

WORKING PAPER #553  
PRINCETON UNIVERSITY  
INDUSTRIAL RELATIONS SECTION  
November 25, 2009  
<http://www.irs.princeton.edu/pubs/pdfs/553.pdf>

## Quasi-Experimental Identification and Estimation in the Regression Kink Design<sup>1</sup>

David Card  
UC Berkeley and NBER

David S. Lee  
Princeton University and NBER

Zhuan Pei  
Princeton University

November 2009

### Abstract

We consider nonparametric identification of the average marginal effect of a continuous endogenous regressor in a generalized nonseparable model when the regressor of interest is a known, deterministic, but kinked function of an observed continuous assignment variable. This design arises in many institutional settings where a policy variable of interest (such as weekly unemployment benefits) is mechanically related to an observed but potentially endogenous variable (like previous earnings). We characterize a broad class of models in which a “Regression Kink Design” (RKD) provides valid inferences for the underlying marginal effects. Importantly, this class includes cases where the assignment variable is endogenously chosen. Under suitable conditions we show that the RKD estimand identifies the “treatment on the treated” parameter (Florens et al., 2009) or the “local average response” (Altonji and Matzkin, 2005) that is identified in an ideal randomized experiment. As in a regression discontinuity design, the required identification assumption implies strong and readily testable predictions for the pattern of pre-determined covariates around the kink point. Standard local linear regression techniques can be easily adapted to obtain “nonparametric” RKD estimates. We illustrate the RKD approach by examining the effect of unemployment insurance benefits on the duration of benefit claims, using rich microdata from the state of Washington.

---

<sup>1</sup>Diane Alexander and Pauline Leung provided excellent research assistance. We also thank Ken Kline at the Upjohn Institute.

# 1 Introduction

Reflecting the widespread concern over the potential fragility of parametric methods, a growing literature considers the identification and estimation of nonparametric regression models with endogenous regressors (e.g., Blundell and Powell (2003); Chesher (2003); Altonji and Matzkin (2005); Florens et al. (2009); Imbens and Newey (2009)). The most general of these models allow both the observed regressors and the unobserved errors to enter an underlying structural function in an arbitrary way. Building on methods for additive models (e.g., Heckman and Robb (1985)) most recent studies follow a control function approach. Conditioning on a suitable control function, the regressor of interest is independent of the unobserved errors, and a variety of non-parametric methods can be used to estimate the associated causal effects. The control function approach relies on the existence of one or more “instruments”— variables that are assumed to be independent of the errors in the regression function. Identification hinges on the validity of the independence assumption, in much the same way that identification in a linear simultaneous equations model depends on orthogonality between the instrumental variables and the additive structural error terms.

In some applied contexts, however, 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 a behaviorally endogenous assignment variable. The level of unemployment benefits, for example, is typically set by a formula that depends on previous earnings. In such settings it is arguably impossible to identify individual characteristics that affect the level of the policy variable and yet are independent of underlying heterogeneity in preferences and/or opportunities. Nevertheless, a common feature of many policy rules is the existence of a “kink”, or series of kinks, in the formula that relates the assignment variable to the policy variable. In the case of unemployment benefits, for example, a typical formula provides a fixed fraction of pre-job-loss earnings, subject to a maximum rate. Likewise, the income tax system in most countries is piece-wise linear, with progressively higher tax rates at each kink point. As has been noted in recent studies by Guryan (2003), Nielsen et al. (forthcoming), and Simonsen et al. (2009), the existence of a kinked policy rule holds out the possibility for identification of the effect of the policy variable, even in the absence of traditional instruments. In essence, the idea is to look for an induced kink in the outcome variable that coincides with the kink in the policy rule, and relate the relative magnitudes of the two kinks. While this “regression kink design” (RKD) is potentially attractive, an important concern is the endogeneity of the assignment variable. As noted by Saez (forthcoming), for

example, a kink in the marginal tax schedule would be expected to lead to “bunching” of taxpayers at the level of income associated with the kink. Such endogenous sorting could lead to a non-smooth distribution of unobserved heterogeneity around the kink-point, confounding inferences based on a regression kink design.

This paper establishes the conditions under which the behavioral response to a formulaic policy variable like unemployment benefits or the marginal tax rate can be identified within a general class of nonparametric regression models. We show that, in the context of a fully nonparametric regression with non-additive errors, the regression kink design can identify what Altonji and Matzkin (2005) have called the “local average response” to the policy variable, or equivalently the “treatment on the treated” parameter characterized by Florens et al. (2009). The key condition for identification is that conditional on the unobservables, the density of the assignment variable is smooth – continuously differentiable – at the kink-point in the policy rule. We show that this “smooth density” condition rules out extreme forms of endogenous sorting, which might arise when agents can deterministically manipulate the value of the assignment variable used in the policy formula, while allowing for many other forms of endogeneity in the assignment variable. 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. In particular, the conditional distribution functions of observed covariates that are determined prior to the policy variable should have continuous derivatives with respect to the assignment variable at the kink point. Thus, we show that, as in a regression discontinuity design (see Lee and Lemieux (forthcoming)), the validity of the regression kink design can be tested.

Using administrative data from the Unemployment Insurance (UI) system in the state of Washington in 1988, we apply a regression kink approach to estimate the average impact of a marginal increase in weekly UI benefits on total benefits paid, and the duration of UI receipt. We find little evidence that the density of base period earnings (the assignment variable) is discontinuous at the threshold associated with the maximum benefit rate; we also find that the means of key covariates generally have smooth derivatives with respect to the assignment variable, consistent with a valid RK Design. Our estimates suggest that a \$1 increase in the benefit amount leads to a 0.04 week increase in the duration of UI receipt and a \$18 rise in total benefits paid. These numbers translate into a 1.6 week increase in the duration of insured unemployment in response to a 10 percentage point increase in the UI replacement rate – an estimate that is similar in magnitude to the estimate by Meyer (1990), but somewhat larger in magnitude than estimates reported by Hamermesh (1977) or Moffitt and Nicholson (1982).

The paper is organized as follows. Section 2 discusses parameters of interest and identification in the Regression Kink Design. Section 3 then describes the institutional details of the UI system in the state of Washington, as background for our empirical analysis, which we present in Section 4. Section 5 concludes.

## 2 Nonparametric Regression and the Regression Kink Design

### 2.1 Background

We begin with some background on the existing literature. Consider the model

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

where  $Y$  is an outcome,  $B$  is a continuous regressor of interest,  $V$  is another covariate that enters the model, and  $W$  is an unobservable, non-additive error term. 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  $Y$ .<sup>1</sup> 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  $Y_1 - Y_0 = y(1, V, W) - y(0, V, W)$ .

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

$$ASF(b, v) = \int y(b, v, w) dG(w),$$

where  $G(\cdot)$  is the c.d.f. of  $W$ . This gives the average value of  $Y$  that would occur if the entire population (as represented by the unconditional distribution of  $W$ ) was assigned to a particular value of the pair  $(b, v)$ . Florens et al. (2009) 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.

---

<sup>1</sup>In Imbens and Newey (2009),  $W$  is considered to have unknown dimension. We, too, can allow for that generality.

A closely related construct is the “treatment on the treated” (TT) parameter of Florens et al. (2009):

$$TT_b(b, v) = \int \frac{\partial y(b, v, w)}{\partial b} dG(w|b, v).$$

As noted by Florens et al. 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 at  $G(\cdot|b, v)$ .

Recent studies, including Florens et al. (2009) 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 considers using 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 policy variable that is a deterministic function of a behaviorally endogenous variable – where no instruments can plausibly satisfy the independence assumption, either globally or locally. In the framework of equation (1), consider the case where  $B$  represents the level of unemployment benefits available to a newly unemployed worker,  $Y$  represents the duration of unemployment, and  $V$  represents pre-job-loss earnings. Assume (as in many institutional settings) that unemployment benefits are computed as a fixed fraction of  $V$  up to some maximum weekly benefit. Conditional on  $V$  there is no variation in the benefit level, so model (1) is not non-parametrically identified. One could try to get around this fundamental non-identification by treating  $V$  as an error component that is 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 circumstance, it may still be possible to exploit the kinked benefit rule to identify a 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  $v = v_0$  (the lowest level of earnings at which the individual receives the maximum benefit rate) then we should expect to see an induced kink in the relationship between  $Y$  and  $V$  at  $v = v_0$ . 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.<sup>2</sup> More recently, Simonsen et al. (2009) use a kinked relationship between the total expenditure on prescription drugs and the marginal price to study the price sensitivity of demand for prescription drugs. Nielsen et al. (2009), who introduce the term “Regression Kink Design” for this approach, use a kinked student aid scheme to identify the effect of direct costs on college enrollment. Nielsen et al. (2009) make precise the assumptions needed to identify the causal effects in the additive model

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

where  $B = b(V)$  is a deterministic (and continuous) function of  $V$  with a kink at  $v = 0$ . Nielsen et al. (2009) show that if  $g(\cdot)$  and  $E[\varepsilon|V = v]$  have derivatives that are continuous in  $v$  at  $v = 0$ , then

$$\tau = \frac{\lim_{v \rightarrow 0^+} \frac{\partial E[Y|V=v]}{\partial v} - \lim_{v \rightarrow 0^-} \frac{\partial E[Y|V=v]}{\partial v}}{\lim_{v \rightarrow 0^+} \frac{\partial b(v)}{\partial v} - \lim_{v \rightarrow 0^-} \frac{\partial b(v)}{\partial v}}.$$

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.<sup>3</sup>

## 2.2 Necessary and Sufficient Conditions for RKD in a Non-Separable Model

As background for our main results we first specify the necessary and sufficient conditions for an RKD to identify marginal effects in a general nonseparable model, in the same way Hahn et al. (2001) formally established the identifying assumptions for Thistlethwaite and Campbell (1960)’s RD design in the heterogeneous treatment effects, potential outcomes framework.

**Proposition 1.** *For the nonseparable model (1),  $B = b(V)$ , where  $b(\cdot)$  is continuous but has discontinuous derivative at 0, let  $g(w|v)$  be the density of  $w$  conditional on  $v$ . If  $\forall w$  i)  $\frac{\partial y(b, v, w)}{\partial b}$  is continuous in  $v$  at  $v = 0$ , ii)  $\frac{\partial y(b, v, w)}{\partial v}$  is continuous in  $v$  except possibly at  $v = 0$ , iii)  $\frac{\partial g(w|v)}{\partial v}$  is continuous in  $v$  except possibly at  $v = 0$ ,*

<sup>2</sup>Guryan (2003) 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.”

<sup>3</sup>In an earlier working paper version, Nielsen et al. (2008) provide similar conditions for identification for a less restrictive, additive model,  $Y = g(B, V) + \varepsilon$ .

then a necessary and sufficient condition for

$$\frac{\lim_{v \rightarrow 0^+} \frac{\partial E[Y|V=v]}{\partial v} - \lim_{v \rightarrow 0^-} \frac{\partial E[Y|V=v]}{\partial v}}{\lim_{v \rightarrow 0^+} b'(v) - \lim_{v \rightarrow 0^-} b'(v)} = \int y_1(b(0), 0, w) g(w|0) dw \equiv TT_b(b(0), 0)$$

is that  $\left. \frac{\partial E[y(b, V, W)|V=v]}{\partial v} \right|_{b=b(v)}$  is continuous in  $v$  at  $v = 0$

To see this, consider the partial derivative of the conditional expectation function  $E[Y|V=v] \equiv E[y(B, V, W)|V=v]$  at  $v \neq 0$ ,

$$\begin{aligned} \frac{\partial E[y(b(V), V, W)|V=v]}{\partial v} &= \int \frac{\partial (y(b(v), v, w) g(w|v))}{\partial v} dw \\ &= \int (y_1(b(v), v, w) b'(v) + y_2(b(v), v, w)) g(w|v) + y(b(v), v, w) \frac{\partial g(w|v)}{\partial v} dw \\ &= E[y_1(b(V), V, W)|V=v] b'(v) + \int y_2(b(v), v, w) g(w|v) + y(b(v), v, w) \frac{\partial g(w|v)}{\partial v} dw \\ &= E[y_1(b(V), V, W)|V=v] b'(v) + \left. \frac{\partial E[y(b, V, W)|V=v]}{\partial v} \right|_{b=b(v)} \end{aligned}$$

By taking the difference between the right and left limits of this expression, we obtain the result that

$$\frac{\lim_{v \rightarrow 0^+} \frac{\partial E[y(b(V), V, W)|V=v]}{\partial v} - \lim_{v \rightarrow 0^-} \frac{\partial E[y(b(V), V, W)|V=v]}{\partial v}}{\lim_{v \rightarrow 0^+} b'(v) - \lim_{v \rightarrow 0^-} b'(v)} = E[y_1(b(0), 0, W)|V=0]$$

if and only if  $\left. \frac{\partial E[y(b, V, W)|V=v]}{\partial v} \right|_{b=b(v)}$  is continuous in  $v$  at  $v = 0$ .<sup>4</sup>

Intuitively, a marginal increase in  $V$  induces an effect on  $Y$  through  $b$ , but also via the functional dependence of  $y(B, V, W)$  on  $V$ , and via the changing distribution of unobserved heterogeneity (reflected in  $g(w|v)$ ). Only when the latter two effects evolve smoothly as  $V$  reaches the kink point will the change in the derivative  $\frac{\partial E[y(b(V), V, W)|V=v]}{\partial v}$  isolate the causal effect of  $b$  on  $Y$ . Formally, condition for the smoothness of  $\left. \frac{\partial E[y(b, V, W)|V=v]}{\partial v} \right|_{b=b(v)}$  can be seen in the equation

$$\left. \frac{\partial E[y(b, V, W)|V=v]}{\partial v} \right|_{b=b(v)} = \int y_2(b(v), v, w) g(w|v) + y(b(v), v, w) \frac{\partial g(w|v)}{\partial v} dw.$$

It is clear that continuity of  $y_2(b(v), v, w)$  and  $\frac{\partial g(w|v)}{\partial v}$  in  $v$  at  $v = 0$  will satisfy the condition for identification.

That is, it is sufficient that the “direct” marginal impact of  $v$  on  $Y$  is continuous in  $v$ , and that the conditional

<sup>4</sup>Note that i) is required for the marginal effect to be well-defined at the point  $b(0)$ , ii) and iii) are required in order to evaluate the integrand  $\frac{\partial (y(b(v), v, w) g(w|v))}{\partial v}$ .

density of the unobservables  $W$  change smoothly with respect to  $V$ .

While the assumption that the derivative of the density  $g(w|v)$  is continuous in  $v$  may seem like a mild restriction, it is important to emphasize that in many potential applications of the RKD idea, the assignment variable  $V$  is an endogenous behavioral outcome. If agents can strategically select a value of  $V$ , then we might be concerned that the density of the unobservables may exhibit non-smooth behavior near the kink point.

Arguably the most important development in recent applications of the regression discontinuity design is the recognition that when the assignment variable is endogenously chosen, inferences from an RD design may be rendered invalid (see e.g., the discussion in Lee and Lemieux (forthcoming), and the theoretical treatments in McCrary (2008) and Lee (2008)). Recent RD analyses (for example, Urquiola and Verhoogen (2009)) therefore devote much attention to the possibility of endogenous sorting on the assignment variable. Nevertheless, Lee (2008) shows that when agents have only imprecise control over the assignment variable, a regression discontinuity design can still deliver valid inferences. Our goal is to illuminate the same set of issues for the regression kink design. Thus, our primary contribution is to characterize a broad class of models for which the RK design will isolate quasi-experimental variation in the treatment of interest, and – like a randomized experiment or an RD design – allow for empirical tests of the model’s validity. As an illustration of the general principles involved, we sketch two different economic models of behavioral responses to unemployment benefits: one that satisfies the conditions for a valid RKD, and one that does not. Finally, we apply these ideas to evaluate the validity of a RKD analysis of the effect of unemployment benefits on the duration of unemployment among job losers in the state of Washington. Given the results of this analysis we go on to use an RKD to estimate the impact of marginal increases in benefit generosity on employment outcomes and UI claim behavior.

### **2.3 Defining Causal Effects in an Ideal Randomized Setting**

Since our goal is to characterize a class of models for which the RKD isolates quasi-experimental or “as good as randomized” variation in a regressor of interest, we first briefly review the statistical implications of running an idealized experiment in which the regressor is randomly assigned. This benchmark will serve as a basis for comparison in the sections to follow when we investigate the RK estimator in the context of non-random treatment assignment.

To envision a randomized scenario that corresponds to a regression kink design setting, a *continuous*

treatment variable is called for. This is different from a classical randomized experiment in which agents are assigned to either a treatment group or a control group.

