labor&welfareSTATE

Nonlinear Policy Rules and the Identification and Estimation of Causal Effects in a Generalized Regression Kink Design

by

David CARD
David S. LEE
Zhuan PEI
Andrea WEBER

Working Paper No. 1214

November 2012

Corresponding author: a.weber@uni-mannheim.de
Phone: +49 (0)621 181 1928

Nonlinear Policy Rules and the Identification and Estimation of Causal Effects in a Generalized Regression Kink Design[1]

David Card
UC Berkeley, NBER and IZA

Zhuan Pei
W. E. Upjohn Institute

David S. Lee
Princeton University and NBER

Andrea Weber
University of Mannheim and IZA

November 2012

Abstract

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. This design arises in many institutional settings where a policy variable (such as weekly unemployment benefits) is determined by an observed but potentially endogenous assignment variable (like previous earnings). We provide new results on identification and estimation for these settings, and apply our results to obtain estimates of the elasticity of joblessness with respect to UI benefit rates. We characterize a broad class of models in which a "Regression Kink Design" (RKD, or RK Design) provides valid inferences for the treatment-on-the-treated parameter (Florens et al. (2008)) that would be identified in an ideal randomized experiment. We show that the smooth density condition that is sufficient for identification rules out extreme sorting around the kink, but is compatible with less severe forms of endogeneity. It also places testable restrictions on the distribution of predetermined covariates around the kink point. We introduce a generalization of the RKD – the "fuzzy regression kink design" – that allows for omitted variables in the assignment rule, as well as certain types of measurement errors in the observed values of the assignment variable and the policy variable. We also show how standard local polynomial regression techniques can be adapted to obtain nonparametric estimates for the sharp and fuzzy RKD. We then use a fuzzy RKD approach to study the effect of unemployment insurance benefits on the duration of joblessness in Austria, where the benefit schedule has kinks at the minimum and maximum benefit level. Our estimates suggest that the elasticity of joblessness with respect to the benefit rate is on the order of 1.5.

---

1 Introduction

There is now an extensive literature that generalizes the additive/linear structure of first-generation econometric models to accommodate nonseparable models (including duration models, censored regression models, and discrete choice models) within semiparametric (e.g., Han (1987); Sherman (1993); Cavanagh and Sherman (1998); Khan and Tamer (2007)) and nonparametric frameworks (e.g., Roehrig (1988); Matzkin (1991); Matzkin (1992); Matzkin (2003)).[1] In particular, a growing body of research considers identification and estimation in a nonseparable model when the continuous regressors of interest are endogenous (e.g., Lewbel (1998); Lewbel (2000); Blundell and Powell (2003); Chesher (2003); Altonji and Matzkin (2005); Florens et al. (2008); Imbens and Newey (2009)). So far, existing approaches rely on the existence of one or more instrumental variables that are assumed to be independent of the errors in the model. Unfortunately, in many applied contexts it is difficult to find candidate instruments that satisfy the necessary independence assumptions. The problem is particularly acute when the regressor of interest is a policy variable that is mechanically determined by an endogenous assignment variable. The level of unemployment benefits, for example, is typically set by a formula that depends on previous earnings. In this situation any instrument that is correlated with benefits is likely to be correlated with the unobserved determinants of previous earnings, and hence also correlated with unobserved determinants of unemployment duration – a violation of independence.

Nevertheless, a common feature of many policy rules is the presence of a kink, or multiple kinks, in the formula that relates the assignment variable to the policy variable. As has been noted by Guryan (2001), Dahlberg et al. (2008), Nielsen et al. (2010) and Simonsen et al. (2011), a kinked assignment rule holds out the possibility for identification of the policy variable effect, even in the absence of traditional instruments. The basic 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. An obvious concern with this "regression kink design" (RKD) is the endogeneity of the assignment variable. Saez (2010), for example, notes that a kink in the tax schedule will typically cause taxpayers to "bunch" at the kink. Such behavior could lead to a non-smooth distribution of unobserved heterogeneity around the kink point, confounding inferences based on a regression kink design.[2]

This paper establishes conditions under which the behavioral response to a formulaic policy variable

---

[1] See Blundell and Powell (2003) and Matzkin (2007) for literature surveys.

[2] A similar concern is often raised in regression discontinuity designs – see Urquiola and Verhoogen (2009), for example.

like unemployment benefits can be identified within a general class of nonparametric and nonseparable regression models. Specifically, we establish conditions for the RKD to identify the same "local average response" (Altonji and Matzkin (2005)) or "treatment-on-the-treated" parameter (Florens et al. (2008)) that would be identified in a randomized experiment. The key assumption is that 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. We show that this smooth density condition rules out deterministic sorting, while allowing for other 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 strong 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 is testable.

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 thus 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.[3]

We then discuss nonparametric estimation of the RKD and fuzzy RKD, using local polynomial estimation (Fan and Gijbels (1992)). Specifically, under a set of regularity conditions we establish consistency and asymptotic normality of local linear and local quadratic estimators for the sharp and fuzzy kink designs, and provide conditions under which nonparametric inference can be conducted using estimates and sampling errors from standard regression routines.

Finally, we use a fuzzy RKD approach to analyze the effect of unemployment insurance (UI) benefits on the duration of benefit claims and joblessness in Austria. As in the U.S., the Austrian UI system specifies

---

[3]The "marginal effect" to which we refer in this paper – the derivative of the outcome with respect to the continuous endogenous regressor – should not be confused with the marginal treatment effect (MTE) of Heckman and Vytlacil (2005), where the treatment is binary.

a benefit level that is proportional to earnings in a "base period" prior to job loss, subject to a minimum and maximum. We study the effects of the kinks at the minimum and maximum benefit levels, using data on a large sample of unemployment spells from the Austrian Social Security Database (see Zweimüller et al. (2009)). An important advantage of these data is that we can measure both the duration of UI benefit claims and the duration of joblessness (i.e., the length of time from the end of the old job and the start of the new job).[4] Simple plots of the data show strong visual evidence of kinks in the relationship between base period earnings and the durations of benefit claims and joblessness at both kink points. We also examine the relationship between base period earnings and various predetermined covariates (such as gender, occupation, age, and industry) around the kink points. The conditional distributions of the covariates evolve smoothly around the "bottom kink" (at the earnings threshold for minimum benefits) but are less smooth around the "top kink" (at the threshold for maximum benefits).

Our fuzzy RKD estimates of the elasticity of the duration of claims and the duration of joblessness are relatively large and reasonably precise, particularly around the bottom kink, where the density of observations is higher. In particular, the estimated elasticity of joblessness using a local linear regression is 1.73 (standard error = 0.44), while the estimated elasticity of the duration of benefit claims is 1.25 (standard error = 0.41). Both estimates are robust to wide variation in the bandwidth choice. The estimated elasticities around the top kink are more sensitive to alternative bandwidth choices, but generally comparable in magnitude. Overall, we conclude that increases in UI benefits in Austria appear to exert a relatively large effect on both low-wage and high-wage individuals, with behavioral elasticities that are somewhat larger than have been found in recent U.S. studies using difference-in-difference designs (e.g., Chetty (2010)).

2   Nonparametric Regression and the Regression Kink Design

2.1   Background

Consider the generalized nonseparable model

$$Y = y(B, V, U) \tag{1}$$

---

[4]Card et al. (2007) point out that it may be important to distinguish between the impacts of UI policy variables on the duration of the UI claim versus impacts on the duration of joblessness.

where Y is an outcome, B is a continuous regressor of interest, V is another observed covariate, and U is an error term that potentially enters the model in a non-additive 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 is understood in the literature, this formulation allows for completely unrestricted heterogeneity in the responsiveness of Y to B. In the case where B is binary, the model is equivalent to a potential outcomes framework where the "treatment effect" of B for a particular individual is given by $Y_1 - Y_0 = y(1, V, U) - y(0, V, U)$. Note that in contexts with discrete outcomes, Y could be defined as an individual-specific probability of a particular outcome (as in a binary response model) or an individual-specific expected value (e.g. an expected duration) that depends on B, V, and U, where the structural function of interest is the relation between B and the probability or expected value.[5]

One natural benchmark object of interest in this setting is the "average structural function" (ASF), as discussed in Blundell and Powell (2003):

$$ASF(b, v) = \int y(b, v, u) \, dF_U(u)$$

where $F_U(\cdot)$ is the c.d.f. of U. This gives the average value of Y that would occur if the entire population (as represented by the unconditional distribution of U) were assigned to a particular value of the pair (b, v). Florens et al. (2008) call the derivative of the ASF with respect to the continuous treatment of interest the "average treatment effect" (ATE), which is a natural extension of the average treatment effect familiar in the binary treatment context.

A closely related construct is the "treatment on the treated" (TT) parameter of Florens et al. (2008):

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

where $F_{U|B=b, V=v}(u)$ is the c.d.f. of U conditional on B, V equal to b, 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, $F_{U|B=b, V=v}(\cdot)$.

---

[5]In these cases, one would use the observed outcome Y^* (0 or 1 outcome, or an observed duration), and use the fact that the expectation of Y^* and Y would be equivalent given the same conditioning statement, in applying all of the identification results below.

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, at the same time, independent of the non-additive errors in the model. Chesher (2003) observes that such independence assumptions may be "strong and unpalatable", and hence proposes the use of local independence of Z to identify local effects.

As mentioned in the introduction, there are some important contexts – particularly when the regressor of interest is a deterministic function of another behaviorally endogenous variable – where no instruments can plausibly satisfy the independence assumption, either globally or locally. For example, consider the case where Y represents the 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 is a linear function of pre-job-loss earnings up to some maximum: $B = b(V) = \rho \min(V, T)$. Conditional on V there is no variation in the benefit level, so model (1) is not nonparametrically identified. One could try to get around this fundamental non-identification 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.

Despite this conclusion, it may be possible to exploit the kink in the benefit rule to identify the causal effect of B on Y, in a similar spirit to the regression discontinuity design of Thistlethwaite and Campbell (1960). 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 $v = T$.[6] This identification strategy has been employed in a few empirical studies. Guryan (2001), for example, uses kinks in state education aid formulas as part of an instrumental variables strategy to study the effect of public school spending.[7] Dahlberg et al. (2008) use the same approach to estimate the impact of intergovernmental grants on local spending and taxes. More recently, Simonsen et al. (2011) use a kinked relationship between total expenditure on prescription drugs and their marginal price to study the price sensitivity of demand for prescription drugs. Nielsen et al. (2010), who introduce the term "Regression

---

[6]Without loss of generality, we normalize the kink threshold T to 0 in the remainder of the paper.

[7]Guryan (2001) describes 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."

Kink Design" for this approach, use a kinked student aid scheme to identify the effect of direct costs on college enrollment.

Nielsen et al. (2010) make precise the assumptions needed to identify the causal effects in the constant-effect, additive model

$$Y = \tau B + g(V) + \varepsilon, \tag{2}$$

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

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

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.[8]

Below we provide the following new identification results. First, we establish conditions for identification of causal effects using the RKD for the most general regression model possible – namely, the nonseparable model in (1). By allowing the error term to enter nonseparably, we are allowing for 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 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). The addition of multiplicative unobservable heterogeneity (see Meyer (1990)) in the baseline hazard poses an even greater challenge to the justification of parametric specifications such as (2). The nonseparable model (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 that has been discussed in the nonseparable regression literature. Given that the identified effect is an average of marginal effects across a heterogeneous population, we also make precise

---

[8]In an earlier working paper version, Nielsen et al. (2010) provide similar conditions for identification for a less restrictive, additive model, $Y = g(B,V) + \varepsilon$.

how the RKD estimand implicitly weights these heterogeneous marginal effects. The weights are intuitive, and correspond to the weights that would be obtained using data generated by 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 (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, or measurement errors in V or B. This "fuzzy RKD" generalization may have broader applicability than the "Sharp RKD".[9]

Finally, we provide testable implications for the 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.[10] 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 same $TT_{b|v}$ parameter that is identified by a randomized experiment.

Sharp RK Design: Let (V, U) be a pair of random variables (with V observable, U unobservable). To emphasize that the distribution of U can be arbitrary and need not be continuous or smooth in any way, and also to ease exposition in the proofs, let U be discrete with an arbitrary number of points of support.[11] Denote the c.d.f. and p.d.f. of V conditional on $U = u$ by $F_{V|U=u}(v)$ and $f_{V|U=u}(v)$. Let $B \equiv b(V)$, let $Y \equiv y(B, V, U)$. Also, define $y_1(b, v, u) \equiv \frac{\partial y(b,v,u)}{\partial b}$ and $y_2(b, v, u) \equiv \frac{\partial y(b,v,u)}{\partial v}$.

Assumption 1. (Regularity) $y(\cdot, \cdot, \cdot)$ is a continuous function with $y_1(b, v, u)$ continuous in b for all b, v and u.

Assumption 2. (Smooth effect of V) $y_2(b, v, u)$ is continuous in v for all b, v and u.

Assumption 3. (First stage) $b(\cdot)$ is a known function, everywhere continuous and continuously differ-

---

[9]The sharp/fuzzy distinction in the RKD is analogous to that for the RD Design (see Hahn et al. (2001)).

[10]Lee (2008) shows that a similar condition is necessary in a regression discontinuity design.

[11]That is, similar results could be derived for continuously distributed U.

7

entiable on $(-\infty, 0)$ and $(0, \infty)$, but $\lim_{v \to 0^+} b'(v) \neq \lim_{v \to 0^-} b'(v)$. In addition, $f_{V|U=u}(0)$ is strictly positive for all $u \in A$, where $\sum_{u \in A} \Pr(U = u) > 0$.

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

Assumption 1 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 would dictate that V not enter as an argument, because here V is allowed to affect Y, as long as its marginal effect is continuous. Assumption 3 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 $V = 0$. It is additionally necessary to assume a positive density of V around the threshold for a non-trivial subpopulation.

Assumption 4 is the key identifying assumption for a valid RK design. As in Lee (2008), this condition rules out precise manipulation of the assignment variable. But whereas continuity of $f_{V|U=u}(v)$ in v is sufficient for identification, 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 Subsection 4.1 below we discuss 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) $\Pr(U \leqslant u | V = v)$ is continuously differentiable in v at $v = 0 \ \forall u$.

$$(b) \ \frac{\lim_{v_0 \to 0^+} \frac{dE[Y|V=v]}{dv}\big|_{v=v_0} - \lim_{v_0 \to 0^-} \frac{dE[Y|V=v]}{dv}\big|_{v=v_0}}{\lim_{v_0 \to 0^+} \frac{db(v)}{dv}\big|_{v=v_0} - \lim_{v_0 \to 0^-} \frac{db(v)}{dv}\big|_{v=v_0}} = E[y_1(b_0, 0, U)|V=0] = \sum_u y_1(b_0, 0, u) \frac{f_{V|U=u}(0)}{f_V(0)} \Pr(U = u) = TT_{b_0|0}$$

where $b_0 = b(0)$.

The proof is in the Supplemental Appendix.

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 $V = 0$.[12] 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 $TT_{b_0|0}$.[13]

---

[12]Note also that (a) implies Proposition 2(a) in Lee (2008), i.e., the continuity of $\Pr(U \leqslant u | V = v)$ 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.

[13]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

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 units with $V = 0$. This would yield

$$
\begin{aligned}
\left. \frac{\partial E[Y|B = b, V = 0]}{\partial b} \right|_{b=b_0} &= \left. \frac{\partial \left(\sum_u y(b,0,u) \Pr(U = u | V = 0, B = b) \right)}{\partial b} \right|_{b=b_0} \\
&= \left. \frac{\partial \left(\sum_u y(b,0,u) \frac{f_{B|V=0,U=u}(b)}{f_{B|V=0}(b)} \frac{f_{V|U=u}(0)}{f_V(0)} \Pr(U = u) \right)}{\partial b} \right|_{b=b_0} \\
&= \left. \frac{\partial \left(\sum_u y(b,0,u) \frac{f_{V|U=u}(0)}{f_V(0)} \Pr(U = u) \right)}{\partial b} \right|_{b=b_0} \\
&= \sum_u y_1(b_0,0,u) \frac{f_{V|U=u}(0)}{f_V(0)} \Pr(U = u).
\end{aligned}
$$

Even though B is randomized in this hypothetical experiment, V is not. Intuitively, even if randomization allows one to identify marginal effects of B, it cannot do anything about the fact that units with $V = 0$ will in general have a particular distribution of U. Of course, the advantage of this hypothetical randomized

---

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 *LAR* 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 $\lim_{b \uparrow b_0} TT_b$.

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 $V = 0$.

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 because the formula – while deterministic – depends on other (unknown) variables in addition to the primary assignment variable, or alternatively because of measurement errors in the available data set. This motivates the extension of fuzzy RKD.[14]

Specifically, assume now that $B = b(V, U)$, where the presence of U in the formula for B allows for unobserved determinants that are potentially correlated with the outcome variable. In addition, assume that the observed values of B and V, B^* and V^* respectively, differ from their true values as follows:

$$B^* \equiv B + U_B; \qquad V^* \equiv V + U_V$$

$$U_V \equiv G \cdot U_{V'},$$

where U_B and $U_{V'}$ are continuously distributed, and that their joint density conditional on U is continuous and supported on an arbitrarily large compact rectangle in R^2; G is an indicator variable that is equal to zero with probability $\pi(V, U, U_B, U_{V'})$. Note that the error in the observed value of V is assumed to be a mixture of a conventional (continuously-distributed) measurement error and a point mass at 0, so with probability $\pi > 0$ we observe the true value of V, and with probability $1 - \pi$ we observe $V + U_{V'}$. The random variables $(V, U, U_B, U_{V'}, G)$ determine (B, B^*, V^*, Y) and we observe (B^*, V^*, Y).

Assumption 3a. (First stage) For each u, $b(\cdot, u)$ is a function that is everywhere continuous and continuously differentiable on $(-\infty, 0)$ and $(0, \infty)$ with respect to the first argument. $b_1^+(u) \equiv \lim_{v \to 0^+} b_1(v, u) \neq b_1^-(u) \equiv \lim_{v \to 0^-} b_1(v, u)$ for all u in some subset A of the population, $\sum_{u \in A} \Pr(U = u) > 0$.

Assumption 4a. (Smooth density) Let $V, U_B, U_{V'}$ have a well-defined joint probability density function conditional on each u, $f_{V, U_B, U_{V'}|U=u}(v, u_B, u_{V'})$, and let $\frac{\partial f_{V, U_B, U_{V'}|U=u}(v, u_B, u_{V'})}{\partial v}$ be continuous for all $v, u_B, u_{V'}, u$.

Assumption 5. (Smooth probability of no error in V) $\frac{\partial \pi(v, u, u_B, u_{V'})}{\partial v}$ is continuous for all $v, u_B, u_{V'}, u$.

Assumption 6. (Monotonicity) Either $b_1^+(u) \geq b_1^-(u)$ for all u or $b_1^+(u) \leq b_1^-(u)$ for all u.

---

[14] See Hahn et al. (2001) for a definition of the fuzzy regression discontinuity design.

Assumption 7. (Non-negligible population at the kink) $\sum_{u \in A} \Pr[U_V = 0 | V = 0, U = u] \left(b_1^+(u) - b_1^-(u) \right) f_{V|U=u}(0) \Pr(U = u) > 0.$

Assumption 3a modifies Assumption 3: $b(\cdot, u)$ is known to have a kink for a non-neglible subset of the population, but there is heterogeneity in the magnitude of the kinks, which are unobservable at the individual level. Assumption 4a modifies Assumption 4: for each u, 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 $V, U_B, U_{V'}$ can be arbitrarily correlated. 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, further assumptions must be made about the measurement error to achieve identification (as in the case with the RD design). Assumption 5 states that the mass point probability, while potentially dependent on all other variables, is smooth with respect to V.

Assumption 6 states that the direction of the kink is either non-negative or non-positive for the entire population. This is analogous to the monotonicity condition of Imbens and Angrist (1994), but is potentially easier to justify, since it is *not* a behavioral assumption about compliance behavior, but rather a restriction on the nature of the policy formula. In particular, Assumption 6 rules out situations where some individuals experience a positive kink at $V = 0$, but others experience a negative kink at $V = 0$. We anticipate that in many contexts, specific knowledge of the policy rule will imply that this condition is met.[15]

Assumption 7 states that we must have a non-negligible 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.

Proposition 2. *In a valid Fuzzy RK Design, that is, when Assumptions 1, 2, 3a, 4a, 5, 6, and 7 hold:*

(a) $\Pr(U \leq u | V^* = v^*)$ is continuously differentiable in v^* at $v^* = 0 \; \forall u.$

$$(b) \; \frac{\lim_{v_0 \to 0^+} \left. \frac{dE[Y|V^*=v^*]}{dv^*} \right|_{v^*=v_0} - \lim_{v_0 \to 0^-} \left. \frac{dE[Y|V^*=v^*]}{dv^*} \right|_{v^*=v_0}}{\lim_{v_0 \to 0^+} \left. \frac{dE[B^*|V^*=v^*]}{dv^*} \right|_{v^*=v_0} - \lim_{v_0 \to 0^-} \left. \frac{dE[B^*|V^*=v^*]}{dv^*} \right|_{v^*=v_0}} = \sum_u y_1 \left(b(0, u), 0, u \right) \varphi(u) \Pr(U = u)$$

where $\varphi(u) = \dfrac{\Pr[U_V=0|V=0,U=u] \left(b_1^+(u) - b_1^-(u) \right) \frac{f_{V|U=u}(0)}{f_V(0)}}{\sum_\omega \Pr[U_V=0|V=0,U=\omega] \left(b_1^+(\omega) - b_1^-(\omega) \right) \frac{f_{V|U=\omega}(0)}{f_V(0)} \Pr(U=\omega)}.$

The proof is in the Supplemental Appendix.

Remark 3. The fuzzy RKD continues to estimate a weighted average of marginal effects of B on Y, but the weight is now given by $\varphi(u)$.

---

[15]For example, in our application below we have a situation where $b_1^+(u) > b_1^-(u)$ for some subset of the population, and $b_1^+(u) = b_1^-(u)$ for the remainder.

Remark 4. The weight $\varphi(u)$ has three components. The first component, $\frac{f_{V|U=u}(0)}{f_V(0)}$, is the same weight as in the sharp RKD. The second component, $(b_1^+(u) - b_0^-(u))$, reflects the size of the kink in the benefit schedule at $V = 0$ for an individual of type u. Analogous to the LATE interpretation, 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. Finally, the third component $\Pr[U_V = 0|V = 0, U = u]$ 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. Note that if π is constant across individuals then this component of the weight is just a constant.

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 RKD is that the density of V or V^* is sufficiently smooth for every individual, ruling out situations where an individual precisely manipulates V (and hence V^*). This smoothness condition cannot be true if we observe a kink in the density of either V or V^*. That is, evidence that there is "deterministic sorting" in V (or V^*) at the kink point implies a violation of the key identifying RKD assumption. This is analogous to the test of manipulation of the assignment variable for RD designs, discussed in McCrary (2008).

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

Assumption 8. There exists an observable random vector $X = x(U)$ determined prior to V. $x(\cdot)$ 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 8 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 8 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 $V^* = 0$. This test of design 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 necessary conditions for a valid RDD.

3 Nonparameteric Estimation and Inference in a Regression Kink Design

In this section, we provide the theory for estimation and inference in the RKD. In Subsection 3.1.1, following the RD literature, we define our proposed Sharp RKD (SRKD) estimator as the difference in the estimated derivative of the conditional expectation of the outcome scaled by the known magnitude of the kink in the endogenous regressor. The difference in derivatives is estimated by the difference in slopes of local polynomial regressions fit to data on either side of the kink point and evaluated at the kink point. For the Fuzzy RKD (FRKD), the magnitude of the "first stage" kink is also estimated by local regressions. In Subsection 3.1.2, we provide the regularity conditions under which the estimators are consistent and asymptotically normal, and derive their asymptotic variance. We show that, even though a local quadratic regression is expected to have smaller bias than a local linear regression under the same bandwidth sequence (see for example, Hahn et al. (2001)), the asymptotic variance for the quadratic is 16 times that for the linear specification. We further show that an alternative estimator that uses a symmetric kernel around the kink point and restricts the regression function to be continuous leads to no reduction in asymptotic variance, even when the restriction is true. Finally, in Subsection 3.2 we provide the conditions under which inference can be conducted using robust standard errors (White (1980)) reported in common statistical software packages.

3.1 Definitions and Asymptotic Properties of RKD Estimators

3.1.1 Definition of the Sharp and Fuzzy RKD Estimators

In our exposition, we follow the notation in Fan and Gijbels (1996) and Hahn et al. (2001) and define

$$m(v) = E[Y_i|V_i = v]$$

$$r(v) = E[B_i|V_i = v]$$

to be the expectation of the outcome and endogenous regressor, respectively, conditional on a given value v of the assignment variable V_i. For both the sharp and fuzzy cases, the RKD estimand is the ratio of $\lim_{v \to 0^+} m'(v) - \lim_{v \to 0^-} m'(v)$ to $\lim_{v \to 0^+} r'(v) - \lim_{v \to 0^-} r'(v)$.

We employ a local polynomial regression framework to estimate these function limits and hence the causal effects of interest. Specifically, we split the data into two subsamples – observations to the left and right of the kink point – and estimate separate local polynomial regressions for each subsample. For the sharp RK design, $r(v) = b(v)$ is a known function and we only need to solve the following least squares problems

$$\min_{\{\tilde{\beta}_j^-\}} \sum_{i=1}^{n^-} \{Y_i^- - \sum_{j=0}^{p} \tilde{\beta}_j^- (V_i^-)^j\}^2 K(\frac{V_i^-}{h}) \tag{3}$$

$$\min_{\{\tilde{\beta}_j^+\}} \sum_{i=1}^{n^+} \{Y_i^+ - \sum_{j=0}^{p} \tilde{\beta}_j^+ (V_i^+)^j\}^2 K(\frac{V_i^+}{h}) \tag{4}$$

where the $-$ and $+$ superscripts denote quantities in the regression on the left and right side of the kink point respectively, p is the order the polynomial, K the kernel, and h the bandwidth. Denote the solutions to the minimization problems of (3) and (4) by $\{\hat{\beta}_j^-\}$ and $\{\hat{\beta}_j^+\}$ respectively, and note that $j!\hat{\beta}_j^-$ and $j!\hat{\beta}_j^+$ are estimators of $\lim_{v \to 0^-} m^{(j)}(v)$ and $\lim_{v \to 0^+} m^{(j)}(v)$, the left and right limits of the j-th derivative of the function $m(v)$ at 0. Since $\kappa_1^+ = \lim_{v \to 0^+} r'(v)$ and $\kappa_1^- = \lim_{v \to 0^-} r'(v)$ are known quantities in a sharp design, the sharp RKD estimator is defined as

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

In a fuzzy RKD, the first stage relationship is no longer deterministic, and $r(\cdot)$ is an unknown function.

So we estimate κ_1^+ and κ_1^- by solving

$$\min_{\tilde{\kappa}_j^-} \sum_{i=1}^{n^-} \{B_i^- - \sum_{j=0}^{p} \tilde{\kappa}_j^- (V_i^-)^j\}^2 K(\frac{V_i^-}{h}) \tag{5}$$

$$\min_{\tilde{\kappa}_j^+} \sum_{i=1}^{n^+} \{B_i^+ - \sum_{j=0}^{p} \tilde{\kappa}_j^+ (V_i^+)^j\}^2 K(\frac{V_i^+}{h}). \tag{6}$$

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

$$\hat{\tau}_{FRKD} = \frac{\hat{\beta}_1^+ - \hat{\beta}_1^-}{\hat{\kappa}_1^+ - \hat{\kappa}_1^-}$$

where $\hat{\kappa}_1^+$ and $\hat{\kappa}_1^-$ are estimators of κ_1^+ and κ_1^- obtained from (5) and (6).

3.1.2   Asymptotic Distributions of the RKD Estimators

In this subsection, we adapt results from Fan and Gijbels (1996), Hahn et al. (1999) and Hahn et al. (2001) to derive the asymptotic properties of the local polynomial estimators. When implementing the RKD estimators 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 et al. (2001) propose local linear regression ($p = 1$) due to the large bias of a local constant fit at the boundary; this has since become the standard of the literature. Analogously, in estimating a derivative, a local quadratic ($p = 2$) approach can be expected to lead to an asymptotically smaller bias compared to a local linear ($p = 1$) fit. But as we demonstrate below, the asymptotic variance of a local quadratic regression is significantly larger than its local linear counterpart, by a factor of 16. For this reason, we provide results for both local linear and local quadratic specifications. Following Imbens and Lemieux (2008) and the common practice in the RD literature, we adopt a uniform kernel (Assumption 13 below), and establish the necessary rate of bandwidth shrinkage for valid inference. We defer the discussion of our bandwidth choice in practice to Subsection 4.5.1, and leave a full theoretical treatment of optimal bandwidth determination for future research.[16]

We make the following assumptions, which are similar to those invoked in Hahn et al. (2001) and Imbens and Lemieux (2008).[17] In the remainder of this section, we denote the left and right limit of any function

---

[16] See Imbens and Kalyanaraman (2012) for a full theoretical analysis of an optimal bandwidth for the case of the regression discontinuity design.

[17] Assumption 9 is implicit in Hahn et al. (1999). Assumptions 10L, 11L, 12, 14 and 15 explicitly mirror Assumptions 1, 2, 3, 5

$g(\cdot)$ at 0 by $g(0^-)$ and $g(0^+)$, respectively, and we use the notation $g(0^\pm)$ in a statement to mean that the statement holds true for both $g(0^+)$ and $g(0^-)$. We use $g^{(j)}(\cdot)$ to denote the j-th derivative of the function $g(\cdot)$.

Assumption 9. The triplets (Y_i, B_i, V_i), $i = 1,...,n$, are i.i.d.

Assumption 10L. $m(v)$ and $r(v)$ are twice continuously differentiable except possibly at $v = 0$. There exists an $M > 0$ such that $|m^{(j)}(v)|$ and $|r^{(j)}(v)|$ are bounded on $[-M,0)$ and $(0,M]$ for $j = 0,1,2$.

Assumption 10Q. $m(v)$ and $r(v)$ are three times continuously differentiable except possibly at $v = 0$. There exists an $M > 0$ such that $|m^{(j)}(v)|$ and $|r^{(j)}(v)|$ are bounded on $[-M,0)$ and $(0,M]$ for $j = 0,1,2,3$.

Assumption 11L. The limits $m^{(j)}(0^\pm)$ and $r^{(j)}(0^\pm)$ exist for $j = 0,1,2$.

Assumption 11Q. The limits $m^{(j)}(0^\pm)$ and $r^{(j)}(0^\pm)$ exist for $j = 0,1,2,3$.

Assumption 12. The density of V_i, $f(v)$, is continuous and bounded near 0 with $f(v) > 0$.

Assumption 13. The kernel is uniform: $K(u) = \frac{1}{2}1_{\{|u|<1\}}$.

Assumption 14. $\sigma_Y^2(v) \equiv Var(Y_i|V_i = v)$ and $\sigma_B^2(v) \equiv Var(B_i|V_i = v)$ are bounded in a neighborhood around $v = 0$. Furthermore, the limits $\sigma_Y^2(0^\pm)$, $\sigma_B^2(0^\pm)$ and $\sigma_{BY}(0^\pm) \equiv \lim_{e \to 0^+} Cov(Y_i, B_i|V_i = \pm e)$ exist.

Assumption 15. $\psi_Y(v) = E[|Y_i - m(V_i)|^3|V_i = v]$ and $\psi_B(V_i) = E[|B_i - r(V_i)|^3|V_i = v]$ are bounded in a neighborhood around 0 and the limits $\psi_Y(0^\pm)$ and $\psi_B(0^\pm)$ are well-defined.

