., AND T. LEMIEUX (2010): "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, 48, 281–355. [2454]

LEWBEL, A. (1998): "Semiparametric Latent Variable Model Estimation With Endogenous or Mismeasured Regressors," *Econometrica*, 66, 105–121. [2453]

——— (2000): "Semiparametric Qualitative Response Model Estimation With Unknown Heteroscedasticity or Instrumental Variables," *Journal of Econometrics*, 97, 145–177. [2453]

MCCRARY, J. (2008): "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," *Journal of Econometrics*, 142, 698–714. [2469]

MEYER, B. D. (1990): "Unemployment Insurance and Unemployment Spells," *Econometrica*, 58, 757–782. [2458]

NIELSEN, H. S., T. SØRENSEN, AND C. R. TABER (2010): "Estimating the Effect of Student Aid on College Enrollment: Evidence From a Government Grant Policy Reform," *American Economic Journal: Economic Policy*, 2, 185–215. [2454,2457,2479]

ROUSSAS, G. G. (2004): *An Introduction to Measure-Theoretic Probability*. Burlington, MA: Academic Press. [2461]

RUPPERT, D., AND M. P. WAND (1994): "Multivariate Locally Weighted Least Squares Regression," *The Annals of Statistics*, 22, 1346–1370. [2471]

SAEZ, E. (2010): "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy*, 2, 180–212. [2480]

SIMONSEN, M., L. SKIPPER, AND N. SKIPPER (2015): "Price Sensitivity of Demand for Prescription Drugs: Exploiting a Regression Kink Design," *Journal of Applied Econometrics* (forthcoming). [2454,2457,2479]

THISTLETHWAITE, D. L., AND D. T. CAMPBELL (1960): "Regression-Discontinuity Analysis: An Alternative to the ex-post facto Experiment," *Journal of Educational Psychology*, 51, 309–317. [2457]

TURNER, L. J. (2013): "The Road to Pell Is Paved With Good Intentions: The Economic Incidence of Federal Student Grant Aid," Technical Report, University of Maryland. [2464]

VYTLACIL, E. J. (2002): "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 70, 331–341. [2454]

WELCH, F. (1977): "What Have We Learned From Empirical Studies of Unemployment Insurance?" *Industrial and Labor Relations Review*, 30, 451–461. [2454]

ZWEIMÜLLER, J., R. WINTER-EBMER, R. LALIVE, A. KUHN, J.-P. WUELLRICH, O. RUF, AND S. BÜCHI (2009): "Austrian Social Security Database, University of Zurich," Working Paper iewwp410, Institute for Empirical Research in Economics. [2474]

Dept. of Economics, UC Berkeley, 549 Evans Hall #3880, Berkeley, CA 94720-3880, U.S.A., NBER, and IZA; card@berkeley.edu,

Dept. of Economics, Princeton University, 3 Nassau Hall, Princeton, NJ 08544, U.S.A. and NBER; davidlee@princeton.edu,

Dept. of Policy Analysis and Management, Cornell University, 134 MVR Hall, Ithaca, NY 14853, U.S.A.; zhuan.pei@cornell.edu,

and

Dept. of Economics, University of Mannheim, L 7, 3-5, D-68131 Mannheim, Germany and IZA; a.weber@uni-mannheim.de.