Let  $(Y, X, B, V)$  be observable random variables, where  $Y$  denotes the outcome of interest,  $B$  is the treatment assignment variable, and  $X$  (a vector) and  $V$  (a scalar) are variables that are determined prior to  $B$ , in that order. The random variable  $V$  will play the role of the assignment variable in the RK Design described below. We denote the density of  $B$  conditional on  $V, W$  as  $f_{B|V,W}(\cdot|\cdot, \cdot)$  and the marginal density of  $B$  as  $f_B(\cdot)$ .

**Definition.** Let  $W$  characterize the “type” of the individual, with c.d.f.  $G(w)$ . All individuals with the same value of  $W$  are identical. Conditional on  $W$ , however, the distribution of  $V$  may be non-degenerate. Let  $Y \equiv y(B, V, W)$ ,  $y_1(b, v, w) \equiv \frac{\partial y(b, v, w)}{\partial b}$ , and  $X \equiv x(W)$ , where  $y(\cdot, \cdot, \cdot)$  and  $x(\cdot)$  are real-valued functions.

As in Lee (2008), there is no loss of generality in assuming that  $W$  is one-dimensional. To give a concrete example,  $W$  could represent potential earnings capacity,  $X$  could be the level of schooling at the time of applying for UI benefits, and  $V$  could represent pre-job-loss earnings. Non-degeneracy of  $V$  could arise from pure “randomness” in the determination of pre-job-loss earnings.

Note that there is no loss in generality in excluding  $X$  from the structural function  $y(\cdot, \cdot, \cdot)$ . The variables  $X$  could be included as a separate argument, but we are not interested in the marginal effects of  $X$ . Thus, we consider the “reduced-form” function  $y(\cdot, \cdot, \cdot)$ , for which the impact of  $W$  is defined to include any indirect effects through  $X$ . Note that so far, our setup corresponds to a standard unrestricted non-separable model, except that we have made clear that  $X$  is determined before  $V$ , which is determined before  $B$ . This ordering implies that  $B$  cannot enter the function  $x(\cdot)$ .

**Condition 2a**  $y_1(b, v, w)$  exists for all  $(b, v, w) \in \mathbb{R}^3$ , is integrable with respect to  $dG(w)$  for all  $b \in \mathbb{R}$ , and is continuous in  $b$  for all  $w$ .

**Condition 2b** (Randomized Treatment Assignment)  $f_{B|V,W}(b|v, w) = f_B(b)$  for all  $(b, v, w) \in \mathbb{R}^3$ .

This condition describes a simple assignment mechanism that is the continuous analogue of a classical randomized experiment. In particular, each individual faces the same probability of receiving any level of the treatment variable.

In the context of an unemployment benefit example, we have in mind the following experiment. For each individual, regardless of pre-job-loss earnings, we randomize the weekly benefit amount they can receive and then observe the outcome  $Y$  (for example,  $Y$  could measure the duration of insured unemployment). A natural function of interest in this setting is  $E[y_1(B, V, W) | B = b, V = v]$ , which represents the average response of  $Y$  to a marginal increase in benefits, for a particular pair  $(b, v)$ . This function could be averaged over the distribution of  $V$  (conditional on  $B$ ) to obtain  $E[y_1(B, V, W) | B = b]$ , which is the average response of  $Y$  to a marginal increase in  $B$  at a particular value of  $b$ .

The key implications of this assignment mechanism are summarized in the following proposition.

**Proposition 2.** *If Condition 2a and Condition 2b hold, then*

(a)  $Pr(W \leq w | B = b) = Pr(W \leq w) \forall w \in R$  and  $\forall b$  in the support of  $B$ .

(b)  $\frac{\partial E[Y | B=b, V=v]}{\partial b} = E[y_1(b, V, W) | V = v] \equiv ATE_{b|v} = TT_{b|v} = \int y_1(b, v, w) \frac{f_{V|W}(v|w)}{f_V(v)} dG(w) \forall b$  in the support of  $B$ .

(c)  $Pr(X \leq x_0 | B = b) = Pr(X \leq x_0) \forall x_0 \in R$  and  $\forall b$  in the support of  $B$ .

The proof of Proposition 2 is in the Appendix. Part (a) states the intuitive consequence of randomized experiment: the distribution of “all other pre-determined factors” is the same, regardless of the level of the treatment  $B$ . Part (b) establishes that the partial derivative of  $E[Y | B = b, V = v]$  with respect to  $b$  identifies the  $ATE$  parameter at  $b$  (at  $V = v$ ).  $ATE_{b|v}$  is also equivalent to the  $TT$  at  $b$ , since in this randomized experiment, the distribution of  $W$  conditional on  $B$  is the same as the unconditional distribution of  $W$ , as stated in (a). Part (b) also shows that the average derivative is actually a weighted average across all individuals, where the weights  $\frac{f_{V|W}(v|w)}{f_V(v)}$  reflect the relative likelihood that a particular type of individual (identified by a value of  $w$ ) will have  $V = v$ . It is also possible to average  $ATE_{b|v}$  over the marginal distribution of  $V$  to obtain the unconditional  $ATE$  of  $E[y_1(b, V, W)]$ .

Part (c) of the proposition is the most distinctive implication of randomized variation in  $B$ . If each agent has the same probability law for  $B$ , then the distribution of the pre-determined covariates,  $X$ , will be identical for any value of  $B$ . Thus, the empirical validity of Conditions 2a and 2b can be tested using baseline covariates. This is analogous to the conventional “test for randomization” that is often employed in randomized controlled trials, whereby the analyst tests that the distributions of the baseline covariates among the treatment and control groups are statistically similar.

## 2.4 Local Random Assignment From a Regression Kink Design

### 2.4.1 Identification

We now consider the case where the regressor of interest,  $B$ , is mechanically determined as a function of the assignment variable  $V$ . We show that if the assignment rule has a kink at  $V = 0$  then: (a) RKD can identify the same  $TT_{b|v}$  parameter that is identified by the above randomized experiment; and (b) the validity of the RKD can be tested by examining the properties of the conditional distribution of the predetermined covariates  $\Pr(X \leq x_0|V = v)$  at  $v = 0$ .

The random variables are determined in the same sequence as in the experiment: first the pre-determined variables  $X$ , then  $V$ , and then  $B$ . This time, however,  $B$  is determined according to the deterministic rule  $B = b(V)$ . Formally,

**Definition.** Let  $(V, W)$  be a pair of random variables (with  $W$  unobservable,  $V$  observable), where the distribution of  $W$  is given by the c.d.f  $G(w)$ , and the distribution and density of  $V$  conditional on  $W$  are given by the c.d.f.  $F_{V|W}(v|w)$  and p.d.f.  $f_{V|W}(v|w)$ . Let  $B \equiv b(V)$ , let  $Y \equiv y(B, V, W)$ , and  $X \equiv x(W)$ . Also, define  $y_1(b, v, w) \equiv \frac{\partial y(b, v, w)}{\partial b}$  and  $y_2(b, v, w) \equiv \frac{\partial y(b, v, w)}{\partial v}$ .

**Condition 3a.** (Regularity)  $y(\cdot, \cdot, \cdot)$  and  $x(\cdot)$  are real-valued functions.  $y_1(b, v, w)$  is continuous in  $b$  and  $y_2(b, v, w)$  is continuous in  $v$  for all  $b, v$  and  $w$ .

Relative to the experimental setting, this condition requires that the direct marginal impact of  $V$  on  $Y$  is smooth.

**Condition 3b.** (First Stage)  $b(\cdot)$  is a known function, everywhere continuous and is continuously differentiable on  $(-\infty, 0)$  and  $(0, \infty)$ , but  $\lim_{v \rightarrow 0^+} b'(v) \neq \lim_{v \rightarrow 0^-} b'(v)$ . In addition,  $f_{V|W}(0|w)$  is strictly positive for  $w \in A$ , where  $\int_A dG(w) > 0$ .

**Condition 3c,** (Smooth Density)  $F_{V|W}(v|w)$  is twice continuously differentiable in  $v$  at  $v = 0$  for every  $w$ . That is,  $\frac{\partial f_{V|W}(v|w)}{\partial v}$ , the derivative of the conditional probability density function  $f_{V|W}(v|w)$  is continuous in  $v$  for all  $w$ .

This is the key assumption that is required for a valid RK Design. As in Lee (2008), this condition rules out precise manipulation of the assignment variable. But whereas in an RD context continuity of  $f_{V|W}(v|w)$  in  $v$  is sufficient for valid inferences, in the RK design we need to have a continuous derivative of  $f_{V|W}(v|w)$  with respect to  $v$ . As we show in the next Section, there are reasonable economic models that predict that the smooth density condition will hold, even though agents have some control over the assignment variable  $V$ , and other models where it will not hold. This underscores the need to be able to empirically test the implications of this assumption.

**Proposition 3.** *If Conditions 3a, 3b and 3c hold, then:*

(a)  $Pr(W \leq w|V = v)$  is continuously differentiable in  $v$  at  $v = 0 \forall w$ .

(b)  $\frac{\lim_{v \rightarrow 0^+} \frac{\partial E[Y|V=v]}{\partial v} - \lim_{v \rightarrow 0^-} \frac{\partial E[Y|V=v]}{\partial v}}{\lim_{v \rightarrow 0^+} \frac{\partial b(v)}{\partial v} - \lim_{v \rightarrow 0^-} \frac{\partial b(v)}{\partial v}} = E[y_1(b_0, 0, W)|V = 0] = \int y_1(b_0, 0, w) \frac{f_{V|W}(0|w)}{f_V(0)} dG(w) = TT_{b_0|0}$  where  $b_0 = b(0)$ .

(c)  $Pr(X \leq x_0|V = v)$  is continuously differentiable in  $v$  at  $v = 0 \forall x_0$ .

Proposition 3 is analogous to Proposition 2, and its proof is in the 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  $V = 0$ .<sup>5</sup> This leads directly to part (b): as a consequence of the smoothness in the underlying distribution of types around the kink, the change in the slope of  $E[Y|V = v]$  at  $v = 0$  divided by the change in slope in  $b(V)$  at the kink point delivers identification of  $TT_{b_0|0}$ . This is the same parameter identified by the randomized experiment above, except that it is evaluated at  $B = b_0$  and  $V = 0$ .<sup>6</sup> The weights in the weighted average interpretation  $TT_{b_0|0}$  are also the same as for the experimentally identified  $TT_{b|v}$ . Note that in contrast to the case of the randomized experiment,  $TT_{b|v}$  is in general different from  $ATE_{b|v}$ , due to the potentially systematic relation between  $B$  and  $W$ . It should also be clear that Conditions 3a, 3b, and 3c will satisfy the necessary condition of Proposition 1.

Finally, and most importantly, Proposition 3c, which is analogous to Proposition 2c, states that under the required conditions for a valid RKD, any pre-determined variable  $X$  should have a c.d.f. that is *continuously*

<sup>5</sup>Note also that (a) implies Proposition 2(a) in Lee (2008), i.e., the continuity of  $Pr(W \leq w|V = v)$  at  $v = 0$  for all  $w$ . This is a consequence of the stronger smoothness assumption we have imposed on the conditional distribution of  $V$  on  $W$ .

<sup>6</sup>Technically, the  $TT$  and  $LAR$  parameters do not condition on  $V$ . But in the case where there is a one-to-one relationship between  $B$  and  $V$ , then the trivial integration over the (degenerate) distribution of  $V$  conditional on  $B = b_0$  will imply that  $TT_{b_0|0} = TT_{b_0} \equiv E[y_1(b_0, V, W)|B = b_0]$ , which is literally the  $TT$  and  $LAR$  parameters discussed in Florens et al. (2009) and Altonji and Matzkin (2005), respectively. 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  $b_0$ ; for  $B < b_0$ ,  $TT_b = TT_{b|v}$ , but for  $B = b_0$ ,  $TT_{b_0} = \int TT_{b_0|v} f_{V|B}(v|b_0) dv$ . In this case, the RKD estimand identifies  $\lim_{b \uparrow b_0} TT_b$ .

*differentiable* (i.e., no kink) with respect to  $V$ . It is important to emphasize that this prediction is stronger than the requirement for a valid RD that the distribution of  $X$  is continuous with respect to  $V$ . In particular, showing that there is a similar distribution of baseline covariates in a neighborhood just to the left and just to the right of  $V = 0$  is *not enough* to verify the prediction of Proposition (3c). Instead, what is needed is evidence on the smoothness of the conditional distribution, based on comparisons of the slope of the conditional expectation function (or the conditional quantile function) of  $X$  given  $v$ .

## 2.4.2 Discussion of the Smooth Density Condition—Illustration With a Simple Labor Supply Model

We now use a simple behavioral model of labor supply responses to variation in the unemployment benefit rate to illustrate the substantive content of the smooth density condition (Condition 3c) that is required for valid RK design. We begin with an example where Condition 3c is violated. In this example, sorting is “too extreme” to satisfy the smooth density assumption. Consider a group of workers who are initially employed in a temporary (or seasonal) job. At the beginning of period 1, workers know that the job will end after one period, and that during the second (and final) period they will receive unemployment insurance benefits that depend on their first-period earnings. In period 1 they earn an exogenous hourly wage  $w$  and consume their entire wage income. In period 2, they receive an unemployment benefit  $b$  that is based on their first period income ( $I$ ):  $b(I) = \min(\gamma I, \bar{b})$ , where  $\gamma \in (0, 1)$  represents the replacement rate and  $\bar{b}$  represents a maximum benefit rate (both of which are constant). For simplicity we assume that there is no possibility of finding another job in period 2, and that unemployment benefits are the only source of second period income.

Workers can choose how many hours  $h \in [0, 1]$  to work in period 1. Workers differ by the relative weight that they assign to consumption versus leisure in their within-period utility functions. We denote the utility function for a worker of type  $\alpha$  as  $u_\alpha(\text{consumption}, \text{leisure})$ . We assume that workers have complete knowledge of the mapping  $b(\cdot)$  before starting to work in period 1. A worker of type  $\alpha$  solves the following problem:

$$\max_h u_\alpha(wh, 1 - h) + \beta u_\alpha(b(wh), 1)$$

where  $\beta > 0$  is a discount factor.

Consider a Cobb-Douglas utility function:  $u_\alpha(\text{consumption}, \text{leisure}) \equiv \log(\text{consumption}) + \alpha \log(\text{leisure})$ .

This parametrization leads to the optimal choices:

$$h^* = \begin{cases} \frac{1+\beta}{1+\alpha+\beta} & \text{if } \frac{1+\beta}{1+\alpha+\beta} < k \\ \frac{1}{1+\alpha} & \text{if } \frac{1}{1+\alpha} \geq k \\ k & \text{otherwise} \end{cases}$$

where  $k = \frac{\bar{b}}{\gamma w}$  is the location of the kink when we plot  $b$  (benefits) against  $h$  (labor supplied in period 1). The relationship between  $h^*$  and  $\alpha$  is plotted in Figure 1 for the case where  $w = 1000$ ,  $\gamma = 0.5$ ,  $\bar{b} = 350$  and  $\beta = 0.5$ .

As shown in the graph, individuals with  $\alpha \in [0.43, 0.64]$  will all choose  $h^* = k = 0.7$ . Assuming that  $\alpha$  is smoothly distributed on  $(0,1)$ , this means that a discrete mass of workers will sort to the kink. In general, then, the density of optimal hours,  $f_{h^*}(h)$ , will have a discontinuous derivative at  $h = k$ , reflecting the fact that the derivative of the inverse mapping from  $\alpha$  to  $h^*$  is different from the left and right at  $h^* = k$ .<sup>7</sup> Consequently, the derivative of the density of baseline earnings  $I$  (the assignment variable) conditional on  $\alpha$  (the latent unobservable) will not be the same on the two sides of the kink, violating Condition 3c.

Next we consider a variant of the model in which Condition 3c is satisfied. Suppose that in period 1 workers do not know the precise location of the kink-point  $k$ , but instead have a prior represented by the density  $f_k(k)$  for  $0 < k < 1$ , with associated c.d.f.  $F_K(k)$ . In this case, uncertainty over the location of the kink translates into uncertainty over  $\bar{b}$ , and the worker only knows that his or her benefit will be  $b = \min(\gamma wh, \gamma wk)$  for each potential value of  $k$ . Given these prior beliefs the worker maximizes expected utility:

$$\log(wh) + \alpha \log(1-h) + \beta \int (\log(\gamma wh)1_{[h < k]} + \log(\gamma wk)1_{[h \geq k]}) f_K(k) dk,$$

which can be simplified to:

$$\log(wh) + \alpha \log(1-h) + \beta \left\{ \log(\gamma wh)(1 - F_K(h)) + \int_{-\infty}^h \log(\gamma wk) f(k) dk \right\}.$$

It can be shown that the first order condition for an optimal hours choice in period 1 has the form:

$$1 - h^* - \alpha h^* + \beta(1 - h^*)(1 - F_K(h^*)) = 0.$$

---

<sup>7</sup>It is  $(1 + \beta)\frac{1-c}{c}$  from the left and  $\frac{1-c}{c}$  from the right.