Assumption 16L. The bandwidth sequence satisfies $h \propto n^{-\rho}$ where $\rho \in (\frac{1}{5}, \frac{1}{3})$.

Assumption 16Q. The bandwidth sequence satisfies $h \propto n^{-\rho}$ where $\rho \in (\frac{1}{7}, \frac{1}{3})$.

Proposition 3. *(a) Local Linear ($p = 1$): Under Assumptions 9, 10L, 11L, 12-15 and 16L, the asymptotic distribution of the local linear slope estimator is given by*

$$\sqrt{nh^3}\{\hat{m}'_L(0^\pm) - m'(0^\pm)\} \Rightarrow N(0, \frac{12\sigma_Y^2(0^\pm)}{f(0)}). \tag{7}$$

(b) Local Quadratic ($p = 2$): Under Assumptions 9, 10Q, 11Q, 12-15 and 16Q, the asymptotic distribution of the local quadratic slope estimator is given by

$$\sqrt{nh^3}\{\hat{m}'_Q(0^\pm) - m'(0^\pm)\} \Rightarrow N(0, \frac{192\sigma_Y^2(0^\pm)}{f(0)}). \tag{8}$$

and 6 in Hahn et al. (1999), while Assumption 13 is a specific case of their Assumption 4. We follow Imbens and Lemieux (2008) and choose the bandwidths in Assumptions 16L and 16Q to eliminate the asymptotic bias, which is different from the approach taken by Hahn et al. (2001). We maintain the identification assumptions from Section 2 in the remainder of this section without explicitly mentioning them.

The proof is in the Supplemental Appendix.

Remark 5. In a fuzzy RKD, Proposition 3 can also be applied to derive the asymptotic distribution of the first stage kink, by replacing m with r and Y with B.

Remark 6. Fan and Gijbels (1996) argue that given the same bandwidth sequence, the variance of $\hat{m}^{(v)}$ does not increase when moving from using a polynomial of order $p = v + 2q$ to that of order $p = v + 2q + 1$ (for nonnegative integers q); they suggest that it is thus "costless" to choose p such that $p - v$ is odd. In our case, the variance in (8) is 16 times that in (7), which is due to estimating $\hat{m}'$ at a boundary. Thus, in the RKD design there is a substantial "cost" in variance from using local quadratic polynomials versus local linear polynomials.[18] For the same bandwidth sequence, an improvement in mean squared error from using a quadratic would require the bias to be at least $\sqrt{15}$ times as large as the standard error in the linear specification.[19]

Noting that observations to the left and right of the kink point are independent, we can derive the asymptotic distribution of the sharp RKD estimator from Proposition 3.

Proposition 4. *(a) Local Linear ($p = 1$): Under Assumptions 9, 10L, 11L, 12-15 and 16L, the asymptotic distribution of the local linear sharp RKD estimator is given by*

$$\sqrt{nh^3}(\hat{\tau}_{SRKD}^L - \tau_{SRKD}) \Rightarrow N(0, 12 \cdot \Omega_{SRKD})$$

where $\Omega_{SRKD} = \frac{1}{(\kappa_1^+ - \kappa_1^-)^2} \frac{\sigma_Y^2(0^+) + \sigma_Y^2(0^-)}{f(0)}$.

(b) Local Quadratic ($p = 2$): Under Assumptions 9, 10Q, 11Q, 12-15 and 16Q, the asymptotic distribution of the local quadratic sharp RKD estimator is given by

$$\sqrt{nh^3}(\hat{\tau}_{SRKD}^Q - \tau_{SRKD}) \Rightarrow N(0, 192 \cdot \Omega_{SRKD}).$$

The proof is in the Supplemental Appendix.

---

[18]To see the intuition behind the variance difference between an interior and boundary point, suppose the true regression function was linear, given by $Y = \alpha_0 + \alpha_1 V + \varepsilon$ with ε homoskedastic. The asymptotic variance of the OLS coefficient on the linear term is $\frac{Var(\varepsilon)}{Var(V)}$ in a linear regression and $\frac{Var(\varepsilon)}{Var(\tilde{V})}$ in a quadratic regression, where $\tilde{V}$ is the residual from a linear projection of V on V^2. If V is symmetrically distributed around 0, then in this "interior" case, there is zero covariance between V and V^2, so $Var(V) = Var(\tilde{V})$. In a boundary case (e.g. $V \geq 0$) there will be a nonzero covariance between V and V^2 and hence $Var(\tilde{V}) < Var(V)$.

[19]In the most favorable case for the quadratic, there is zero bias in the quadratic specification. If the square of the bias is less than 15 times the variance in the linear specification, then the linear specification will have a smaller mean squared error. If the quadratic specification has nonzero bias, then an improvement would require the linear specification to be even more biased.

For the fuzzy RKD estimator, we have

Proposition 5. *(a) Local Linear ($p = 1$): Under Assumptions 9, 10L, 11L, 12-15 and 16L, the asymptotic distribution of the local linear fuzzy RKD estimator is given by*

$$\sqrt{nh^3}(\hat{\tau}_{FRKD}^L - \tau_{FRKD}) \Rightarrow N(0, 12 \cdot \Omega_{FRKD})$$

where

$$\Omega_{FRKD} = \frac{1}{f(0)} \left\{ \frac{\left(\sigma_Y^2(0^-) + \sigma_Y^2(0^+)\right)}{\tau_B^2} + \frac{\tau_Y^2\left(\sigma_B^2(0^-) + \sigma_B^2(0^+)\right)}{\tau_B^4} - \frac{2\tau_Y\left(\sigma_{BY}(0^-) + \sigma_{BY}(0^+)\right)}{\tau_B^3} \right\}$$

with $\tau_B = r'(0^+) - r'(0^-)$ and $\tau_Y = m'(0^+) - m'(0^-)$.

(b) Local Quadratic ($p = 2$): Under Assumptions 9, 10Q, 11Q, 12-15 and 16Q, the asymptotic distribution of the local quadratic fuzzy RKD estimator is given by

$$\sqrt{nh^3}(\hat{\tau}_{FRKD}^Q - \tau_{FRKD}) \Rightarrow N(0, 192 \cdot \Omega_{FRKD}).$$

The proof is in the Supplemental Appendix.

Given the identification assumptions above, one expects the conditional expectation of Y to be continuous at the threshold. One natural question is the extent to which imposing continuity in estimation (as opposed to estimating separate local polynomials on either side of the threshold) can reduce variance in the estimator of the kink. The result below indicates that when the kernel is symmetric, a constrained RKD estimator has the same asymptotic properties as the unconstrained estimators described above.

Let $\hat{\tau}_{SRKD}^{LR}$ and $\hat{\tau}_{SRKD}^{QR}$ be the sharp RKD estimator from the local linear and quadratic regressions with the restriction $m_L(0^-) = m_L(0^+)$, respectively. They can be obtained by solving the "pooled" least squares problem

$$\min_{\{\tilde{\gamma}_j^S, \tilde{\delta}_j^S\}} \sum_{i=1}^{n} \{Y_i - \sum_{j=0}^{p} \tilde{\gamma}_j^S (V_i)^j - \sum_{j=1}^{p} \tilde{\delta}_j^S D_i (V_i)^j\}^2 K\left(\frac{V_i}{h}\right) \tag{9}$$

where $D_i = 1_{\{V_i > 0\}}$ and by excluding D_i as a regressor we require that the left and right intercepts be the same. Denoting the solutions to (9) by $\hat{\gamma}_j^S$ and $\hat{\delta}_j^S$,[20] the resulting RKD estimator is $\hat{\tau}_{SRKD}^{pR} = \frac{\hat{\delta}_1^S}{\kappa_1^+ - \kappa_1^-}$ for $p = L, Q$. Similarly, let $\hat{\tau}_{FRKD}^{LR}$ and $\hat{\tau}_{FRKD}^{QR}$ be the fuzzy RKD estimators from the local linear and quadratic

---

[20]For simplicity of notation, we do not include the superscript p in $\hat{\gamma}_j^S$ and $\hat{\delta}_j^S$, but the dependence of the estimators on the polynomial order is implied.

regression with the restrictions that $m_L(0^-) = m_L(0^+)$ *and* $r_L(0^-) = r_L(0^+)$. This amounts to replacing the denominator $\kappa_1^+ - \kappa_1^-$ with an estimated kink in B_i and solving the same problem as in (9). If the same bandwidth is used for the estimation of the kink in B_i and the outcome equation (as in Imbens and Lemieux (2008) and Lee and Lemieux (2010)), this amounts to solving

$$\min_{\{\tilde{\gamma}_j^F, \tilde{\delta}_j^F\}} \sum_{i=1}^n \{Y_i - \sum_{j=0}^p \tilde{\gamma}_j^F (V_i)^j - \tilde{\delta}_1^F \hat{B}_i - \sum_{j=2}^p \tilde{\delta}_j^F D_i(V_i)^j\}^2 K(\frac{V_i}{h}) \tag{10}$$

where $\hat{B}_i$ is the predicted value of B_i from the local polynomial regression of B on V^j ($j = 0, ..., p$) and $D \cdot V^j$ ($j = 1, ...p$).[21] Specifically, $\hat{\tau}_{FRKD}^{pR} = \hat{\delta}_1^F$ for $p = L, Q$; this is analogous to the 2SLS estimator proposed by Hahn et al. (2001).

Proposition 6. *(a) Local Linear ($p = 1$): Under Assumptions 9, 10L, 11L, 12-15 and 16L,*

$$\sqrt{nh^3}(\hat{\tau}_{SRKD}^{LR} - \tau_{SRKD}) \quad \Rightarrow \quad N(0, 12 \cdot \Omega_{SRKD}))$$
$$\sqrt{nh^3}(\hat{\tau}_{FRKD}^{LR} - \tau_{FRKD}) \quad \Rightarrow \quad N(0, 12 \cdot \Omega_{FRKD})).$$

(b) Local Quadratic ($p = 2$): Under Assumptions 9, 10Q, 11Q, 12-15 and 16Q,

$$\sqrt{nh^3}(\hat{\tau}_{SRKD}^{QR} - \tau_{SRKD}) \quad \Rightarrow \quad N(0, 192 \cdot \Omega_{SRKD}))$$
$$\sqrt{nh^3}(\hat{\tau}_{FRKD}^{QR} - \tau_{FRKD}) \quad \Rightarrow \quad N(0, 192 \cdot \Omega_{FRKD})).$$

The proof is in the Supplemental Appendix.

Remark 7. The intuition behind this result can be seen in a simple example. Suppose the conditional expectation can be modeled as a piecewise linear model

$$Y = c_0 + c_1 V + d_1 DV + \varepsilon$$

where $D = 1[V \geq 0]$ and ε is homoskedastic, and V is symmetrically distributed around 0. The restricted RKD estimator is the linear regression of Y on V and DV, while the unrestricted estimator (two separate linear models) additionally (and unnecessarily) includes D as a control. The asymptotic variance for the

---

[21]The last summation $\sum_{j=2}^p \tilde{\delta}_j^F D_i(V_i)^j$ is omitted in the linear case.

restricted and unrestricted estimators for d_1 are $\frac{Var(\varepsilon)}{Var(\tilde{DV})}$ and $\frac{Var(\varepsilon)}{Var(\check{DV})}$, respectively, where $\tilde{DV}$ is the residual from projecting DV on V and $\check{DV}$ is the residual from projecting DV on V and D. Due to symmetry, the coefficient on D in the latter projection is zero, and hence $\tilde{DV} = \check{DV}$.[22]

Remark 8. Despite the asymptotic statement in Proposition 6, we focus on the restricted RKD estimators below in our empirical analysis. We do this because in practice, the bandwidths we use are large enough such that the distribution of V is not symmetric, which implies that if the true function is piecewise linear (or quadratic) within the window, the restricted estimators will be more precise.

3.2 Variance Estimation and Inference in RKD

In this subsection, we discuss variance estimation and inference in an RK design. The restricted SRKD estimator amounts to restricting the data to a symmetric window around the kink point, and regressing Y on linear (or quadratic) terms in V and DV (where D is the indicator for $V \geq 0$), but leaving out the "main effect" D. The restricted FRKD estimator uses the same specification, except that it replaces DV with B and uses DV as the excluded instrument. We show that the corresponding standard heteroskedasticity-robust variance estimators – commonly built into statistical software – can be used for inference.[23]

In a sharp RKD, the variance estimator for the $(2p+1) \times 1$ vector $\hat{\beta}_{sharp}^{R} = (\hat{\gamma}_0^S, ..., \hat{\gamma}_p^S, \hat{\delta}_1^S, ..., \hat{\delta}_p^S)^T$, the solution to (9), is[24]

$$\hat{Var}(\hat{\beta}_{sharp}^{R}) = \mathbf{S}_{XX}^{-1} \hat{\mathbf{S}}_{sharp} \mathbf{S}_{XX}^{-1}$$

where $\mathbf{S}_{XX} = \sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i^T K(\frac{V_i}{h})$ and $\hat{\mathbf{S}}_{sharp} = \sum_{i=1}^{n} (\mathbf{X}_i \mathbf{X}_i^T) \hat{\varepsilon}_i^2 K^2(\frac{V_i}{h})$ with $\mathbf{X}_i = (1, ..., V_i^p, D_i V_i, ..., D_i V_i^p)^T$ and $\hat{\varepsilon}_i = Y_i - \mathbf{X}_i^T \hat{\beta}_{sharp}^{R}$.

In a fuzzy RKD, the variance estimator for the $(2p+1) \times 1$ vector $\hat{\beta}_{fuzzy}^{R} = (\hat{\gamma}_0^F, ..., \hat{\gamma}_p^F, \hat{\delta}_1^F ..., \hat{\delta}_p^F)^T = (\frac{1}{n} \sum_{i=1}^{n} \mathbf{W}_i \mathbf{X}_i^T K(\frac{V_i}{h}))^{-1} (\frac{1}{n} \sum_{i=1}^{n} \mathbf{W}_i Y_i K(\frac{V_i}{h}))$, the solution to (10), is

$$\hat{Var}(\hat{\beta}_{fuzzy}^{R}) = \mathbf{S}_{XW}^{-1} \hat{\mathbf{S}}_{fuzzy} \mathbf{S}_{WX}^{-1}$$

---

[22]Consider using the Frisch-Waugh theorem to compute the coefficient on D. Under symmetry $\frac{Cov(DV,V)}{Var(V)} = \frac{1}{2}$, and the $DV - \frac{1}{2}V$ is orthogonal to V by construction and also orthogonal to D due to the symmetry of the distribution of V. Hence $DV - \frac{1}{2}V$ is orthogonal to the residual from projecting D on V.

[23]An alternative, and asymptotically equivalent, approach for the FRKD would be (analogous to Imbens and Lemieux (2008)) a plug-in variance estimator that separately estimates each component of the asymptotic variance.

[24]Since we are concerned with their consistency, the heteroskedasticity-robust variance estimators defined here abstract away from the finite sample adjustments that may appear in statistical software packages.

where $\mathbf{S}_{XW} = \sum_{i=1}^{n} \mathbf{X}_i \mathbf{W}_i^T K(\frac{V_i}{h})$, $\mathbf{S}_{WX} = \mathbf{S}_{XW}^T$ and $\hat{\mathbf{S}}_{fuzzy} = \sum_{i=1}^{n} (\mathbf{X}_i \mathbf{X}_i^T) \hat{u}_i^2 K^2(\frac{V_i}{h})$ and $\hat{u}_i = Y_i - \mathbf{W}_i^T \hat{\beta}_{fuzzy}^R$ with $\mathbf{W}_i = (1, ..., V_i^p, B_i, D_i V_i^2 ..., D_i V_i^p)^T$.[25]

Since we are interested in the scalar RKD estimator $\hat{\delta}_1^S$ in the sharp case and $\hat{\delta}_1^F$ in the fuzzy case, we define $\mathbf{e}_k$ to be the unit vector of length $(2p+1)$ whose k-th entry is equal to 1, and let

$$\hat{Var}(\hat{\tau}_{SRKD}^{LR}) = \frac{1}{(\kappa_1^+ - \kappa_1^-)^2} \mathbf{e}_3^T \hat{Var}(\hat{\beta}_{sharp}^{LR}) \mathbf{e}_3; \qquad \hat{Var}(\hat{\tau}_{FRKD}^{LR}) = \mathbf{e}_3^T \hat{Var}(\hat{\beta}_{fuzzy}^{LR}) \mathbf{e}_3$$

$$\hat{Var}(\hat{\tau}_{SRKD}^{QR}) = \frac{1}{(\kappa_1^+ - \kappa_1^-)^2} \mathbf{e}_4^T \hat{Var}(\hat{\beta}_{sharp}^{QR}) \mathbf{e}_4; \qquad \hat{Var}(\hat{\tau}_{FRKD}^{QR}) = \mathbf{e}_4^T \hat{Var}(\hat{\beta}_{fuzzy}^{QR}) \mathbf{e}_4$$

where the L and Q superscripts denote whether a given estimator is obtained from a linear or quadratic local regression.

In order for these variance estimators to be consistent, we require the additional regularity condition:

Assumption 17. $\kappa_Y(v) = E[|Y_i - m(V_i)|^4 | V_i = v]$ and $\kappa_B(V_i) = E[|B_i - r(V_i)|^4 | V_i = v]$ are bounded in a neighborhood around 0 and the limits $\kappa_Y(0^{\pm})$ and $\kappa_B(0^{\pm})$ are well-defined.

Proposition 7. *Consistency of the variance estimators for the restricted RKD estimators.*[26]

(a) Local Linear ($p = 1$): Under Assumptions 9, 10L, 11L, 12-14, 16L, and 17,

$$nh^3 \hat{Var}(\hat{\tau}_{SRKD}^{LR}) \to 12 \cdot \Omega_{SRKD}; \qquad nh^3 \hat{Var}(\hat{\tau}_{FRKD}^{LR}) \to 12 \cdot \Omega_{SRKD}.$$

(b) Local Quadratic ($p = 2$): Under Assumptions 9, 10Q, 11Q, 12-14, 16Q, and 17,

$$nh^3 \hat{Var}(\hat{\tau}_{SRKD}^{QR}) \to 192 \cdot \Omega_{SRKD}; \qquad nh^3 \hat{Var}(\hat{\tau}_{FRKD}^{QR}) \to 192 \cdot \Omega_{FRKD}.$$

The Proof is in the Supplemental Appendix.

Combining Proposition 4, Proposition 5 and Proposition 7, we have

Corollary 3. *For p=L,Q,*

$$\frac{1}{\sqrt{\hat{Var}(\hat{\tau}_{SRKD}^{pR})}} (\hat{\tau}_{SRKD}^{pR} - \tau_{SRKD}) \Rightarrow N(0,1); \qquad \frac{1}{\sqrt{\hat{Var}(\hat{\tau}_{FRKD}^{pR})}} (\hat{\tau}_{FRKD}^{pR} - \tau_{FRKD}) \Rightarrow N(0,1).$$

---

[25] Note that in the linear case, $\mathbf{W}_i$ does not include any interaction terms.

[26] We can also prove that the heteroskedasticity-robust variance estimators for the unrestricted RKD estimators defined in the previous subsection are consistent.

4 The Effect of UI Benefits on the Duration of Benefit Claims and Joblessness

In this section, we use a fuzzy RKD approach to estimate the effect of higher unemployment benefits on the duration of 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.[27] Obtaining credible estimates of this effect is difficult, however, because UI benefits are determined by previous earnings, and are therefore correlated with unobserved characteristics of workers that influence their expected duration of unemployment. Since the UI benefit formula in Austria has both a minimum and maximum, a regression kink approach can provide new evidence on the impact of higher UI benefits at two different points in the benefit schedule. Before presenting our empirical analysis, we briefly discuss the underlying model of job search behavior that we use to frame our analysis, focusing on the plausibility of the density smoothness condition that is sufficient for a valid RKD. We then describe the benefit system in Austria, our data sources, and our main estimation results.

4.1 Theoretical background

In a standard search model, higher UI benefits reduce the incentives for search and raise the reservation wage, leading to an increase in the expected duration of joblessness.[28] More subtly, higher benefits can also exert a general equilibrium effect on steady state distribution of wages. Christensen et al. (2005), for example, derive the equilibrium distribution of wages, given a fixed UI benefit and a latent distribution of wage offers. In their model, a twice continuously differentiable wage offer distribution function ensures that the cumulative distribution of wages – and the associated distribution of previous wages for newly laid-off workers – is twice continuously differentiable.

In the Supplemental Appendix we present an extension of this model with one new feature: a UI benefit

---

[27]The original analysis in Baily (1978) has been generalized to allow for liquidity constraints (Chetty (2010)) and variable takeup (Kroft (2008)).

[28]In a standard model, workers receive wage offers from an offer distribution $F(w)$, and the arrival rate of offers is λs, where s is the (endogenous) level of search intensity. Assuming an unemployed worker receives an indefinite benefit b, and that λ is the same for employed and unemployed searchers, a currently unemployed worker accepts any offer with $w \geq b$, and unemployment spells are exponentially distributed with mean $[\lambda s_u^*(1 - F(b))]^{-1}$, where s_u^* is the optimal search intensity of unemployed workers. An increase in b raises the reservation wage and leads to lower value of s, implying an increase in the average duration of search.

that is proportional to the wage on the preceding job, up to some maximum benefit level (for simplicity we ignore any minimum benefit provision, though such a feature could be easily incorporated). Relative to the canonical model with a fixed UI benefit, a wage-dependent benefit introduces heterogeneity among the unemployed. In particular, the value of unemployment is increasing in the level of the previous wage (w_{-1}), up to the threshold for the maximum benefit (T^{max}), and is then constant for all higher wages. Consequently, the value function for an unemployed worker contains a kink at $w_{-1} = T^{max}$. A wage-dependent benefit also enhances the value of a higher-wage job, since it entitles the worker to higher benefits when the job ends. This incremental benefit stops once the wage (w) hits the threshold for the maximum UI benefit, creating a kink at $w = T^{max}$ in the value function for employed workers. This kink induces a kink in the mapping from wages to optimal on-the-job search effort by currently employed workers, which in turn leads to a kink in the density of wages at T^{max} (see the Supplemental Appendix for details). Since the rate of job destruction is assumed to be constant across jobs, the same kink appears in the density of previous wages for new job-losers. Such a kink – at precisely the threshold for the maximum benefit rate – violates the smooth density condition (i.e., Assumptions 4/4a) necessary for a valid regression kink design based on the change in the slope of the benefit function at T^{max}.[29]

The prediction of a kink in the equilibrium density of wages depends on individuals being able to perfectly forecast the location of the kink in the UI benefit schedule. In our context – the Austrian labor market during the period from 2001-2008 – we suspect that this is unrealistic, since the kink point is adjusted each year depending on lagged rates of wage inflation (see below). In the Supplemental Appendix we show that if a given worker chooses search intensity assuming the wage threshold is $T^{max} + \varepsilon$, rather than T^{max}, and ε has a continuous distribution across workers, then the equilibrium density of wages (and the density of previous wages among new job-losers) will be smooth at T^{max}. On the other hand, search effort (the focus of interest) among job-losers will still contain a kink at the actual threshold T^{max}, so a regression kink design based on the kink in the UI schedule will still identify the incentive effect of higher UI benefits, even in the presence of small and continuously distributed errors.

Given these theoretical possibilities, we consider it important to examine the actual distribution of pre-displacement wages among job seekers and test for the presence of kinks around the minimum and maximum

---

[29]In the model as written all workers are identical: hence a kink in the density of wages at T^{max} will not actually invalidate an RK design. More realistically, however, workers differ in their cost of search (and in other dimensions) and the kink at T^{max} will be larger for some types than others, causing a discontinuity in the conditional distribution of unobserved heterogeneity at T^{max} that will lead to bias in an RKD.

benefit thresholds, as well as for kinks in the conditional distributions of any predetermined covariates. While we do not necessarily expect to find kinks in our setting (given the difficulty of precisely forecasting the benefit thresholds), a kink may well exist in other settings – e.g., in the U.S., where maximum UI benefits are often fixed for several years at the same nominal value.

4.2 The Unemployment Insurance System in Austria

Job-losers in Austria who have worked at least 52 weeks in the past 24 months are eligible for unemployment insurance benefits, with a rate that depends on their average daily earnings in the "base calendar year" for their benefit claim (the previous calendar year for claims filed in the second half of the year; the second most recent calendar year for claims filed in the first half of the year).[30] Average daily earnings in the base year are adjusted to the current year by a nominal wage growth factor, and converted to *net* daily earnings, using the Social Security and Income tax schedules for a single white-collar worker.[31] The daily UI benefit is calculated as 55% of net daily earnings, subject to a maximum benefit level that is adjusted each year.[32] During our sample period the maximum increased from €36.05 (for claims in 2001) to €41.77 (for claims in 2008).

Unemployed individuals with dependent family members are also eligible for a family allowance that is added to their basic UI benefit. During the years 2001 - 2008 the allowance was €0.97 per day per dependent. Finally, individuals whose total benefit (base plus family allowance) is below a specified minimum receive a supplement that tops-up their total benefit to the minimum.[33] However, total UI benefit cannot exceed 60% of net daily base year earnings for a single individual, or 80% for a claimant with dependents, so the minimum benefit provision is not binding for individuals with very low earnings.

These rules create a piecewise linear relationship between base calendar year earnings and UI benefits that depends on the Social Security and income tax rates as well as the replacement rate and the minimum

---

[30]Specifically, for claims filed from January to June of year t the base calendar year is year $t - 2$, while for claims filed from July to December of year t the base calendar year is year $t - 1$. The back-dating of the base year for claims filed early in the year reflects the fact that earnings information is extracted from the Social Security database (the same system used to record earnings information for state-run pensions) which is typically updated with a lag.

[31]This calculation assumes that the individual works the entire year at the average daily wage, and ignores the actual tax deductions for a given claimant.

[32]The maximum is linked to the maximum earnings level for Social Security contributions, which is adjusted each year based on the past changes in average Social Security-taxable wages. Specifically, the earnings level at which the maximum UI benefit is reached in year t is set by the Social Security contribution cap in year $t - 3$ multiplied by $12/14$ (to reflect the fact that employees in Austria receive 14 months of salary each year). Thus the lowest earnings level at which maximum UI benefits apply is about 20% below the contemporaneous contribution cap.

[33]The minimum benefit is linked to the minimum income amount for individuals receiving social insurance benefits such as public pensions.

24

and maximum benefit amounts. As an illustration, Figure 1a plots *actual* daily UI benefits against annual base year earnings for a sample of UI claimants in 2004. The presence of a high fraction of observations with benefits that lie precisely on the benefit schedule leads to a series of clearly discernible lines in the figure, though there are also many observations scattered above and below the lines, which we discuss in more detail below.

In the middle of the graph there are 5 distinct upward-sloping linear segments, corresponding to individuals with 0, 1, 2, 3, or 4 dependents. These schedules all reach an upper kink point at approximately €38,700, the threshold for the maximum benefit in 2004 (which is shown in the graph by a solid vertical line).[34] At the lower end, the situation is more complicated: each of the upward-sloping segments reaches the minimum daily benefit of €21.77/day at a different level of earnings, reflecting the fact that the basic benefit includes family allowances, but the minimum does not. Thus, for example, a single claimant reaches the minimum benefit with annual earnings of $\sim$ €19,200, while a claimant with 2 dependents reaches the minimum with earnings of $\sim$ €17,100.

Finally, among the lowest-paid claimants the benefit schedule becomes upward-sloping again. Low-earning claimants with no dependents have a benefit that is 60% of their base earnings (net of taxes), while those with dependents have a benefit that is 80% of their net base year earnings. These rules generate the two lines at the bottom of the graph with benefits below the €21.77/day "minimum".[35]

In addition to the observations that lie on the piecewise linear segments in Figure 1a, there are many other observations that lie off these segments. There are at least three explanations for these observations. The first is that we have incorrectly calculated base year earnings from the available administrative earnings data, due to error in the calculation of the claim start date, for example. The second is that the Social Security earnings records are inaccurate, and were over-ridden by benefit administrators using updated information.[36] A third explanation is that UI benefits are mis-reported. Our belief, based in part on experiences with other administrative files, is that all three types of errors are possible, and routinely occur in many other settings (see e.g., Kapteyn and Ypma (2007)).

---

[34]€38,700 is approximately €105.7 per day gross earnings, or about €67.8 per day in net earnings, after deducting the $\sim 18\%$ employee Social Security contribution, and income taxes. The 55% replacement rate then generates a daily benefit of €37.3, which was the maximum benefit rate in 2004.

[35]The line for low-earning single claimants actually bends at earnings of around €15,000, reflecting the fact at this point a single claimant begins paying income taxes, which are deducted from gross earnings in the calculation of UI benefits.

[36]Errors in the recording or calculation of base earnings will lead to points that lie to the left or right of the benefit schedule. Such errors may account for the presence of substantial numbers of claimants who receive the minimum benefit but have relatively high or relatively low recorded benefits.

Our RKD analysis exploits the kinks in the UI benefit schedule induced by the minimum and maximum benefit levels. Unfortunately, as we discuss below, we do not have access to information on the number of dependents reported by a UI claimant, so we adopt a fuzzy RKD approach in which the number of dependents is treated as an unobserved determinant of benefits. This does not affect the location of the "top kink" associated with the maximum benefit, since claimants with different numbers of dependents all have the same threshold earnings level T^{max} for reaching the maximum (though the level of maximum benefits differs by family size). For the "bottom kink" associated with the minimum benefit, however, there is a different threshold T^{min} for each family size. Moreover, there is a only a limited range of earnings for which T^{min} actually applies: for single claimants, for example, this is the set of people with net daily base period wages between $1/0.60$ and $1/0.55$ of the minimum benefit.[37]

We deal with this complication by focusing on the location of the kink in the benefit schedule for single claimants, and again relying upon the fuzzy RKD framework. Specifically, we define T^{min} as the bottom kink point for a *single* claimant: this is the level of annual earnings – approximately €19,200 – shown in the figure by a solid vertical line. To the right of T^{min} the benefit schedules for *all* claimant groups slope upward. To the left, benefits for claimants with dependents continue to fall, but benefits for single claimants are constant with respect to earnings: thus we expect to measure a kink in the average benefit function at T^{min}, with a magnitude that is (roughly) $0.55p^s$, where p^s is the (local) fraction of single claimants. To avoid the "subminimum" portion of the benefit schedule, we limit our analysis to claimants with daily net earnings above $1/0.60$ times the minimum daily benefit. For 2004 claims, this cutoff point is shown by the dashed line in Figure 1 at approximately €17,100. We also limit attention to claimants whose annual earnings are below the Social Security contribution cap. This cutoff is shown by the dashed line at approximately €48,000 in Figure 1.

Regular UI benefits in Austria have a relatively short duration. The basic entitlement is 20 weeks of benefits, with an additional 10 weeks (30 weeks total) for claimants with more than 3 years of employment in the last 5 years, and another 9 weeks (39 weeks total) for claimants over the age of 40 with at least 6 years of employment in the last 10 years.[38] When regular benefits are exhausted, however, claimants who satisfy a family-income-based means test can receive Unemployment Assistance (UA) benefits, which are linked to

---

[37]Single claimants receive the minimum benefit if and only if it is between 55% and 60% of their net daily earnings. That is, their net daily earnings must exceed $1/0.60$ of the minimum benefit, but fall below $1/0.55$ of the minimum benefit in order to receive the minimum.

[38]There are further extensions for claimants over 50, but we limit our sample to claimants under the age of 50.

UI. In particular, during the first 6 months of UA, the benefit amount (for those who satisfy the means test) is 92% of a claimant's "regular UI benefit" (i.e., the benefit that is calculated as 55% of net daily earnings) if his/her regular benefit is above the minimum level, or 95% if regular benefits are below the minimum.[39] After those first 6 months, the UA benefit falls to the minimum UI amount, with dependent allowances and an additional premium for those with longer UI entitlement.