The derivative of the optimal hours choice as a function of the worker's type is:

$$\frac{dh^*}{d\alpha} = -\frac{h^*}{(1 + \alpha) + \beta(1 - F_K(h^*)) + \beta(1 - h^*)f_K(h^*)}.$$

The denominator of this expression is strictly positive, and  $\frac{dh^*}{d\alpha}$  is continuous in  $\alpha$  at every  $\alpha$  and  $h^*$  that satisfy the first order condition. Assuming that  $\alpha$  is smoothly distributed,  $h^*$  will then be smoothly distributed, as will baseline earnings  $I = wh^*$ . Thus, under the assumption that agents have a smooth prior on the location of the kink, Condition 3c will hold. The lack of information among workers rules out the extreme form of sorting in the first example and ensures that there is a smooth mapping from the underlying heterogeneity to the assignment variable.

### 3 Unemployment Insurance in Washington State: Background and Data

#### 3.1 The Washington Reemployment Bonus Experiment

In this section we use a regression kink approach to estimate the marginal effect of unemployment benefits on claimant behavior for a sample of individuals in the Washington Reemployment Bonus Experiment (WREB). The WREB was a randomized experiment conducted during 1988 to study the responses of unemployment insurance (UI) claimants to alternative incentive schemes. Specifically, the experiment provided different lump sum bonus amounts to claimants who found a job within alternative time-limits. A total of 15,534 claimants participated in the experiment, of whom 12,452 were assigned to one of the six treatment groups (3 different bonus amounts  $\times$  2 different time-limits) and 3,082 were assigned to the control group. All participants were subject to the same (standard) provisions of the Washington UI system for determining their benefit amounts. To the extent that the bonus provisions may have interacted with the effects of the benefit level, our analysis – which ignores the bonus aspect of the experiment – provides estimates of the marginal impact of higher UI benefit levels averaged across the 7 experimental regimes (6 treatment regimes and 1 control regime).

## 3.2 UI Institutions in Washington State

### 3.2.1 Benefit Determination Rules

In Washington, as in other US states, UI entitlement is based on a formula that depends on labor market activities in the period before the start of the claim. The Washington formula uses earnings in a “base year”, defined as either (i) the first four of the last five completed calendar quarters, or (ii) the last four completed calendar quarters, immediately preceding the start of the claim. Provided that an individual had enough earnings (and, uniquely to the Washington system, worked a minimum of 680 hours) in the base year, he or she is eligible to draw benefits over a year-long period (the so-called “benefit year”). The individual’s benefit amount is  $1/50$  of total wages earned in the two highest-earning quarters of the base period. Thus, for an individual with constant weekly earnings of  $\$I$  over the base year, the benefit amount is  $26/50 \times I \approx 0.52I$ . The benefit amount is subject to a maximum (which was  $\$205/\text{week}$  in the first 6 months of the experimental period, and  $\$209/\text{week}$  in the second half) as well as a minimum (which we ignore by eliminating the small fraction of claimants affected by this provision). Claimants also face a ceiling on the total amount of benefits claimed, which cannot exceed the lesser of  $30 \times$  their weekly benefit amount, or one-third of their base-year earnings. This ceiling determines the maximum number of weeks of UI they can draw.

Formally, the benefit rules can be summarized as follows. Let  $b$  denote “weekly benefit amount”,  $Totalbenefits$  denote “total benefits payable” and  $Maxduration$  denote “weeks of unemployment benefit entitlement”. Let  $Q_1, Q_2, Q_3$  and  $Q_4$  be the quarterly earnings in the four quarters of the base period, ranked in order of earnings, so  $Q_1$  and  $Q_2$  are the two highest among the four. The rules that determine  $b$ ,  $Totalbenefits$  and  $Maxduration$  in the first half of 1988 (the definition is analogous for the last two quarters of 1988) are summarized as follows:

$$b = \begin{cases} \frac{Q_1+Q_2}{50} & \text{if } \frac{Q_1+Q_2}{50} < 205 \\ 205 & \text{if } \frac{Q_1+Q_2}{50} \geq 205 \end{cases}$$

$$Totalbenefits = \min\left(30 \cdot b, \frac{Q_1 + Q_2 + Q_3 + Q_4}{3}\right)$$

$$Maxduration = \frac{Totalbenefits}{b}$$

Define  $r \equiv \frac{Q_3+Q_4}{Q_1+Q_2} \in [0, 1]$  and  $V \equiv Q_1 + Q_2$ . Then with some simplification we can re-write the rules as:

$$b = \begin{cases} \frac{V}{50} & \text{if } V < 10,250 \\ 205 & \text{if } V \geq 10,250 \end{cases}$$

$$Totalbenefits = \begin{cases} \frac{3}{5}V & \text{if } V < 10,250 \text{ and } r \geq \frac{4}{5} \\ \frac{1}{3}V(1+r) & \text{if } V < 10,250 \text{ and } r < \frac{4}{5} \\ 6,150 & \text{if } V \geq 10,250 \text{ and } r \geq \frac{4}{5} \\ \frac{1}{3}V(1+r) & \text{if } V \geq 10,250 \text{ and } r < \frac{4}{5} \end{cases}$$

$$Maxduration = \begin{cases} 30 & \text{if } V < 10,250 \text{ and } r \geq \frac{4}{5} \\ \frac{1}{3}(1+r) \cdot 50 & \text{if } V < 10,250 \text{ and } r < \frac{4}{5} \\ 30 & \text{if } V \geq 10,250 \text{ and } r \geq \frac{4}{5} \\ \frac{1}{3}(1+r) \cdot \frac{V}{205} & \text{if } V \geq 10,250 \text{ and } r < \frac{4}{5} \end{cases}$$

Notice that for claimants with  $r \geq \frac{4}{5}$ , *Totalbenefits* “top out” at the same point as *b*; as a result *Maxduration* is exactly 30 weeks, regardless of *V*. For claimants with  $r < \frac{4}{5}$ , however, *Totalbenefits* =  $\frac{1}{3}V(1+r)$  for all values of *V*. Since  $Maxduration = \frac{Totalbenefits}{b}$ , and *b* flattens out once *V* reaches 10,250, there is an “upward” kink in the relation between *Maxduration* and *V* at  $V = 10,250$  for people with  $r < \frac{4}{5}$ .

Naturally, having two endogenous regressors that are kinked at the same point will make it impossible to distinguish between the independent effects of *b* and *Maxduration*. But since the kink in *Maxduration* is entirely driven by the kink in *b*, we can still obtain estimates of the “reduced-form” effect of *b*. This reduced form effect arises through two channels: the direct effect of *b* on the outcome of interest, and an indirect effect through *Maxduration*. That is, the identified marginal effect is from manipulating *b*, while holding constant the other aspects of the rules determining *Totalbenefits* and *Maxduration*.

### 3.3 Data Issues

#### 3.3.1 Data Source and Variables

Our data are derived from the public use file of the WREB, which combines information from several administrative data sources. Most of the variables, including wage earnings, UI benefits, and demographic characteristics, come from the Benefit Automated System and the WAGE database provided by the Washington State Employment Security Department (WSESD). Also included in the file are variables on local labor market conditions provided by the Labor Market and Economic Analysis Branch of the WSESD. The assignment variables that are crucial to the evaluation of the re-employment bonus experiment come from the Participant Tracking System of the WREB, and will not be used in our analysis. Details on the construction of the data file can be found in Spiegelman et al. (1992).

The variables relevant for our study are:

- Earnings: quarterly earnings from the first quarter of 1985 to the last quarter of 1989.
- Unemployment Insurance: date of UI claim;  $b$ ; *Totalbenefits*; *Maxduration*; Net UI payment for every week in the benefit year.
- Baseline covariates: age, gender, race, education, one-digit SIC code of base year employer, Job Service Center where the claim was filed.

Table 1 reports summary statistics for these variables for the analysis sample that we describe below.

#### 3.3.2 Adjusting Base Period Earnings for Changes in the Maximum Weekly Benefit

The maximum weekly benefit amount increased from \$205 to \$209 beginning July 1, 1988. To facilitate a pooled analysis using data for the entire year, we adjust  $V$  (the sum of earnings in the two highest quarters) for claimants who filed after July 1, 1988 by subtracting \$200 (= \$4 per week  $\times$  50 weeks) from the sum of their two highest-quarter earnings. With this adjustment, the kink in the benefit rule is at the same point ( $V = 10,250$ ) for all claimants in the sample. The formulaic relationship between  $b$  and the adjusted value of  $V$  is shown in Figure 2. (The Figure also shows the minimum weekly benefit rate, which we ignore in our empirical analysis by eliminating very low-earning claimants). Note that although the relationship has the same kink point for claimants in the two halves of the year, the benefit rates (as a function of  $V$ ) are slightly higher for claimants from the second half. Estimated treatment effects from the pooled sample therefore

represent an average of the marginal effects for these two levels of benefits. If the marginal effects are the same, then the pooled data will yield more efficient estimates.

### 3.3.3 Measurement Error in Base Period Earnings

In principle, a plot of the actual weekly benefit amounts received by claimants against their normalized base period earnings should replicate Figure 2. In practice, the empirical relationship (depicted in Figure 3) shows deviations from the UI benefit rules. Out of 15,534 claimants in our overall sample, some 8% (1,249 cases) have benefit amounts that appear to deviate from the formula (this group also includes a small number of claimants with missing data for three or four quarters in the baseline). Figure 4a plots the histogram of the differences between the actual and predicted values of the weekly benefit amount  $b$ . Since 92% of observations have a deviation of precisely 0, the figure is not very informative: Figure 4b shows the histogram after excluding the 0's, and suggests that the deviations are slightly left-skewed (i.e. actual benefits tend to be a little lower than predicted, on average). Further investigation revealed that the likely source of the discrepancy between actual and predicted benefits arises because the benefit system data files incorporate an unedited version of quarterly earnings in the base period, whereas the benefit formula uses a verified measure of earnings.<sup>8</sup>

Under the presumption that the benefit rate  $b$  was correctly computed from actual earnings, it is possible to correct the measure of  $V$  by inverting the benefit formula. After this correction procedure the only remaining deviations arise from the 90 observations whose actual weekly benefit is at the maximum level (\$205 for the claims filed in the first half of 1988, \$209 for those filed in the second half of 1988) but whose reported value of  $V$  is below the kink point (\$10,250). For simplicity we drop these cases from our main analysis sample. The relationship between  $b$  and  $V$  in the analysis sample is shown in Figure 3b and (by construction) follows the predicted pattern of Figure 2 exactly.<sup>9</sup> A minor complication arises because there is an unusual mass of claimants with filing dates in July 1988 who have values of  $b = \$205$ , even though the maximum benefit rate had increased to \$209 effective July 1. These claims were most likely processed according to the rules for the first half of the year. Accordingly, we assume that claimants from July 1988 whose actual benefit amount is \$205/week, but whose predicted benefit exceeds \$205, were

---

<sup>8</sup>According to Ken Kline at the Upjohn Institute, if an applicant's earnings do not match the amount in the system database, then the employer is contacted for verification in order to calculate UI benefit. But the system database, which we use, is not subsequently updated with the correct information.

<sup>9</sup>We have performed empirical analyses on both the "raw" data and on the corrected sample, and the results do not differ substantially. We report results from the corrected sample below.

processed according to the rules before June 30, 1988.

## 4 Empirical Results from the RKD Analysis

### 4.1 Outcomes of Interest

We focus on three main outcomes associated with the effect of changes in weekly UI benefits: the total amount of UI payments received over the benefit year; the total number of weeks of UI claimed; and the duration of the initial unemployment (claim) spell. Our definition of the initial spell is borrowed directly from Spiegelman et al. (1992): this spell starts with the so-called waiting week (a week during which no payments are received) and ends when there is a gap of at least two weeks in the receipt of benefits.

The effect of an increase in the weekly benefit amount  $b$  on total UI system costs (per claimant) is of great policy interest in itself. It is important to emphasize, however, that a finding of a significant effect for this outcome does not necessarily imply that higher benefits induce a behavioral response among UI claimants. Even in the absence of any behavioral response, an increase in weekly benefits paid to some group will lead to an increase in total UI benefits that is proportional to the average number of weeks of UI claimed. Thus, a kink in the relationship between total UI payments and base earnings at the point where weekly benefits are capped will in part reflect the purely mechanical relationship between the benefit amount and the cost of payments.

It is tempting to interpret the effect of the weekly benefit amount on the number of weeks of UI claimed, or on the length of the initial spell, as a purely behavioral response. In the case of Washington's UI system, however, a straightforward RKD analysis will not isolate a purely behavioral effect on either of these outcomes. The reason is that for claimants with  $r = \frac{Q_3+Q_4}{Q_1+Q_2} < \frac{4}{5}$  there is a kink in the relationship between *Maxduration* (the maximum number of weeks of UI available) and base period earnings  $V$  (see Section 3.2.1). For this subgroup (who represent approximately 60% of claimants in the WREB), a small increase in  $V$  when  $V$  is to the left of the kink-point in the benefit formula leads to a higher benefit rate  $b$ , but no increase in *Maxduration*. In contrast, a small increase in  $V$  to the right of the kink-point leads to both an increase in  $b$  and an increase in *Maxduration*. Thus, a comparison of the slopes of the relationship between  $V$  and the duration of UI claims on either side of the kink-point combines a behavioral response and a mechanical entitlement effect. To isolate the behavioral component, we consider artificially censoring the data on the right so that the relationship between the censored maximum duration and  $V$  is smooth. Specifically,

consider a maximum potential benefit duration measure  $Maxduration_{smooth}$  that is constructed by applying the formula for  $Maxduration$  that prevails on the left side of the kink to all claimants:

$$Maxduration_{smooth} = \begin{cases} 30 & r \geq \frac{4}{5} \\ \frac{1}{3}(1+r) \cdot 50 & r < \frac{4}{5} \end{cases}.$$

If we conduct an analysis using the number of weeks of benefits claimed (or the duration of the initial UI claim) censored at  $Maxduration_{smooth}$ , we potentially eliminate the mechanical entitlement effect and isolate the behavioral impact of the change in UI benefits.<sup>10</sup> To evaluate the validity of this simple approach we plotted mean weeks of actual UI entitlement against the value of  $V$  for a set of discrete bins (30 different bins of \$500 each, equally distributed on the two sides of the kink point). We then compared this to a plot of censored entitlements, censoring actual entitlements at  $Maxduration_{smooth}$ . Whereas uncensored mean entitlements show a pronounced kink at  $V = 10,250$ , the relationship between censored entitlements and base period earnings is smooth (see Figure 5), suggesting that the censoring approach will work. We therefore focus on five “outcomes of interest” in our empirical analysis: total UI payment received, weeks of UI claimed, the duration of the initial UI spell, and censoring-adjusted versions of the latter two outcomes.

## 4.2 Graphical Presentation

An attractive feature of a Regression Kink Design is that the results from the analysis can be summarized graphically, in a fashion similar to the way that results from a Regression Discontinuity Design are typically presented. In particular, one can plot the means of the outcomes of interest, as well as the means for the predetermined covariates, against the assignment variable  $V$  (base period earnings) and look for potential kinks around the kink-point in the formula that maps the assignment variable to the regressor of interest. For such a presentation we need to divide the range of  $V$  into suitable “bins.” Given the modest sample sizes available, we use \$500 bins.<sup>11</sup> We also limit attention to observations with  $V \in [2750, 17750]$ , resulting in a

---

<sup>10</sup>In practice we use  $Maxduration_{smooth} = \begin{cases} 30 & r > 0.74 \\ \text{ceiling}[\frac{1}{3}(1+r) \cdot 50] & r \leq 0.74 \end{cases}$  to incorporate the effect of the way that rounding

is implemented in the Washington State UI system. For example, when  $Totalbenefits/b = 29.1$ , a claimant is entitled to 30 weeks of benefits: he or she can receive full weekly benefits for the first 29 weeks of claim and one tenth of the weekly benefit amount in the 30th week. There are also claimants for whom weeks of UI received are greater than  $Maxduration$ . This happens when a claimant receives partial benefits during a week when he or she is working part-time while on claim. For simplicity we do not cap the weeks of UI claimed (or initial spell length) at  $Maxduration_{smooth}$  for these claimants.