Figure 1b shows weekly initial UA benefit amounts actually received by 2004 UI claimants who exhausted their regular benefits, as a function of their base year earnings. As in Figure 1a, we display solid lines at T^{min} and T^{max}, and dashed lines at the upper and lower earnings limits for our sample. Note that UA benefits, like UI benefits, reach a plateau once base year wages exceed T^{max}. In contrast, the two benefit formulas display very different behavior around the lower kink point in the UI formula, T^{min}. Approaching T^{min} from the left, UA benefits for a single filer are calculated as $0.95 \times 0.55 \times net\,daily\,wage$, whereas approaching from the right they are calculated as $0.92 \times 0.55 \times net\,daily\,wage$. The change in the formula generates a small "flat spot" which is (barely) visible in the graph (for people with gross annual earnings between T^{min} and $1.03 \times T^{min}$), and a small *decrease* in the slope of the relationship between base year earnings and UA benefits (from 52.25% of net daily wages to 50.60%).[40]

The relationship between UA and UI benefits impacts the interpretation of the marginal effects estimated in our RKD analysis. Specifically, around the top kink, where UI and initial UA benefits both "top out" at the same point, the estimated marginal effect of higher UI benefits from our analysis represents a combined impact of higher UI benefits and higher UA benefits for the subset of claimants who are eligible for UA. Around the bottom kink in the UI benefit schedule, however, UA benefits are (nearly) smooth, so the estimated marginal effect of higher UI benefits in our analysis is (implicitly) holding constant UA benefits.

4.3 Data and Analysis Sample

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 et al. (2009)). In addition to starting and ending event dates, the ASSD contains information on total 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

---

[39]A claimant receives 95% of the minimum in the region where the second formula exceeds the first.

[40]Recall that T^{min} is the level of gross annual earnings that yields a net daily wage that is $1/0.55$ of the minimum daily benefit.

benefit actually received by each claimant, and the duration of the benefit spell. We use the UI claim dates to assign the base calendar year for each claim, and then calculate base year earnings for each claim, which is the observed assignment variable for our RKD analysis (i.e., V^* in the notation of Section 2).

The ASSD allows us to construct a number of different outcomes associated with each new unemployment claim. We focus on two main ones: the duration of the initial spell of UI; and the time between the end of the old job and the start of any new job (which we censor at 1 year, the maximum time we can observe for spells in the final year of our claim sample). The duration of the initial UI spell is comparable to the outcome variable that has been analyzed in many previous studies, which rely exclusively on claim records from the UI system (e.g., Meyer (1990)).[41] The duration of time between jobs is a broader measure of inactivity that may better reflect the moral hazard costs of UI. This is particularly true in the Austrian context, where people who have exhausted regular UI benefits can receive UA benefits that are closely linked to their UI benefits. In addition to these two key outcomes, the ASSD provides data on such characteristics as age, gender, education, and marital status, as well as detailed employment histories and the industry and location of each job.

Our analysis sample focuses on individuals who leave a job with minimum of one year of job tenure (thus ensuring eligibility for UI) and initiate a UI claim within four weeks of the job ending date (eliminating job-quitters, who face a four-week waiting period). Starting with a sample of 792,054 unemployment spells that meet these two criteria, we make three additional restrictions to arrive at our analysis sample. First, we drop 34,352 spells with either zero earnings in the base year, or with no valid UI claim.[42] Second, we eliminate individuals older than 50, reducing the sample by 115,576 spells. This exclusion is motivated by the strong interaction between the UI system with early retirement regulations in Austria, which results in an extremely low rate of re-entry into employment for job-losers who are over 50 years of age. Finally, we restrict the sample to individuals with baseline earnings high enough to receive the minimum UI benefit but less than the Social Security contribution cap (i.e., between the dashed lines in Figures 1a and 1b). The resulting sample includes 369,566 unemployment spells initiated between 2001 and 2008.

We pool observations from different years in the following way. First, we divide the claimants in each

---

[41]Some claimants end their initial UI spell then restart the claim after a period of time, without an intervening period of (covered) employment. Thus, the duration of the initial spell is sometimes less than the total number of days of UI payments received. We use this outcome variable because it corresponds to "time to first exit", which is the main outcome in studies that use a hazard-modeling framework.

[42]People without a valid claim could be receiving disability insurance payments, or could have some other reason for not being eligible to receive UI.

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, the cutoff threshold is closer to T^{min} than T^{max}, implying a narrower support for our observed assignment variable V^* (observed annual base year earnings) around the bottom kink than the top kink. Next we re-center base year earnings for observations in the bottom kink subsample around T^{min}, and re-center base year earnings for observation in the top kink subsample around T^{max}, so both kinks occur at $V^* = 0$. Finally, we pool the yearly re-centered subsamples.

Table 1 reports basic summary statistics for the bottom and top kink samples. Mean base year earnings for the bottom kink group are about €21,000, with a relatively narrow range of variation (standard deviation = €2,300), while mean earnings in the top kink group are higher (mean = €31,000) and more disperse (standard deviation = €5,700). Mean daily UI benefits are €24 for the bottom kink group (implying an annualized benefit of €8,800, about 46% of T^{min}), while mean benefits for the top kink sample are €31.7 (implying an annualized benefit of €11,600, about 30% of T^{max}). Looking down the table, people in the bottom kink sample are more likely to be female (38% versus 21% in the top kink sample), 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 two main outcome variables are not too different in the bottom kink and top kink samples. The average duration of the first UI spell is around 80 days, while the mean duration of joblessness is much longer (150 days). Overall, job-losers receive UI payments for about 84% of the time between the previous and next jobs. Only about 10 percent of claimants exhaust their regular UI benefits, and only 6-7 percent of all claimants end up receiving UA. Conditional on exhaustion, however, the fraction of UI claimants who initiate UA is around 60%.

As we emphasized in Section 2, a critical concern for valid inference based on an RKD is that the density of the observed assignment variable (in our case, base year earnings) is smooth at the kink point. Figures 2a and 2b show the frequency distributions of base year earnings around the bottom and top kink points in our two subsamples. For the bottom kink sample we use 100-Euro bins, reflecting the relatively narrow range for V^* around T^{min}, and the relatively high density of our sample in this range (about 2,500-2,800 observations per 100-Euro bin). For the top kink we use a coarser 300-Euro bin, reflecting the wider range for V^* around T^{max}, and the lower density of our sample in this range (about 1,400 observations per 300-Euro bin in a neighborhood of T^{max}).

An examination of these figures suggests that the empirical frequency distributions are relatively smooth around the two kink points, with no obvious discontinuities in the derivatives of the underlying density functions. To test this more formally, we fit a series of simple models for the empirical histograms shown in Figures 2a and 2b, using a minimum chi-squared approach.[43] These models assume that the mean fraction of the population in each bin around the kink point is a continuous polynomial function of the bin midpoint, but allow the first and higher-order derivatives of the mean function to jump at the kink point. We test for a smooth density by testing for a jump in the linear term of the polynomial at the kink. Supplemental Appendix Table 1 summarizes the estimated kinks in the density for models with polynomials of order 2, 3, 4, or 5, as well as the goodness of fit of each model and the associated Akaike model selection statistic. For the lower kink sample the third-order model has the lowest Akaike criterion, and the fitted values are shown in Figure 2a. This model and the higher order models all show no evidence of a kink in the density of observations around T^{min}. For the upper kink sample the 4th and 5th order models have very similar Akaike statistics and we plot the fitted values from the fourth order model. Again, there is no evidence of a kink in the density around T^{mas}. We provide additional evidence on validity of an RKD approach in the next subsection, based on the conditional distribution functions of various predetermined covariates around the two kink points.

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

As a starting point for our RKD analysis, Figures 3 and 4 show the relationships between base years earnings and actual UI benefits around the bottom and top kinks. Note that in these (and subsequent) figures we have narrowed our focus to a range of $\pm€2,000$ around the bottom kink (using 100-Euro bins, with 20 bins on each side of the kink), and $\pm€6,000$ around the top kink (using 300-Euro bins, with 20 bins on each side of the kink). Given the upper and lower earnings cutoffs for our sample, these are (approximately) the largest symmetric ranges we can use around the two kink points.[44]

As expected given the UI benefit formula, Figures 3 and 4 show very clear kinks in the empirical relationship between average benefits and base year earnings, with a sharp increase in slope as earnings pass through the lower threshold T^{min}, and a sharp decrease as they pass through the upper threshold T^{max}. De-

---

[43]As shown in Lindsay and Qu (2003), the minimum chi-squared objective can be interpreted as a optimally weighted minimum distance objective for the multinomial distribution of histogram frequencies.

[44]These ranges are approximately $\pm\sigma$, where σ is the standard deviation of the observed assignment variable in the corresponding lower or upper kink samples.

spite the fixity of the basic UI benefit schedule for single claimants to the left of T^{min} and to the right of T^{max}, average benefits actually paid are increasing in both tails, albeit more slowly than in the intervening interval. The explanation is family allowances. Moving left from T^{min} the average number of dependent allowances is falling, as claimants with successively higher numbers of dependents hit the minimum benefit level (see Figure 1). Likewise, moving right from T^{max} the average number of allowances is rising, reflecting a positive correlation between earnings and family size. In addition, the rising slope in average benefits to the right of T^{max} reflects a shifts in the earnings distribution. In later years, when the maximum benefit level is higher, there is a larger mass of observations farther to the right of T^{max}.

Figures 5-8 present a parallel set of figures for the log of time to next job and the log of the initial UI spell duration, which we refer to for simplicity as "claim duration" (we use the logs of these duration variables to facilitate the calculation of elasticities). Figures 5 and 6 show that there are well-defined kinks in the relationship between base year earnings and the duration of time to next job at both the bottom and top thresholds, though the data to the right of T^{max} in Figure 6 are a little noisy. Figures 7 and 8 show similar kinks for claim duration, though again the data to the right of the top threshold appear relatively dispersed.

Given the relatively short duration of UI benefits in our sample (20 - 39 weeks), it is also interesting to look at the probability a claimant exhausts benefits, and the probability of initiating a UA benefit spell. Supplemental Appendix Figures 1a and 1b show these outcomes around the bottom kink. Both the probability of exhaustion and the probability of initiating a UA claim show evidence of a discrete increase in the slope with respect to base year earnings, suggesting that higher UI benefits increase the probability of exhaustion and increase the probability of starting a UA benefit spell. Figures 1c and 1d present parallel graphs around the top kink. Both the probability of exhaustion and the probability of starting UA exhibit kinks in the expected direction, though as with our main outcome variables, the probabilities are relatively noisy in the range of earnings just above T^{max}.

Finally, we turn to the patterns of the predetermined covariates around T^{min} and T^{max}. Supplemental Appendix Figures 2 and 3 show the conditional means of four main covariates around the two kink points: age, fraction female, fraction blue-collar, and fraction who were recalled to the previous job.[45] Most of these graphs show little indication of a kink in the conditional mean function, though mean age appears to have a slight kink at T^{min}, while the fraction recalled to the previous job shows a mild kink at T^{max}. To increase the

---

[45] Many seasonal jobs in Austria lay off workers at the end of the season and re-hire them again at the start of the next season. Having been recalled from unemployment to the recently lost job is a good indicator that the present spell may end with recall to that job again – see Del Bono and Weber (2008).

power of this analysis we construct a pair of "covariate indices" – predicted outcomes using only baseline covariates as regressors – from simple linear regression models relating our two outcome variables to a vector of 59 covariates, including dummies for gender, blue collar occupation, and being recalled to the previous job, decile of age (9 dummies), decile of previous job tenure (9 dummies), quintile of previous daily wage (4 dummies), major industry (6 dummies), region (3 dummies), year of claim (7 dummies), decile of previous firm size (9 dummies), and decile of previous firm's recall rate (9 dummies).[46] These estimated covariate index functions can be interpreted as "best linear predictions" of mean log time to next job and mean log claim duration, given the vector of predetermined variables.

Figure 9 plots the conditional mean values of the estimated covariate indices around the upper and lower kinks. The graphs in the left panel of the Figure show that the mean predicted outcomes evolve smoothly through the bottom kink. In contrast, the graphs on the right suggest that the conditional means of predicted time to next job and predicted claim duration both kink slightly at T^{max}. As we show below, this visual impression is confirmed by more formal comparisons of the estimated slopes of the conditional mean functions for covariate indices to the right and left of T^{max}.

4.5 RKD Estimation Results

4.5.1 Bandwidth Choice

The estimators proposed and discussed in Section 3 require a choice of bandwidth. In our empirical analysis, we use a "rule-of-thumb" bandwidth based on Equation (3.20) of Fan and Gijbels (1996) and refer to our bandwidth as the "FG bandwidth":

$$h \quad = \quad \mathrm{C}_p \cdot \big[\frac{\hat{\sigma}^2(0)}{\{\hat{m}^{(p+1)}(0)\}^2 \hat{f}(0)} \big]^{\frac{1}{2p+3}} n^{-\frac{1}{2p+3}}$$

where the constants are computed to be $\mathrm{C}_1 = 2.35$ for the local linear and $\mathrm{C}_2 = 3.93$ for the local quadratic case.[47] The quantities $\hat{\sigma}^2(0)$ and $\hat{m}^{(p+1)}(0)$ are estimated from a global polynomial regression of Y on the assignment variable V allowing for a kink at the threshold, and the order of the polynomial is chosen by the

---

[46]We fit a single prediction model for each outcome using the pooled bottom kink and top kink samples.

[47]Following Fan and Gijbels (1996), the constant is computed by $\mathrm{C}_p = \big(\frac{3\mathbf{e}_2^T \mathbf{S}^{+-1} \mathbf{S}^{*+} \mathbf{S}^{+-1} \mathbf{e}_2}{2p\{\mathbf{e}_2^T \mathbf{S}^{+-1} \mathbf{c}_p^+ \frac{1}{(p+1)!}\}^2} \big)^{\frac{1}{2p+3}}$ as a result of minimizing the asymptotic mean squared error of the derivative estimator. The quantities $\mathbf{e}_2$, $\mathbf{S}^+$, $\mathbf{S}^{*+}$ and $\mathbf{c}_p^+$ are defined in the Supplemental Appendix, and plugging in the "−" (minus) counterparts of these quantities results in the same bandwidth. The constant we compute is appropriate for boundary estimation.

32

Akaike Information Criterion (AIC).[48] $\hat{f}(0)$ is estimated from a global polynomial fit to the histogram of V (the base year earnings variable). Following the convention in the RD literature, we use the same h on both sides of the kink. We do not attempt to derive an optimal bandwidth based on the mean squared error of the RKD estimators as Imbens and Kalyanaraman (2012) have done in the RDD context.[49] We do, however, explore the sensitivity of the estimates to a range of bandwidths.

4.5.2   Reduced Form Kinks in Assignment and Outcome Variables

Table 2 presents reduced form estimates of the kinks in our endogenous covariate of interest (log daily benefits) and our two main outcome variables (log of time to next job and log of initial UI spell duration) around T^{min} and T^{max}. For each variable we show the estimated symmetric bandwidths (columns 1 and 3) and the associated kink estimates (columns 2 and 4) from local linear and local quadratic models.

Looking first at log daily benefits, notice that the estimated kinks are relatively precisely estimated, and quite similar using either the local linear or local quadratic models, though as expected given the results in Section 3, the estimated standard errors from the local quadratic model are substantially larger. At the bottom kink, the FG bandwidth choice for the local linear model is slightly larger than the maximum available symmetric bandwidth in the data (€2,010), while the FG bandwidth for the local quadratic model is substantially above the maximum. We therefore use the largest possible symmetric bandwidth for our point estimate of the kink. At the top kink a similar situation arises: the FG bandwidths for the linear and quadratic models are both somewhat larger than the maximum available symmetric bandwidth (€6,125), so again we use the maximum available symmetric bandwidth.

Turning to the outcome variables, the estimated kinks are somewhat less precisely estimated than the kinks in log benefits, but all are significantly different from 0 at conventional levels and of the expected sign. The point estimates are also somewhat sensitive to choice of a local linear versus local quadratic approximation, with the latter models yielding 50 - 100% larger estimates of the kinks in both outcomes. In terms of magnitudes, the estimates of the kink in log time to next job are 1.7 - 3.5 times larger than the corresponding estimates of the kink in log benefits, suggesting that the elasticity of time out of work with

---

[48]$\hat{\sigma}^2(0)$ is taken to be the estimated error variance assuming homoskedasticity. When a polynomial of order lower than $(p+1)$ is selected by the AIC, we choose the polynomial to be of order $(p+1)$.

[49]In Section 3, we did not discuss the behavior of the local polynomial estimator associated with a bandwidth growing at the rate of $n^{-\frac{1}{2p+3}}$, which is $n^{-\frac{1}{5}}$ for the local linear and $n^{-\frac{1}{7}}$ for the local quadratic case. However, we could have chosen the rate of shrinkage to be slightly above $n^{\frac{1}{5}}$ for the linear and $n^{\frac{1}{7}}$ for the quadratic specifications, which correspond to the propositions in Section 3 but make little difference empirically. Therefore, we present the FG bandwidth for simplicity.

respect to benefits is above 1. The estimates of the kink in log claim duration are a little smaller and imply elasticities in the range of 1.2 - 2.8.

A potential concern with the estimates in Table 2 is their sensitivity to the choice of bandwidth. To address this concern, we estimated the kinks in log benefits, log time to next job, and log duration of the initial UI spell at T^{min} and T^{max} using local linear models with a range of potential bandwidths, and then plotted the estimated kink against the bandwidth choice for each variable. The results are summarized in Supplemental Appendix Figures 4-9. Supplemental Appendix Figure 4, for example, shows the estimated kink in log daily UI benefits at the lower kink point for bandwidth choices between €600 and €4,000. Since the maximal bandwidth to the left of the T^{min} is only €2,133, for wider bandwidth choices we use all the available data to the left and expand the range of base year earnings to the right according to the bandwidth. We also indicate the FG symmetric bandwidth reported in Table 2 by a vertical line in the graph (which in this case is €2,133).

Examination of Supplemental Appendix Figure 4 shows that the estimated kink in log daily benefits at the lower threshold is quite stable at around 0.020 - 0.023 for all possible bandwidths. Supplemental Appendix Figure 5 shows a similar figure for the estimated kink in log daily benefits at the upper threshold. Here, the range of possible bandwidths is higher, although the maximal range of data to the right of the upper kink is €6,125, so for higher bandwidths only the range to the left of the kink point is expanded. Again the estimated kink is relatively stable at -0.015 to -0.016 in a range of bandwidths between €3,000 and €8,000.

The effects of different bandwidth choices on the estimated kinks in log of time to next job at T^{min} and T^{max} are shown in Supplemental Appendix Figures 6 and 7. At the bottom threshold the estimated kinks are fairly stable at between 0.03 and 0.04 in a range of bandwidths between €1,200 and €3,500. At the top threshold the estimated kinks are a little more sensitive to the choice of bandwidth, with a range from -0.05 (at a bandwidth of €3,000) to -0.025 (at a bandwidth of €7,000). The FG symmetric bandwidth (€4,148) is near the lower end of this range and yields an estimated kink that is somewhat larger in absolute value than one would obtain from wider bandwidths. Supplemental Appendix Figures 8 and 9 show the estimated kinks in log claim duration for different bandwidth choices. These appear to be relatively stable, once the bandwidth is over €1,600 (for the bottom threshold) or €4,000 (for the top threshold).

Overall, we conclude that the estimated reduced form kinks in the assignment variable and the outcome variable are not particularly sensitive to even a relatively wide range of alternative choices for the bandwidth used to calculate the slopes on each side of the kink points T^{min} and T^{max}.

4.5.3 Kinks in Conditional Means of Predetermined Covariates

While the reduced form estimates in Table 2 appear to be relatively robust, an important question is how smoothly the characteristics of workers vary with their base period wages around the two kink points. As noted in Subsection 2.3, the smooth density assumption that ensures a valid RK design despite unobserved heterogeneity, or endogeneity in the covariate of interest has the following implication – the conditional distribution of any predetermined covariate will evolve smoothly around the benefit thresholds. Table 3 presents tests of this smoothness prediction for the covariates indices introduced in Figure 9, as well as for four main components of the indices. The format is the same as in Table 2: we show the estimated kinks using local linear and local quadratic models with FG symmetric bandwidths for each covariate.

The estimates in Panel A of Table 3 confirm the impression from Figure 9 that despite a small (but statistically significant) kink in the conditional distribution of age around T^{min}, there is no corresponding kink in predicted time to next job or predicted claim duration. The estimates in Panel B are more problematic. The fraction of blue collar workers and the fraction of workers who were recalled to their last job both exhibit kinks at T^{max}, as do the two covariate indices. The magnitudes of the kinks in the predicted durations of joblessness and claim duration are also relatively large. For example, using a local linear specification, the estimated kink in predicted log time to next job is -0.0138, 35% as big as the corresponding kink in actual log time to next job (-0.0396). Similarly, the estimated kink in predicted log claim duration from a local linear model (-0.0096) is 43% as big as the kink in actual log claim duration. These estimates suggest that 35-45% of the kinks in the two outcome variables at T^{max} could be explained by kinks in the distribution of observed characteristics of workers.

Given the relatively smooth density of UI claims around T^{max} shown in Figure 2b, and the institutional fact that actual value of T^{max} is hard to forecast, we think it is unlikely that the kinks in predicted joblessness and predicted claim duration are due to deterministic sorting around T^{max}. We think a more likely explanation is the inherent heterogeneity in the upper tail of the earnings distribution, which is evident in the noisiness of the conditional means of covariate indices to the right of T^{max} in Figure 9. Regardless of the explanation, however, the presence of these kinks means that RKD estimates from the top kink sample must be interpreted cautiously.

4.5.4 Fuzzy RKD Estimates

As a final step in our empirical analysis we present fuzzy RKD estimates of the effect of higher UI benefits on the duration of UI claims and joblessness. To obtain point estimates for the elasticities we first choose an FG symmetric bandwidth for the outcome variable (either time to next job or claim duration) around either the bottom or top kink. We then use this bandwidth to estimate the kinks in both the unemployment benefit function and the outcome variable, and take the ratio of these estimates as our fuzzy RKD estimate. Except for the fact that we choose a different bandwidth in estimating the kink in log UI benefits, the estimated elasticities in Table 4 would be equivalent to dividing the reduced-form kinks in the outcome variables in Table 2 by the corresponding kinks in log benefits. And since estimates of the kink in log benefits are relatively stable across bandwidth choices (Supplemental Appendix Figures 3 and 4), the actual estimates are not very different from an "indirect two-stage" approach based on the reduced-form kink estimates.

Using local linear specifications, the estimated elasticities of log time to next job and log claim duration around the bottom kink are 1.73 and 1.25, respectively, while the corresponding elasticities around the top kink are 2.64 and 1.31, respectively. The elasticities from local quadratic models are typically about twice as large (reflecting the larger reduced form kinks in the outcome variables from a local quadratic specification), but substantially less precise. A potential problem for the quadratic models is that the FG bandwidths are in all cases wider than our maximum available symmetric bandwidths (€2,010 around the bottom kink or €6,125 around the top kink), so we are not able to utilize as much data as would be optimal, given the apparent smoothness of the target functions we are trying to approximate. In view of this issue, and the superior asymptotic performance of a local linear specification for the same bandwidth sequence, we focus on the estimates from the local linear models.

As with the reduced-form kink estimates, a concern with the structural estimates from the fuzzy RKD is that the point estimates may be sensitive to the bandwidth choice. Figures 10-13 plot the structural elasticity estimates for our two outcome variables around the bottom and top kinks associated with a range of potential bandwidths. Figure 10 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 1.75 for a very wide range of bandwidths. Figure 11 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. Figures 12 and 13 show that the elasticities of the duration of the initial UI spell with respect to UI

benefits for the bottom and top kink samples are both quite stable and in a relatively narrow range between 1.1 and 1.4.

We also investigate the robustness of the point estimates by calculating the FG bandwidths using a selection procedure that does not allow a change in slopes in the outcome of interest at the kink point. The reduced-form kink estimates from this alternative procedure are presented in Supplemental Appendix Table 2, while the structural RKD estimates are presented in Supplemental Appendix Table 3. Comparisons of these estimates to the corresponding estimates in Tables 2 and 4 suggest that the alternative procedure yields bandwidth choices and point estimates that are relatively close to our baseline procedure. One notable difference is in the bandwidth choice for the log of time to next job at the top kink. Here the alternative procedure selects a substantially wider bandwidth than the baseline procedure, leading to a smaller estimate of the reduced form kink, and a smaller elasticity estimate that is much closer in magnitude to the estimated elasticity of time to next job in the bottom kink sample.

In another robustness check, we investigate the effect of our decision to censor the time to next job at one year in our main analysis.[50] Supplemental Appendix Table 4 presents fuzzy RKD estimates of the benefit elasticity of time to next job when the outcome variable is censored at 52 weeks, 39 weeks, 30 weeks, and 20 weeks. The choice of a lower censoring point leads to somewhat smaller elasticity estimates in both the bottom kink and top kink samples. For example, lowering the censoring point to 39 weeks (which is the longest duration of regular UI benefits for any job-losers in our sample) leads to a 10% smaller estimated elasticity at the bottom kink, and a 30% smaller elasticity at the top kink. These results suggest that our main results are qualitatively unaffected by alternative choices for the censoring point.

Finally, we also investigate potential differences in benefit responsiveness between the duration of the initial UI benefit spell and the total number of days of UI benefits received by a claimant. Supplemental Appendix Table 5 shows RKD estimates of the structural benefit elasticities for both measures of UI utilization. Using local linear models, the estimated elasticities for total days of UI are 10-20% larger than the corresponding elasticities for the duration of the first spell. These estimates suggest that duration of second spells of UI may be more sensitive to the benefit level than the duration of the initial spells, though it should be noted that estimates from the local quadratic models suggest the opposite. Given the standard errors and

---

[50]A small fraction of job-losers (3-5%) never return to jobs in the private sector, even after a decade. These individuals may move to self-employment or government jobs (which are not included in the ASSD), or they may find a job outside Austria, or withdraw from the labor force. In principle an RKD approach can be applied in a Tobit or hazard model framework, though the estimation methods presented in Section 3 cannot be directly used in such models.

the conflicting relative magnitudes of the elasticities from local linear and local quadratic models, however, we cannot draw definitive conclusions.

Overall, the estimates from our main specifications and the various alternatives lead us to two main conclusions. First, for job-losers from the lower part of the earnings distribution (around T^{min}), the elasticity of the duration of the initial UI claim spell with respect to unemployment benefits is around 1.25, while the elasticity of the time to next job with respect to UI benefits is probably a little larger, in the range of 1.5 to 2.0. Our confidence in these estimates is strengthened by their relative robustness to alternative bandwidth choices, and by the fact that tests for the validity of an RK design around the lower kink show little evidence that the design is compromised by sorting.

A second, more tentative conclusion is that the corresponding elasticities for job-losers in the upper part of the earnings distribution (around T^{max}) are of a roughly similar magnitude. Two factors lead us to a more cautious assessment of the benefit responsiveness for higher-wage job-losers. Most importantly, tests for the validity of an RKD approach show that the conditional distributions of observed worker characteristics change slopes around T^{max}. These changes are not associated with any discernible "bunching" at the kink point in the benefit schedule, but they are large enough to cause a 30-40% upward bias in the estimated elasticities of our two main outcomes. A second factor is that the point estimates of the reduced-form kinks in the outcome variables and in the structural RKD coefficients are both somewhat more sensitive to alternative bandwidth choices than the estimates for lower-wage workers.

How do our estimated benefit elasticities compare to those in the existing literature? Supplemental Appendix Table 6 contains a brief summary of the existing literature, drawing on the survey by Krueger and Meyer (2002) for the earlier U.S.-based literature, all of which use administrative records on unemployment insurance claims and estimate the effect of UI benefits on the duration of the initial spell of insured unemployment. These studies point to a benefit elasticity in the range of 0.3 to 0.8.[51] A recent study by Landais (2012) applies a regression kink design to some of the same data used in these earlier studies and obtains estimates of the elasticity of the initial UI benefit spell that range from 0.40 to 0.70. Another recent study by Chetty (2010) uses retrospective interview data from the Survey of Income and Program Participation and obtains an average benefit elasticity of about 0.5. Taken together these U.S. studies suggest a benchmark

---

[51] Krueger and Meyer attribute an estimated elasticity of 1.0 to Solon (1985) who studies the effect of making UI benefits taxable on the unemployment duration of high-earning claimants. He finds that the introduction of taxation caused a 22% reduction in the average duration of initial UI claims by higher-earning claimants (with no effect on low-earners). Assuming an average tax rate of 30%, this implies an elasticity of 0.73, which is our preferred interpretation of Solon (1985)'s results.

of around 0.5 for the elasticity of initial UI claim duration on the UI benefit. Most of the European studies included in Supplemental Appendix Table 6 estimate the effect of benefits on the time to first exit from the UI system, and obtain benefit elasticities that are similar to the U.S. studies. An exception is Carling et al. (2011), who study the effect of a reduction in the benefit replacement rate in Sweden in 1996 on the exit rate from unemployment recipiency to employment. Their estimate of the elasticity of time to next job with respect to the benefit level is 1.6, which is very similar to our point estimate of 1.73 for the bottom kink sample.

Our finding that the estimated benefit elasticity of the time to next job for the bottom kink sample is as high or higher than the estimated elasticity of the initial UI claim duration is interesting in light of relatively smooth UA benefit formula at T^{min}. Specifically, since time to next job (N) is approximately equal to the duration of the first UI spell (D_1) plus "uncovered" time (D_2) during which the job seeker is off UI:

$$Kink_{T^{min}}E[N|v^*] \approx Kink_{T^{min}}E[D_1|v^*] + Kink_{T^{min}}E[D_2|v^*]$$

where $Kink_{T^{min}}g(v^*)$ denotes the kink in the function $g(v^*)$ at $v^* = T^{min}$, i.e., the difference between the left and right derivatives of $g(v^*)$ at $v^* = T^{min}$. This says that the reduced form kink in time to next job at T^{min} is (approximately) the sum of the kink in the duration of the first claim and the kink in uncovered time. In terms of the implied RKD benefit elasticity estimates,

$$\varepsilon_{NB} \approx \frac{E[D_1|v^* = T^{min}]}{E[N|v^* = T^{min}]}\varepsilon_{D_1B} + \frac{E[D_2|v^* = T^{min}]}{E[N|v^* = T^{min}]}\varepsilon_{D_2B},$$

where ε_{NB} is the benefit elasticity of time to next job, ε_{D_1B} is the benefit elasticity of the duration of the initial UI spell, and ε_{D_2B} is the benefit elasticity of uncovered time. Our estimates suggest $\varepsilon_{NB} \geqslant \varepsilon_{D_1B}$, which implies that there must be a kink in the relationship between uncovered time and base year earnings at T^{min}, despite the fact that UA benefits available to exhaustees with lower family income during the uncovered period have (almost) no kink at T^{min}.

We checked this conclusion directly using data on the duration of uncovered time (i.e., time to next job minus total time on UI) for people in the bottom kink sample. Local linear regression models confirm that there is a significant positive kink at T^{min} in the relation between uncovered time and base year earnings (estimated kink = 0.69, standard error = 0.21).[52] We conjecture that this kink may be due in part to lags

---

[52]There is a larger negative kink in the relation between uncovered time and base year earnings at T^{max} (estimated kink = -1.79,

in the search process. People who search less while on UI because of more generous benefits will have a smaller inventory of prior contacts at the expiration of benefits, causing lower job-starting rates even after benefits expire. In any case, our findings underscore the potential value in being able to measure time to next job in assessing the incentive effects of the UI system.

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 et al. (2010); Simonsen et al. (2011)). 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 causal effects in a general nonseparable model (e.g., Blundell and Powell (2003)). The 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

standard error = 0.83). Given the kink in UA benefits, however, this is expected.