<sup>11</sup>Lee and Lemieux (forthcoming) present a formal procedure for choosing bin size, based on goodness-of-fit tests which evaluate the fit of simple models with bin dummies. A bin size of 500 is the largest that passes the two tests suggested by Lee and Lemieux

graphical analysis with 15 bins on each side of the kink-point ( $V = 10250$ ).<sup>12</sup>

Proposition 3 establishes that there are two key testable implications of a valid RK design. First, the density of the assignment variable  $V$  has to be continuously differentiable at the kink-point. Second, the conditional expectations (and conditional quantile functions) of any baseline covariates have to be continuously differentiable at the threshold. As in a RDD, these testable conditions can be visually examined. We proceed by plotting the number of observations in each bin, and the mean values of the covariates for the claimants in each bin, against base period earnings.

Figure 6 presents a plot of the “density” of  $V$  (i.e., the histogram across the 30 bins).<sup>13</sup> The histogram is somewhat bumpy, with a drop between 15<sup>th</sup> bin (to the left of the kink) and the 16<sup>th</sup> bin (to the right), although the drops at other points (e.g., between the 9<sup>th</sup> and 10<sup>th</sup> bins, or between the 17<sup>th</sup> and 18<sup>th</sup> bins) are similar in magnitude.

Next we examine plots of the conditional means of age, education, gender, race, region, and industry for different values of  $V$ . All of these covariates were presumably determined *before* the claimant’s base period earnings. The results are shown in Figures 7(a) through 6(e). As a simple indicator of region we use the fraction of claimants who filed for UI at a Job Service Center in Western Washington.<sup>14</sup>

Inspection of the pattern of the “dots” in Figures 7(a)-7(e) leads us to conclude that the conditional means of the covariates evolve smoothly across the kink-point in the benefit determination schedule. The evolution of mean education (Figure 7(b)) shows some evidence of a kink in the neighborhood of  $V = 7250$ , but around the critical kink-point ( $V = 10250$ ) it appears to evolve relatively smoothly. On balance there is no strong visual evidence of discontinuities in  $\partial E[X|V = v]/\partial v$  at the kink-point. We evaluate the smoothness of the distributions of the covariates more formally in the next section.

The relative smoothness in the conditional means of the covariates around the kink point becomes even more apparent when compared with the patterns for the outcome variables in Figures 9(a)-9(d). For all five outcomes there is a clearly discernible change in the slope of the relationship with  $V$  at the kink-point in the benefit formula. In each case the outcome variable is increasing in  $V$  to the left of the threshold and decreasing in  $V$  to the right of the threshold. Notice that this is true for total UI benefits (Figure 8(a)) which incorporates both a mechanical and behavioral effect of weekly benefits, as well as in the censored versions

---

(forthcoming) for all the dependent variables in our analysis.

<sup>12</sup>Outside of the range [2750, 17750] the sample sizes per bin are very small.

<sup>13</sup>Since our interest is in evaluating the smoothness of the density function we do not display the more conventional “smoothed” histogram.

<sup>14</sup>At the time of the WREB the state had 21 Job Service Centers, 14 of which are located in Western Washington.

of weeks of UI claimed (Figure 8(d)) and the duration of the initial claim (Figure 8(e)), which incorporate only a behavioral component. In the following section, we present the numerical estimates of the change in slopes based on simple parametric specifications.

### 4.3 Estimation Results

#### 4.3.1 Empirical Specification

In empirically implementing the RKD estimator, we follow two complementary approaches. For our first approach we follow Lee and Lemieux (forthcoming) and estimate parametric polynomial models of the form:

$$E[Y|V = v] = \alpha_0 + \sum_{p=1}^{\bar{p}} [\alpha_p(v-k)^p + \beta_p(v-k)^p \cdot D] \text{ where } |v-k| \leq h \quad (2)$$

where  $k = \$10250$  (the kink point),  $D = 1_{[V \geq k]}$ , an indicator for the event that base period earnings exceeds the kink-point, the  $\alpha$ 's and the  $\beta$ 's are polynomial coefficients,  $\bar{p}$  is the maximum polynomial order, and  $h$  is the bandwidth that determines the window  $[k-h, k+h]$  within which the sample is selected.<sup>15</sup> In this approach the change in the derivative of the conditional expectation function – the numerator of the RK estimand – is given by the coefficient  $\beta_1$ . Since the slope of the benefit function  $b(V)$  changes from  $\frac{1}{50}$  to 0 at the kink point, the denominator of the RK estimand is  $-\frac{1}{50}$  and so we multiply  $\hat{\beta}_1$  by  $-50$  to obtain the  $TT_b$  effect of  $b$  on  $Y$ .

We present a sensitivity analysis, choosing several levels of  $h$  ranging from  $h = 1000$  to  $h = 7500$ . For each bandwidth choice we vary  $\bar{p}$  from 1 to 3 and report the order of the polynomial preferred by the Aikaike Information Criterion (AIC). As suggested in Lee and Lemieux (forthcoming), we also run an additional “unrestricted” model in which we include a set of dummy variables which indicate consecutive intervals (of width 500) in  $V$ , and compute a goodness-of-fit test that compare the polynomial model to the dummy variable specification. Within this framework we can also easily probe the sensitivity of the results to inclusion of the baseline covariates, by adding these as additional regressors.

<sup>15</sup>Note that this specification imposes continuity in the conditional expectation function at the kink-point. We have also estimated all models allowing for a potential discontinuity (by including  $D$  as a separate regressor). Estimates of the change in the slope at the kink-point are very similar, and the implied “jumps” in the conditional expectation function are never statistically significant.

### 4.3.2 RKD Estimates

Table 2 reports RKD estimates using the baseline specification (2) and Table 3 shows that the results are robust to the inclusion of baseline covariates. For each table, we include point estimates of  $\hat{\beta}_1$  and robust standard errors for each of the regressions. We also report the p-values from the Goodness-of-Fit tests including the bin dummies. Results for bandwidths of  $h = 7500$ ,  $h = 2500$  and  $h = 1000$  are reported. For each bandwidth, we report the coefficient from regressions up to a third order polynomial.

In general, within the bandwidths we consider, the AIC suggests a linear specification, with the exception being that a quadratic is chosen by the AIC for “total weeks claimed”. It is also true that the linear specification is not considered too restrictive relative to a model that includes bin dummies: none of the p-values are less than 0.05 in any of the specifications. The most precise estimates of the effect of a dollar increase in benefit on total UI received are around \$17 to \$18, while for the point estimates for “Total Weeks Claimed” are around 0.04 of a week. Note that for the bandwidth of 7500, the inclusion of quadratic terms causes the point estimate to fall and become statistically insignificant, although a 0.04 effect could not be ruled out at conventional levels of significance. By comparison, the effects are less sensitive to the inclusion of second order terms for the “Initial Spell Length” variable. As expected, the magnitude of the point estimates increase slightly when we artificially censor the data to isolate the purely behavioral impact of the benefit on the weeks claimed and initial spell variables.

As we might expect, the estimates become less precise as we shrink the bandwidth, and when it becomes 1000, the standard errors are much larger relative to the point estimates, even for “Total UI Received”, which is the most striking kink displayed in our figures. Although we report the point estimates and corresponding p-values for the specification tests for all of these permutations, we focus on the first, second, and fourth rows of Table 2 as our preferred specifications, and believe at a minimum that these models are the least likely to be over-fit. The pattern of results in Table 3 – where we include baseline covariates – are similar both qualitatively and quantitatively, with the estimated standard errors being slightly smaller, and the point estimates being sometimes higher and sometimes lower, but not by a significant amount, depending on the outcome and specification.

We benchmark our estimates against three studies from the UI literature. Hamermesh (1977) concludes that “the best estimate—if one chooses a single figure—is that a 10-percentage point increase in the gross replacement rate leads to an increase in the duration of insured unemployment of about half a week when

labor markets are tight." Moffitt and Nicholson (1982) find that that a 10-percentage point increase in the replacement rate was associated with about a one week increase in the average length of unemployment spells, while the estimate of Meyer (1990) is around one and a half weeks in response to a 10-percentage point increase in replacement rate.

In our setting, a \$1 increase in the weekly benefit amount for the population near  $V = 10,250$  corresponds to about a 0.25 percentage point increase in the UI replacement rate<sup>16</sup>. Since our estimates indicate roughly a 0.04 increase in insured spells, this implies that the response to a 10 percentage point increase in replacement rate would be an increase in insured unemployment duration by about  $0.04 * (10/0.25) = 1.6$  weeks, which is of a similar magnitude as the estimate in Meyer (1990).

### 4.3.3 Testing for Kinks in the Density of Baseline Earnings and Conditional Expectation of Covariates

To provide an estimate of a potential kink in the density of  $V$  at the threshold, we follow the approach of McCrary (2008) and first collapse the data into equal-sized bins of width 500. The collapsed data set contains 30 observations as we restrict the sample to  $V \in [2750, 17750]$ . The two key variables in the collapsed data set are: the number of original observations in each bin  $N_{bin}$  and the baseline earnings amount each bin is centered around,  $V_{bin}$ . We then regress  $N_{bin}$  on polynomials of  $(V_{bin} - k)$  and the interaction term  $1_{[V_{bin} \geq k]}(V_{bin} - k)$  where  $k = 10,250$  is the kink-point. Because of the small number of observations, we do not interact  $1_{[V_{bin} \geq k]}$  with higher order polynomial terms of  $(V_{bin} - k)$ . A fourth order polynomial does a good job fitting the data with an  $R^2 = 0.99$ . As suggested by Figure 7, the coefficient on the interaction term is statistically insignificant (a t-statistic of -0.78).

Table 4 is the analogy to Table 2, except that the dependent variables are the baseline covariates. If the RKD is valid, and the assumption of a smooth density of  $V$  is reasonable, then we expect not to see systematic evidence of kinks. If we again focus on the first, second, and fourth rows, as we did for Tables 2 and 3, we find that most of the point estimates are statistically insignificant at conventional levels. For the dummy variable indicating the Job Service Center was in western Washington, the point estimates are significant for the linear and quadratic specifications. On the other hand, the specification tests clearly reject those polynomial orders, and the AIC suggests a third order polynomial (third row), and in that row the point

<sup>16</sup>The weekly earnings for the claimants with  $V = 10,250$  are about  $\$10,250/26 = \$394$ . So a \$1 increase is equivalent to an increase in the replacement rate of  $\$1/\$394 = 0.25$  percentage point.

estimate is insignificant. In a similar way, the goodness-of-fit statistics for mean Education – for which the point estimates are statistically significant – are rejected at the .10 level. Finally, the point estimates for White are statistically significant in the first and second rows, but not in the third and fourth rows. On balance, if we evaluate the specifications in Tables 2 and 3 and Table 4 with a similar standard of needing to both pass the goodness-of-fit statistic at the .10 level and be chosen by the AIC, we see that almost all of the covariates exhibit no significant kink, while the effects on the outcomes are most striking for “Total UI Received” and “Initial Spell Length”.

## 5 Conclusion

This paper considers the identification of marginal effects in nonparametric models of endogenous regressors with nonseparable errors, using the Regression Kink Design. In this context, we establish the necessary condition for identification, which is, loosely speaking, that “all other factors” are evolving smoothly – in the sense of a continuous derivative – with respect to the assignment variable. The problem with such an assumption, is that it involves a statement about the distribution of unobservables, which can be difficult to justify and – given the unobservability of these factors – impossible to test.

Our main contribution is to characterize a class of models that are instead based on an assumption about the distribution of an *observable* variable,  $V$ . In particular, the assumption is that for each agent, the density of  $V$  is continuously differentiable at the kink-point. This assumption may follow naturally from models of the underlying behavior. In our context, we have outlined an illustrative model that would suggest the smooth density condition would be violated, and another model in which it would be satisfied. Most importantly, our characterization of a valid RK Design also generates the testable prediction that pre-determined covariates will have a distribution that is “smooth” with respect to  $V$  in the kink-point.

Applying these ideas to a study of UI claimant behavior in the state of Washington, we find evidence consistent with a valid RK Design, and estimates of the impact of a marginal increase in the benefit level on insured unemployment spells that are in the higher range of magnitudes found in the existing literature. In ongoing research, we are investigating the asymptotic properties of non-parametric estimators of the RK Design.

## References

- Altonji, J.G. and R.L. Matzkin**, “Cross section and panel data estimators for nonseparable models with endogenous regressors,” *Econometrica*, 2005, 73:4, 1053–1102.
- 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.
- Chesher, A.**, “Identification in Nonseparable Models,” *Econometrica*, 2003, 71(5), 1405–1441.
- 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*, 2009, 76, 1191–1206.
- Guryan, Jonathan**, “Does Money Matter? Regression-Discontinuity Estimates from Education Finance Reform in Massachusetts,” Working Paper WP8269, National Bureau of Economic Research 2001.
- , “Does Money Matter? Regression-Discontinuity Estimates from Education Finance Reform in Massachusetts,” Working Paper WP8269, National Bureau of Economic Research 2003.
- Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw**, “Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design,” *Econometrica*, January 2001, 69 (1), 201–209.
- Hamermesh, Daniel S.**, *Jobless Pay and the Economy*, Baltimore: The Johns Hopkins University Press, 1977.
- Heckman, J. J. and R. Robb**, “Alternative Methods for Evaluating the Impact of Interventions,” *Journal of Econometrics*, 1985, 30, 239–267.
- Imbens, Guido W. and Whitney K. Newey**, “Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity,” *Econometrica*, 2009, 77(5), 1481–1512.
- 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*, forthcoming.
- 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.
- Moffitt, Robert and Walter Nicholson**, “The Effect of Unemployment Insurance on Unemployment: The Case of Federal Supplemental Benefits,” *The Review of Economics and Statistics*, 1982, 64:1, 1–11.
- Nielsen, Helena S., Torben Sorensen, and Christopher R. Taber**, “Estimating the Effect of Student Aid on College Enrollment: Evidence from a Government Grant Policy Reform,” NBER Working Paper WP14535, National Bureau of Economic Research 2009.

- Nielsen, Helena Skyt, Torben Sørensen, and Christopher R. Taber**, “Estimating the Effect of Student Aid of College Enrollment: Evidence from a Government Grant Policy Reform,” Technical Report, NBER WP 14535 2008.
- , —, **and Christopher Taber**, “Estimating the Effect of Student Aid on College Enrollment: Evidence from a Government Grant Policy Reform,” *American Economic Journal: Economic Policy*, forthcoming.
- Saez, Emmanuel**, “Do Taxpayers Bunch at Kink Points?,” *American Economic Journal: Economic Policy*, forthcoming.
- Simonsen, Marianne, Lars Skipper, and Niels Skipper**, “Price Sensitivity of Demand of Prescription Drugs: Exploiting a Kinked Reimbursement Scheme,” Technical Report November 2009.
- Spiegelman, Robert G., Christopher J. O’Leary, and Kenneth J. Kline**, “The Washington Reemployment Bonus Experiment Final Report,” Technical Report, W.E. Upjohn Institute for Employment Research 1992.
- Thistlethwaite, Donald L. and Donald T. Campbell**, “Regression-Discontinuity Analysis: An Alternative to the Ex-Post Facto Experiment,” *Journal of Educational Psychology*, December 1960, 51, 309–317.
- Urquiola, Miguel and Eric Verhoogen**, “Class-Size Caps, Sorting, and the Regression-Discontinuity Design,” *American Economic Review*, 2009, 99(1), 179–215.

# Appendix

## Proofs

Proof of Proposition 2:

(a) From Condition 2a, we have that  $\frac{f_{B,W,V}(b,w,v)}{f_{W,V}(w,v)} = f_B(b) \forall b, w, v$ . Therefore,  $f_{B,W,V}(b, w, v) = f_B(b)f_{W,V}(w, v)$ .

Integrating both sides with respect to  $v$  gives us  $f_{B,W}(b, w) = f_B(b)f_W(w)$ , and consequently  $f_{B|W}(b, w) = f_B(b)$ . Thus,

$$\begin{aligned}\Pr(W \leq w | B = b) &= \int_{-\infty}^w f_{W|B}(w'|b)dw' \\ &= \int_{-\infty}^w \frac{f_{B|W}(b|w')}{f_B(b)} dG(w') \\ &= \int_{-\infty}^w dG(w') \\ &= \Pr(W \leq w).\end{aligned}$$

(b) Condition 2b allows the interchange of differentiation and integration in

$$\frac{\partial E[Y|B=b]}{\partial b} = \int \int \frac{\partial}{\partial b} [y(b, v, w) f_{V,W|B}(v, w|b)] dv dw.$$

Applying the Bayes Rule and invoking Condition 2a, we have

$$\begin{aligned}\int \int \frac{\partial}{\partial b} [y(b, v, w) f_{V,W|B}(v, w|b)] dv dw &= \int \int \frac{\partial}{\partial b} [y(b, v, w) \frac{f_{B|V,W}(b|v, w)}{f_B(b)}] dF_{V,W}(v, w) \\ &= \int \int \frac{\partial}{\partial b} [y(b, v, w)] dF_{V,W}(v, w) \\ &= E[y_1 | B = b]\end{aligned}$$

(c) The proof is analogous to (a).