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.

Our final methodological contribution is to show how standard local polynomial regression techniques (Fan and Gijbels (1992)) can be adapted to obtain nonparametric estimators for the sharp and fuzzy RKD, and to characterize their asymptotic behavior.

We illustrate the use 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 level. We present a variety of simple graphical evidence showing that these kinks induce kinks in the duration of benefit receipt, and in the duration of total joblessness between the end of the previous job and the start of the next job. We also present a variety of tests of the smooth density assumption around the thresholds for the minimum and maximum benefit amounts. We then present local polynomial-based estimates of the "reduced form" kinks in these outcome variables, and of the "structural" RKD estimates of the elasticities of the duration of UI recipiency and duration of joblessness with respect to the UI benefit. Our estimates point to elasticities that are in the range between 1 and 2 – somewhat higher than most of the estimates from previous studies in the U.S. and Europe.

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.

Amemiya, Takeshi, *Advanced Econometrics*, Harvard University Press, 1985.

Arraz, Jose M., Fernando Munoz-Bullon, and Juan Muro, "Do Unemployment Benefit Legislative Changes Affect Job Finding?," Working Paper, Universidad de Alcalä 2008.

Baily, Martin N., "Some Aspects of Optimal Unemployment Insurance," *Journal of Public Economics*, 1978, *10* (3), 379–402.

Blundell, Richard and James L. Powell, "Endogeneity in Nonparametric and Semiparametric Regression Models," in Mathias Dewatripont, Lars Peter Hansen, and Stephen J. Turnovsky, eds., *Advances in economics and econometrics theory and applications : Eighth World Congress.*, Vol. II of *Econometric Society monographs no. 36*, Cambridge: Cambridge University Press, 2003, pp. 312–357.

Card, David, 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.

Carling, Kenneth, Bertil Holmlund, and Altin Vejsiu, "Do Benefit Cuts Boost Job Finding? Swedish Evidence from the 1990s," *Economic Journal*, 2011, *111* (474), 766–790.

Cavanagh, Christopher L. and Robert P. Sherman, "Rank Estimators for Monotonic Index Models," *Journal of Econometrics*, 1998, *84* (2), 351–381.

Chesher, Andrew, "Identification in Nonseparable Models," *Econometrica*, 2003, *71* (5), 1405–1441.

Chetty, Raj, "Moral Hazard versus Liquidity and Optimal Unemployment Insurance," *Journal of Political Economy*, 2010, *116* (2), 173–234.

_ , "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," *Econometrica*, 2012, *80* (3), 969–1018.

Christensen, Bent Jesper, Rasmus Lenz, Dale T. Mortensen, George R. Neumann, and Axel Werwatz, "On the Job Search and the Wage Distribution," *Journal of Labor Economics*, 2005, *23* (1), 181–221.

Classen, Kathleen P., "Unemployment Insurance and Job Search," in S.A. Lippman and J.J. McCall, eds., *Studies in the Economics of Search*, Amsterdam: North-Holland, 1977, pp. 191–219.

Dahlberg, Matz, Eva Mork, Jorn Rattso, and Hanna Agren, "Using a Discontinuous Grant Rule to Identify the Effect of Grants on Local Taxes and Spending," *Journal of Public Economics*, 2008, *92* (12), 2320–2335.

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 E. and David S. Lee, "Program Evaluation and Research Designs," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 4A, Elsevier, 2011.

Fan, Jianqing and Irene Gijbels, "Variable Bandwidth and Local Linear Regression Smoothers," *Annals of Statistics*, 1992, *20* (4), 2008–2036.

_ **and** _ , *Local Polynomial Modelling and Its Applications*, Chapman and Hall, 1996.

Florens, J. P., J. J. Heckman, C. Meghir, and E. Vytlacil, "Identification of Treatment Effects Using Control Functions in Models With Continuous, Endogenous Treatment and Heterogeneous Effects," *Econometrica*, 2008, *76* (5), 1191–1206.

Guryan, Jonathan, "Does Money Matter? Regression-Discontinuity Estimates from Education Finance Reform in Massachusetts," Working Paper 8269, National Bureau of Economic Research 2001.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw, "Evaluating the Effect of an Antidiscrimination Law Using a Regression-Discontinuity Design," Working Paper 7131, National Bureau of Economic Research May 1999.

_ , _ , **and** _ , "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 2001, *69* (1), 201–209.

Han, Aaron K., "Non-Parametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimator," *Journal of Econometrics*, 1987, *35*, 303–316.

Heckman, James J. and Edward Vytlacil, "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 2005, *73* (3), 669–738.

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

Imbens, Guido W. and Joshua D. Angrist, "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 1994, *62* (2), 467–475.

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

_ **and Whitney K. Newey**, "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," *Econometrica*, 2009, *77* (5), 1481–1512.

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.

Katz, Lawrence H. and Bruce D. Meyer, "The Impact of the Potential Duration of Unemployment Benefits on the Duration of Unemployment," *Journal of Public Economics*, 1990, *41* (1), 45–72.

Khan, Shakeeb and Elie Tamer, "Partial Rank Estimation of Duration Models with General Forms of Censoring," *Journal of Econometrics*, 2007, *136* (1), 251 – 280.

Kroft, Kory, "Takeup, Social Multipliers and Optimal Social Insurance," *Journal of Public Economics*, 2008, *92* (3-4), 722–737.

Krueger, Alan B. and Bruce D. Meyer, "Labor Supply Effects of Social Insurance," in Alan J. Auerbach and Martin S Feldstein, eds., *Handbook of Public Economics*, Amsterdam and New York: Elsevier, 2002, pp. 2327–2392.

Lalive, Rafael, Jan C. Van Ours, and Josef Zweimüller, "How Changes in Financial Incentives Affect the Duration of Unemployment," *Review of Economic Studies*, 2006, *73* (4), 1009–1038.

Landais, Camille, "Assessing the Welfare Effects of Unemployment Benefits Using the Regression Kink Design," Working Paper, Stanford Institute for Economic Policy Research 2012.

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

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

Lewbel, Arthur, "Semiparametric Latent Variable Model Estimation with Endogenous or Mismeasured Regressors," *Econometrica*, 1998, *66* (1), 105–121.

＿ , "Semiparametric Qualitative Response Model Estimation with Unknown Heteroscedasticity or Instrumental Variables," *Journal of Econometrics*, 2000, *97*, 145–177.

Lindsay, Bruce G. and Annie Qu, "Inference Functions and Quadratic Score Tests," *Statistical Science*, 2003, *18* (3), 394–410.

Matzkin, Rosa L., "A Nonparametric Maximum Rank Correlation Estimator," in W.A. Barnett, J.L. Powell, and G.E. Tauchen, eds., *Nonparametric and Semiparametric Methods in Economics and Statistics*, Cambridge University Press, 1991.

＿ , "Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and The Binary Choice Models," *Econometrica*, 1992, *60(2)*, 239–270.

＿ , "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, 2003, *71* (5), 1339–1375.

＿ , "Chapter 73 Nonparametric identification," in James J. Heckman and Edward E. Leamer, eds., *James J. Heckman and Edward E. Leamer, eds.*, Vol. 6, Part B of *Handbook of Econometrics*, Elsevier, 2007, pp. 5307 – 5368.

McCrary, Justin, "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," *Journal of Econometrics*, 2008, *142* (2), 698–714. The regression discontinuity design: Theory and applications.

Meyer, Bruce D., "Unemployment Insurance and Unemployment Spells," *Econometrica*, 1990, *58* (4), 757–782.

＿ **and Wallace K. C. Mok**, "Quasi-Experimental Evidence on the Effects of Unemployment Insurance from New York State," NBER Working Paper No.12865 2007.

Moffitt, Robert, "Unemployment Insurance and the Distribution of Unemployment Spells," *Journal of Econometrics*, 1985, *28* (1), 85–101.

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.

Roed, Knut and Tao Zhang, "Does Unemployment Compensation Affect Unemployment Duration?," *Economic Journal*, 2003, *113* (484), 190–206.

Roehrig, Charles S., "Conditions for Identification in Nonparametric and Parametric Models," *Econometrica*, 1988, *56* (2), 433–447.

Rudin, Walter, *Principles of Mathematical Analysis* International Series in Pure and Applied Mathematics, 3rd ed., McGraw-Hill, 1976.

Saez, Emmanuel, "Do Taxpayers Bunch at Kink Points?," *American Economic Journal: Economic Policy*, 2010, *2* (3), 180–212.

Sherman, Robert P., "The Limiting Distribution of the Maximum Rank Correlation Estimator," *Econometrica*, 1993, *61* (1), pp. 123–137.

Simonsen, Marianne, Lars Skipper, and Niels Skipper, "Price Sensitivity of Demand for Prescription Drugs: Exploiting a Regression Kink Design," Working Paper 2010-3, University of Aarhus Department of Economics 2011.

Solon, Gary, "Work Incentive Effects of Taxing Unemployment Benefits," *Econometrica*, 1985, *53* (2), 295–306.

Thistlethwaite, Donald L. and Donald T. Campbell, "Regression-Discontinuity Analysis: An Alternative to the Ex-Post Facto Experiment," *Journal of Educational Psychology*, 1960, *51* (6), 309–317.

Urquiola, Miguel and Eric Verhoogen, "Class-Size Caps, Sorting, and the Regression-Discontinuity Design," *American Economic Review*, 2009, *99* (1), 179–215.

Vytlacil, Edward, "Independence, Monotonicity, and Latent Index Models: An Equivalence Result," *Econometrica*, 2002, *70* (1), 331–341.

White, Halbert, "A Heteroskedasticty-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica*, 1980, *48* (4), 817–838.

Zweimüller, Josef, Rudolf Winter-Ebmer, Rafael Lalive, Andreas Kuhn, Jean-Philippe Wuellrich, Oliver Ruf, and Simon Büchi, "Austrian Social Security Database, University of Zurich," Working Paper iewwp410, Institute for Empirical Research in Economics 2009.

Figure 1a: UI Benefits in 2004
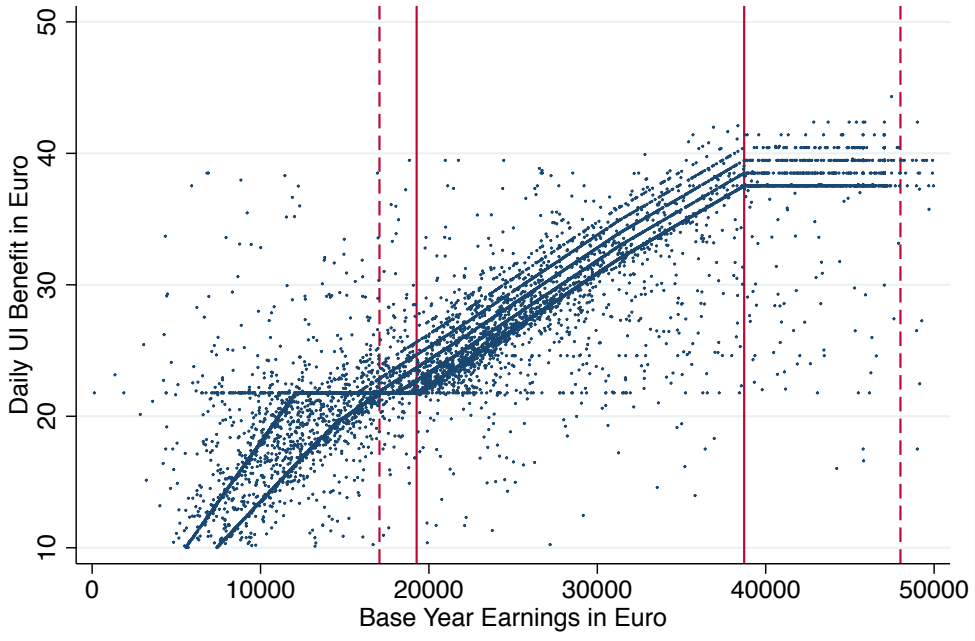
Figure 1b: UA Benefits in 2004
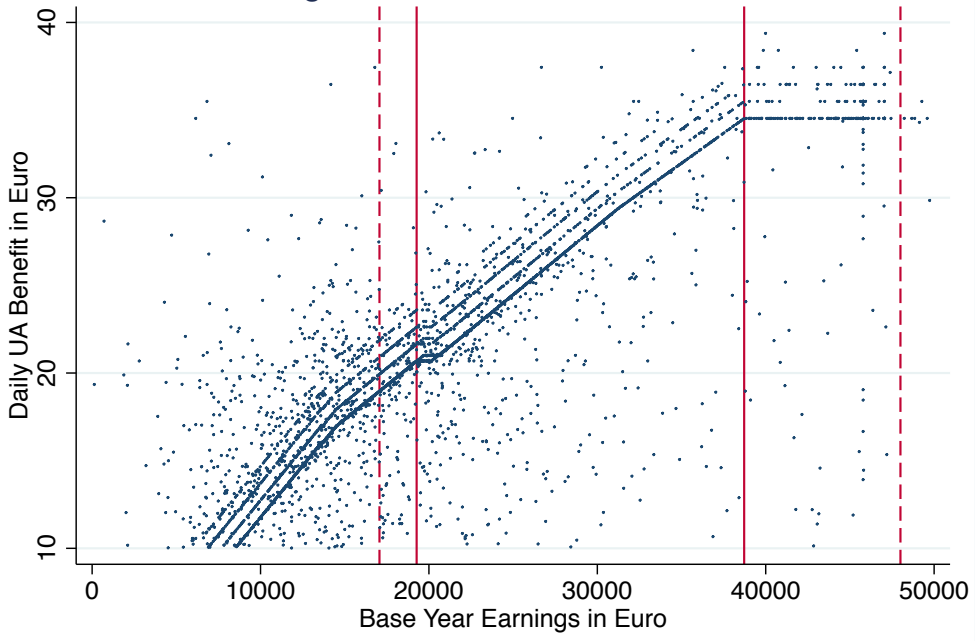
Figure 2a: Density in Bottom Kink Sample

Note: restricted cubic polynomial fit shown.
T-statistic for continuous slope at kink-point = -0.71
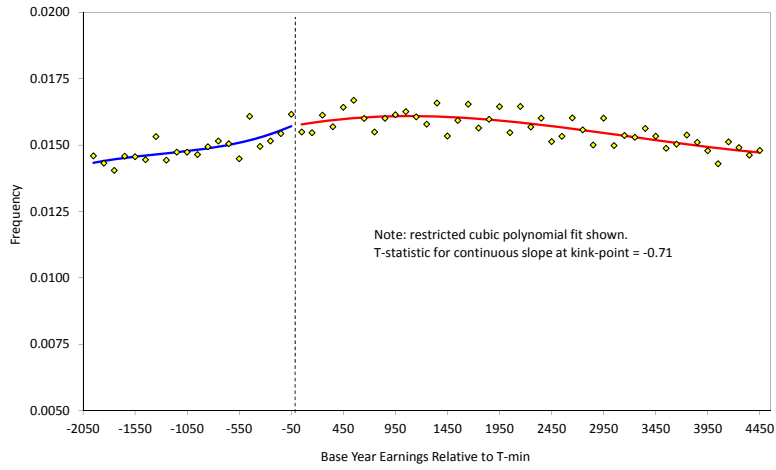
Figure 2b: Density in Top Kink Sample

Note: restricted quartic polynomial fit shown.
T-statistic for continuous slope at kink-point=1.22
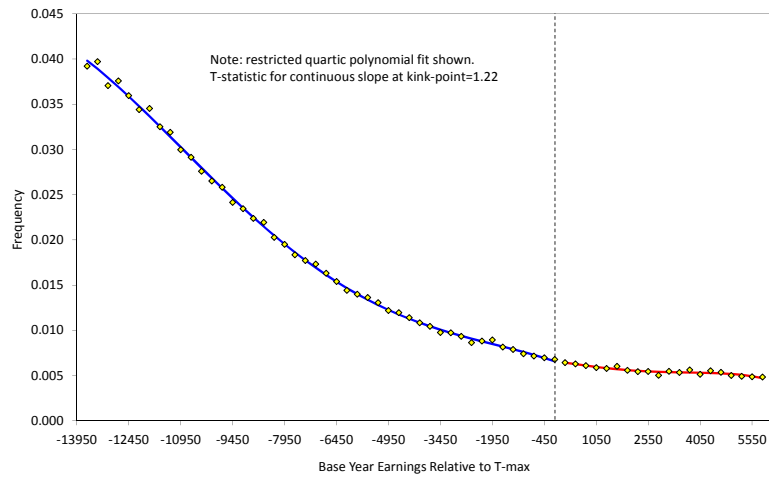
Figure 3: Daily UI Benefits
Bottom Kink Sample

Figure 4: Daily UI Benefits
Top Kink Sample
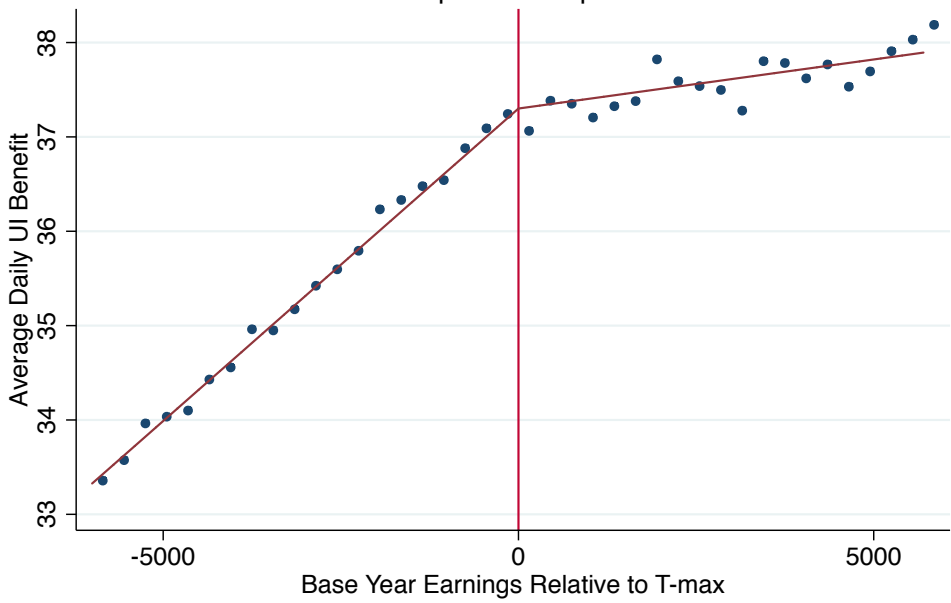
Figure 5: Log Time to Next Job
Bottom Kink Sample

Figure 6: Log Time to Next Job
Top Kink Sample

Figure 7: Log Claim Duration
Bottom Kink Sample

Figure 8: Log Claim Duration
Top Kink Sample
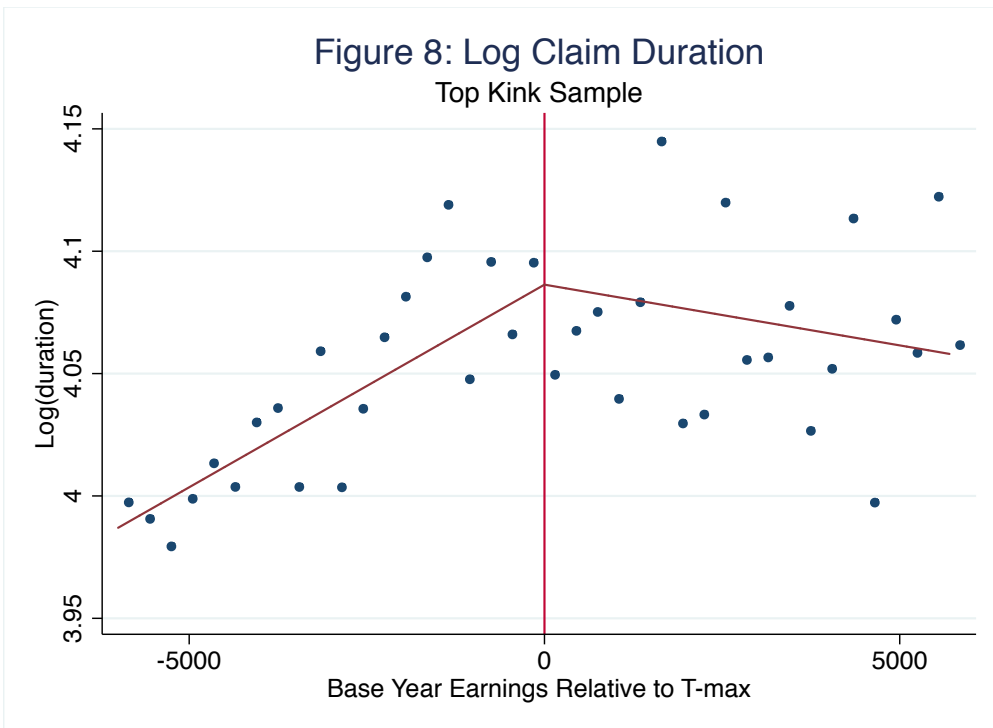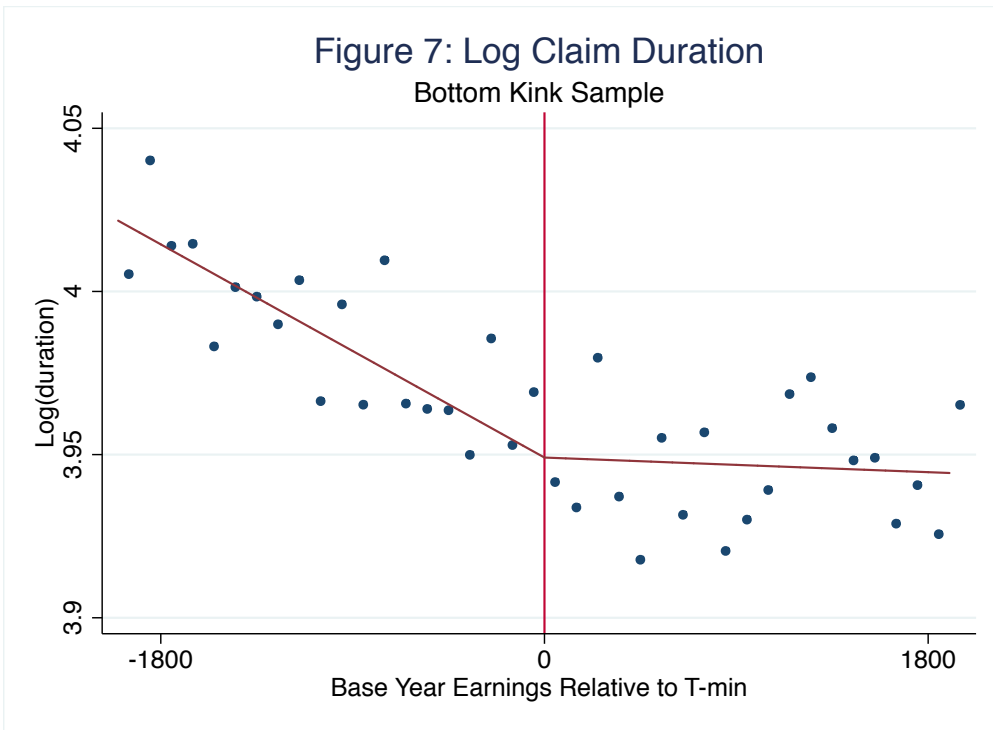
Figure 9: Covariate Indices
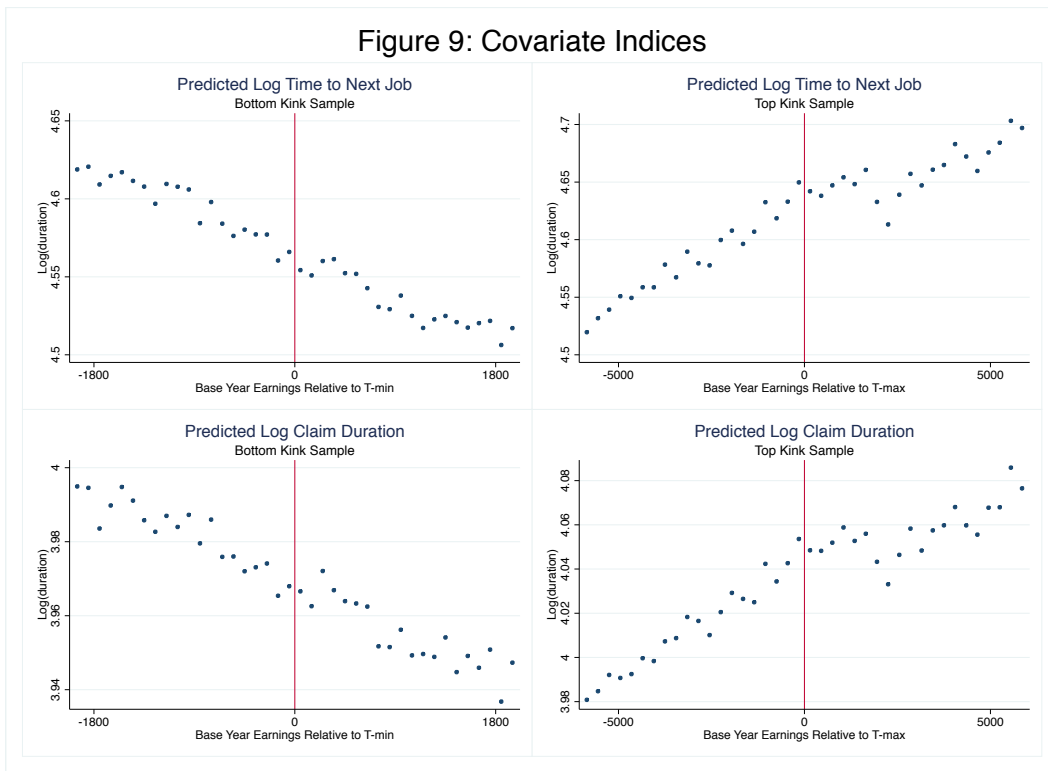
Figure 10: Fuzzy RKD Estimation with Varying Bandwidth

Log Time to Next Job, Bottom Kink Sample
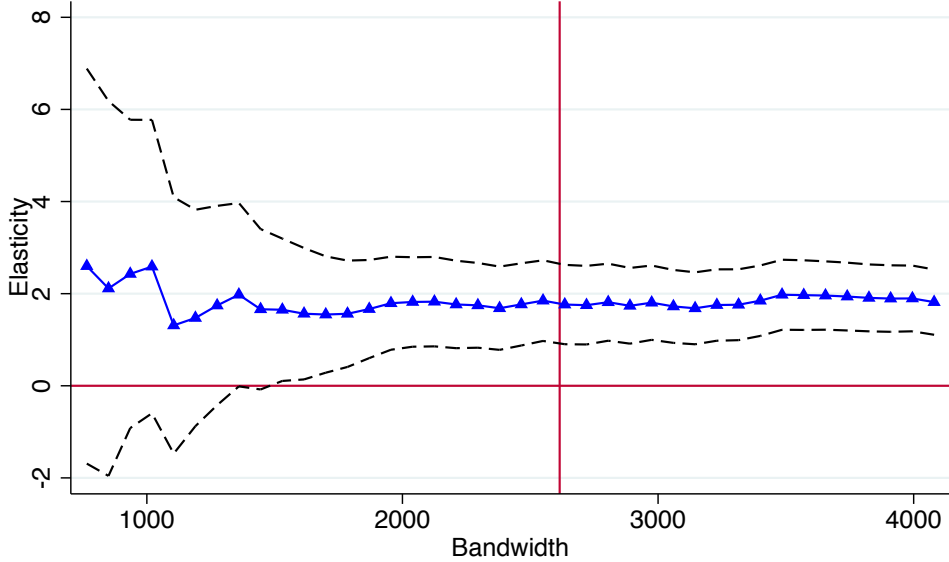
Notes: local linear estimation, vertical bar denotes FG bandwidth

Figure 11: Fuzzy RKD Estimation with Varying Bandwidth

Log Time to Next Job, Top Kink Sample
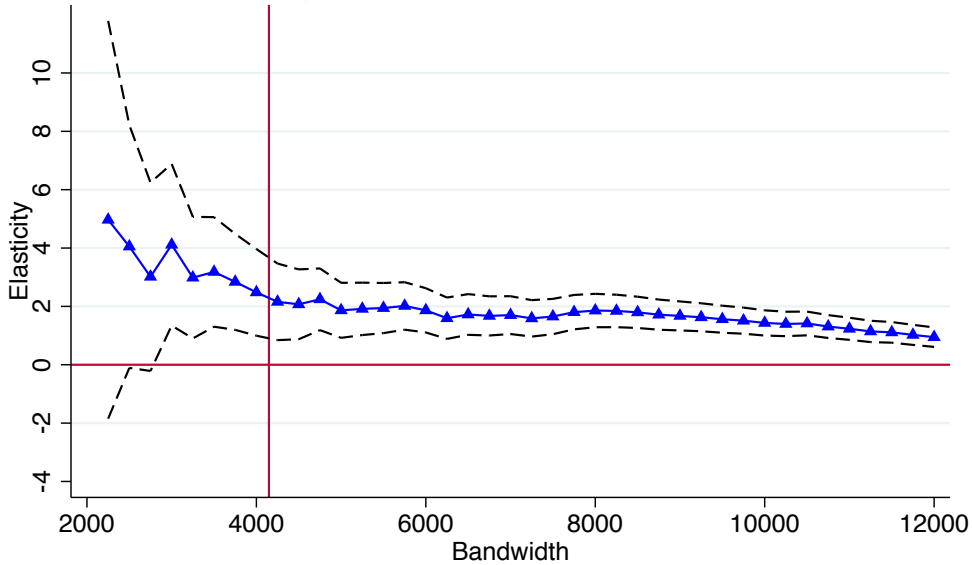
Notes: local linear estimation, vertical bar denotes FG bandwidth

Figure 12: Fuzzy RKD Estimation with Varying Bandwidth

Log Claim Duration, Bottom Kink Sample
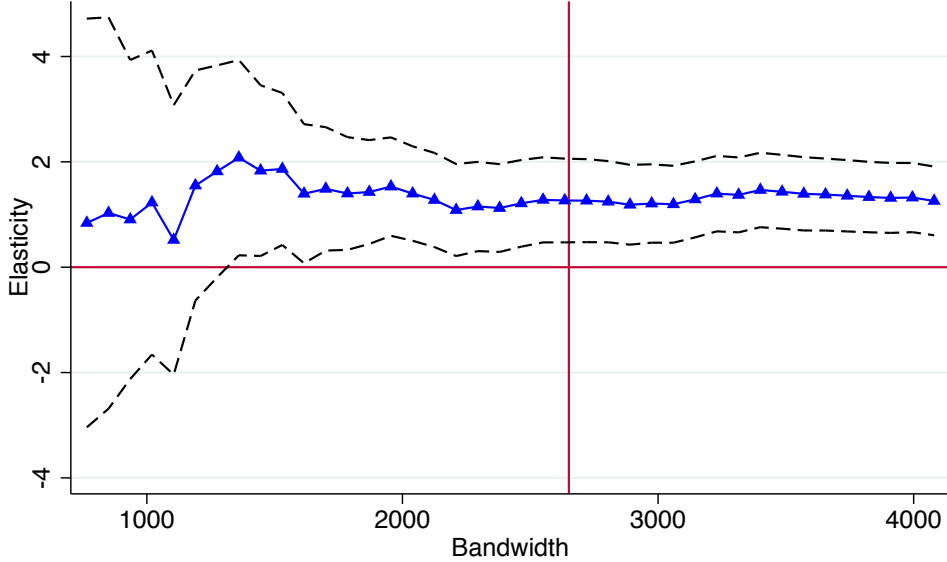
Notes: local linear estimation, vertical bar denotes FG bandwidth

Figure 13: Fuzzy RKD Estimation with Varying Bandwidth

Log Claim Duration, Top Kink Sample
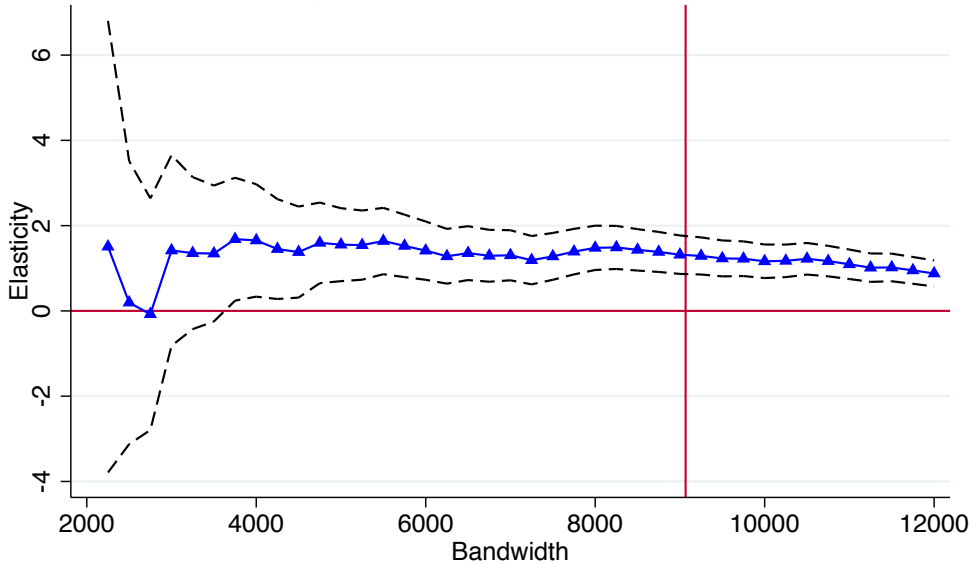
Notes: local linear estimation, vertical bar denotes FG bandwidth

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)	20,970	2,324	31,384	5,666
Daily UI benefit (euros)	24.0	2.5	31.7	4.6
Time to next job (days)*	150.4	129.1	149.3	131.6
Duration of initial UI spell (days)**	78.9	67.0	80.8	70.0
Total days of UI received	101.0	75.9	101.0	78.7
Fraction exhausted benefits***	0.10	0.30	0.09	0.29
Fraction of days to next job covered by UI	0.83	0.25	0.84	0.25
Fraction claiming UA after UI	0.07	0.26	0.07	0.25
Fraction with time to next job censored	0.19	0.40	0.21	0.40
Fraction eligible for extended benefits	0.78	0.42	0.90	0.29
Max. weeks of UI eligibility	29.88	6.30	31.92	5.56
Fraction female	0.38	0.48	0.21	0.41
Mean Age	33.1	8.4	35.9	7.4
Fraction Austrian nationals	0.84	0.36	0.90	0.30
Fraction married	0.38	0.48	0.45	0.50
Fraction bluecollar occupation	0.65	0.48	0.56	0.50
Fraction with higher education	0.11	0.32	0.19	0.39
Fraction in Vienna	0.21	0.41	0.24	0.43
Tenure in most recent job (Years)	3.42	3.35	4.25	4.20
Recalled to last job	0.21	0.41	0.26	0.44
Industry:				
Construction	0.14	0.35	0.25	0.44
Manufacturing	0.24	0.42	0.25	0.43
Trade	0.21	0.41	0.16	0.37
Services	0.27	0.44	0.24	0.43
Number observations	183,479		186,087	

Notes: sample contains UI claimants under the age of 50 with claims in 2001-2008, 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 2: Reduced Form Estimates of Kink Effects in Benefits and Durations

	Local Linear Models		Local Quadratic Models	
	FG Bandwidth (1)	Estimated Kink (2)	FG Bandwidth (3)	Estimated Kink (4)
A. Bottom Kink:				
Log daily UI benefit	2,133	0.0222 (0.0010)	4,564	0.0192 (0.0025)
Log time to next job	2,615	0.0375 (0.0093)	4,328	0.0598 (0.0280)
Log claim duration	2,651	0.0269 (0.0085)	4,564	0.0541 (0.0254)
B. Top Kink:				
Log daily UI benefit	7,064	-0.0154 (0.0006)	6,577	-0.0166 (0.0027)
Log time to next job	4,148	-0.0396 (0.0100)	7,521	-0.0577 (0.0191)
Log claim duration	9,067	-0.0221 (0.0038)	9,355	-0.0363 (0.0151)

Notes: standard errors in parentheses. See text for a description of the FG bandwidth determination.