Proof of Proposition 3:

(a)

$$\begin{aligned}\frac{\partial}{\partial v} \Pr(W \leq w|V = v) &= \frac{\partial}{\partial v} \int_{-\infty}^w \frac{f_{V|W}(v|w')}{f_V(v)} dG(w') \\ &= \int_{-\infty}^w \frac{\partial}{\partial v} \frac{f_{V|W}(v|w')}{f_V(v)} dG(w')\end{aligned}$$

Since we assume that  $f_{V|W}(v|w')$  is continuously differentiable in  $v$  at 0 for all  $w'$ ,  $\frac{\partial}{\partial v} \Pr(W \leq w|V = v)$  is continuous at 0.

(b) On the numerator,

$$\begin{aligned}\lim_{v \rightarrow 0^+} \frac{\partial E[Y|V = v]}{\partial v} &= \lim_{v \rightarrow 0^+} \frac{\partial}{\partial v} \int y(b(v), v, w) \frac{f_{v|w}(v|w)}{f(v)} dG(w) \\ &= \lim_{v \rightarrow 0^+} \int \frac{\partial}{\partial v} y(b(v), v, w) \frac{f_{v|w}(v|w)}{f(v)} dG(w) \\ &= \lim_{v \rightarrow 0^+} \int [y_1(b(v), v, w) \frac{\partial b(v)}{\partial v} + y_2(b(v), v, w)] \frac{f_{v|w}(v|w)}{f(v)} dG(w) + \\ &\quad \lim_{v \rightarrow 0^+} \int y(b(v), v, w) \frac{\partial}{\partial v} \frac{f_{v|w}(v|w)}{f(v)} dG(w) \\ &= \lim_{v \rightarrow 0^+} \frac{\partial b(v)}{\partial v} \int y_1(b(v), v, w) \frac{f_{v|w}(v|w)}{f(v)} dG(w) + \\ &\quad \lim_{v \rightarrow 0^+} \int y_2(b(v), v, w) \frac{f_{v|w}(v|w)}{f(v)} + y(b(v), v, w) \frac{\partial}{\partial v} \frac{f_{v|w}(v|w)}{f(v)} dG(w)\end{aligned}$$

Similarly,

$$\begin{aligned}\lim_{v \rightarrow 0^-} \frac{\partial E[Y|V = v]}{\partial v} &= \lim_{v \rightarrow 0^-} \frac{\partial b(v)}{\partial v} \int y_1(b(v), v, w) \frac{f_{v|w}(v|w)}{f(v)} dG(w) + \\ &= \lim_{v \rightarrow 0^-} \int y_2(b(v), v, w) \frac{f_{v|w}(v|w)}{f(v)} + y(b(v), v, w) \frac{\partial}{\partial v} \frac{f_{v|w}(v|w)}{f(v)} dG(w)\end{aligned}$$

By Conditions 3a and 3c, we have

$$\begin{aligned}&\lim_{v \rightarrow 0^+} \int y_2(b(v), v, w) \frac{f_{v|w}(v|w)}{f(v)} + y(b(v), v, w) \frac{\partial}{\partial v} \frac{f_{v|w}(v|w)}{f(v)} dG(w) \\ &= \lim_{v \rightarrow 0^-} \int y_2(b(v), v, w) \frac{f_{v|w}(v|w)}{f(v)} + y(b(v), v, w) \frac{\partial}{\partial v} \frac{f_{v|w}(v|w)}{f(v)} dG(w)\end{aligned}$$

and therefore,

$$\begin{aligned} & \lim_{v \rightarrow 0^+} \frac{\partial E[Y|V=v]}{\partial v} - \lim_{v \rightarrow 0^-} \frac{\partial E[Y|V=v]}{\partial v} \\ &= \lim_{v \rightarrow 0^+} \frac{\partial b(v)}{\partial v} \int y_1(b(v), v, w) \frac{f_{v|w}(v|w)}{f(v)} dG(w) - \lim_{v \rightarrow 0^-} \frac{\partial b(v)}{\partial v} \int y_1(b(v), v, w) \frac{f_{v|w}(v|w)}{f(v)} dG(w) \end{aligned}$$

Conditions 3a, 3b and 3c together guarantee that we can write

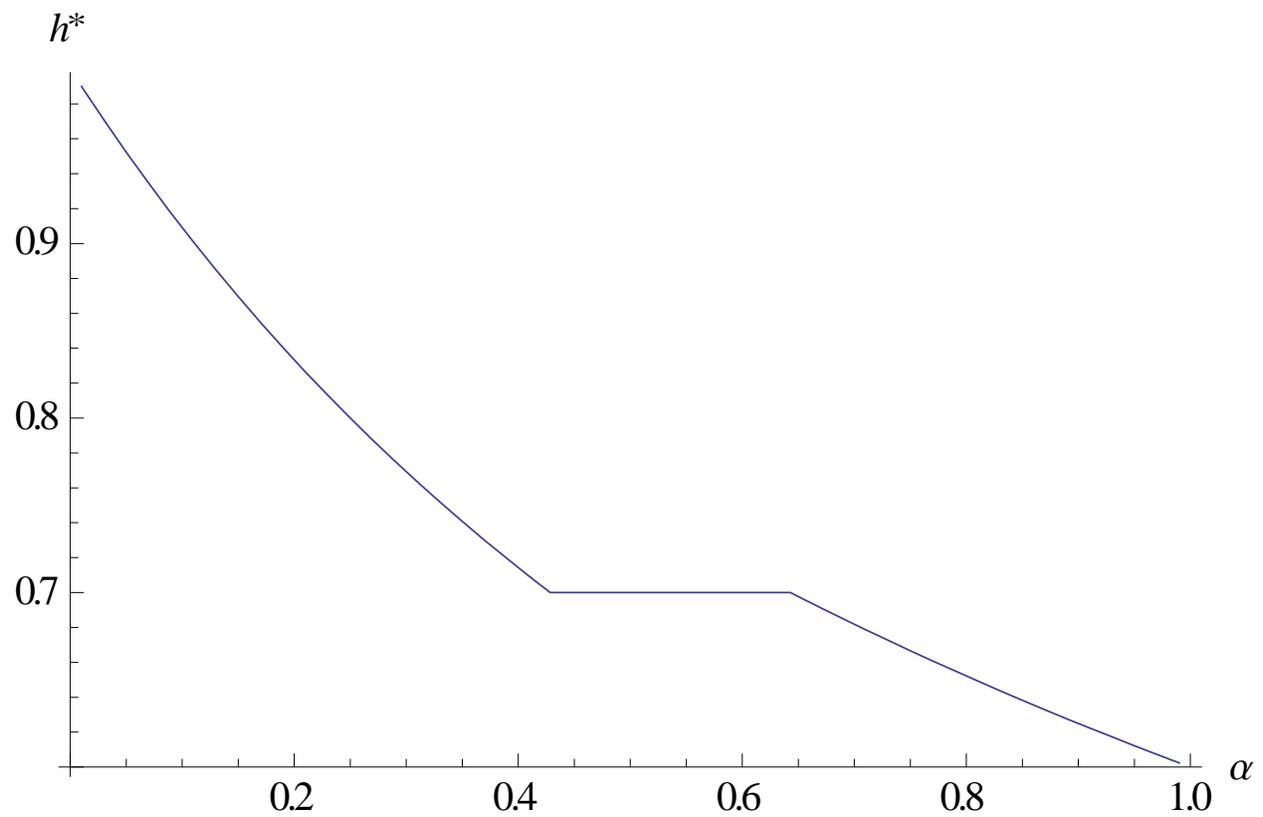
$$\begin{aligned} & \lim_{v \rightarrow 0^+} \frac{\partial b(v)}{\partial v} \int y_1(b(v), v, w) \frac{f_{v|w}(v|w)}{f(v)} dG(w) - \lim_{v \rightarrow 0^-} \frac{\partial b(v)}{\partial v} \int y_1(b(v), v, w) \frac{f_{v|w}(v|w)}{f(v)} dG(w) \\ &= \left( \lim_{v \rightarrow 0^+} \frac{\partial b(v)}{\partial v} - \lim_{v \rightarrow 0^-} \frac{\partial b(v)}{\partial v} \right) \int y_1(b(0), 0, w) \frac{f_{v|w}(0|w)}{f(0)} dG(w). \end{aligned}$$

Condition 3b guarantees that  $\lim_{v \rightarrow 0^+} \frac{\partial b(v)}{\partial v} - \lim_{v \rightarrow 0^-} \frac{\partial b(v)}{\partial v}$  is nonzero, and hence we have

$$\frac{\lim_{v \rightarrow 0^+} \frac{\partial E[Y|V=v]}{\partial v} - \lim_{v \rightarrow 0^-} \frac{\partial E[Y|V=v]}{\partial v}}{\lim_{v \rightarrow 0^+} \frac{\partial b(v)}{\partial v} - \lim_{v \rightarrow 0^-} \frac{\partial b(v)}{\partial v}} = E[y_1(b(0), 0, w) | V = 0].$$

(c) The proof is analogous to (a).

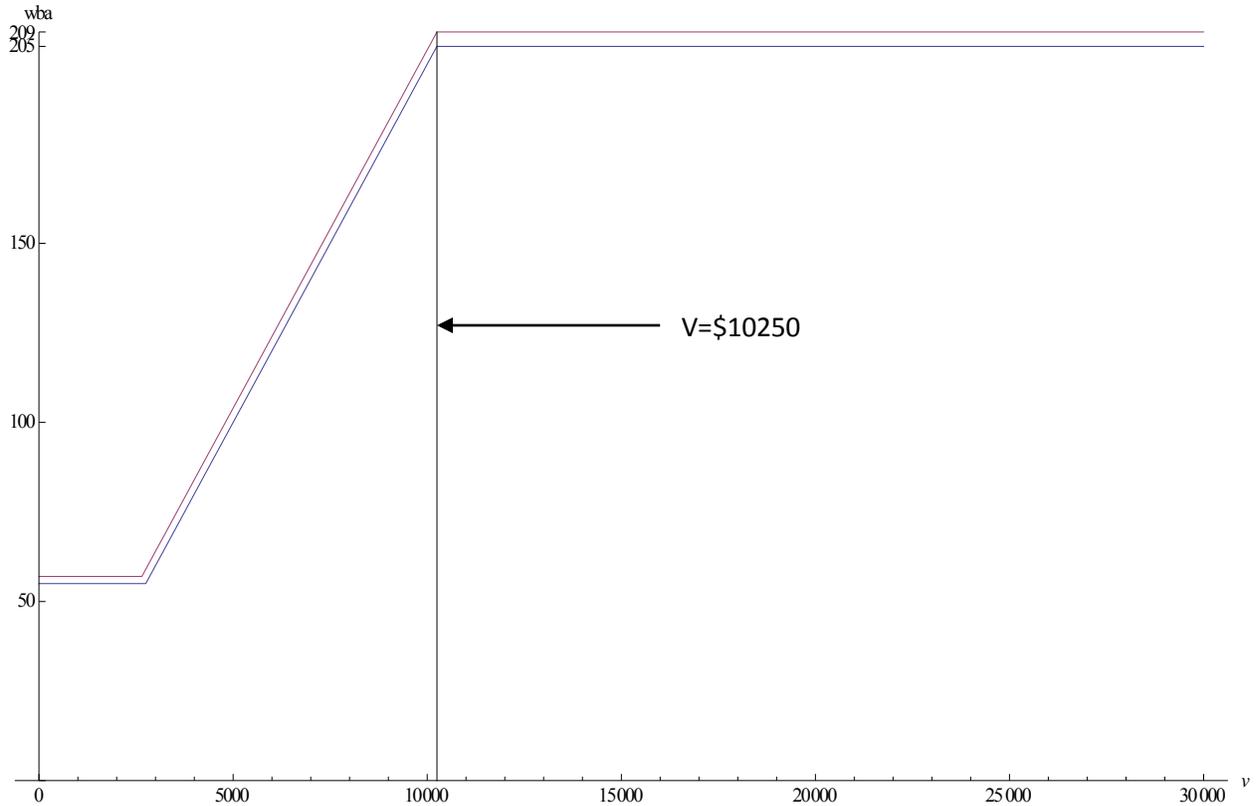
**Figure 1: The Relationship between  $h^*$  and  $\alpha$  in the Behavioral Model in Section 2.4.2.**



In the behavioral model,  $h^*$  is the labor supply in the first period and  $\alpha$  indicates the relative share of leisure in the Cobb-Douglas utility function. This figure indicates that for a range of  $\alpha$ , agents will “bunch” on  $h^*=0.7$ .

**Figure 2: Weekly Benefit Amount Determination Rule in Washington State in 1988**

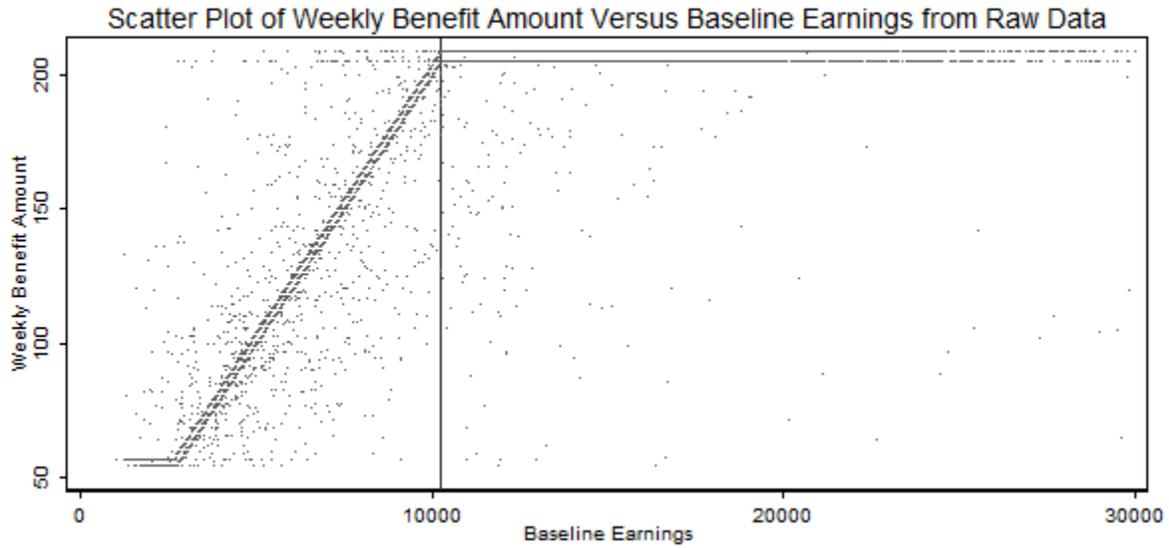
**WBA as a Function of Normalized Baseline Wage**



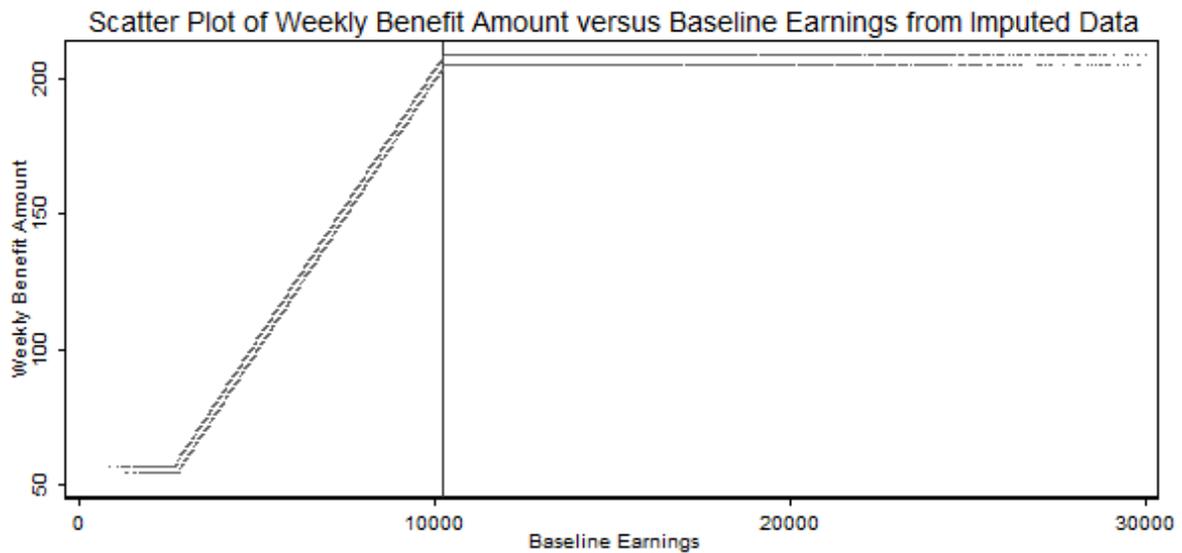
The rule that determines weekly benefit amount is different in the two halves of 1988. The top line corresponds to the rule after June 30, 1988, and the bottom line corresponds to the rule before June 30, 1988.