Table 3: Estimates of Kink Effects in Distribution of Covariates

	Local Linear Models		Local Quadratic Models	
	FG Bandwidth (1)	Estimated Kink (2)	FG Bandwidth (3)	Estimated Kink (4)
A. Bottom Kink:				
Predicted log time to next job	1,989	0.0040 (0.0037)	2,409	0.0045 (0.0129)
Predicted log claim duration	2,764	0.0021 (0.0021)	2,736	0.0012 (0.0086)
Female	1,381	0.0089 (0.0088)	2,573	0.0059 (0.0170)
Age	2,363	0.3103 (0.0784)	3,652	0.4373 (0.2408)
Blue collar occupation	3,289	-0.0072 (0.0036)	3,284	-0.0354 (0.0143)
Recalled to last job	1,846	-0.0055 (0.0046)	3,322	-0.0046 (0.0038)
B. Top Kink:				
Predicted log time to next job	4,230	-0.0138 (0.0033)	6,722	-0.0251 (0.0069)
Predicted log claim duration	4,083	-0.0096 (0.0023)	7,060	-0.0176 (0.0045)
Female	2,129	-0.0222 (0.0103)	5,884	-0.0129 (0.0088)
Age	5,737	-0.0393 (0.0358)	10,128	-0.0936 (0.0936)
Blue collar occupation	3,867	0.0175 (0.0047)	7,370	0.0350 (0.0080)
Recalled to last job	7,218	0.0032 (0.0017)	10,437	0.0122 (0.0053)

Notes: standard errors in parentheses. See text for a description of the FG bandwidth determination. Predicted log time to next job and predicted log claim duration are estimated covariate indexes for these two outcomes, fit on the pooled bottom and top kink samples. The vector of 59 covariates includes dummies for gender, blue collar occupation, and being recalled to the previous job, decile of age (9 dummies), decile of previous job tenure (9 dummies), quintile of previous daily wage (4 dummies), major industry (6 dummies), region (3 dummies), year of claim (7 dummies), decile of previous firm size (9 dummies), and decile of previous firm's recall rate (9 dummies).

Table 4: Estimated Structural Coefficients from Fuzzy Regression Kink Design

	Local Linear Models		Local Quadratic Models	
	FG Bandwidth (1)	Estimated Elasticity (2)	FGl Bandwidth (3)	Estimated Elasticity (4)
A. Bottom Kink:				
Log time to next job	2,615	1.726 (0.440)	4,328	3.024 (1.501)
Log claim duration	2,651	1.250 (0.406)	4,564	2.816 (1.401)
B. Top Kink:				
Log time to next job	4,148	2.643 (0.715)	7,521	3.497 (1.278)
Log claim duration	9,067	1.312 (0.228)	9,355	2.500 (1.103)

Notes: Standard errors in parentheses. FG bandwidth is based on determination for outcome variable. Estimated elasticities in columns 2 and 4 are obtained from a 2SLS procedure, instrumenting the treatment variable B with the interaction term D × V, see text section 3.2.

Supplemental Appendix–Identification

Proof of Proposition 1: For part (a), we take the derivative

$$
\begin{aligned}
\frac{d}{dv}\Pr(U \leqslant u|V = v) &= \frac{d}{dv}\left(\frac{f_{V|U\leq u}(v)}{f_V(v)}\Pr[U \leq u]\right)\\[2mm]
&= \frac{d}{dv}\left(\sum_{u'\leqslant u}\frac{f_{V|U=u'}(v)}{f_V(v)}\Pr(U = u')\right)\\[2mm]
&= \sum_{u'\leqslant u}\frac{d}{dv}\frac{f_{V|U=u'}(v)}{f_V(v)}\Pr(U = u') \qquad (11)
\end{aligned}
$$

where the first line follows by Bayes' Rule. Since $f_{V|U=u}(v)$ is continuously differentiable in v at 0 for all u by Assumption 4 and that (11) involves a finite summation, $\frac{d}{dv}\Pr(U \leqslant u|V = v)$ is continuous at 0.

For part (b), in the numerator,

$$
\begin{aligned}
\lim_{v_0\to0^+}\frac{dE[Y|V = v]}{dv}\bigg|_{v=v_0} &= \lim_{v_0\to0^+}\frac{d}{dv}\sum_u y(b(v),v,u)\Pr(U = u|V = v)\bigg|_{v=v_0}\\[2mm]
&= \lim_{v_0\to0^+}\frac{d}{dv}\sum_u y(b(v),v,u)\frac{f_{V|U=u}(v)}{f_V(v)}\Pr(U = u)\bigg|_{v=v_0}\\[2mm]
&= \lim_{v_0\to0^+}\sum_u \frac{\partial}{\partial v}y(b(v),v,u)\frac{f_{V|U=u}(v)}{f_V(v)}\Pr(U = u)\bigg|_{v=v_0}\\[2mm]
&= \lim_{v_0\to0^+}b'(v_0)\sum_u y_1(b(v_0),v_0,u)\frac{f_{V|U=u}(v_0)}{f_V(v_0)}\Pr(U = u) +\\[2mm]
&\quad\lim_{v_0\to0^+}\sum_u\left\{y_2(b(v_0),v_0,u)\frac{f_{V|U=u}(v_0)}{f_V(v_0)}+y(b(v_0),v_0,u)\frac{\partial}{\partial v}\frac{f_{V|U=u}(v_0)}{f_V(v_0)}\right\}\Pr(U = u).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\lim_{v_0\to0^-}\frac{dE[Y|V = v]}{dv}\bigg|_{v=v_0} &= \lim_{v_0\to0^-}b'(v_0)\sum_u y_1(b(v_0),v_0,u)\frac{f_{V|U=u}(v_0)}{f(v_0)}\Pr(U = u) +\\[2mm]
&\quad\lim_{v_0\to0^-}\left\{\sum_u y_2(b(v_0),v_0,u)\frac{f_{V|U=u}(v_0)}{f(v_0)}+y(b(v_0),v_0,u)\frac{\partial}{\partial v}\frac{f_{V|U=u}(v_0)}{f(v_0)}\right\}\Pr(U = u).
\end{aligned}
$$

The continuity assumed in Assumptions 1-4 implies that terms can be eliminated and combined to obtain

$$
\begin{aligned}
&\lim_{v_0\to0^+}\frac{dE[Y|V = v]}{dv}\bigg|_{v=v_0} - \lim_{v_0\to0^-}\frac{dE[Y|V = v]}{dv}\bigg|_{v=v_0}\\[2mm]
&=(\lim_{v\to0^+}b'(v_0) - \lim_{v\to0^-}b'(v_0))\sum_u y_1(b(0),0,u)\frac{f_{V|U=u}(0)}{f_V(0)}\Pr(U = u).
\end{aligned}
$$

Assumption 3 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

$$\frac{\lim_{v_0 \to 0^+} \frac{dE[Y|V=v]}{dv}\Big|_{v=v_0} - \lim_{v_0 \to 0^-} \frac{dE[Y|V=v]}{dv}\Big|_{v=v_0}}{\lim_{v_0 \to 0^+} b'(v_0) - \lim_{v_0 \to 0^-} b'(v_0)} = E[y_1(b(0),0,U)|V=0] = \sum_u y_1(b(0),0,u) \frac{f_{V|U=u}(0)}{f_V(0)} \Pr(U=u),$$

which completes the proof.

In order to prove Proposition 2, we first present and prove the following Lemma's.

Lemma 1. *Let S be a sub-vector of the random vector $(U, U_B, U_{V'})$ that at least includes U, and S^* the vector of the remaining random variables. Then $f_{V|S=s}(v)$ is continuously differentiable in v for all s.*

To see this, note that

$$f_{V|S=s,S^*=s^*}(v) = f_{V,U_B,U_{V'}|U=u}(v,u_B,u_{V'}) \frac{\Pr(U=u)}{\sum_\omega f_{U_B,U_{V'}|U=\omega}(u_B,u_{V'})\Pr(U=\omega)}$$

$$\int f_{V|S=s,S^*=s^*}(v) f_{S^*|S=s}(s^*) ds^* = \int \left[f_{V,U_B,U_{V'}|U=u}(v,u_B,u_{V'}) \frac{\Pr(U=u)}{\sum_\omega f_{U_B,U_{V'}|U=\omega}(u_B,u_{V'})\Pr(U=\omega)} \right] f_{S^*|S=s}(s^*) ds^*$$

$$f_{V|S=s}(v) = \int \left[f_{V,U_B,U_{V'}|U=u}(v,u_B,u_{V'}) \frac{\Pr(U=u)}{\sum_\omega f_{U_B,U_{V'}|U=\omega}(u_B,u_{V'})\Pr(U=\omega)} \right] f_{S^*|S=s}(s^*) ds^*,$$

where the first line follows by Bayes' rule, and we integrate both sides over $f_{S^*|S=s}(s^*)$ in the second line. Taking derivatives on both sides of the third line, interchanging differentiation and integration (permitted by Assumption 4a and the continuity $f_{U_B,U_{V'}|U=u}$ over a compact rectangle as per Theorem 9.42 of Rudin (1976)) and using Assumption 4a again, we obtain the result. QED.

Lemma 2. $\frac{\partial f_{V^*|S=s}(v^*)}{\partial v^*}$ is continuous for all v^* and s.

To see this, note that after applying Bayes' Rule, and re-arranging, we obtain

$$f_{V^*|S=s,S^*=s^*}(v^*) = \Pr[U_V=0|S=s,S^*=s^*] f_{V|S=s,S^*=s^*,U_V=0}(v^*)$$

$$+ \Pr[U_V \neq 0|S=s,S^*=s^*] f_{V^*|S=s,S^*=s^*,U_V \neq 0}(v^*)$$

$$= \Pr[U_V=0|S=s,S^*=s^*] \Pr[U_V=0|V=v^*,S=s,S^*=s^*] \frac{f_{V|S=s,S^*=s^*}(v^*)}{\Pr[U_V=0|S=s,S^*=s^*]}$$

$$+ \Pr[U_V \neq 0|S=s,S^*=s^*] \Pr[U_V \neq 0|V=v^*-u_{V'},S=s,S^*=s^*] \frac{f_{V|S=s,S^*=s^*}(v^*-u_{V'})}{\Pr[U_V \neq 0|S=s,S^*=s^*]}$$

$$= \Pr[U_V=0|V=v^*,S=s,S^*=s^*] f_{V|S=s,S^*=s^*}(v^*)$$

$$+ \Pr[U_V \neq 0|V=v^*-u_{V'},S=s,S^*=s^*] f_{V|S=s,S^*=s^*}(v^*-u_{V'})$$

Multiplying both sides of the last line by $f_{S^*|S=s}(s^*)$ and integrating over s^*, taking the partial derivative with respect to v^*, and applying Assumptions 4a and 5 and the reasoning in Lemma 1, we have the desired result. QED.

Lemma 3. $\frac{\partial \Pr[U_V=0|V^*=v^*,S=s]}{\partial v^*}$ *is continuous for each* v^* *and s.*

To see this,

$$
\begin{aligned}
\Pr[U_V = 0|V^* = v^*, S = s] \;&=\; \Pr[V = V^*|V^* = v^*, S = s] \\
&=\; f_{V^*|S=s,V=V^*}(v^*)\frac{\Pr[V = V^*|S = s]}{f_{V^*|S=s}(v^*)} \\
&=\; f_{V|S=s,V=V^*}(v^*)\frac{\Pr[V = V^*|S = s]}{f_{V^*|S=s}(v^*)} \\
&=\; f_{V|S=s,V=V^*}(v^*)\frac{\Pr[V = V^*|S = s]}{f_{V|S=s}(v^*)}\frac{f_{V|S=s}(v^*)}{f_{V^*|S=s}(v^*)} \\
&=\; \Pr[V = V^*|V = v^*, S = s]\frac{f_{V|S=s}(v^*)}{f_{V^*|S=s}(v^*)} \\
&=\; \int \pi(v^*, u, u_B, u_{V'})f_{S^*|V=v^*,S=s}(s^*)\,ds^*\frac{f_{V|S=s}(v^*)}{f_{V^*|S=s}(v^*)} \\
&=\; \int \pi(v^*, u, u_B, u_{V'})f_{V^*|S^*=s^*,S=s}(v^*)\frac{f_{S^*|S=s}(s^*)}{f_{V^*|S=s}(v^*)}\,ds^*\frac{f_{V|S=s}(v^*)}{f_{V^*|S=s}(v^*)}.
\end{aligned}
$$

where the partial derivative of the right hand side in the last line is continuous for all v^* and s by Assumption 5 and Lemmas 1 and 2. QED.

Proof of Proposition 2

For part (a), the proof is the same as for part (a) in Proposition 1, replacing V with V^* and using Lemma 2.

For part (b), we can write

$$
\begin{aligned}
E[Y|V^* = v^*] \;&=\; \sum_u E[Y|V^* = v^*, U = u]\Pr(U = u|V^* = v^*) & (12) \\
&=\; \sum_u (E[Y|U_V = 0, V^* = v^*, U = u]\Pr[U_V = 0|V^* = v^*, U = u] + \\
&\quad\; E[Y|U_V \neq 0, V^* = v^*, U = u]\Pr[U_V \neq 0|V^* = v^*, U = u])\Pr(U = u|V^* = v^*) \\
&=\; \sum_u \left(z_1 z_2 z_3 + \left[\int z_4 z_5 \, du_{V'}\right]\cdot[1 - z_2 z_3]\right)z_6. & (13)
\end{aligned}
$$

where the second line follows from the law of iterated expectations, and to ease exposition below, we use

the notation:

$$z_1 \equiv y\left(b\left(v^*,u\right),v^*,u\right)$$

$$z_2 \equiv \Pr\left[V = V^* | V = v^*, U = u\right]$$

$$z_3 \equiv \frac{f_{V|U=u}\left(v^*\right)}{f_{V^*|U=u}\left(v^*\right)}$$

$$z_4 \equiv y\left(b\left(v^* - u_{V'},u\right),v^* - u_{V'},u\right)$$

$$z_5 \equiv f_{U_{V'}|U_V \neq 0,V^*=v^*,U=u}\left(u_{V'}\right)$$

$$z_6 \equiv \Pr(U = u | V^* = v^*).$$

Note that we have used the fact that $\Pr\left[U_V = 0 | V^* = v^*, U = u\right] = z_2 z_3$ because

$$
\begin{aligned}
\Pr\left[U_V = 0 | V^* = v^*, U = u\right] &= \Pr\left[V = V^* | V^* = v^*, U = u\right] \\
&= f_{V^*|U=w,V=V^*}\left(v^*\right)\frac{\Pr\left[V = V^* | U = u\right]}{f_{V^*|U=u}\left(v^*\right)} \\
&= f_{V|U=u,V=V^*}\left(v^*\right)\frac{\Pr\left[V = V^* | U = u\right]}{f_{V^*|U=u}\left(v^*\right)} \\
&= f_{V|U=u,V=V^*}\left(v^*\right)\frac{\Pr\left[V = V^* | U = u\right]}{f_{V|U=u}\left(v^*\right)}\frac{f_{V|U=u}\left(v^*\right)}{f_{V^*|U=u}\left(v^*\right)} \\
&= \Pr\left[V = V^* | V = v^*, U = u\right]\frac{f_{V|U=u}\left(v^*\right)}{f_{V^*|U=u}\left(v^*\right)}
\end{aligned}
$$

where the second and fifth lines follow from Bayes' rule.

The derivative of $E\left[Y | V^* = v^*\right]$ in equation (12) with respect to v^* is

$$\frac{dE\left[Y | V^* = v^*\right]}{dv^*} = \sum_u z_1' z_2 z_3 z_6 + \sum_u z_1 \frac{\partial\left(z_2 z_3 z_6\right)}{\partial v^*} \tag{14}$$
$$+ \sum_u \frac{\partial\left[\left(\int z_4 z_5 du_{V'}\right)\left[1 - z_2 z_3\right] z_6\right]}{\partial v^*}$$

where z_j' denotes the partial derivative of z_j with respect to v^*.

In a parallel fashion, we can write

$$E\left[B^* | V^* = v^*\right] = \sum_u \left(\left[\int z_7 z_8 du_B\right] z_2 z_3 + \left[\int \int z_9 z_{10} du_{V'} du_B\right] \cdot \left[1 - z_2 z_3\right]\right) z_6$$

with

$$z_7 \equiv b(v^*, u) + u_B$$

$$z_8 \equiv f_{U_B|U_V=0, V^*=v^*, U=u}(u_B)$$

$$z_9 \equiv b(v^* - u_{V'}, u) + u_B$$

$$z_{10} \equiv f_{U_B, U_{V'}|U_V \neq 0, V^*=v^*, U=u}(u_B, u_{V'}).$$

And the analogous derivative with respect to v^* is

$$\frac{dE[B^*|V^*=v^*]}{dv^*} = \sum_u \left(\left[\int z_7' z_8 du_B \right] z_2 z_3 \right) z_6 + \sum_u \int z_7 \frac{\partial (z_8 z_2 z_3 z_6)}{\partial v^*} du_B \qquad (15)$$

$$+ \sum_u \frac{\partial \left[\left(\int \int z_9 z_{10} du_{V'} du_B \right) [1 - z_2 z_3] z_6 \right]}{\partial v^*}.$$

The proof of part (b) follows from showing that the partial derivatives of $z_2, z_3, \int z_4 z_5 du_{V'}, z_6, z_8$, as well as $\int \int z_9 z_{10} du_{V'} du_B$ with respect of v^* are continuous (or we say interchangeably that these functions are "continuously differentiable in v^*"), and noting that that z_1 and z_7 are continuous by Assumptions 1, 2, and 3a. From this it follows that there is no discontinuity in the second and third terms of the right hand sides of (14) and (15) at $v^* = 0$. The RKD estimand can then be shown to be the ratio of the discontinuities in the first terms of those two equations.

As shown by Lemma 3, z_2 is continuously differentiable in v^*.

The quantity z_3 is continuously differentiable by Lemmas 1 and 2 and Assumption 7, and z_5 is continuously differentiable in v^* because

$$f_{U_{V'}|U_V \neq 0, V^*=v^*, U=u}(u_{V'}) = \frac{\Pr[U_V \neq 0|U_{V'}=u_{V'}, V^*=v^*, U=u] f_{U_{V'}|V^*=v^*, U=u}(u_{V'})}{\Pr[U_V \neq 0|V^*=v^*, U=u]}$$

$$= \frac{(1 - \Pr[U_V = 0|U_{V'}=u_{V'}, V^*=v^*, U=u]) \frac{f_{V^*|U_{V'}=u_{V'}, U=u}(v^*) f_{U_{V'}|U=u}(u_{V'})}{f_{V^*|U=u}(v^*)}}{1 - z_2 z_3},$$

and the derivative of the last line is continuous by Lemmas 2 and 3.

The expression $\int z_4 z_5 du_{V'}$ is continuously differentiable because, by a change of variables where $q = v^* - u_{V'}$, it is equal to $- \int y(q, b(q, u), u) f_{U_{V'}|U_V \neq 0, V^*=v^*, U=u}(v^* - q) dq$, which is continuously differentiable in v^* by the continuous differentiability of z_5.

We see that z_6 is continuously differentiable in v^* because $\Pr(U=u|V^*=v^*) = f_{V^*=v^*|U=u}(v^*)\frac{\Pr(U=u)}{\sum_\omega f_{V^*|U=\omega}(v^*)\Pr(U=\omega)}$; z_7 is continuously differentiable by Assumption 3a, and z_8 is continuously differentiable in v^* because

$$
\begin{aligned}
f_{U_B|U_V=0,V^*=v^*,U=u}(u_B) &= \frac{\Pr[U_V=0|V^*=v^*,U=u,U_B=u_B]}{\Pr[U_V=0|V^*=v^*,U=u]}f_{U_B|V^*=v^*,U=u}(u_B)\\
&= \frac{\Pr[U_V=0|V^*=v^*,U=u,U_B=u_B]}{\Pr[U_V=0|V^*=v^*,U=u]}f_{V^*|U_B=u_B,U=u}(v^*)\frac{f_{U_B|U=u}(u_B)}{f_{V^*|U=u}(v^*)}
\end{aligned}
$$

where the derivative of the last line is continuous in v^* by Lemmas 2 and 3.

The quantity z_{10} is continuously differentiable in v^* because

$$
\begin{aligned}
f_{U_B,U_{V'}|U_V\neq 0,V^*=v^*,U=u}(u_B,u_{V'}) &= \Pr[U_V\neq 0|U_{V'}=u_{V'},U_B=u_B,V^*=v^*,U=u]\frac{f_{U_{V'},U_B|V^*=v^*,U=u}(u_{V'},u_B)}{\Pr[U_V\neq 0|V^*=v^*,U=u]}\\
&= \Pr[U_V\neq 0|U_{V'}=u_{V'},U_B=u_B,V^*=v^*,U=u]\frac{f_{V^*|U_{V'}=u_{V'},U_B=u_B,U=u}(v^*)}{\Pr[U_V\neq 0|V^*=v^*,U=u]}\\
&\quad \frac{f_{U_{V'},U_B|U=u}(u_{V'},u_B)}{f_{V^*|U=u}(v^*)},
\end{aligned}
$$

and the derivative of the last line is continuous by Lemmas 2 and 3.

The expression $\int\int z_9 z_{10}\,du_{V'}\,du_B$ is continuously differentiable because with a change of variables ($q=v^*-u_{V'}$), it is equal to $-\int\int(b(q,u)+u_B)f_{U_B,U_{V'}|U_V\neq 0,V^*=v^*,U=u}(u_B,v^*-q)\,dq\,du_B$, which is continuously differentiable in v^* by Assumption 3a and the continuous differentiability of z_{10}.

As a result of the smoothness of these above terms, we can write

$$
\lim_{v_0\to 0^+}\frac{dE[Y|V^*=v^*]}{dv^*}\Big|_{v^*=v_0} - \lim_{v_0\to 0^-}\frac{dE[Y|V^*=v^*]}{dv^*}\Big|_{v^*=v_0} \tag{16}
$$

$$
= \lim_{v_0\to 0^+}\sum_u z_1' z_2 z_3 z_6 - \lim_{v_0\to 0^-}\sum_u z_1' z_2 z_3 z_6 \tag{17}
$$

$$
= \sum_u\left(\lim_{v_0\to 0^+}z_1' - \lim_{v_0\to 0^-}z_1'\right)z_2 z_3 z_6\big|_{v^*=v_0} \tag{18}
$$

$$
= \sum_u y_1\left(b(0,u),0,u\right)\left(b_1^+(u)-b_1^-(u)\right)z_2 z_3 z_6\big|_{v^*=v_0} \tag{19}
$$

where the last line follows from Assumptions 1, 2, and 3a.

Similarly, we can write

$$\lim_{v_0 \to 0^+} \frac{dE\left[B^*|V^*=v^*\right]}{dv^*}\bigg|_{v^*=v_0} - \lim_{v_0 \to 0^-} \frac{dE\left[B^*|V^*=v^*\right]}{dv^*}\bigg|_{v^*=v_0} \tag{20}$$

$$= \lim_{v_0 \to 0^+} \sum_u \left(\left[\int z_7' z_8 \, du_B\right] z_2 z_3\right) z_6 \, du - \lim_{v_0 \to 0^-} \int \left(\left[\int z_7' z_8 \, du_B\right] z_2 z_3\right) z_6 \tag{21}$$

$$= \sum_u \left(\lim_{v_0 \to 0^+} \left[\int z_7' z_8 \, du_B\right] - \lim_{v_0 \to 0^-} \left[\int z_7' z_8 \, du_B\right]\right) z_2 z_3 z_6|_{v^*=v_0} \tag{22}$$

$$= \sum_u \left(\left[\int z_8|_{v^*=v_0} \, du_B\right]\right) \left(b_1^+(u) - b_1^-(u)\right) z_2 z_3 z_6|_{v^*=v_0} \tag{23}$$

$$= \sum_u \left(b_1^+(u) - b_1^-(u)\right) z_2 z_3 z_6|_{v^*=v_0} . \tag{24}$$

Finally, consider the term $z_2 z_3 z_6|_{v^*=v_0}$, which, after applying Bayes' rule and re-arranging, is equal to

$$z_2 z_3 z_6|_{v^*=v_0} = \Pr\left[V = V^*|V = 0, U = u\right] \frac{f_{V|U=u}(0)}{f_{V^*|U=u}(0)} \Pr(U = u|V^* = 0)$$

$$= \Pr\left[V = V^*|V = 0, U = u\right] \frac{f_{V|U=u}(0)}{f_V(0)} \Pr(U = u) \frac{f_V(0)}{\Pr(U = u)} \frac{\Pr(U = u|V^* = 0)}{f_{V^*|U=u}(0)}$$

$$= \Pr\left[V = V^*|V = 0, U = u\right] \frac{f_{V|U=u}(0)}{f_V(0)} \Pr(U = u) \frac{f_V(0)}{f_{V^*}(0)} \frac{\Pr(U = u|V^* = 0) f_{V^*}(0)}{\Pr(U = u) f_{V^*|U=u}(0)}$$

$$= \Pr\left[V = V^*|V = 0, U = u\right] \frac{f_{V|U=u}(0)}{f_V(0)} \Pr(U = u) \frac{f_V(0)}{f_{V^*}(0)} .$$

Because $\frac{f_V(0)}{f_{V^*}(0)}$ can be pulled out of the integral in both (19) and (24), we have the result

$$\frac{\lim_{v_0 \to 0^+} \frac{dE[Y|V^*=v^*]}{dv^*}\big|_{v^*=v_0} - \lim_{v_0 \to 0^-} \frac{dE[Y|V^*=v^*]}{dv^*}\big|_{v^*=v_0}}{\lim_{v_0 \to 0^+} \frac{dE[B^*|V^*=v^*]}{dv^*}\big|_{v^*=v_0} - \lim_{v_0 \to 0^-} \frac{dE[B^*|V^*=v^*]}{dv^*}\big|_{v^*=v_0}} = \sum_u y_1\left(b\left(0, u\right), 0, u\right) \varphi\left(u\right) \Pr(U = u)$$

where $\varphi\left(u\right) = \dfrac{\Pr[U_V = 0|V = 0, U = u]\left(b_1^+(u) - b_1^-(u)\right) \frac{f_{V|U=u}(0)}{f_V(0)}}{\sum_u \Pr[U_V = 0|V = 0, U = \omega]\left(b_1^+(\omega) - b_1^-(\omega)\right) \frac{f_{V|U=\omega}(0)}{f_V(0)} \Pr(U = u)}$. Note that Assumptions 6 and 7 guarantee non-negative, finite weights. QED.

Supplemental Appendix–Estimation Theory

Asymptotic Distribution Theory of the RKD Estimators

Notations: We follow Fan and Gijbels (1996) and Hahn et al. (1999) and define

- $\mu_j^+ \equiv \int_0^\infty u^j K(u) du$, $\mu_j^- \equiv \int_{-\infty}^0 u^j K(u) du$ and $\mathbf{S}^\pm \equiv (\mu_{j+l}^\pm)_{0 \leqslant j,l \leqslant p}$; $\mathbf{c}_p^\pm \equiv (\mu_{p+1}^\pm, ..., \mu_{2p+1}^\pm)$;

- $v_j^+ \equiv \int_0^\infty u^j K^2(u) du$, $v_j^- \equiv \int_{-\infty}^0 u^j K^2(u) du$, and $\mathbf{S}^{*\pm} \equiv (v_{j+l}^\pm)_{0 \leqslant j,l \leqslant p}$; the $\mathbf{S}^\pm$ and $\mathbf{S}^{*\pm}$ defined here are not related to the S and S^* in the proof of Proposition 2.

- Let $\mathbf{H}$ be the $(p+1) \times (p+1)$ matrix $\mathbf{Diag}(1, h, ...h^p)$.

- Define

$$\varsigma(v) \equiv m(v) - \sum_{l=0}^p \beta_l v^l - \frac{1}{(p+1)!} m^{(p+1)}(0) v^{p+1}$$

$$\xi(v) \equiv r(v) - \sum_{l=0}^p \kappa_l v^l - \frac{1}{(p+1)!} r^{(p+1)}(0) v^{p+1}$$

for a local polynomial regression of order p. Note that Taylor's Theorem and Assumption 10 imply

$$\sup_{0 < v \leqslant Mh} |\varsigma(v)| = o(h^2) \qquad \sup_{0 < v \leqslant Mh} |\xi(v)| = o(h^2) \, for \, p = 1$$

$$\sup_{0 < v \leqslant Mh} |\varsigma(v)| = o(h^3) \qquad \sup_{0 < v \leqslant Mh} |\xi(v)| = o(h^3) \, for \, p = 2.$$

Proof of Proposition 3, 4 and **5:** The proofs are analogous to those in Hahn et al. (1999) and are available upon request.

Proof of Proposition 6: We prove the proposition for the local linear case, and the L superscript in most of the estimators is omitted. Proof for the local quadratic case is analogous.

1) Asymptotic distribution of the restricted sharp RKD estimator.

Extending the Hahn et al. (1999) framework, we let $\mathbf{Y} = (Y_1^-, ..., Y_{n^-}^-, Y_1^+ ..., Y_{n^+}^+)^T$ be the "stacked" $n \times 1$ outcome variable vector where the first n^- entries are observations to the left of the threshold and the last n^+ entries are those to the right of the threshold. Let Z be the $n \times 4$ matrix whose i^{th} row is $(1, \frac{V_i^-}{h}, 0, 0)$ for $i \leqslant n^-$ or $(0, 0, 1, \frac{V_i^+}{h})$ for $i > n^-$. Also let $\mathbf{W}_K = \begin{pmatrix} \mathbf{W}_K^- & 0 \\ 0 & \mathbf{W}_K^+ \end{pmatrix}$ with $\mathbf{W}_K^\pm$ being the diagonal matrices

$\mathbf{Diag}(K(\frac{V_1^\pm}{h}),...,K(\frac{V_{n^\pm}^\pm}{h}))$. The restricted sharp RKD estimator can be obtained from solving the constrained least squares problem

$$\min_{\tilde{\beta}^R} (\mathbf{Y} - \mathbf{Z}\tilde{\beta}^R)^T \mathbf{W}_K (\mathbf{Y} - \mathbf{Z}\tilde{\beta}^R)$$

$$\text{s.t.} \quad \mathbf{R}\tilde{\beta}^R = 0 \tag{25}$$

where $\mathbf{R} = (1,0,-1,0)$.

Denote the resulting estimator by $\hat{\beta}^R = (\hat{\beta}_0^{R-}, h\hat{\beta}_1^{R-}, \hat{\beta}_0^{R+}, h\hat{\beta}_1^{R+})^T$ and define $\mathbf{Y}^* = \mathbf{Y} - \mathbf{Z}\beta$. It is a standard result (e.g. a generalization of (1.4.5) of Amemiya (1985)) that

$$(\hat{\beta}^R - \beta) \quad = \quad [(\mathbf{Z}^T\mathbf{W}_K\mathbf{Z})^{-1} - (\mathbf{Z}^T\mathbf{W}_K\mathbf{Z})^{-1}\mathbf{R}^T(\mathbf{R}(\mathbf{Z}^T\mathbf{W}_K\mathbf{Z})^{-1}\mathbf{R}^T)^{-1}\mathbf{R}(\mathbf{Z}^T\mathbf{W}_K\mathbf{Z})^{-1}]\mathbf{Z}^T\mathbf{W}_K\mathbf{Y}^*.$$

It is straightforward to show that $\hat{\beta}_1^{R+} - \hat{\beta}_1^{R-} = \hat{\delta}_1^S$ and hence $\frac{\hat{\beta}_1^{R+} - \hat{\beta}_1^{R-}}{\kappa_1^+ - \kappa_1^-} = \hat{\tau}_{SRKD}^{LR}$ is the restricted sharp RKD estimator defined in the text.

Under Assumptions 9, 10L, 11L, 12-15 and 16L, the same reasoning in Hahn et al. (1999) gives

$$\frac{1}{\sqrt{nh}} \left(\mathbf{Z}^T\mathbf{W}_K\mathbf{Y}^* \right) \Rightarrow N(0, f(0)\Sigma_{sharp}) \tag{26}$$

with

$$\Sigma_{sharp} = \begin{pmatrix} \sigma_Y^2(0^-)\mathbf{S}^{*-} & 0 \\ 0 & \sigma_Y^2(0^+)\mathbf{S}^{*+} \end{pmatrix}.$$

Next, we examine the asymptotic behavior of

$$(\mathbf{Z}^T\mathbf{W}_K\mathbf{Z})^{-1} - (\mathbf{Z}^T\mathbf{W}_K\mathbf{Z})^{-1}\mathbf{R}^T(\mathbf{R}(\mathbf{Z}^T\mathbf{W}_K\mathbf{Z})^{-1}\mathbf{R}^T)^{-1}\mathbf{R}(\mathbf{Z}^T\mathbf{W}_K\mathbf{Z})^{-1}.$$

Lemma 1 of Hahn et al. (1999) implies that

$$\frac{1}{nh}\mathbf{Z}^T\mathbf{W}_K\mathbf{Z} \to f(0) \begin{pmatrix} \mathbf{S}^- & 0 \\ 0 & \mathbf{S}^+ \end{pmatrix},$$

and as a consequence,

$$nh\{(\mathbf{Z}^T\mathbf{W}_K\mathbf{Z})^{-1} - (\mathbf{Z}^T\mathbf{W}_K\mathbf{Z})^{-1}\mathbf{R}^T(\mathbf{R}(\mathbf{Z}^T\mathbf{W}_K\mathbf{Z})^{-1}\mathbf{R}^T)^{-1}\mathbf{R}(\mathbf{Z}^T\mathbf{W}_K\mathbf{Z})^{-1}\}$$

$$= (\frac{1}{nh}\mathbf{Z}^T\mathbf{W}_K\mathbf{Z})^{-1} - (\frac{1}{nh}\mathbf{Z}^T\mathbf{W}_K\mathbf{Z})^{-1}\mathbf{R}^T(\mathbf{R}(\frac{1}{nh}\mathbf{Z}^T\mathbf{W}_K\mathbf{Z})^{-1}\mathbf{R}^T)^{-1}\mathbf{R}(\frac{1}{nh}\mathbf{Z}^T\mathbf{W}_K\mathbf{Z})^{-1}$$

$$= \frac{1}{f(0)}\Gamma^{-1} - \Gamma^{-1}\mathbf{R}^T(\mathbf{R}\Gamma^{-1}\mathbf{R}^T)^{-1}\mathbf{R}\Gamma^{-1} + o_p(1) \qquad (27)$$

where $\Gamma = \begin{pmatrix} \mathbf{S}^- & 0 \\ 0 & \mathbf{S}^+ \end{pmatrix}$.

Putting together (26) and (27), and plugging in $\mathbf{S}^- = \begin{pmatrix} \frac{1}{2} & -\frac{1}{4} \\ -\frac{1}{4} & \frac{1}{6} \end{pmatrix}$, $\mathbf{S}^+ = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{6} \end{pmatrix}$, $\mathbf{S}^{*-} = \begin{pmatrix} \frac{1}{4} & -\frac{1}{8} \\ -\frac{1}{8} & \frac{1}{12} \end{pmatrix}$,

$\mathbf{S}^{*+} = \begin{pmatrix} \frac{1}{4} & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{12} \end{pmatrix}$ we have

$$nh^3 Var(\hat{\beta}_1^{R+} - \hat{\beta}_1^{R-}) = 12f(0)(\sigma_Y^2(0^+) + \sigma_Y^2(0^-)),$$

and it follows that the asymptotic distribution of $\hat{\tau}_{SRKD}^{LR} = \frac{\hat{\beta}_1^{R+} - \hat{\beta}_1^{R+}}{\kappa_1^+ - \kappa_1^-}$ is

$$\sqrt{nh^3}(\hat{\tau}_{SRKD}^{LR} - \tau) \Rightarrow N(0, 12 \cdot \Omega_{SRKD}).$$

(2) Asymptotic distribution of the restricted fuzzy RKD estimator.

Similar to the proof for the sharp case, let $\mathbf{B} = (B_1^-, ..., B_{n^-}^-, B_1^+ ..., B_{n^+}^+)^T$ be the "stacked" $n \times 1$ vector of the treatment/policy variables, and the minimization problem for the first stage equation is defined analogously to (25). Denote the resulting estimator by $\hat{\kappa}^R = (\hat{\kappa}_0^{R-}, h\hat{\kappa}_1^{R-}, \hat{\kappa}_0^{R+}, h\hat{\kappa}_1^{R+})$ for the first stage and define $\mathbf{B}^* = \mathbf{B} - \mathbf{Z}\kappa$. As argued in the sharp case above, $\hat{\kappa}_1^{R+} - \hat{\kappa}_1^{R-}$ can be equivalently obtained from regressing B on 1, V and DV. It follows that $\frac{\hat{\beta}_1^{R+} - \hat{\beta}_1^{R-}}{\hat{\kappa}_1^{R+} - \hat{\kappa}_1^{R-}}$ is equal to the (2SLS) fuzzy RKD estimator $\hat{\tau}_{FRKD}^{LR}$. Similar to the sharp case, we have:

$$(\hat{\beta}^R - \beta) = [(\mathbf{Z}^T\mathbf{W}_K\mathbf{Z})^{-1} - (\mathbf{Z}^T\mathbf{W}_K\mathbf{Z})^{-1}\mathbf{R}^T(\mathbf{R}(\mathbf{Z}^T\mathbf{W}_K\mathbf{Z})^{-1}\mathbf{R}^T)^{-1}\mathbf{R}(\mathbf{Z}^T\mathbf{W}_K\mathbf{Z})^{-1}]\mathbf{Z}^T\mathbf{W}_K\mathbf{Y}^*$$

$$(\hat{\kappa}^R - \kappa) = [(\mathbf{Z}^T\mathbf{W}_K\mathbf{Z})^{-1} - (\mathbf{Z}^T\mathbf{W}_K\mathbf{Z})^{-1}\mathbf{R}^T(\mathbf{R}(\mathbf{Z}^T\mathbf{W}_K\mathbf{Z})^{-1}\mathbf{R}^T)^{-1}\mathbf{R}(\mathbf{Z}^T\mathbf{W}_K\mathbf{Z})^{-1}]\mathbf{Z}^T\mathbf{W}_K\mathbf{B}^*.$$

Under Assumptions 9, 10L, 11L, 12-15 and 16L, the same reasoning in Hahn et al. (1999) gives the joint

distribution

$$\frac{1}{\sqrt{nh}} \begin{pmatrix} \mathbf{Z}^T \mathbf{W}_K \mathbf{Y}^* \\ \mathbf{Z}^T \mathbf{W}_K \mathbf{B}^* \end{pmatrix} \Rightarrow N(0, f(0)\Sigma_{fuzzy}) \tag{28}$$

with

$$\Sigma_{fuzzy} = \begin{pmatrix} \sigma_Y^2(0^-)\mathbf{S}^{*-} & 0 & \sigma_{BY}(0^-)\mathbf{S}^{*-} & 0 \\ 0 & \sigma_Y^2(0^+)\mathbf{S}^{*+} & 0 & \sigma_{BY}(0^+)\mathbf{S}^{*+} \\ \sigma_{BY}(0^-)\mathbf{S}^{*-} & 0 & \sigma_B^2(0^-)\mathbf{S}^{*-} & 0 \\ 0 & \sigma_{BY}(0^+)\mathbf{S}^{*+} & 0 & \sigma_B^2(0^+)\mathbf{S}^{*+} \end{pmatrix}.$$

Combining (27) and (27) and plugging in the values for $\mathbf{S}^{\pm}$ and $\mathbf{S}^{*\pm}$, we have

$$
\begin{aligned}
nh^3 Var(\hat{\beta}_1^{R+} - \hat{\beta}_1^{R-}) &= 12f(0)(\sigma_Y^2(0^+) + \sigma_Y^2(0^-)) \\
nh^3 Var(\hat{\kappa}_1^{R+} - \hat{\kappa}_1^{R-}) &= 12f(0)(\sigma_B^2(0^+) + \sigma_B^2(0^-)) \\
nh^3 Cov(\hat{\beta}_1^{R+} - \hat{\beta}_1^{R-}, \hat{\kappa}_1^{R+} - \hat{\kappa}_1^{R-}) &= 12f(0)(\sigma_{BY}^2(0^+) + \sigma_{BY}^2(0^-)),
\end{aligned}
$$

and applying the Delta Method gives that the asymptotic distribution of $\hat{\tau}_{FRKD}^{LR} = \frac{\hat{\beta}_1^{R+} - \hat{\beta}_1^{R+}}{\hat{\kappa}_1^{R+} - \hat{\kappa}_1^{R-}}$ is

$$\sqrt{nh^3}(\hat{\tau}_{FRKD}^{LR} - \tau) \Rightarrow N(0, 12 \cdot \Omega_{FRKD}).$$

Proof of Proposition 7: Again, we prove the proposition for the local linear case because the proof for the local quadratic case is analogous, and we omit the L superscript in most of the estimators for simplicity of exposition.

1) Consistency of the sharp RKD variance estimator: First, Lemma 1 of Hahn et al. (1999) gives that

$$\mathbf{S}_{XX} = nhf(0)\left\{ \begin{pmatrix} \mathbf{H} & 0 \\ 0 & h \end{pmatrix} \begin{pmatrix} \mu_0^- + \mu_0^+ & \mu_1^- + \mu_1^+ & \mu_1^+ \\ \mu_1^- + \mu_1^+ & \mu_2^- + \mu_2^+ & \mu_2^+ \\ \mu_1^+ & \mu_2^+ & \mu_2^+ \end{pmatrix} \begin{pmatrix} \mathbf{H} & 0 \\ 0 & h \end{pmatrix} + o_p(1) \right\}$$

For the other term in the variance estimator,

$$\hat{\mathbf{S}}_{sharp} = \sum_{i=1}^{n}(\mathbf{X}_i\mathbf{X}_i^T)(Y_i - \mathbf{X}_i^T\beta_{sharp}^R + \mathbf{X}_i^T\beta_{sharp}^R - \mathbf{X}_i^T\hat{\beta}_{sharp}^R)^2 K^2(\frac{V_i}{h})$$

$$= \underbrace{\sum_{i=1}^{n}(\mathbf{X}_i\mathbf{X}_i^T)(Y_i - \mathbf{X}_i^T\beta_{sharp}^R)^2 K^2(\frac{V_i}{h})}_{(i)}$$

$$\underbrace{+2\sum_{i=1}^{n}(\mathbf{X}_i\mathbf{X}_i^T)(Y_i - \mathbf{X}_i^T\beta_{sharp}^R)\mathbf{X}_i^T(\beta_{sharp}^R - \hat{\beta}_{sharp}^R)K^2(\frac{V_i}{h})}_{(ii)}$$

$$\underbrace{+\sum_{i=1}^{n}(\mathbf{X}_i\mathbf{X}_i^T)\{\mathbf{X}_i^T(\beta_{sharp}^R - \hat{\beta}_{sharp}^R)\}^2 K^2(\frac{V_i}{h})}_{(iii)}.$$

where $\beta_{sharp}^R = (\gamma_0^S, \gamma_1^S, \delta_1^S)^T$. We examine the three terms separately

$$(i) = \sum_{i=1}^{n}\{(1-D_i)\begin{pmatrix} 1 & V_i & 0 \\ V_i & V_i^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}(Y_i - \gamma_0^S - \gamma_1^S V_i)^2 K^2(\frac{V_i}{h}) +$$

$$D_i\begin{pmatrix} 1 & V_i & V_i \\ V_i & V_i^2 & V_i^2 \\ V_i & V_i^2 & V_i^2 \end{pmatrix}(Y_i - \gamma_0^S - (\gamma_1^S + \delta_1^S)V_i)^2 K^2(\frac{V_i}{h})\}.$$

We examine the matrix expression above entry by entry: let $\Upsilon_i^{l-} \equiv (1-D_i)V_i^l(Y_i - \gamma_0^S - \gamma_1^S V_i)^2 K^2(\frac{V_i}{h})$ and $\Upsilon_i^{l+} \equiv D_i V_i^l(Y_i - \gamma_0^S - (\gamma_1^S + \delta_1^S)V_i)^2 K^2(\frac{V_i}{h})$. Note that

$$E[\sum_{i=1}^{n}\Upsilon_i^{l+}] = nE[D_i V^l\{\varepsilon_i + \frac{1}{2}m''(0^+)V_i^2 + \varsigma(V_i)\}^2 K^2(\frac{V_i}{h})]$$

$$= n\int_0^{\infty} v^l \sigma_Y^2(v)K^2(\frac{v}{h})f(v)dv + n\int_0^{\infty} v^l[\frac{1}{2}m''(0^+)v + \varsigma(v)]^2 K^2(\frac{v}{h})f(v)dv$$

$$= n\int_0^{\infty}(uh)^l \sigma_Y^2(uh)K^2(u)h(f(uh)du +$$

$$n\int_0^{\infty}(uh)^l[\{\frac{1}{2}m''(0^+)\}^2 u^2 h^2 + o(h^2)]K^2(u)f(uh)hdu$$

$$= nh^{l+1}\sigma_Y^2(0^+)f(0)(v_l^+ + o(1))$$

where the second last line follows from Assumption 10L and Taylor's Theorem and the last line from the

Dominated Convergence Theorem as well as Assumptions 12, 13 and 14. Also note that

$$
\begin{aligned}
Var(\sum_{i=1}^{n}\Upsilon_{i}^{l+}) \;&\leqslant\; nE[D_{i}V_{i}^{2l}\{\varepsilon_{i}+\frac{1}{2}m''(0^{+})V_{i}^{2}+\varsigma(V_{i})\}^{4}K^{4}(\frac{V_{i}}{h})] \\
&\leqslant\; 27nE[\{D_{i}V_{i}^{2l}\varepsilon_{i}^{4}+D_{i}V_{i}^{2l}(|\frac{1}{2}m''(0^{+})|V_{i}^{2})^{4}+D_{i}V_{i}^{2l}\varsigma^{4}(V_{i})\}K^{4}(\frac{V_{i}}{h})] \\
&=\; 27n\int_{0}^{\infty}\{(uh)^{2l}\kappa_{Y}(uh)+\frac{1}{2}|m''(0^{+})|(uh)^{2l+8}+(uh)^{2l}\varsigma^{4}(uh)\}K^{4}(u)f(u)hdu \\
&=\; O(nh^{2l+1})
\end{aligned}
$$

where the second line follows from Jensen's inequality and the last line from the Dominated Convergence Theorem as well as Assumptions 12, 13 and 17. By Chebyshev's Inequality,

$$
\begin{aligned}
\sum_{i=1}^{n}\Upsilon_{i}^{l+} \;&=\; E[\sum_{i=1}^{n}\Upsilon_{i}^{l+}]+O_{p}(\sqrt{Var(\sum_{i=1}^{n}\Upsilon_{i}^{l+})}) \\
&=\; nh^{l+1}\sigma_{Y}^{2}(0^{+})f(0)(v_{l}^{+}+o(1))+O_{p}(n^{\frac{1}{2}}h^{l+\frac{1}{2}}) \\
&=\; nh^{l+1}\sigma_{Y}^{2}(0^{+})f(0)(v_{l}^{+}+o_{p}(1)).
\end{aligned}
$$

Similarly,

$$
\sum_{i=1}^{n}\Upsilon_{i}^{l-}=nh^{l+1}\sigma_{Y}^{2}(0^{-})f(0)(v_{l}^{-}+o_{p}(1)).
$$

Using matrix notations, we have

$$
(i)=nhf(0)\begin{pmatrix} \mathbf{H} & 0 \\ 0 & h \end{pmatrix}\begin{pmatrix} \sigma_{Y}^{2}(0^{-})v_{0}^{-}+\sigma_{Y}^{2}(0^{+})v_{0}^{+} & \sigma_{Y}^{2}(0^{-})v_{1}^{-}+\sigma_{Y}^{2}(0^{+})v_{1}^{+} & \sigma_{Y}^{2}(0^{+})v_{2}^{+} \\ \sigma_{Y}^{2}(0^{-})v_{1}^{-}+\sigma_{Y}^{2}(0^{+})v_{1}^{+} & \sigma_{Y}^{2}(0^{-})v_{2}^{-}+\sigma_{Y}^{2}(0^{+})v_{2}^{+} & \sigma_{Y}^{2}(0^{+})v_{2}^{+} \\ \sigma_{Y}^{2}(0^{+})v_{1}^{+} & \sigma_{Y}^{2}(0^{+})v_{2}^{+} & \sigma_{Y}^{2}(0^{+})v_{2}^{+} \end{pmatrix}\begin{pmatrix} \mathbf{H} & 0 \\ 0 & h \end{pmatrix}(1+o_{p}(1)).
$$

Next we investigate the asymptotic properties of (ii) and (iii). We write (ii) as

$$
\begin{aligned}
(ii) \;=\; 2\sum_{i=1}^{n}[(1-D_{i})\begin{pmatrix} 1 & V_{i} & 0 \\ V_{i} & V_{i}^{2} & 0 \\ 0 & 0 & 0 \end{pmatrix}+D_{i}\begin{pmatrix} 1 & V_{i} & V_{i} \\ V_{i} & V_{i}^{2} & V_{i}^{2} \\ V_{i} & V_{i}^{2} & V_{i}^{2} \end{pmatrix}] \\
(Y_{i}-\mathbf{X}_{i}^{T}\beta_{sharp}^{R})\mathbf{X}_{i}^{T}(\beta_{sharp}^{R}-\hat{\beta}_{sharp}^{R})K^{2}(\frac{V_{i}}{h}).
\end{aligned}
$$

We examine the matrix expression above entry by entry, and the basic building blocks are terms of the form

$$\sum_{i=1}^{n} D_i V_i^l (Y_i - \mathbf{X}_i^T \beta_{sharp}^R) \mathbf{X}_i^T (\beta_{sharp}^R - \hat{\beta}_{sharp}^R) K^2 \left(\frac{V_i}{h}\right)$$

$$= \underbrace{\{\sum_{i=1}^{n} D_i V_i^l (Y_i - \mathbf{X}_i^T \beta_{sharp}^R) \mathbf{X}_i^T \begin{pmatrix} \mathbf{H}^{-1} & 0 \\ 0 & h^{-1} \end{pmatrix} K^2 \left(\frac{V_i}{h}\right)\}}_{\text{ii(a)}}$$

$$\underbrace{\begin{pmatrix} \mathbf{H} & 0 \\ 0 & h \end{pmatrix} (\beta_{sharp}^R - \hat{\beta}_{sharp}^R)}_{\text{ii(b)}}.$$

Note that

$$\text{ii(a)} = \{\sum_{i=1}^{n} D_i V_i^l (\varepsilon_i + \frac{1}{2} m''(0^+) V_i^2 + \varsigma(V_i))(1, \frac{V_i}{h}, \frac{V_i}{h}) K^2 \left(\frac{V_i}{h}\right)\}$$

and

$$
\begin{aligned}
E[\text{ii(a)}] &= nE[D_i V_i^l (\frac{1}{2} m''(0^+) V_i^2 + \varsigma(V_i))(1, \frac{V_i}{h}, \frac{V_i}{h}) K^2 \left(\frac{V_i}{h}\right)] \\
&= n \int_0^\infty v^l (\frac{1}{2} m''(0^+) v^2 + \varsigma(v))(1, \frac{v}{h}, \frac{v}{h}) K^2 \left(\frac{v}{h}\right) f(v) dv \\
&= n \int_0^\infty (uh)^l [\frac{1}{2} m''(0^+)(uh)^2 + \varsigma(uh)](1, u, u) K^2(u) f(uh) h \, du \\
&= O_p(nh^{l+3})
\end{aligned}
$$

where the integral operator is applied entry-by-entry. Since

$$
\begin{aligned}
Var(\text{ii(a)}) &\leqslant 27nE[D_i V_i^{2l} \{\varepsilon_i^2 + (\frac{1}{2} m''(0^+) V_i^2)^2 + \varsigma^2(V_i)\}(1, (\frac{V_i}{h})^2, (\frac{V_i}{h})^2) K^4 \left(\frac{V_i}{h}\right)] \\
&= 27n \int_0^\infty (uh)^{2l} \{\sigma_Y^2(uh) + (\frac{1}{2} m''(0^+))^2 (uh)^4 + \varsigma^2(uh)\}(1, u^2, u^2) K^4(u) f(u) h \, du \\
&= O_p(nh^{2l+1})
\end{aligned}
$$

where $\leqslant$ is defined entry-wise (for vectors $\mathbf{a}$ and $\mathbf{b}$, $\mathbf{a} \leqslant \mathbf{b} \Rightarrow \mathbf{a}_{(i)} \leqslant \mathbf{b}_{(i)}$), we have by Assumption 16L

$$\text{ii(a)} = E[\text{ii(a)}] + O_p(\sqrt{Var(\text{ii(a)})}) = O_p(h^l \sqrt{nh}).$$

Similarly, we can show that

$$\sum_{i=1}^{n}(1-D_i)V_i^l(Y_i - \mathbf{X}_i^T \beta_{sharp})\mathbf{X}_i^T \begin{pmatrix} \mathbf{H}^{-1} & 0 \\ 0 & h^{-1} \end{pmatrix} K^2(\frac{V_i}{h}) = O_p(h^l\sqrt{nh}).$$

Note that ii(b) $= O_p((nh)^{-\frac{1}{2}})$ from the proof of Proposition **6**, so that

$$\begin{aligned}
(\text{ii}) &= \begin{pmatrix} \mathbf{H} & 0 \\ 0 & h \end{pmatrix} O_p(1) \begin{pmatrix} \mathbf{H} & 0 \\ 0 & h \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{H} & 0 \\ 0 & h \end{pmatrix} o_p(nh) \begin{pmatrix} \mathbf{H} & 0 \\ 0 & h \end{pmatrix}.
\end{aligned}$$

Analogously, the building blocks of (iii) are terms of the form

$$(\hat{\beta}_{sharp} - \beta_{sharp})^T \{\sum_{i=1}^{n} D_i V_i^l \mathbf{X}_i \mathbf{X}_i^T K^2(\frac{V_i}{h})\}(\hat{\beta}_{sharp} - \beta_{sharp}) = O_p(h^l)$$

and it follows that

$$\begin{aligned}
(\text{iii}) &= \begin{pmatrix} \mathbf{H} & 0 \\ 0 & h \end{pmatrix} O_p(1) \begin{pmatrix} \mathbf{H} & 0 \\ 0 & h \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{H} & 0 \\ 0 & h \end{pmatrix} o_p(nh) \begin{pmatrix} \mathbf{H} & 0 \\ 0 & h \end{pmatrix}.
\end{aligned}$$

Therefore,

$$\hat{\mathbf{S}}_{sharp} = nhf(0) \begin{pmatrix} \mathbf{H} & 0 \\ 0 & h \end{pmatrix} \begin{pmatrix} \sigma_Y^2(0^-)v_0^- + \sigma_Y^2(0^+)v_0^+ & \sigma_Y^2(0^-)v_1^- + \sigma_Y^2(0^+)v_1^+ & \sigma_Y^2(0^+)v_2^+ \\ \sigma_Y^2(0^-)v_1^- + \sigma_Y^2(0^+)v_1^+ & \sigma_Y^2(0^-)v_2^- + \sigma_Y^2(0^+)v_2^+ & \sigma_Y^2(0^+)v_2^+ \\ \sigma_Y^2(0^+)v_1^+ & \sigma_Y^2(0^+)v_2^+ & \sigma_Y^2(0^+)v_2^+ \end{pmatrix} \begin{pmatrix} \mathbf{H} & 0 \\ 0 & h \end{pmatrix}(1+o_p(1)).$$

Putting together the expressions for $\mathbf{S}_{XX}^{-1}$ and $\hat{\mathbf{S}}_{sharp}$ and plugging in the numerical values for $\mu_j^{\pm}$ and $v_j^{\pm}$ ($j = 0, 1, 2$), we come to the desired result,

$$nh^3 \mathbf{e}_3^T \hat{Var}(\hat{\beta}_{sharp})\mathbf{e}_3 \rightarrow \frac{12\{\sigma_Y^2(0^+) + \sigma_Y^2(0^-)\}}{f(0)}$$

for the local linear regression.