The baseline earnings,  $v$ , is defined as the sum of the two highest quarterly earnings in the base year. It is normalized here by subtracting \$200 from the baseline earnings of claimants who filed in the second half of 1988. According to the normalized rules, there is a kink in the relationship between WBA and  $V$  at \$10,250 for all UI claimants in 1988.

**Figure 3a: Scatter Plot of Weekly Benefit Amount versus Baseline Earnings from Raw Data**



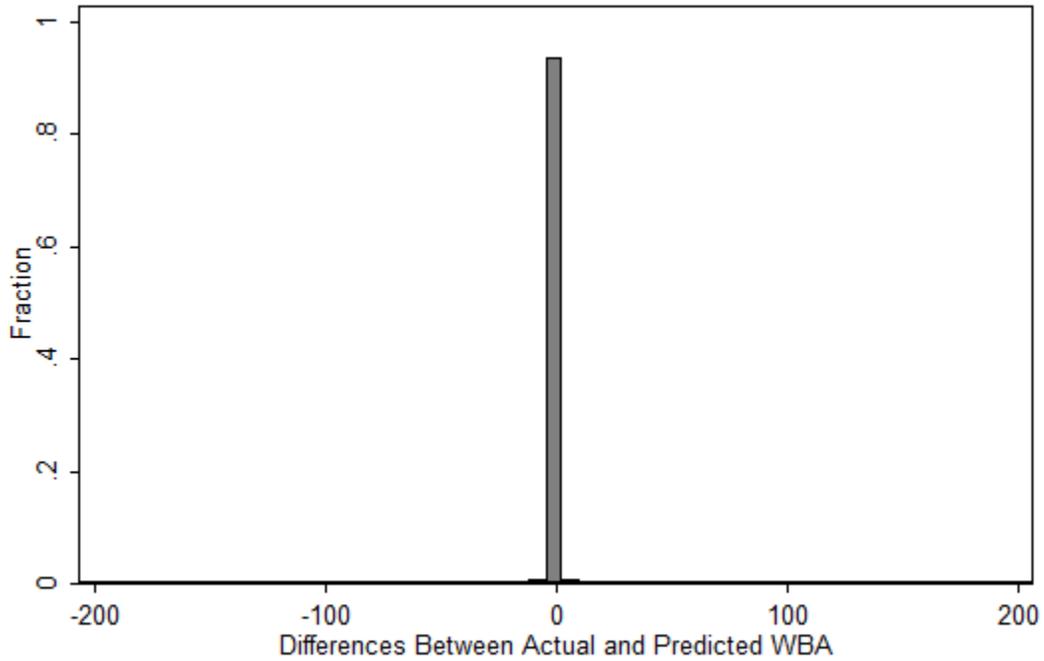
**Figure 3b: Scatter Plot of Weekly Benefit Amount versus Baseline Earnings from Imputed Data**



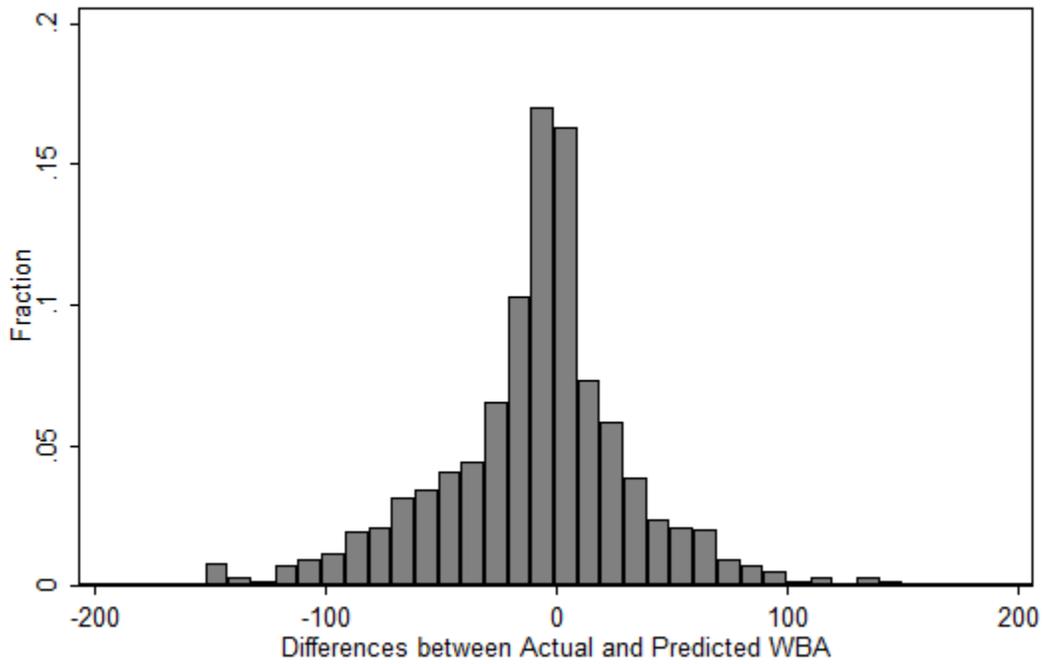
Baseline earnings are the sum of the two highest quarterly earnings in the base year. They are normalized here so that the higher kink takes place at \$10,250 (the vertical line) for all claimants who filed in 1988. The normalization procedure is described in section 3.3.2 and the caption of Figure 2.

The figures above are restricted to the observations whose baseline earnings are less than \$30,000 for ease of visual inspection.

**Figure 4a: Histogram of Differences between Actual and Predicted WBA in Raw Sample**

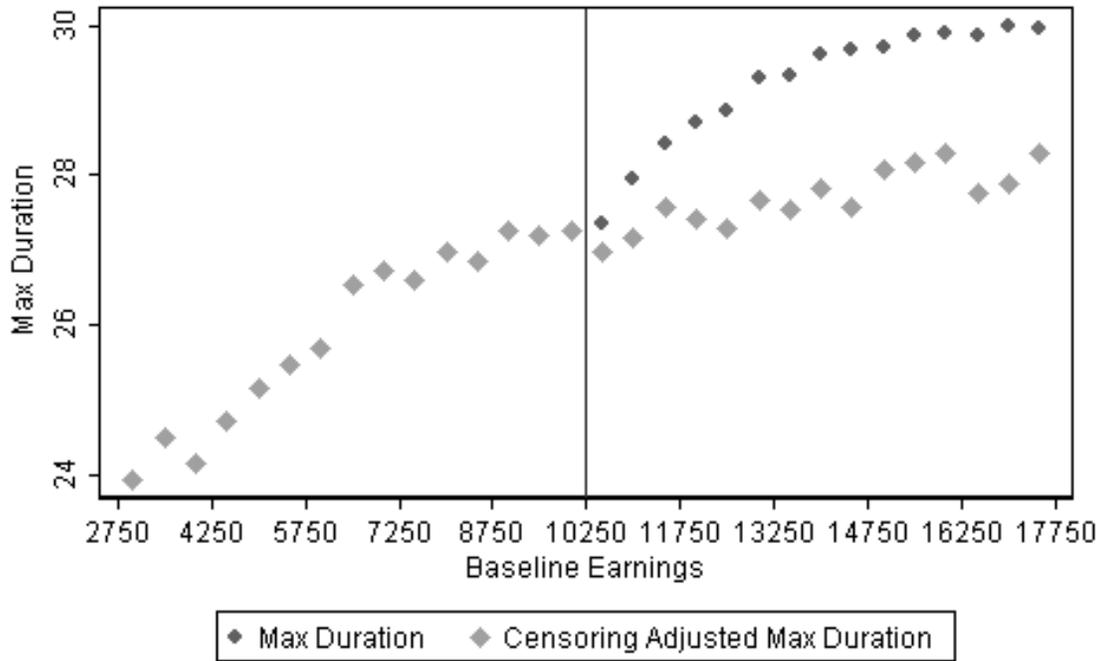


**Figure 4b: Histogram of Differences between Actual and Predicted WBA in Raw Sample Conditional on Differences Not Being Zero**



Note that the histogram of differences for the imputed sample would simply be a point mass at 0.

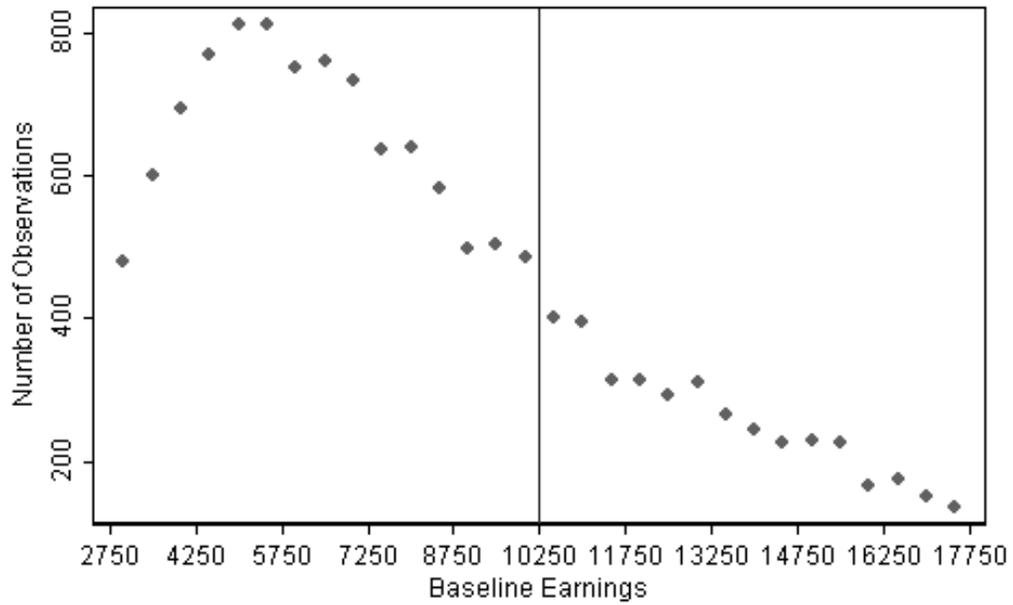
Figure 5: Maximum Duration of UI Payments vs Baseline Earnings



Each black point represents the local average of the actual maximum duration of UI payments over the benefit year in each baseline earnings bin. Each grey point represents the local average of the censoring corrected maximum duration of UI payments over the benefit year in each baseline earnings bin. See Section 4.1 for the construction of the censoring adjusted Max Duration measure.

The baseline earnings bins are of size 500 and are centered at multiples of 500. The vertical line corresponds to the kink point in weekly benefit amount determination: baseline earnings of \$10,250.

Figure 6: Number of Observations in Each Bin of Baseline Earnings



Each point represents the number of observations in each bin of baseline earnings. The bins are of size 500 and are centered at multiples of 500. The vertical line corresponds to the kink point in weekly benefit amount determination: baseline earnings of \$10,250.

Figure 7(a): Age vs Baseline Earnings

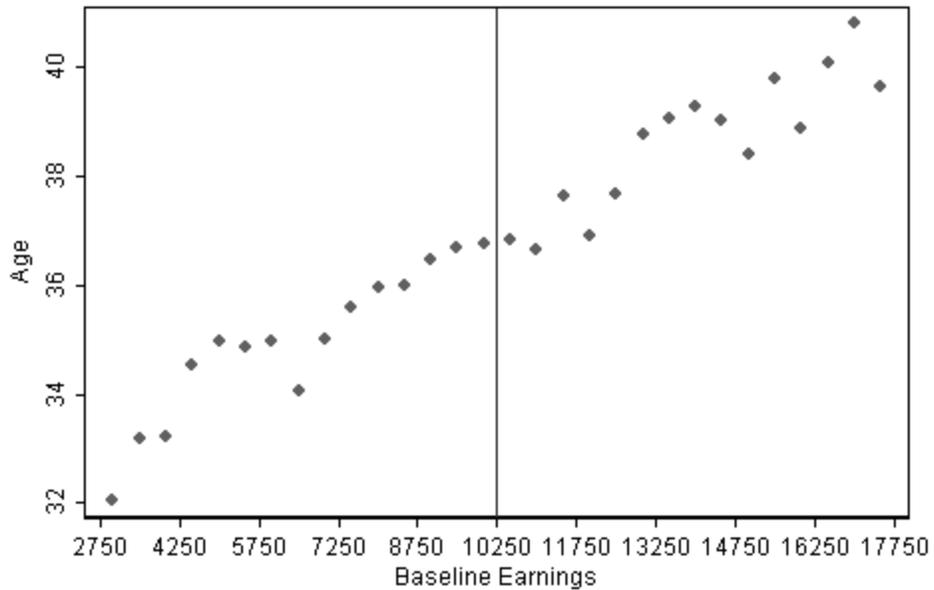
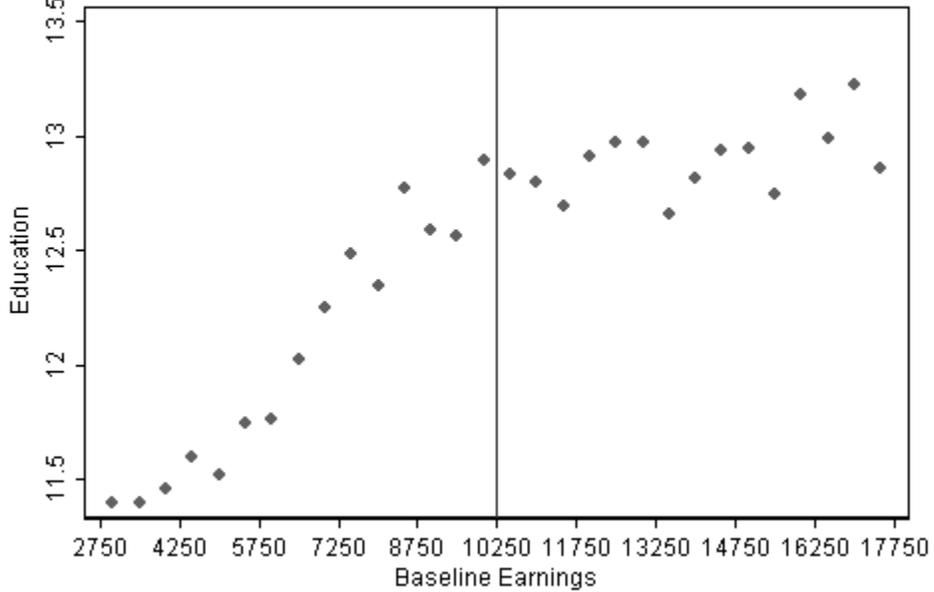


Figure 7(b): Education vs Baseline Earnings



Each point in 7(a) represents the local average of the age for claimants in each baseline earnings bin, and likewise in 7(b) the years of educational attainment. The baseline earnings bins are of size 500 and are centered at multiples of 500. The vertical line corresponds to the kink point in weekly benefit amount determination: baseline earnings of \$10,250.

Figure 7(c): Fraction of Males vs Baseline Earnings

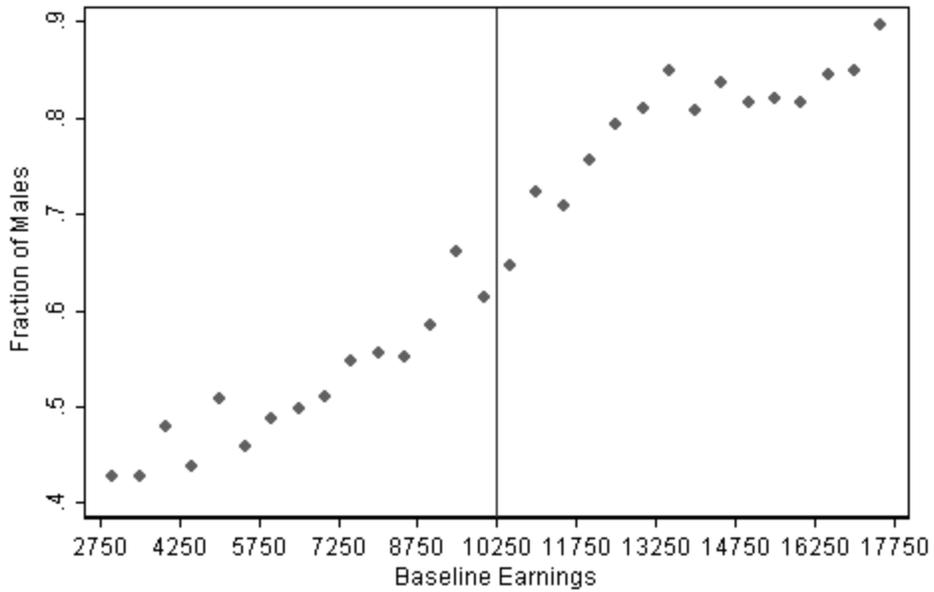
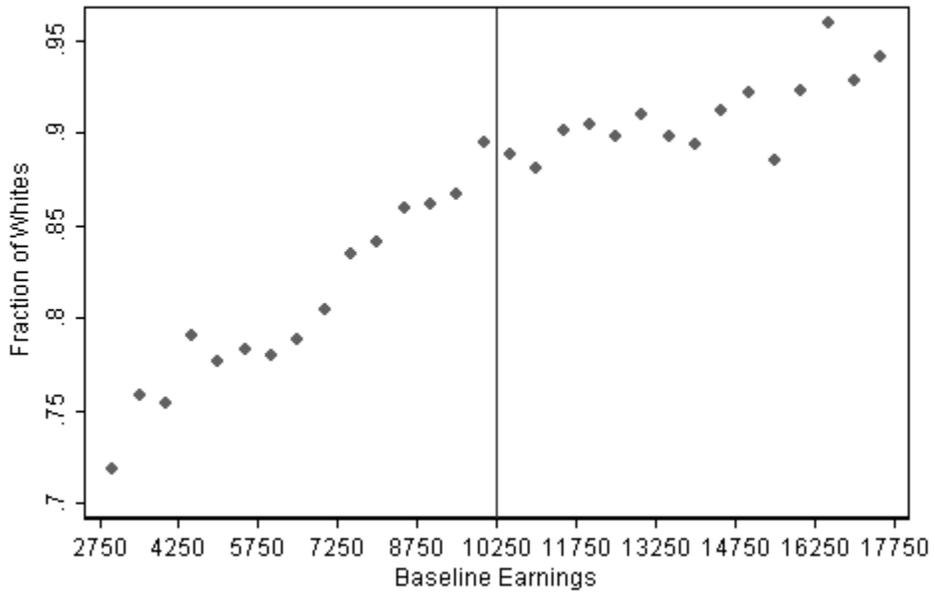
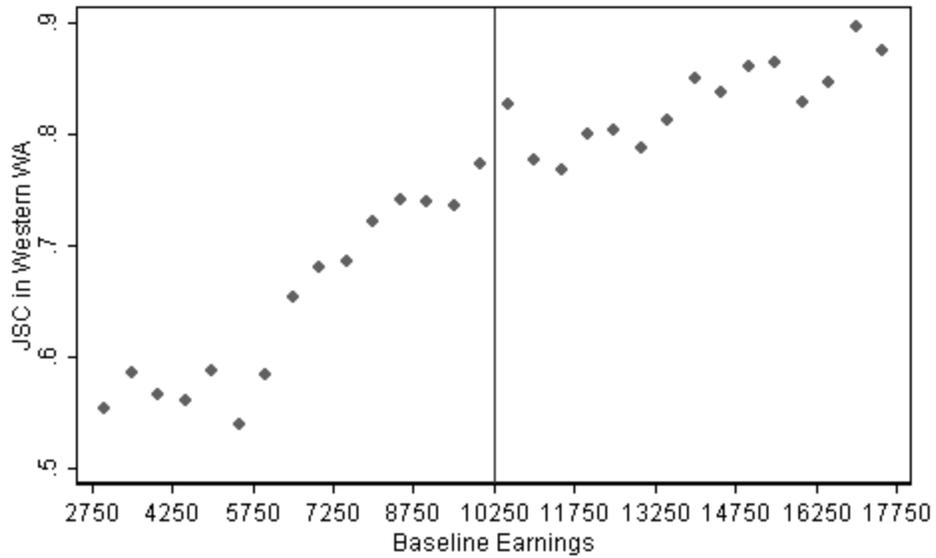


Figure 7(d): Fraction of Whites vs Baseline Earnings



Each point in 7(c) represents the fraction of males for claimants in each baseline earnings bin, and likewise in 7(d) the fraction of whites. The bins are of size 500 and are centered at multiples of 500. The vertical line corresponds to the kink point in weekly benefit amount determination: baseline earnings of \$10,250.

Figure 7(e): Fraction of Claimants Who Filed at Job Service Center in Western WA vs Baseline Earnings



Each point in 7(e) represents the local average of the fraction of claimants who filed at a job service center in Western Washington in each baseline earnings bin. The bins are of size 500 and are centered at multiples of 500. The vertical line corresponds to the kink point in weekly benefit amount determination: baseline earnings of \$10,250.

Figure 8(a): Total UI Benefit Payment vs Baseline Earnings

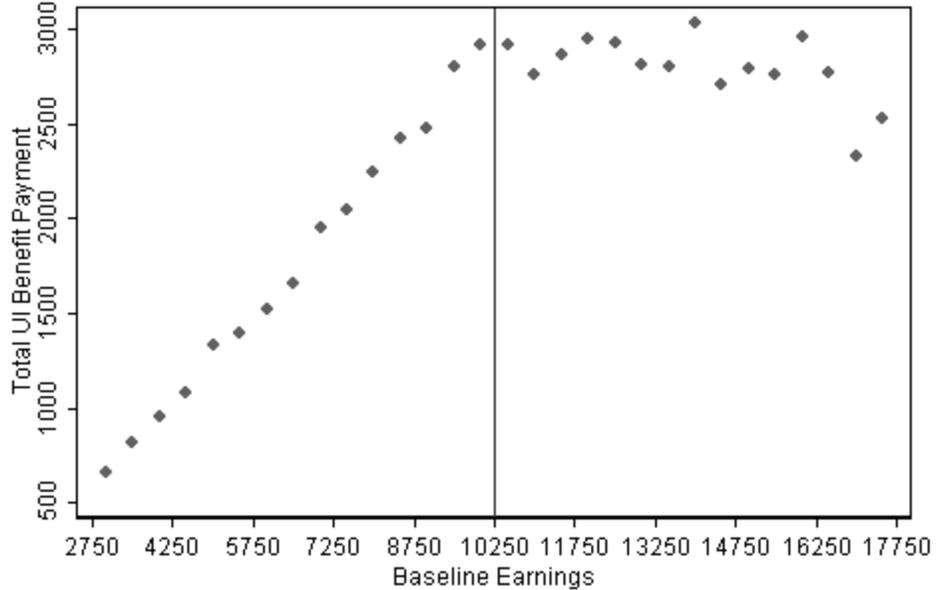
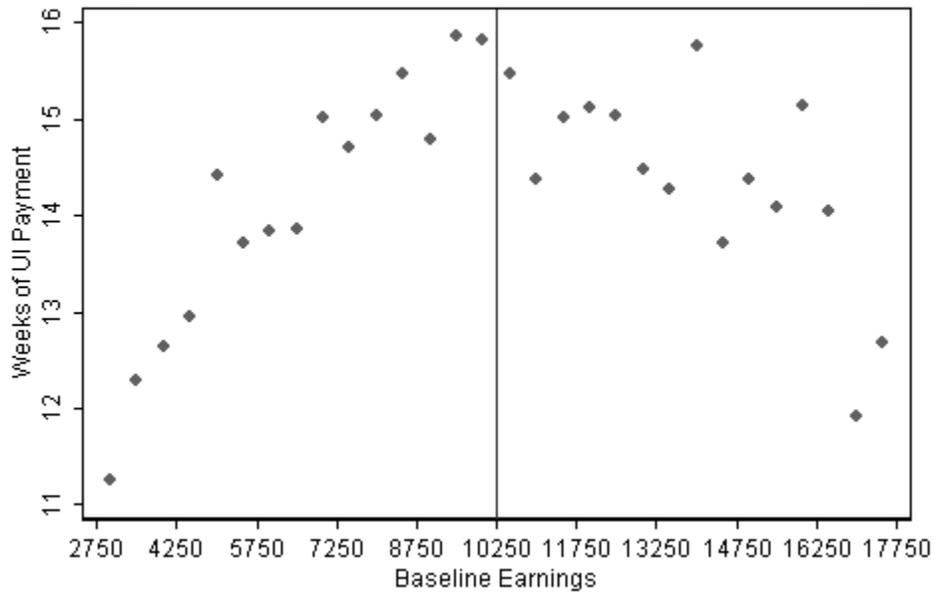


Figure 8(b): Weeks of UI Claimed vs Baseline Earnings



Each point in 8(a) represents the local average of the total UI benefit received over the benefit year for claimants in each baseline earnings bin, and likewise in 8(b) the weeks of UI payment claimed. The bins are of size 500 and are centered at multiples of 500. The vertical line corresponds to the kink point in weekly benefit amount determination: baseline earnings of \$10,250.

Figure 8(c): Length of Initial Spell vs Baseline Earnings

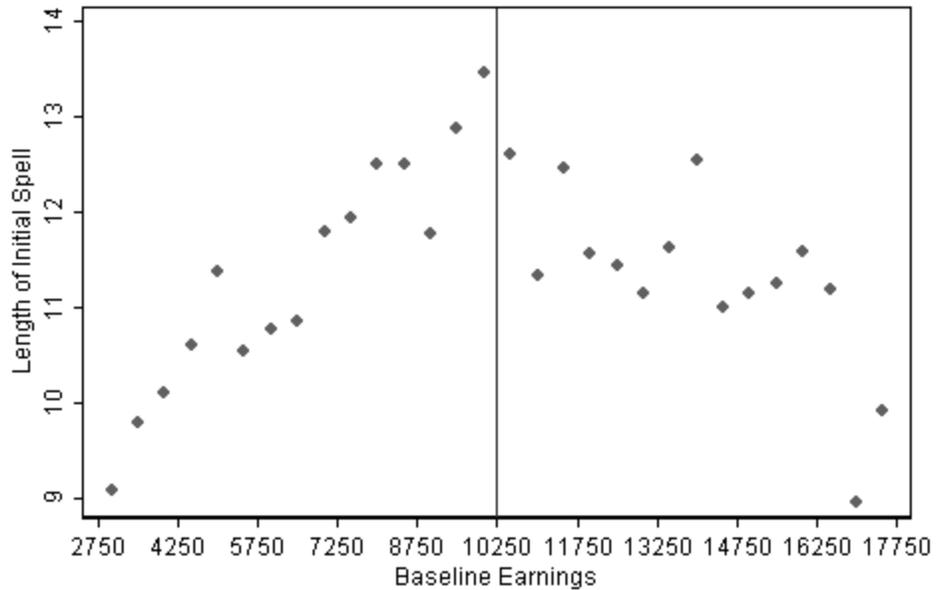
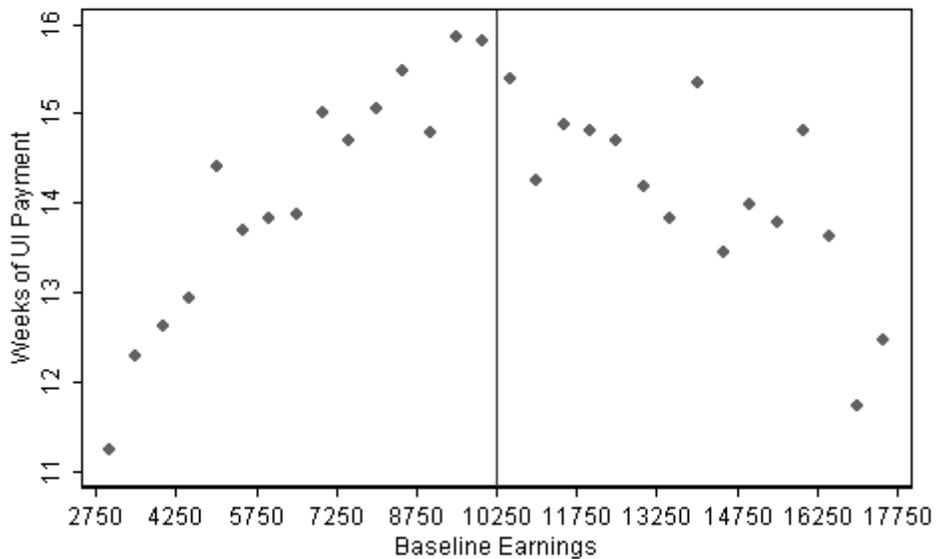
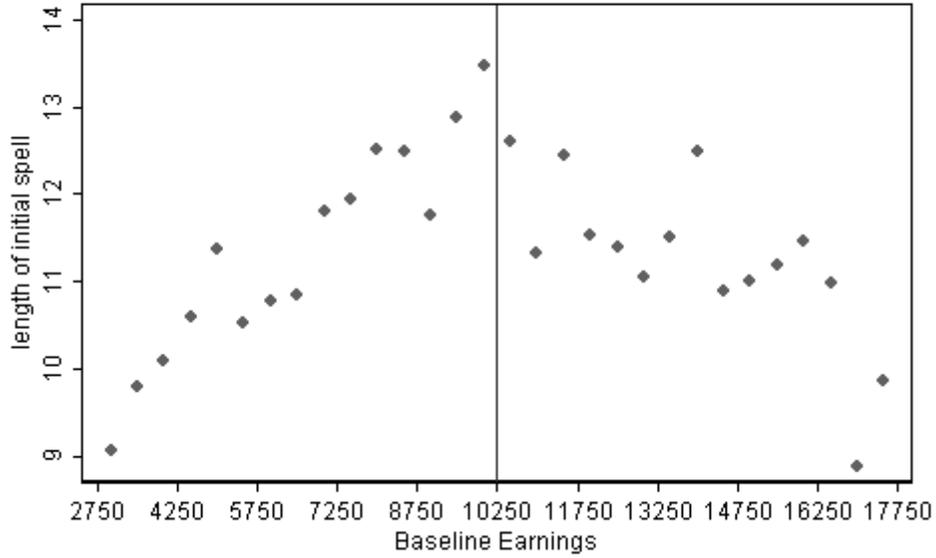


Figure 8(d): Weeks of UI Claimed vs Baseline Earnings  
Censoring Adjusted



Each point in 8(c) represents the local average of the length of the initial spell in the benefit year for claimants in each baseline earnings bin, and likewise in 8(d) the weeks of UI payment claimed, censoring adjusted. See section 4.1 for a discussion of the censoring adjustment procedure. The bins are of size 500 and are centered at multiples of 500. The vertical line corresponds to the kink point in weekly benefit amount determination: baseline earnings of \$10,250.

Figure 8(e): Length of Initial Spell vs Baseline Earnings  
Censoring Adjusted



Each point in 8(e) represents the local average of the length of initial spell, censoring adjusted, for claimants in each baseline earnings bin. See section 4.1 for a discussion of the censoring adjustment procedure. The baseline earnings bins are of size 500 and are centered at multiples of 500. The vertical line corresponds to the kink point in weekly benefit amount determination: baseline earnings of \$10,250.