2) Consistency of the fuzzy RKD variance estimator: we separately examine the asymptotic behavior of $\mathbf{S}_{XW}$ and $\hat{\mathbf{S}}_{fuzzy}$.

$$\mathbf{S}_{XW} = \sum_{i=1}^{n} \mathbf{X}_i \mathbf{W}_i^T K(\frac{V_i}{h})$$

$$= \sum_{i=1}^{n} \begin{pmatrix} 1 & V_i & B_i \\ V_i & V_i^2 & B_i V_i \\ D_i V_i & D_i V_i^2 & D_i B_i V_i \end{pmatrix} K(\frac{V_i}{h})$$

$$= \sum_{i=1}^{n} \{ D_i \begin{pmatrix} 1 & V_i & B_i \\ V_i & V_i^2 & B_i V_i \\ V_i & V_i^2 & B_i V_i \end{pmatrix} + (1-D_i) \begin{pmatrix} 1 & V_i & B_i \\ V_i & V_i^2 & B_i V_i \\ 0 & 0 & 0 \end{pmatrix} \} K(\frac{V_i}{h}).$$

As shown previously,

$$\sum_{i=1}^{n} D_i V_i^l K(\frac{V_i}{h}) = n h^{l+1} f(0)(\mu_l^+ + o_p(1))$$

$$\sum_{i=1}^{n} (1-D_i) V_i^l K(\frac{V_i}{h}) = n h^{l+1} f(0)(\mu_l^- + o_p(1)),$$

and we proceed to investigate the terms $\sum_{i=1}^{n} D_i B_i V_i^l K(\frac{V_i}{h})$ and $\sum_{i=1}^{n} (1-D_i) B_i V_i^l K(\frac{V_i}{h})$. Because

$$E[\sum_{i=1}^{n} D_i B_i V_i^l K(\frac{V_i}{h})] = n \int_0^{\infty} r(v) v^l K(\frac{v}{h}) f(v) dv$$

$$= n \int_0^{\infty} \{ r(0^+) + r'(0^+) uh + \frac{1}{2} r''(0^+)(uh)^2 + \xi(uh) \}$$

$$(uh)^l K(u) f(uh) h du$$

$$= n h^{l+1} f(0) r(0^+) \mu_l^+ + n h^{l+2} f(0)(r'(0^+) \mu_{l+1}^+ + o(1))$$

and

$$Var(\sum_{i=1}^{n} D_i B_i V_i^l K(\frac{V_i}{h})) \leqslant n E[D_i B_i^2 V_i^{2l} K^2(\frac{V_i}{h})]$$

$$= n \int_0^{\infty} (r^2(v) + \sigma_B^2(v)) v^{2l} K^2(\frac{v}{h}) dv$$

$$= n \int_0^{\infty} (r^2(uh) + \sigma_B^2(uh))(uh)^{2l} K^2(u) f(uh) h du$$

$$= n h^{2l+1} f(0) \{ (r^2(0^+) + \sigma_B^2(0^+)) v^{2l} + o(1) \},$$

by Assumption 16L, we have

$$\sum_{i=1}^{n} D_i B_i V_i^l K(\frac{V_i}{h})$$

$$= E[\sum_{i=1}^{n} D_i B_i V_i^l K(\frac{V_i}{h})] + O_p(\sqrt{Var(\sum_{i=1}^{n} D_i B_i V_i^l K(\frac{V_i}{h}))})$$

$$= nh^{l+1} f(0) r(0^+) \mu_l^+ + nh^{l+2} f(0)(r'(0^+) \mu_{l+1}^+ + o_p(1)).$$

Similarly,

$$\sum_{i=1}^{n} (1 - D_i) B_i V_i^l K(\frac{V_i}{h}) = nh^{l+1} f(0) r(0^-) \mu_l^- + nh^{l+2} f(0)(r'(0^-) \mu_{l+1}^- + o_p(1)).$$

Therefore,

$$\mathbf{S}_{XW} = nhf(0)\{ \begin{pmatrix} \mu_0^+ & h\mu_1^+ & r(0^+)\mu_0^+ + hr'(0^+)\mu_1^+ \\ h\mu_1^+ & h^2\mu_2^+ & hr(0^+)\mu_1^+ + h^2 r'(0^+)\mu_2^+ \\ h\mu_1^+ & h^2\mu_2^+ & hr(0^+)\mu_1^+ + h^2 r'(0^+)\mu_2^+ \end{pmatrix}$$

$$+ \begin{pmatrix} \mu_0^- & h\mu_1^- & r(0^-)\mu_0^- + hr'(0^-)\mu_1^- \\ h\mu_1^- & h^2\mu_2^- & hr(0^-)\mu_1^- + h^2 r'(0^-)\mu_2^- \\ 0 & 0 & 0 \end{pmatrix} \}(1 + o_p(1)).$$

We investigate the properties of $\hat{\mathbf{S}}_{fuzzy}$ next:

$$
\begin{aligned}
\hat{\mathbf{S}}_{fuzzy} &= \sum_{i=1}^{n}(\mathbf{X}_i\mathbf{X}_i^T)\hat{u}_i^2 K^2(\frac{V_i}{h}) \\
&= \sum_{i=1}^{n}(\mathbf{X}_i\mathbf{X}_i^T)(Y_i - \mathbf{W}_i^T\beta_{fuzzy} + \mathbf{W}_i^T\beta_{fuzzy} - \mathbf{W}_i^T\hat{\beta}_{fuzzy})^2 K^2(\frac{V_i}{h}) \\
&= \underbrace{\sum_{i=1}^{n}(\mathbf{X}_i\mathbf{X}_i^T)(Y_i - \mathbf{W}_i^T\beta_{fuzzy})^2 K^2(\frac{V_i}{h})}_{(iv)} \\
&\quad \underbrace{+2\sum_{i=1}^{n}(\mathbf{X}_i\mathbf{X}_i^T)(Y_i - \mathbf{W}_i^T\beta_{fuzzy})(\mathbf{W}_i^T\beta_{fuzzy} - \mathbf{W}_i^T\hat{\beta}_{fuzzy})K^2(\frac{V_i}{h})}_{(v)} \\
&\quad \underbrace{+\sum_{i=1}^{n}(\mathbf{X}_i\mathbf{X}_i^T)(\mathbf{W}_i^T\beta_{fuzzy} - \mathbf{W}_i^T\hat{\beta}_{fuzzy})^2 K^2(\frac{V_i}{h}).}_{(vi)}
\end{aligned}
$$

We write

$$
\begin{aligned}
(iv) &= \sum_{i=1}^{n}\{(1-D_i)
\begin{pmatrix} 1 & V_i & 0 \\ V_i & V_i^2 & 0 \\ 0 & 0 & 0 \end{pmatrix}
(Y_i - \gamma_0^F - \gamma_1^F V_i - \delta_1^F B_i)^2 K^2(\frac{V_i}{h}) + \\
&\quad D_i
\begin{pmatrix} 1 & V_i & V_i \\ V_i & V_i^2 & V_i^2 \\ V_i & V_i^2 & V_i^2 \end{pmatrix}
(Y_i - \gamma_0^F - \gamma_1^F V_i - \delta_1^F B_i)^2 K^2(\frac{V_i}{h})\}
\end{aligned}
$$

We examine the matrix expression above entry-by-entry: let $\Theta_i^{l+} \equiv D_i V_i^l (Y_i - \gamma_0^F - \gamma_1^F V_i - \delta_1^F B_i)^2 K^2(\frac{V_i}{h})$ and $\Theta_i^{l-} \equiv (1-D_i)V_i^l(Y_i - \gamma_0^F - \gamma_1^F V_i - \delta_1^F B_i)^2 K^2(\frac{V_i}{h})$.

First note that if the local outcome and first stage equations are

$$
\begin{aligned}
Y &= c_0 + c_1 V + d_1 DV \\
B &= a_0 + a_1 V + b_1 DV
\end{aligned}
$$

then $\gamma_0^F = c_0 - \delta_1^F a_0$, $\gamma_1^F = c_1 - \delta_1^F a_1$ and $\delta_1^F = \frac{d_1}{b_1}$. Plugging in these relations, we have

$$\Theta_i^{l+} = D_i V_i^l \{(\frac{1}{2}m''(0^+)V_i^2 + \varsigma(V_i) + \varepsilon_i) - \delta_1^F(\frac{1}{2}r''(0^+)V_i^2 + \xi(V_i) + \phi_i)\}^2 K^2(\frac{V_i}{h})$$

$$\Theta_i^{l-} = (1 - D_i)V_i^l \{(\frac{1}{2}m''(0^-)V_i^2 + \varsigma(V_i) + \varepsilon_i) - \delta_1^F(\frac{1}{2}r''(0^-)V_i^2 + \xi(V_i) + \phi_i)\}^2 K^2(\frac{V_i}{h})$$

where $\phi_i = B_i - r(V_i)$ is the error term in the first stage relationship. Note that

$$E[\sum_{i=1}^n \Theta_i^{l+}]$$

$$= nE[D_i V_i^l \{(\frac{1}{2}m''(0^+)V_i^2 + \varsigma(V_i) + \varepsilon_i) - \delta_1^F(\frac{1}{2}r''(0^+)V_i^2 + \xi(V_i) + \phi_i)\}^2 K^2(\frac{V_i}{h})]$$

$$= nE[D_i V_i^l \varepsilon_i^2 K^2(\frac{V_i}{h})] + nE[D_i V_i^l(\frac{1}{2}m''(0^+)V_i^2 + \varsigma(V_i))^2 K^2(\frac{V_i}{h})]$$

$$+ nE[D_i V_i^l(\delta_1^F)^2 \phi_i^2 K^2(\frac{V_i}{h})] + nE[D_i V_i^l(\delta_1^F)^2(\frac{1}{2}r''(0^+)V_i^2 + \xi(V_i))^2 K^2(\frac{V_i}{h})]$$

$$- 2nE[D_i V_i^l \varepsilon_i \delta_1^F \phi_i K^2(\frac{V_i}{h})]$$

$$- 2nE[D_i V_i^l(\frac{1}{2}m''(0^+)V_i^2 + \varsigma(V_i))\delta_1^F(\frac{1}{2}r''(0^+)V_i^2 + \xi(V_i))K^2(\frac{V_i}{h})]$$

$$= nh^{l+1}f(0)\sigma_Y^2(0^+)(v_l^+ + o(1)) + o(nh^{l+1})$$

$$+ nh^{l+1}f(0)(\delta_1^F)^2\sigma_B^2(0^+)(v_l^+ + o(1)) + o(nh^{l+1})$$

$$- nh^{l+1}f(0)2\delta_1^F \sigma_{BY}(0^+)(v_l^+ + o(1)) - o(nh^{l+1})$$

$$= nh^{l+1}f(0)(\sigma_Y^2(0^+) - 2\delta_1^F \sigma_{BY}(0^+) + (\delta_1^F)^2\sigma_B^2(0^+))(v_l^+ + o_p(1)).$$

Similarly,

$$E[\sum_{i=1}^n \Theta_i^{l+}] = nh^{l+1}f(0)(\sigma_Y^2(0^-) - 2\delta_1^F \sigma_{BY}(0^-) + (\delta_1^F)^2\sigma_B^2(0^-))(v_l^- + o_p(1)).$$

Also note that

$$Var(\sum_{i=1}^n \Theta_i^{l+}) \leq nE[D_i V_i^{2l}\{\varepsilon_i + \frac{1}{2}m''(0^+)V_i^2 + \varsigma(V_i) - \delta_1^F(\frac{1}{2}r''(0^+)V_i^2 + \xi(V_i) + \phi_i)\}^4 K^4(\frac{V_i}{h})]$$

$$\leq 216nE[\{D_i V_i^{2l}\varepsilon_i^4 + D_i V_i^{2l}(|\frac{1}{2}m''(0^+)|V_i^2)^4 + D_i V_i^{2l}\varsigma^4(V_i)\}K^4(\frac{V_i}{h})]$$

$$+ 216nE[(\delta_1^F)^4\{D_i V_i^{2l}\phi_i^4 + D_i V_i^{2l}(|\frac{1}{2}r''(0^+)|V_i^2)^4 + D_i V_i^{2l}\xi^4(V_i)\}K^4(\frac{V_i}{h})]$$

$$= 216O(nh^{2l+1})$$

where the second line follows from Jensen's inequality and the last line from the Dominated Convergence Theorem as well as Assumptions 12, 13 and 17. As a result,

$$
\begin{aligned}
\sum_{i=1}^{n} \Theta_i^{l+} &= E[\sum_{i=1}^{n} \Theta_i^{l+}] + O_p(\sqrt{Var(\sum_{i=1}^{n} \Theta_i^{l+})}) \\
&= nh^{l+1} f(0)(\sigma_Y^2(0^+) - 2\delta_1^F \sigma_{BY}(0^+) + (\delta_1^F)^2 \sigma_B^2(0^+))(v_l^+ + o_p(1)) + O_p(n^{\frac{1}{2}} h^{l+\frac{1}{2}}) \\
&= nh^{l+1} f(0)(\sigma_Y^2(0^+) - 2\delta_1^F \sigma_{BY}(0^+) + (\delta_1^F)^2 \sigma_B^2(0^+))(v_l^+ + o_p(1)).
\end{aligned}
$$

Similarly,

$$
\sum_{i=1}^{n} \Theta_i^{l-} = nh^{l+1} f(0)(\sigma_Y^2(0^-) - 2\delta_1^F \sigma_{BY}(0^-) + (\delta_1^F)^2 \sigma_B^2(0^-))(v_l^- + o_p(1)).
$$

Using matrix notations, we have

$$
(iv) = nhf(0) \begin{pmatrix} \mathbf{H} & 0 \\ 0 & h \end{pmatrix} \begin{pmatrix} \sigma^2_{Y-\delta_1^F B}(0^-)v_0^- + \sigma^2_{Y-\delta_1^F B}(0^+)v_0^+ & \sigma^2_{Y-\delta_1^F B}(0^-)v_1^- + \sigma^2_{Y-\delta_1^F B}(0^+)v_1^+ & \sigma^2_{Y-\delta_1^F B}(0^+)v_1^+ \\ \sigma^2_{Y-\delta_1^F B}(0^-)v_1^- + \sigma^2_{Y-\delta_1^F B}(0^+)v_1^+ & \sigma^2_{Y-\delta_1^F B}(0^-)v_2^- + \sigma^2_{Y-\delta_1^F B}(0^+)v_2^+ & \sigma^2_{Y-\delta_1^F B}(0^+)v_2^+ \\ \sigma^2_{Y-\delta_1^F B}(0^+)v_1^+ & \sigma^2_{Y-\delta_1^F B}(0^+)v_2^+ & \sigma^2_{Y-\delta_1^F B}(0^+)v_2^+ \end{pmatrix} \begin{pmatrix} \mathbf{H} & 0 \\ 0 & h \end{pmatrix} (1 + o_p(1))
$$

where $\sigma^2_{Y-\delta_1^F B}(0^\pm) = \sigma_Y^2(0^\pm) - 2\delta_1^F \sigma_{BY}(0^\pm) + (\delta_1^F)^2 \sigma_B^2(0^\pm)$.

Similar to the proof in the sharp case, we can show that

$$
(v) \quad, (vi) = \begin{pmatrix} \mathbf{H} & 0 \\ 0 & h \end{pmatrix} o_p(nh) \begin{pmatrix} \mathbf{H} & 0 \\ 0 & h \end{pmatrix}
$$

so that

$$
\hat{\mathbf{S}}_{fuzzy} = nhf(0) \begin{pmatrix} \mathbf{H} & 0 \\ 0 & h \end{pmatrix} \begin{pmatrix} \sigma^2_{Y-\delta_1^F B}(0^-)v_0^- + \sigma^2_{Y-\delta_1^F B}(0^+)v_0^+ & \sigma^2_{Y-\delta_1^F B}(0^-)v_1^- + \sigma^2_{Y-\delta_1^F B}(0^+)v_1^+ & \sigma^2_{Y-\delta_1^F B}(0^+)v_1^+ \\ \sigma^2_{Y-\delta_1^F B}(0^-)v_1^- + \sigma^2_{Y-\delta_1^F B}(0^+)v_1^+ & \sigma^2_{Y-\delta_1^F B}(0^-)v_2^- + \sigma^2_{Y-\delta_1^F B}(0^+)v_2^+ & \sigma^2_{Y-\delta_1^F B}(0^+)v_2^+ \\ \sigma^2_{Y-\delta_1^F B}(0^+)v_1^+ & \sigma^2_{Y-\delta_1^F B}(0^+)v_2^+ & \sigma^2_{Y-\delta_1^F B}(0^+)v_2^+ \end{pmatrix} \begin{pmatrix} \mathbf{H} & 0 \\ 0 & h \end{pmatrix} (1 + o_p(1)).
$$

Plugging in values of $\mu_l^\pm$ and $v_l^\pm$, we have the desired result for the local linear fuzzy estimator:

$$
\begin{aligned}
\mathbf{e}_3^T \hat{Var}(\hat{\beta}_{fuzzy}) \mathbf{e}_3 &= \mathbf{S}_{XW}^{-1} \hat{\mathbf{S}}_{fuzzy} \mathbf{S}_{WX}^{-1} \\
&= \frac{12}{nh^3} (\frac{\sigma^2_{Y-\delta_1^F B}(0^+) + \sigma^2_{Y-\delta_1^F B}(0^-)}{\{r'(0^+) - r'(0^-)\}^2} + o_p(1)) \\
&= \frac{12}{nh^3} (\Omega_{FRKD} + o_p(1)).
\end{aligned}
$$

Supplemental Appendix–A Job Search Model with Wage-Dependent UI Benefits

This Appendix describes an equilibrium wage posting model with a wage-dependent UI benefit, and a maximum benefit level. We ask to what extent the model is consistent with assumptions for identification in the RKD, and reach two main conclusions. First, when there is a kink in the UI benefit formula, a baseline model predicts a kink in the density of wages among job-losers at the level of wages corresponding to the maximum benefit. Second, this prediction relies on complete information about the location of the kink in the benefit schedule and is *not* robust to allowing for small errors in agents beliefs about the location of the kink in the benefit schedule.

Setup. Consider an infinite horizon, discrete-time, posted-wage model of job search with an exogenous distribution of wage offers, and equally efficient search among employed and unemployed agents. With a level of search intensity s the arrival rate of job offers is $\lambda \cdot s$; there is also an exogenous job destruction rate of δ. There is a strictly increasing and convex cost-of-search function $c(\cdot)$ with $c(0) = 0$. Wage offers come from a stationary, twice continuously differentiable c.d.f. $F(\cdot)$. The setup is identical to the model used by Christensen et al. (2005), except that we cast the problem in discrete time (with a discount rate β) and assume a wage-dependent UI benefit.[53]

Specifically, we assume that the UI benefit b is a function of the last wage received before being laid off, w_{-1}, given by the formula $b(w_{-1}) \equiv \underline{b} + \rho \min(w_{-1}, T^{max})$, where $\rho < 1$ and $b(w) < w$ for all w. As in most actual benefit systems, agents with a previous wage above the threshold T^{max} receive a maximum benefit level $\bar{b} = \underline{b} + \rho T^{max}$. The dependence of benefits on previous wages adds two novel considerations to the standard search model: 1) when choosing search intensity and whether to accept a wage offer, an agent must take into account the effect of the wage on future UI benefits; 2) when taking a new job, an unemployed worker resets their benefit level. Because we assume that UI benefits last indefinitely, and that the benefit is reset immediately upon taking a new job, our model arguably over-emphasizes both these considerations relative to a more realistic setting where benefits can expire, and UI entitlement is based on earnings over a previous base period of several quarters' duration.

An agent's choice problem is characterized by two value functions: $W_{em}(w)$, the value function for being

---

[53]To translate the model to our generalized regression setting, note that we can allow for unrestricted heterogeneity and index all the model's elements by U, the unobserved type. The discussion below is conditional on the type U, and we suppress any notation indicating the value of U.

employed with current wage w, and $W_{un}(w_{-1})$, for being unemployed with previous wage w_{-1}:

$$W_{em}(w) = \max_{s \geq 0} \left\{ w - c(s) + \beta \left[(1-\delta) \left[\lambda s \int \max\{W_{em}(x), W_{em}(w)\} dF(x) + (1-\lambda s) W_{em}(w) \right] + \delta W_{un}(w) \right] \right\} \quad (29)$$

$$W_{un}(w_{-1}) = \max_{s \geq 0} \left\{ b(w_{-1}) - c(s) + \beta \left[\lambda s \int \max\{W_{em}(x), W_{un}(w_{-1})\} dF(x) + (1-\lambda s) W_{un}(w_{-1}) \right] \right\}. \quad (30)$$

Note that W_{un} is an increasing function of the previous wage for $w_{-1} < T^{max}$, since a higher previous wage entitles the agent to higher benefits. Once the previous wage reaches the threshold T^{max}, however, there is no further increase in W_{un}: thus the value function W_{un} is kinked at $w_{-1} = T^{max}$, with $W'_{un}(w_{-1}) > 0$ for $w_{-1} < T_{max}$ and $W'_{un}(w_{-1}) = 0$ for $w_{-1} > T^{max}$. Inspection of the value functions shows that this in turn induces a kink in $W_{em}(w)$ at $w = T^{max}$, provided that $\delta > 0$.

Optimal Search Behavior. It can be shown that the optimal behavior is characterized by a reservation wage strategy while employed, another reservation wage strategy while unemployed, and a choice of optimal search intensity $s_{em}(w)$ when employed at wage w and $s_{un}(w_{-1})$ when unemployed with previous wage w_{-1}.[54] Clearly, an employed worker will accept any wage offer that exceeds her current wage. An unemployed worker with previous wage w_{-1} will accept any wage offer w with $W_{em}(w) \geq W_{un}(w_{-1})$, implying a reservation wage $R(w_{-1})$ such that $W_{em}(R(w_{-1})) = W_{um}(w_{-1})$.

It is well known that when the UI benefit is a fixed constant b the optimal strategy for an unemployed worker is to take any job with $w \geq b$, implying $R(w_{-1}) = b$, since there is no extra disutility of work versus unemployment, and the arrival rate and search costs are the same whether working or not. This simple rule is no longer true when the benefits depend on w_{-1}. Consider an unemployed worker with an offer $w = b(w_{-1})$. Taking the job will yield the same flow utility as remaining on unemployment, but when the job ends she will receive a *lower* future UI benefit (assuming that the benefit-replacement rate is less than 1). Thus, a higher wage offer is required for indifference, implying that $R(w_{-1}) > b(w_{-1})$ when $w_{-1} < T^{max}$.

Given the value functions above, and a strictly increasing, convex, and twice continuously differentiable cost function $c(\cdot)$, we can implicitly solve for the optimal search functions $s_{un}(\cdot)$ and $s_{em}(\cdot)$ via the first

---

[54]It can be shown that $W_{em}(w)$ is strictly increasing in w and that $W_{un}(w_{-1})$ is increasing in w_{-1}, which leads to reservation wage strategies in each case.

order conditions for interior solutions for (29) and (30),

$$c'(s_{em}(w)) = \beta(1-\delta)\lambda \int_w^{\bar{w}} [W_{em}(x) - W_{em}(w)] dF(x) \tag{31}$$

$$c'(s_{un}(w_{-1})) = \beta\lambda \int_{R(w_{-1})}^{\bar{w}} [W_{em}(x) - W_{un}(w_{-1})] dF(x),$$

where $\bar{w}$ is the upper bound of the support of the offer distribution. Consideration of these first order conditions shows that the optimal levels of search intensity both have a kink at the wage threshold T^{max}. For example, the right derivative of $s_{em}(w)$ at $w = T^{max}$ is:

$$s'_{em}(T^{max+}) = \frac{-\beta(1-\delta)\lambda(1-F(T))}{c''(s_{em}(T))} W'_{em}(T^{max+}),$$

while the derivative from the left is:

$$s'_{em}(T^{max-}) = \frac{-\beta(1-\delta)\lambda(1-F(T))}{c''(s_{em}(T))} W'_{em}(T^{max-}).$$

Since $W_{em}(w)$ has a kink at $w = T^{max}$, $W'(T^{max+}) \neq W'(T^{max-})$ and the left and right limits of the derivative of $s_{em}(w)$ are different at $w = T^{max}$. A similar argument applies to the derivative of $s_{un}(w_{-1})$ at $w_{-1} = T^{max}$.

Steady State Wage Distribution. A standard wage posting model yields a steady state unemployment rate u and a steady state distribution of wages $G(w)$ that stochastically dominates the distribution of wage offers $F(w)$, reflecting the fact that employed workers are always searching for higher wage offers. When the benefit level varies across unemployed workers, and workers with different benefit levels have different reservation wages, there is also a steady state distribution of previous wages in the stock of unemployed workers, which we denote by $H(w)$.[55]

In the steady state, the inflow into the set of workers employed with a wage of w or less must equal the outflow:

$$u\lambda \int_0^{\bar{w}} s_{un}(x) [\max\{F(w) - F(R(x)), 0\}] dH(x) = \delta G(w)(1-u) + \left[(1-\delta)\lambda(1-u) \int_0^w s_{em}(x) dG(x) \right](1-F(w)) \tag{32}$$

$$Inflow(w) = Layoff(w) + Offer(w) \cdot (1-F(w)). \tag{33}$$

---

[55]It can be shown that:

$$H(w) = \frac{\int_0^w \frac{dG(x)}{\lambda s_{un}(x) - \lambda s_{un}(x) F(R(x))}}{\int_0^{\bar{w}} \frac{dG(x)}{\lambda s_{un}(x) - \lambda s_{un}(x) F(R(x))}}.$$

The quantity $Inflow(w)$ is the fraction of the stock of unemployed workers who receive a wage offer that exceeds their reservation wage, but is less than w. On the right hand side, the proportion with a wage less than w and displaced with probability δ is given by $Layoff(w)$), while the proportion of individuals who will leave jobs that pay less than w for jobs that pay more than w is given by $(Offer(w) \cdot (1 - F(w)))$.

Now consider a w within a neighborhood of the threshold T^{max}.[56] Consider the above flow equation for observed wages between $w + h$ and w. Some re-arrangement yields:

$$
\begin{aligned}
Inflow(w+h) - Inflow(w) + Offer(w)((F(w+h)) - F(w)) &= Layoff(w+h) - Layoff(w) \\
&\quad + (Offer(w+h) - Offer(w))(1 - F(w+h)).
\end{aligned}
$$

Applying a mean value theorem for Stieltjes integrals on the right hand side, re-arranging, and dividing by h, we obtain

$$
\frac{\frac{Inflow(w+h) - Inflow(w)}{h} + Offer(w)\frac{F(w+h) - F(w)}{h}}{\delta(1-u) + (1-\delta)\lambda(1-u)c_O(1 - F(w+h))} = \frac{G(w+h) - G(w)}{h}
$$

where $\inf_{x \in [w, w+h]} s_{em}(x) \le c_O \le \sup_{x \in [w, w+h]} s_{em}(x)$.

By assumption the distribution of wage offers $F(\cdot)$ is differentiable. Moreover, it can be shown that the search intensity choice of employed workers is continuous, and that $Inflow(w)$ is differentiable in a neighborhood of T.[57] Taking the limit as $h \to 0$, we obtain:

$$
\frac{Inflow'(w) + Offer(w)f(w)}{(\delta(1-u) + (1-\delta)\lambda(1-u)s_{em}(w)(1 - F(w)))} = g(w) \tag{34}
$$

which means that the density of wages $g(w)$ is well-defined in this neighborhood.

It can be shown that every function of w on the left-hand side of this equation is continuously differentiable at $w = T^{max}$ except the search intensity function $s_{em}(\cdot)$, which is kinked at T^{max}. As noted above, this arises because of the kinks in the value functions $W_{em}(\cdot)$ and $W_{um}(\cdot)$ at the wage threshold T^{max}. As a consequence, the density of wages among employed workers has a kink at $w = T^{max}$. Assuming that the job destruction rate is constant across all jobs, the population of new UI claimants has the same distribution of previous wages as the pool of employed workers. As a consequence, this model implies that the density of

---

[56] We choose a neighborhood of T^{max} in which $w > R(T^{max})$. Such a neighborhood always exists because $T^{max} > R(T^{max})$—a worker who accepts a wage T^{max} will be strictly better off than remaining unemployed with the maximum benefit $\bar{b}$.

[57] Differentiability of $Inflow(\cdot)$ follows because in a neighborhood of T, $w > R(x)$ for all x. Thus $Inflow(w) = u\lambda \int_0^{\bar{w}} s_{un}(x)[F(w) - F(R(x))]\,dH(x)$. We can differentiate under the integral sign because the derivative of the integrand with respect to w is continuous in the rectangle defined by the neighborhood of T and $[0, \bar{w}]$, and $F(\cdot)$ is differentiable by assumption.

wages among new UI claimants has a kink at $w = T^{max}$.

Model with Imperfect Information About Benefit Schedules. We now consider a variant of the preceding model in which agents have imperfect information on the location of the kink point in the benefit schedules. We show that the prediction of a kinked density is *not* robust to small errors. To proceed, assume that the true kink in the benefit schedule occurs at $w = T^{max}$, but the agent makes choices assuming the kink is at $T^{max} + \varepsilon$. This leads to value functions, indexed by the error ε, , $W_{em}^{\varepsilon}(w)$ and $W_{un}^{\varepsilon}(w)$ parallel to those in equations (29) and (30). In addition, there is another value function defined by:

$$W_{un}^{\varepsilon *}(w_{-1}) = \max_{s \geq 0} \left\{ b(w_{-1}) - c(s) + \beta \left[\lambda s \int \max\{W_{em}^{\varepsilon}(x), W_{un}^{\varepsilon *}(w_{-1})\} dF(x) + (1 - \lambda s)W_{un}^{\varepsilon *}(w_{-1}) \right] \right\}.$$

$W_{un}^{\varepsilon}(w_{-1})$ is the perceived value of unemployment for a worker who is using an incorrect benefit formula, whereas $W_{un}^{\varepsilon *}(w_{-1})$ is the perceived value of unemployment of an unemployed worker who is receiving benefits $b(w_{-1})$ based on the correct formula, but is evaluating the value of potential future employment using W_{em}^{ε}.

The result of this small optimization error is that actual search intensity for an individual (in the employed and unemployed state) will be given by the first-order conditions

$$c'(s_{em}^{\varepsilon}(w)) = \beta(1 - \delta)\lambda \int_{w}^{\bar{w}} [W_{em}^{\varepsilon}(x) - W_{em}^{\varepsilon}(w)] dF(x)$$

$$c'(s_{un}^{\varepsilon *}(w_{-1})) = \beta\lambda \int_{R^{\varepsilon *}(w_{-1})}^{\bar{w}} [W_{em}^{\varepsilon}(x) - W_{un}^{\varepsilon *}(w_{-1})] dF(x).$$

Moreover, the reservation wage for employed agents is still their current wage, while the reservation wage if unemployed is $R^{\varepsilon *}(w_{-1})$, implicitly defined by $W_{em}^{\varepsilon}(R^{\varepsilon *}(w_{-1})) = W_{un}^{\varepsilon *}(w_{-1})$.

With an error in the perceived kink the steady state flow equation for the wage density $G(w)$ is the same as in equation (32), after replacing $s_{em}(\cdot), s_{un}(\cdot)$ with $s_{em}^{\varepsilon}(\cdot)$, $s_{un}^{\varepsilon *}$. As a result, the steady state density for a population of agents of type "ε" exhibits a kink at $T^{max} + \varepsilon$. If the true population contains a mixture of agents with different values of ε, drawn from a density $\phi_{\varepsilon}(\cdot)$, then the steady state flow equation for the density of wages is the same as in equation (32), after replacing $s_{em}(\cdot), s_{un}(\cdot)$ with $E[s_{em}^{\varepsilon}(\cdot)]$, $E[s_{un}^{\varepsilon *}(\cdot)]$ where expectations are taken with respect to $\phi_{\varepsilon}(\cdot)$. It can be shown that if ε is continuously distributed, then $E[s_{em}^{\varepsilon}(x)]$ will be continuously differentiable, leading to a continuously differentiable steady state density $g(\cdot)$.[58] Thus, a continuous distribution of errors in agents' beliefs about the location of the kink point will

---

[58] Specifically, $E[s_{em}^{\varepsilon}(w)]$ is continuously differentiable because $\frac{ds_{em}^{\varepsilon}(w)}{dw}$ is continuous at $w = T^{max}$ almost everywhere (it is non-differentiable only when $\varepsilon = 0$, a measure zero event).

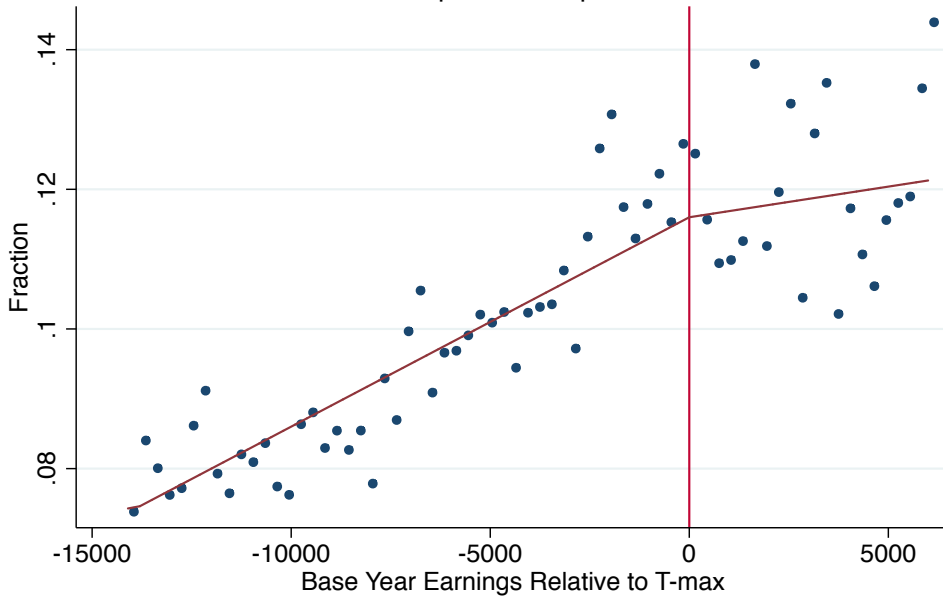smooth out the kink that arises with full information.

Appendix Figure 1a: Fraction Exhausting UI Benefits
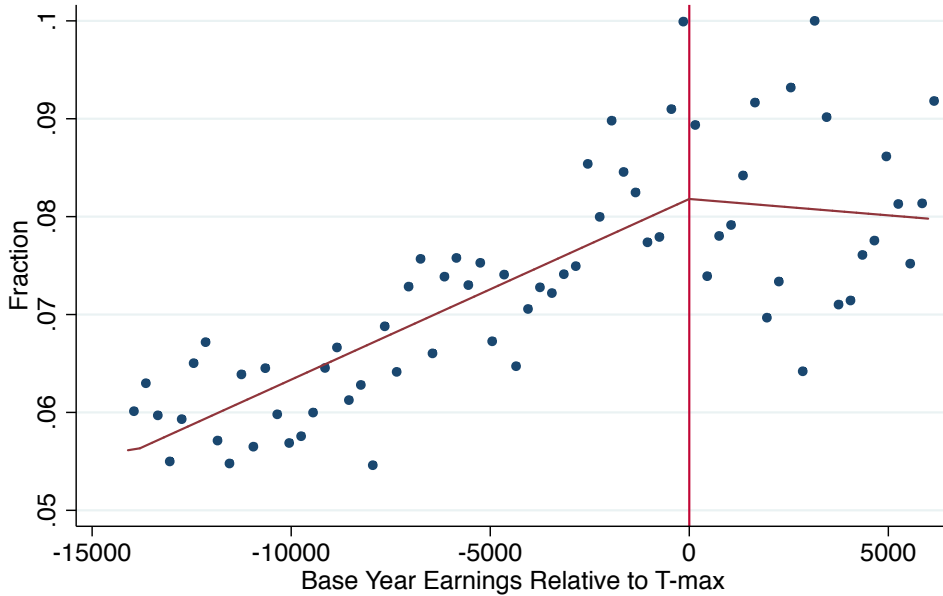
Bottom Kink Sample
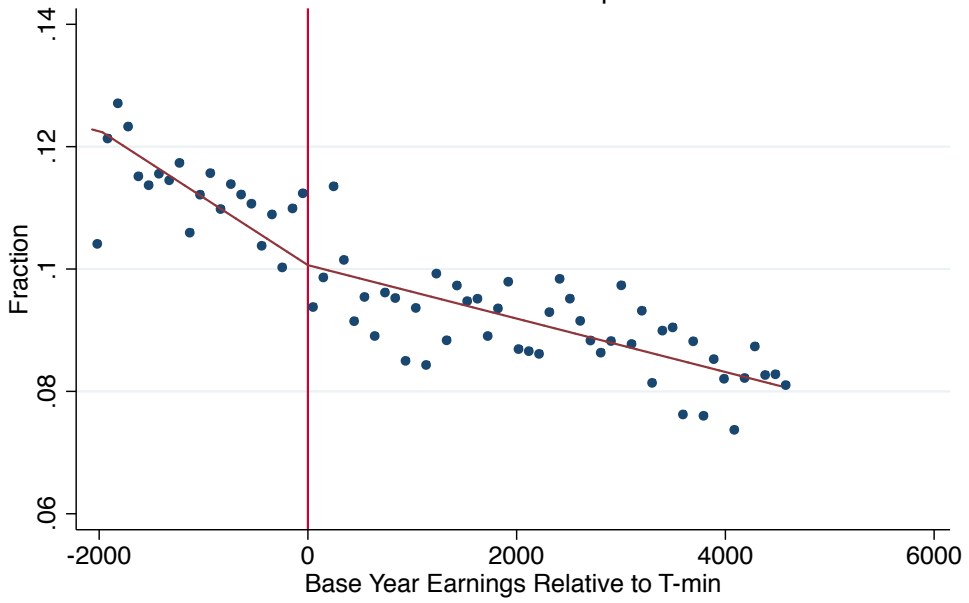
Appendix Figure 1b: Fraction Claiming UA Benefits
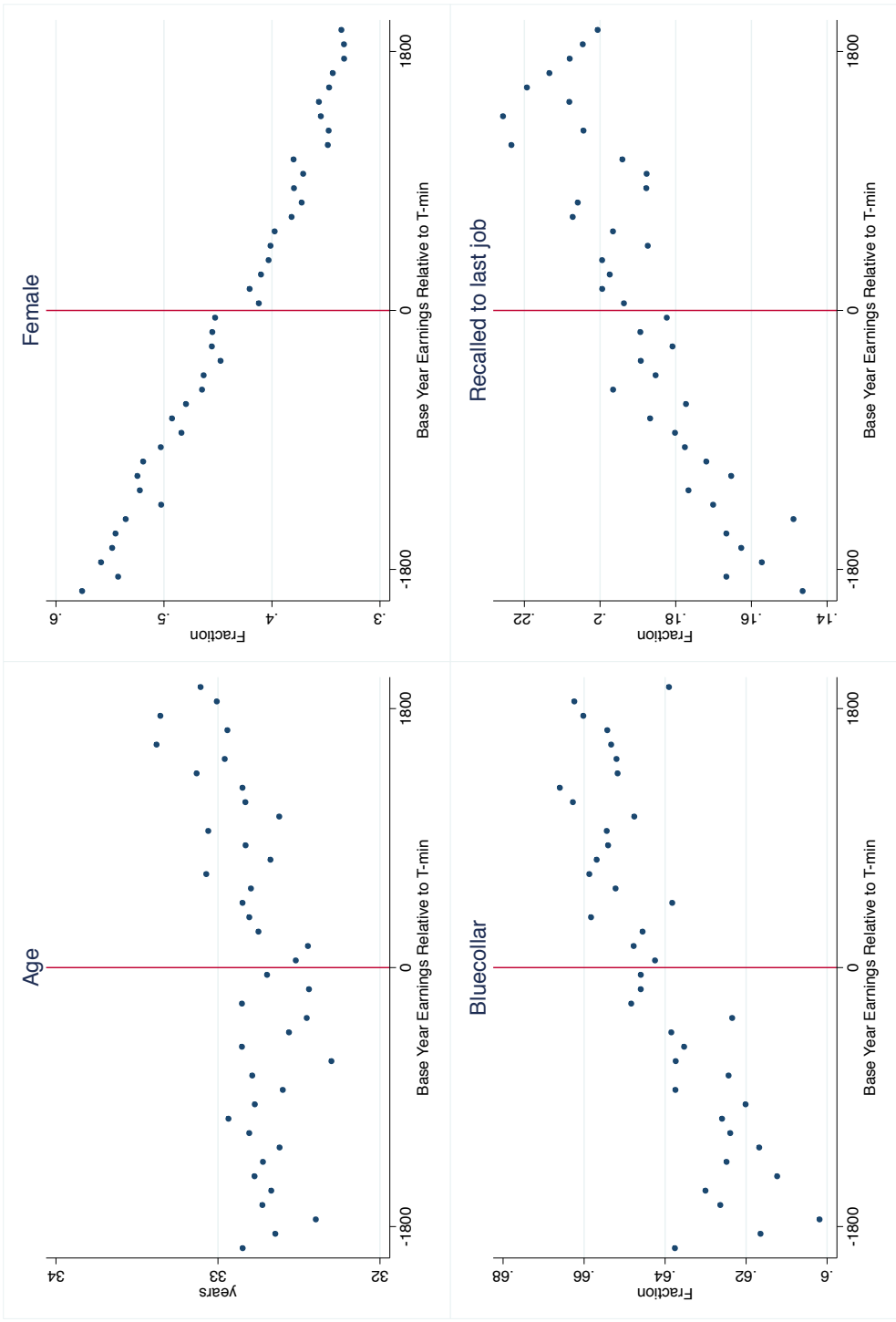
Bottom Kink Sample
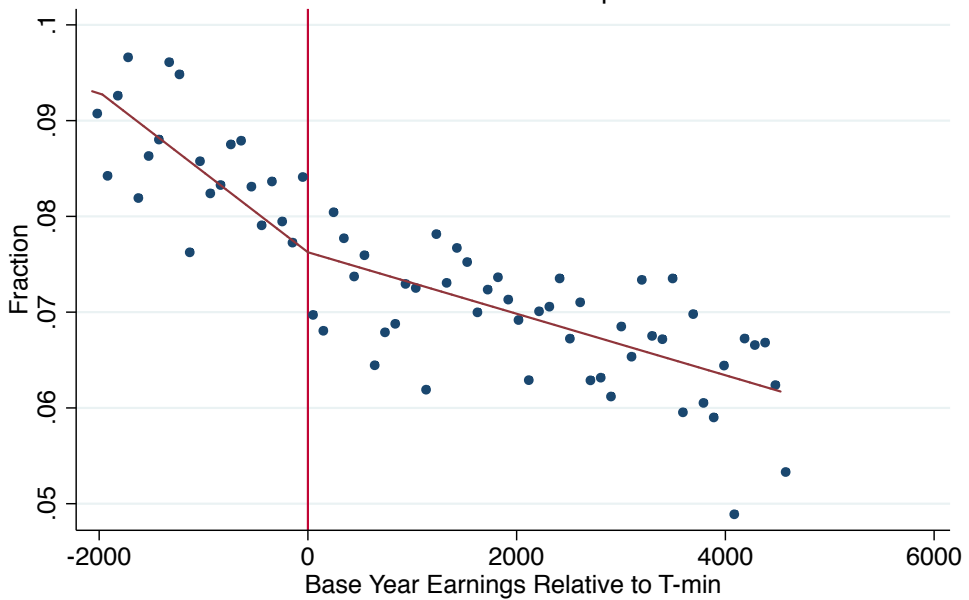
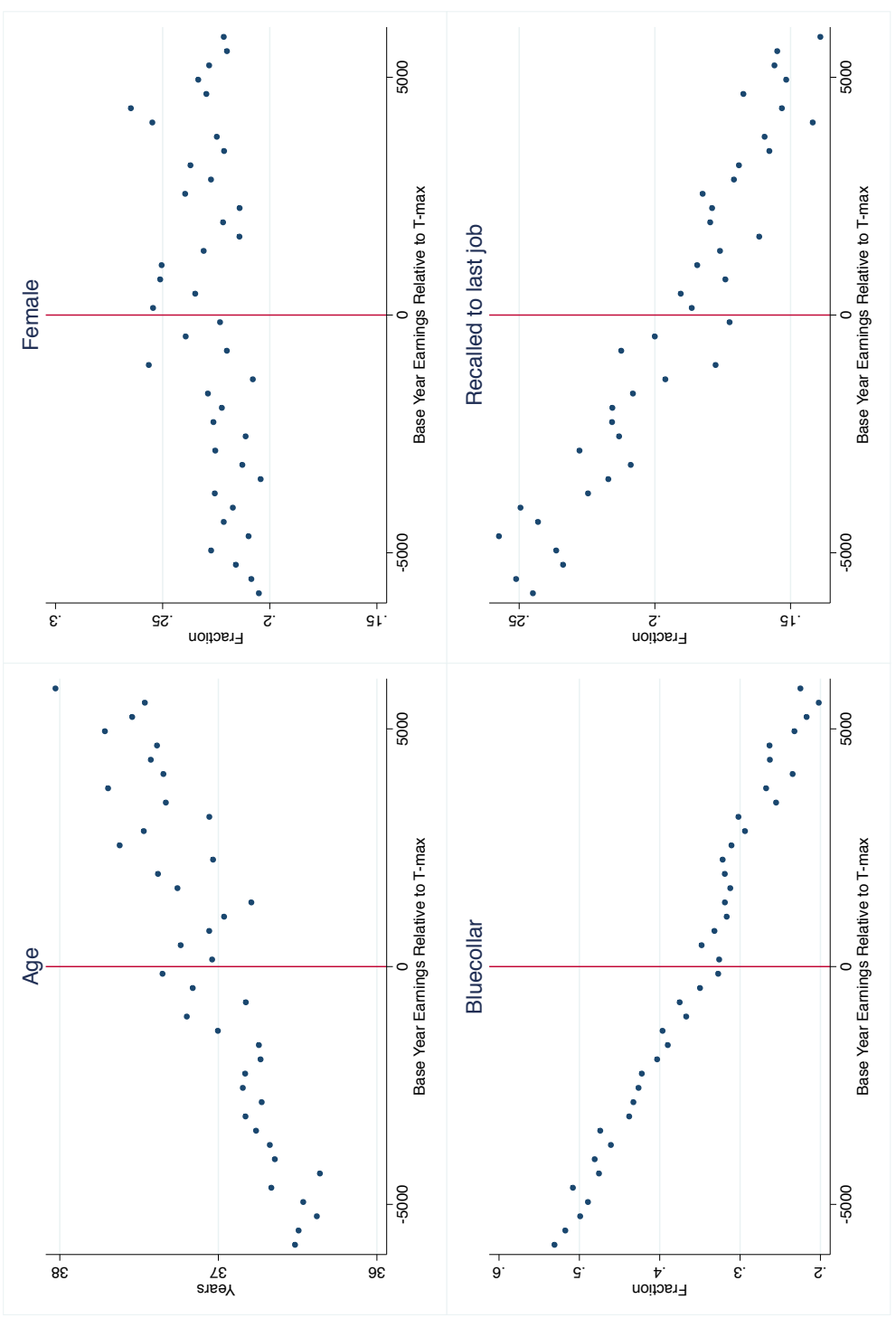Appendix Figure 1c: Fraction Exhausting UI Benefits

Top Kink Sample

Appendix Figure 1d: Fraction Claiming UA Benefits
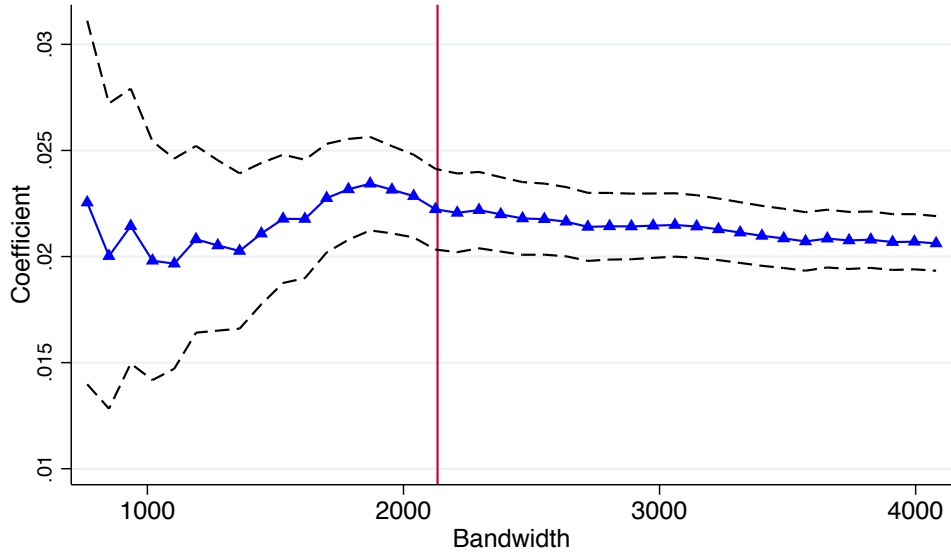
Top Kink Sample

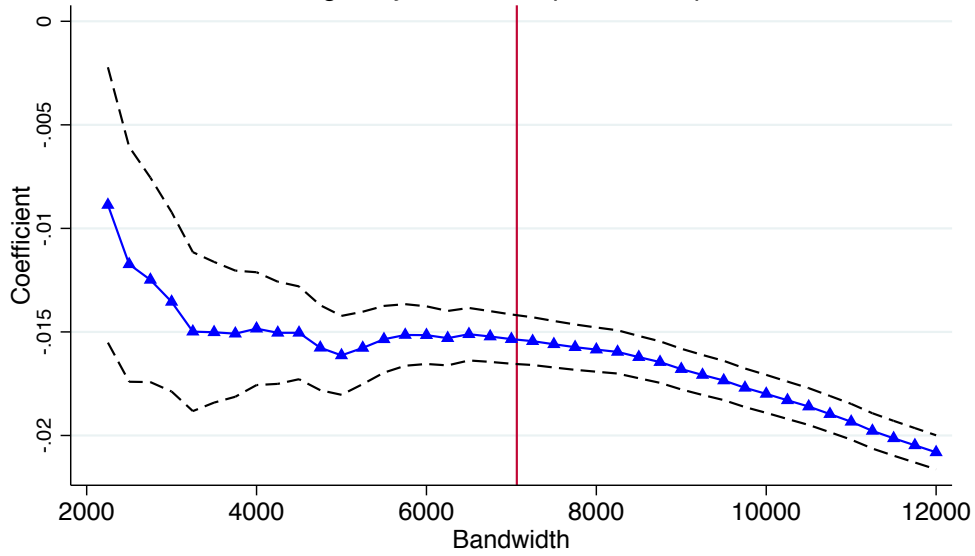# Appendix Figure 2: Covariates Bottom Kink Sample

Appendix Figure 3: Covariates Top Kink Sample

Appendix Figure 4: Reduced Form Estimation with Varying Bandwidth
Log Daily Benefit, Bottom Kink Sample

Notes: local linear estimation, estimated coefficients (blue) with confidence bounds (dash), vertical line denotes FG bandwidth

Appendix Figure 5: Reduced Form Estimation with Varying Bandwidth
Log Daily Benefit, Top Kink Sample

Notes: local linear estimation, estimated coefficients (blue) with confidence bounds (dash), vertical line denotes FG bandwidth

Appendix Figure 6: Reduced Form Estimation with Varying Bandwidth
Log Time to Next Job, Bottom Kink Sample
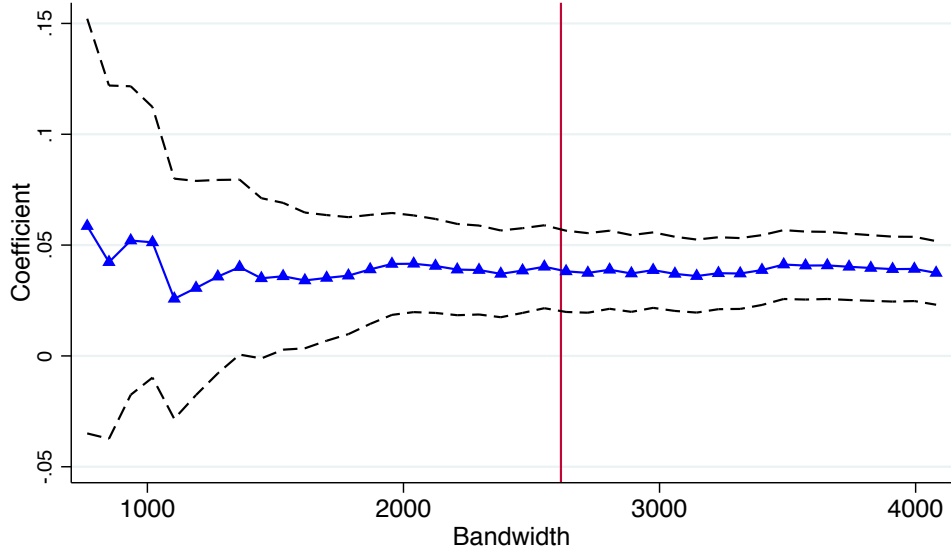
Notes: local linear estimation, estimated coefficients (blue) with confidence bounds (dash),
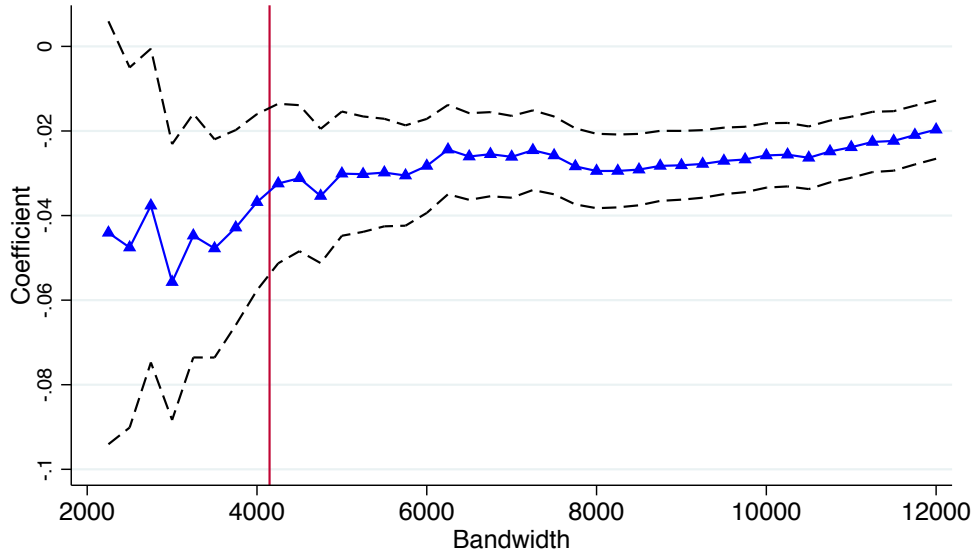vertical line denotes FG bandwidth

Appendix Figure 7: Reduced Form Estimation with Varying Bandwidth
Log Time to Next Job, Top Kink Sample

Notes: local linear estimation, estimated coefficients (blue) with confidence bounds (dash),
vertical line denotes FG bandwidth

Appendix Figure 8: Reduced Form Estimation with Varying Bandwidth
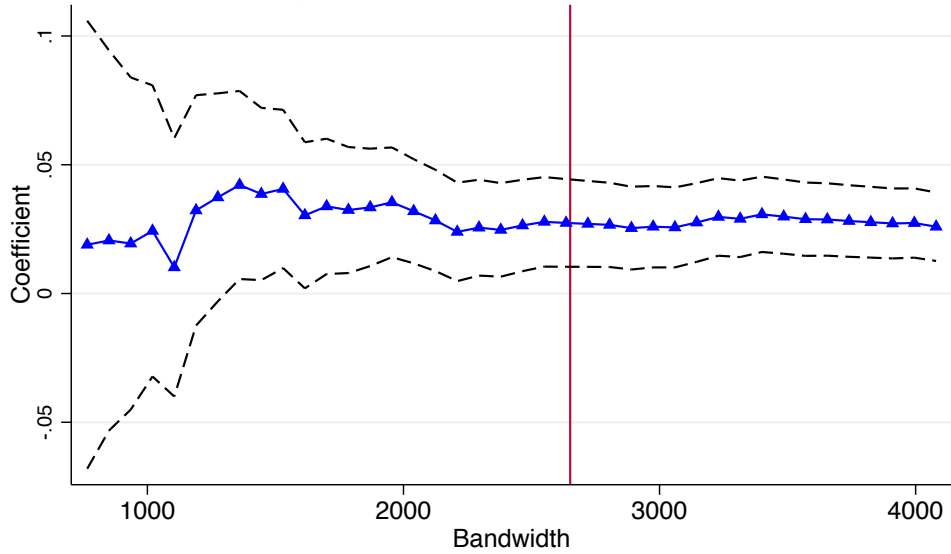Log Claim Duration, Bottom Kink Sample

Notes: local linear estimation, estimated coefficients (blue) with confidence bounds (dash),
vertical line denotes FG bandwidth

Appendix Figure 9: Reduced Form Estimation with Varying Bandwidth
Log Claim Duration, Top Kink Sample

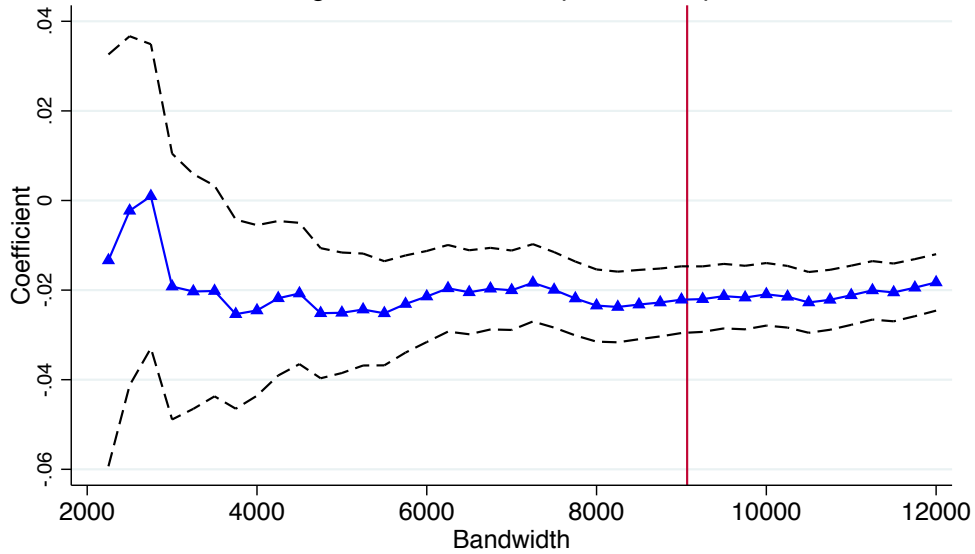Notes: local linear estimation, estimated coefficients (blue) with confidence bounds (dash),
vertical line denotes FG bandwidth

Appendix Table 1: Test for Kinks in Density of Previous Wage Based on Models for Frequency
Distribution of Observations Estimated by Minimum Chi-Squared

Polynomial Order in Model for Density (1)	Estimated Kink x 1000 (Standard Error) (2)	Pearson χ^2 (P-value) (3)	Akaike Criterion (4)
A. Bottom Kink Sample (65 bins of width = 100 Euro/year)			
2	-0.120 (0.051)	100.52 (0.08%)	105.52
3	-0.090 (0.127)	96.04 (0.13%)	103.04
4	-0.200 (0.255)	95.33 (0.08%)	104.33
5	-0.080 (0.444)	95.10 (0.05%)	106.10
B. Top Kink Sample (66 bins of width = 300 Euro/year)			
2	-0.060 (0.038)	124.80 (0.00%)	129.80
3	-0.280 (0.087)	104.99 (0.00%)	111.99
4	0.158 (0.129)	60.08 (36.48%)	69.08
5	0.149 (0.219)	57.80 (37.22%)	68.80

Notes: See text. Bottom kink sample includes 181,623 observations (1856 observations in left-most and right-most bins deleted). Top kink sample includes 183,056 observations (3031 observations in left-most and left-most bins deleted). Models are estimated by minimum chi-square estimation. Model for fraction of observations in a bin is a polynomial function of the bin counter of the order indicated in column 1, with interactions of poynomial terms with indicator for bins to the right of the kink point (main effect of indicator is excluded). Estimated kink in column 2 is coefficient of interaction between linear term and indicator for bin to right of kinkpoint. Akaike criterion in column 4 is chi-squared model fit statistic plus 2 times the number of parameters in the model.

Appendix Table 2: Reduced Form Estimates of Kink Effects in Benefits and Durations, based on Alternative Bandwidth Determination

	Local Linear Models		Local Quadratic Models	
	FG Bandwidth (1)	Estimated Kink (2)	FG Bandwidth (3)	Estimated Kink (4)
A. Bottom Kink:				
Log daily UI benefit	550	0.0251 (0.0070)	1,806	0.0148 (0.0048)
Log time to next job	1,547	0.0366 (0.0166)	3,482	0.0335 (0.0316)
Log claim duration	1,762	0.0317 (0.0127)	3,799	0.0437 (0.0281)
B. Top Kink:				
Log daily UI benefit	1,606	-0.0191 (0.0059)	5,753	-0.0154 (0.0032)
Log time to next job	3,510	-0.0462 (0.0131)	8,329	-0.0525 (0.0178)
Log claim duration	3,790	-0.0259 (0.0106)	9,647	-0.0355 (0.0149)

Notes: see Table 2. In this table, the FG bandwidth determination is based on a first step global polynomial with a continuous first derivative at kink.

Appendix Table 3: Estimated Structural Coefficients from Fuzzy Regression Kink Design, Alternative Bandwidth Determination

	Local Linear Models		Local Quadratic Models	
	FG Bandwidth (1)	Estimated Elasticity (2)	FG Bandwidth (3)	Estimated Elasticity (4)
A. Bottom Kink:				
Log time to next job	1,547	1.673 (0.772)	3,482	1.604 (1.534)
Log claim duration	1,762	1.383 (0.562)	3,799	2.186 (1.450)
B. Top Kink:				
Log time to next job	3,510	3.098 (0.954)	8,329	3.186 (1.177)
Log claim duration	3,790	1.727 (0.728)	9,647	2.565 (1.137)

Notes: FG bandwidth determination is based on a first step global polynomial with a continuous first derivative at kink. FG bandwidth is based on determination for outcome variable. Estimated elasticities in columns 2 and 4 are obtained from a 2SLS procedure. See note to Table 4.

Appendix Table 4: Fuzzy RKD Estimates of Benefit Elasticities of Time to Next Job with Alternative
Censoring Points

	Local Linear Models		Local Quadratic Models	
	FG Bandwidth (1)	Estimated Elasticity (2)	FG Bandwidth (3)	Estimated Elasticity (4)
A. Bottom Kink Sample				
Censored at 52 weeks	2,615	1.726	4,328	3.024
		(0.440)		(1.501)
Censored at 39 weeks	2,616	1.549	4,142	2.468
		(0.410)		(1.396)
Censored at 30 weeks	2,579	1.521	4,029	1.749
		(0.382)		(1.315)
Censored at 20 weeks	2,509	1.254	3,925	1.259
		(0.337)		(1.134)
B. Top Kink Sample				
Censored at 52 weeks	4,148	2.643	7,521	3.497
		(0.715)		(1.278)
Censored at 39 weeks	4,219	2.039	7,613	3.012
		(0.625)		(1.173)
Censored at 30 weeks	4,262	1.839	7,687	2.589
		(0.559)		(1.056)
Censored at 20 weeks	4,314	1.506	7,838	2.174
		(0.471)		(0.906)

Notes: see notes to Table 4. Dependent variable in all models is log of time to next job, censored at maximum value indicated in row heading.

Appendix Table 5: Fuzzy RKD Estimates of Benefit Elasticities of Alternative Measures of UI Claim Duration

	Local Linear Models		Local Quadratic Models	
	FG Bandwidth (1)	Estimated Elasticity (2)	FG Bandwidth (3)	Estimated Elasticity (4)
A. Bottom Kink Sample				
Duration of First UI Spell	2,651	1.250 (0.406)	4,564	2.816 (1.401)
Total Days of UI Claimed	2,654	1.462 (0.390)	4,203	1.984 (1.324)
B. Top Kink Sample				
Duration of First UI Spell	9,067	1.313 (0.228)	9,355	2.500 (1.103)
Total Days of UI Claimed	5,301	1.568 (0.401)	8,447	2.347 (1.009)

Notes: see notes to Table 4.

Appendix Table 6: Summary of Estimated Benefit Elasticities in Existing Literature

Authors (date)	Data	Design	Elasticity Estimate or Range
U.S. Studies			
Classen (1977)	CWBH*, Arizona	Pre/post	0.6 - 1.0
Moffitt (1985)	CWBH, 13 states	Cross-sectional	0.4
Solon (1985)	CWBH, Georgia	Diff-in-diff, tax policy change	0.7
Meyer (1990)	CWBH, all states	State-by-year	0.8
Katz and Meyer (1990)	CWBH, all states	State-by-year	0.8
Meyer and Mok (2007)	New York State UI Records	Pre/post	0.3
Chetty (2010)	SIPP (retrospective interviews)	State-by-year	0.5
Landais (2012)	CWBH, Louisiana/Washington	RKD, maximum benefit	0.4 - 0.7
European Studies			
Carling et al. (2011)	Sweden, register data (outcome = time to next job)	Diff-in-diff, repl. rate change	1.6
Roed and Zhang (2003)	Norway, register data, previous job < 2 years	Calendar vs. spell dating	0.3 (female) 0.9 (male)
Lalive et al. (2006)	Austria, register/Social Security data	Diff-in-diff, repl. rate change	0.15
Arraz et al. (2008)	Spain, register data	Pre/post	0.8

* Note: CWBH is the Continuous Work and Benefit History data set, based on employment and unemployment records.