Figure 9(a): Total UI Benefit Payment vs Baseline Earnings  
Controlling for Covariates

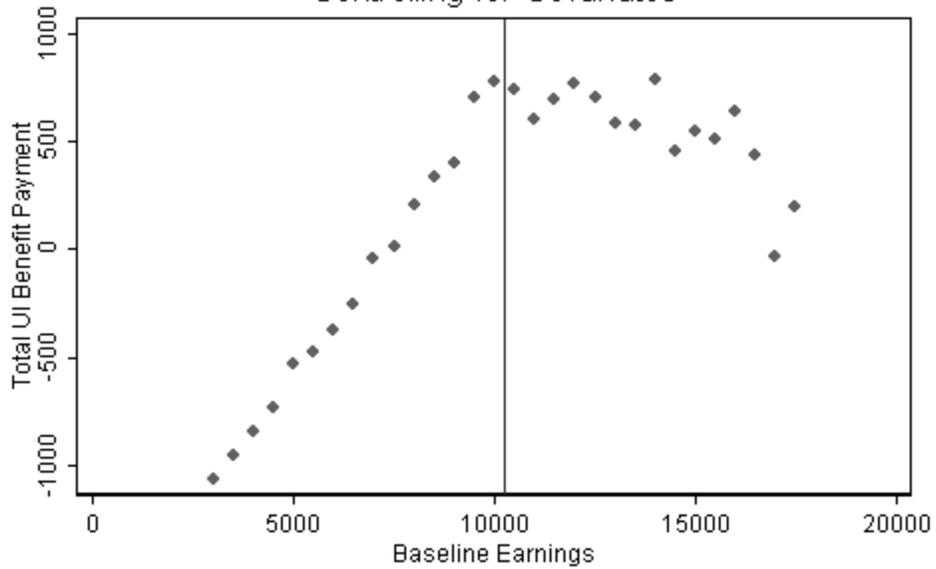
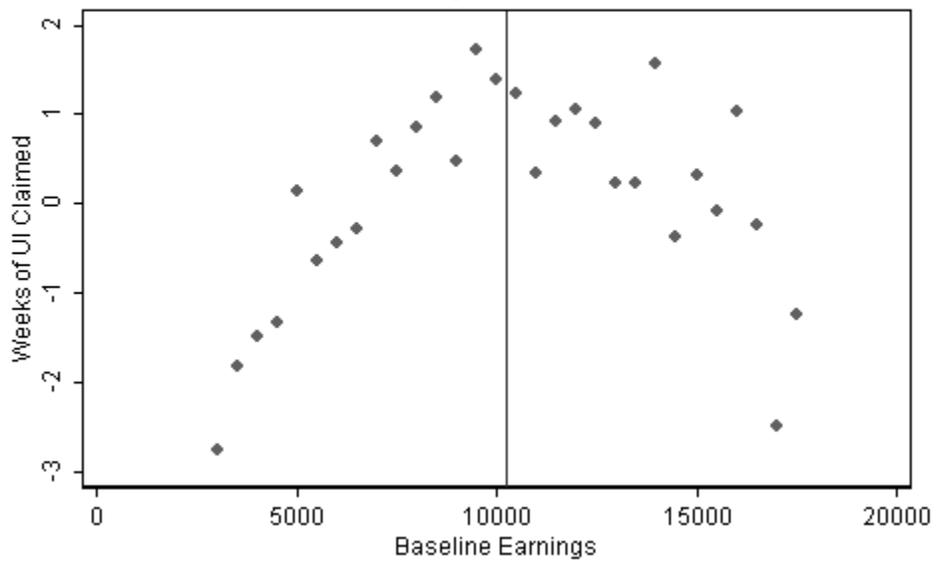


Figure 9(b): Weeks of UI Claimed vs Baseline Earnings  
Controlling for Covariates



Each point in 9(a) represents the local average of the total UI benefit received over the benefit year in each bin of baseline earnings, controlling for covariates, and likewise in 9(b) the weeks of UI claimed. The earnings bins are of size 500 and are centered at multiples of 500. The vertical line corresponds to the kink point in weekly benefit amount determination: baseline earnings of \$10,250.

Figure 9(c): Length of Initial Spell vs Baseline Earnings  
Controlling for Covariates

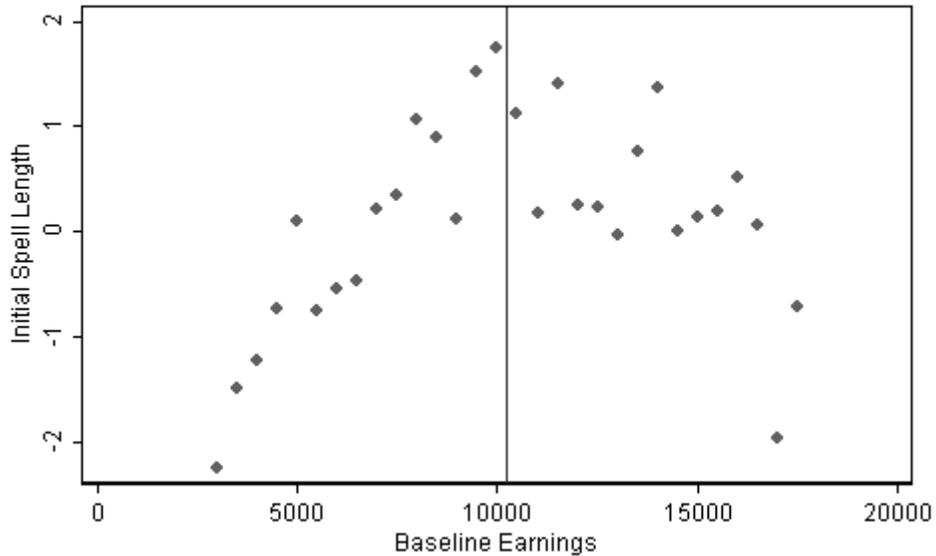
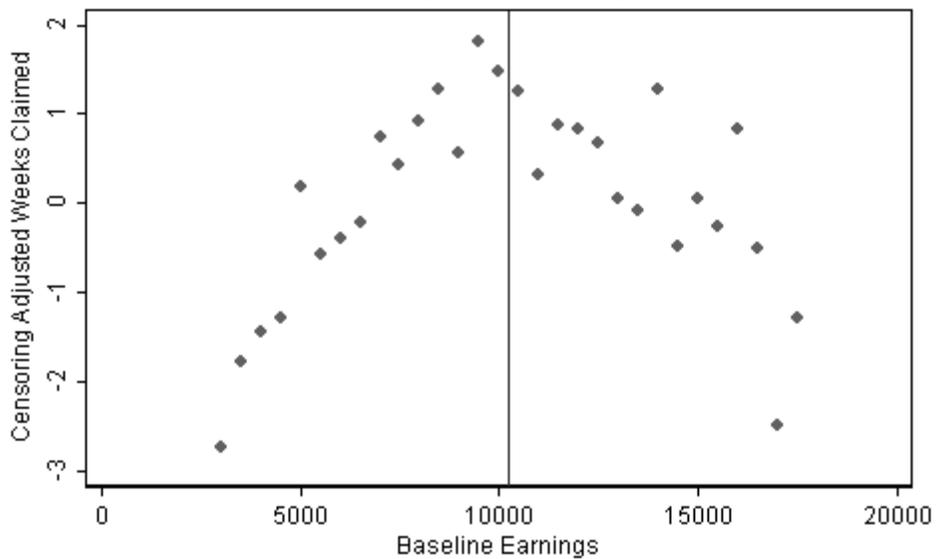
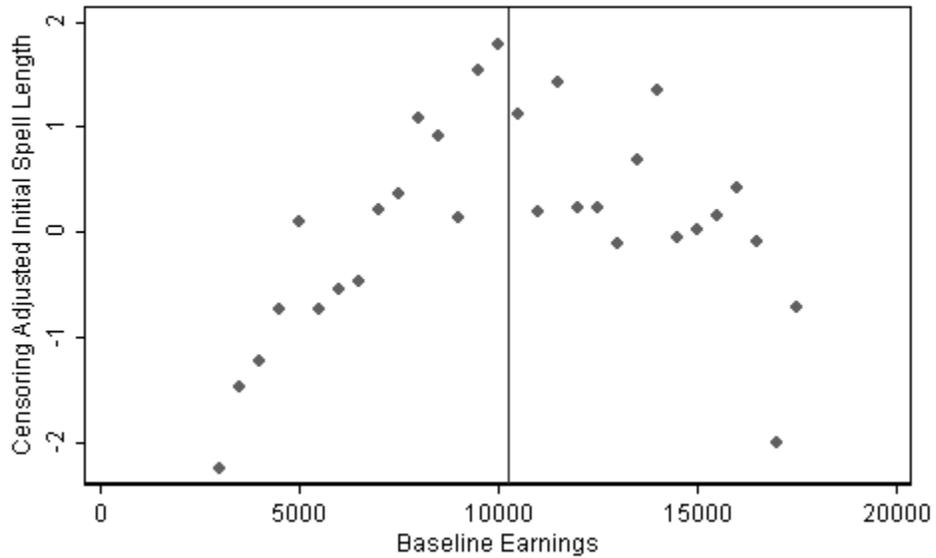


Figure 9(d): Censoring Adjusted Weeks Claimed  
vs Baseline Earnings: Controlling for Covariates



Each point in 9(c) represents the local average of the initial spell length controlling for covariates in each bin of baseline earnings, and likewise in 9(d) the length of the weeks of UI claim, censoring adjusted, controlling for covariates. The earnings bins are of size 500 and are centered at multiples of 500. The vertical line corresponds to the kink point in weekly benefit amount determination: baseline earnings of \$10,250.

Figure 9(e): Censoring Adjusted Length of Initial Spell vs Baseline Earnings: Controlling for Covariates



Each point in 9(e) represents the local average of the censoring adjusted length of the initial spell, and controlling for covariates, in the benefit year in each bin of baseline earnings. See section 4.1 for a discussion of the censoring adjustment procedure. The earnings bins are of size 500 and are centered at multiples of 500. The vertical line corresponds to the kink point in weekly benefit amount determination: baseline earnings of \$10,250.

Table 1: Summary Statistics

Variable	Mean	SD	Min	Median	Max	N
Baseline Earnings	9610.58	7176.64	828.53	7900.15	173406.10	15444
Weekly Benefit Amount	152.92	51.90	55	160	209	15444
Maximum Insured Unemployment Duration	26.86	4.17	10	30	30	15444
Total Unemployment Benefit Received During Benefit Year	2044.34	1868.74	0	1512	6488	15444
Number of Weeks claimed for Unemployment Benefit	14.09	10.80	0	13	51	15444
Length of Initial Spell in Weeks	11.28	10.72	0	7	52	15444
No. of Weeks claimed for Unemployment Benefit, Censoring Adjusted	13.98	10.69	0	13	51	15444
Length of Initial Spell in Weeks, Censoring Adjusted	11.26	10.69	0	7	52	15444
Age	36.24	11.27	15.7	34.2	87.9	15444
Male	0.608	0.488	0	1	1	15444
White	0.839	0.367	0	1	1	15444
Education	12.355	2.674	1	12	19	15444
Claimant filed at a Job Service Center in Western Washington	0.701	0.458	0	1	1	15444

Note that the initial spell includes the waiting week, but the total number of weeks claimed for unemployment benefit does not.

Table 2: RK estimates for Outcomes of Interest: Baseline Specification

Bandwidth	Polynomial of Order	Total UI Received	Total Weeks Claimed	Initial Spell Length	Censoring Adjusted	
					Total Weeks Claimed	Initial Spell Length
7500 N=13605	One	18.1 (0.9) [0.396]	0.0441 (0.0049) [0.091]	0.0415 (0.0049) [0.301]	0.0461 (0.0048) [0.080]	0.0423 (0.0048) [0.302]
	Two	16.8 (3.3) [0.514]	0.0050 (0.0187) [0.394]	0.0343 (0.0188) [0.378]	0.0138 (0.0185) [0.417]	0.0365 (0.0187) [0.385]
	Three	23.9 (8.4) [0.675]	0.0867 (0.0462) [0.602]	0.1196 (0.0467) [0.711]	0.0968 (0.0458) [0.642]	0.1197 (0.0466) [0.702]
	Optimal Poly Order (AIC)	1	2	1	2	1
2500 N=4426	One	17.0 (4.3) [0.732]	0.0296 (0.0234) [0.700]	0.0489 (0.0238) [0.490]	0.0370 (0.0232) [0.690]	0.0496 (0.0238) [0.489]
	Two	36.1 (17.6) [0.847]	0.1397 (0.0947) [0.844]	0.2041 (0.0973) [0.542]	0.1398 (0.0940) [0.848]	0.2044 (0.0972) [0.533]
	Three	74.2 (43.2) [0.940]	0.3718 (0.2317) [0.932]	0.4215 (0.2382) [0.716]	0.3803 (0.2303) [0.942]	0.4180 (0.2380) [0.704]
	Optimal Poly Order (AIC)	1	1	1	1	1
1000 N=1786	One	28.9 (17.2) [0.984]	0.1133 (0.0909) [0.985]	0.1603 (0.0936) [0.909]	0.1181 (0.0905) [0.986]	0.1605 (0.0935) [0.908]
	Two	143.9 (68.3) [0.495]	0.6556 (0.3644) [0.735]	0.9708 (0.3746) [0.867]	0.6420 (0.3631) [0.774]	0.9676 (0.3745) [0.871]
	Three	53.6 (173.6) [0.207]	0.1747 (0.9067) [0.401]	0.0046 (0.9301) [0.425]	0.1665 (0.9038) [0.422]	0.0116 (0.9300) [0.426]
	Optimal Poly Order (AIC)	1	1	1	1	1

For each polynomial-bandwidth specification, the first row reports the RKD estimate, the second row the robust standard error, and the third row the p-value from the Goodness of Fit test. For each bandwidth, the optimal order of polynomial is indicated.

Table 3: RK estimates for Outcomes of Interest: Specification with Covariates

Bandwidth	Polynomial of Order	Total UI Received	Total Weeks Claimed	Initial Spell Length	Censoring Adjusted	
					Total Weeks Claimed	Initial Spell Length
7500 N=13605	One	18.1 (0.8) [0.325]	0.0450 (0.0048) [0.139]	0.0370 (0.0047) [0.332]	0.0469 (0.0048) [0.139]	0.0378 (0.0047) [0.334]
	Two	16.7 (3.3) [0.570]	0.0082 (0.0182) [0.374]	0.0255 (0.0181) [0.382]	0.0170 (0.0180) [0.397]	0.0277 (0.0180) [0.383]
	Three	22.2 (8.1) [0.698]	0.0709 (0.0450) [0.515]	0.1050 (0.0447) [0.627]	0.0807 (0.0445) [0.553]	0.1052 (0.0446) [0.612]
	Optimal Poly Order (AIC)	1	2	1	2	1
2500 N=4426	One	16.0 (4.2) [0.818]	0.0234 (0.0227) [0.745]	0.0390 (0.0226) [0.421]	0.0307 (0.0225) [0.729]	0.0397 (0.0226) [0.417]
	Two	31.1 (17.1) [0.786]	0.1167 (0.0921) [0.728]	0.1740 (0.0923) [0.396]	0.1169 (0.0913) [0.737]	0.1743 (0.0922) [0.384]
	Three	50.8 (42.1) [0.847]	0.2472 (0.2257) [0.779]	0.2098 (0.2263) [0.479]	0.2563 (0.2242) [0.798]	0.2061 (0.2261) [0.464]
	Optimal Poly Order (AIC)	1	1	1	1	1
1000 N=1786	One	21.0 (16.9) [0.864]	0.0696 (0.0890) [0.785]	0.0872 (0.0891) [0.957]	0.0741 (0.0886) [0.801]	0.0872 (0.0890) [0.956]
	Two	130.1 (66.8) [0.264]	0.5439 (0.3544) [0.401]	0.7429 (0.3544) [0.841]	0.5290 (0.3527) [0.442]	0.7397 (0.3543) [0.845]
	Three	45.1 (168.7) [0.077]	0.1718 (0.8740) [0.147]	-0.0188 (0.8740) [0.280]	0.1608 (0.8704) [0.160]	-0.0107 (0.8738) [0.281]
	Optimal Poly Order (AIC)	1	1	1	1	1

For each polynomial-bandwidth specification, the first row reports the RKD estimate, the second row the robust standard error, and the third row the p-value from the Goodness of Fit test. For each bandwidth, the optimal order of polynomial is indicated.

Table 4: RK Estimates for Covariates

Bandwidth	Polynomial of Order	Age	Education	Male	White	West
7500 N=13605	One	0.0017 (0.0047) [0.145]	0.0108 (0.0011) [0.074]	-0.0002 (0.0002) [0.000]	0.0007 (0.0001) [0.891]	0.0011 (0.0002) [0.003]
	Two	-0.0249 (0.0177) [0.196]	0.0146 (0.0041) [0.099]	-0.0009 (0.0008) [0.202]	0.0014 (0.0006) [0.916]	0.0029 (0.0007) [0.011]
	Three	0.0320 (0.0429) [0.203]	-0.0035 (0.0100) [0.372]	-0.0025 (0.0019) [0.320]	0.0014 (0.0014) [0.904]	0.0002 (0.0017) [0.074]
	Optimal Poly Order (AIC)	1	1	2	1	3
	One	-0.0021 (0.0215) [0.733]	0.0068 (0.0049) [0.115]	-0.0009 (0.0010) [0.281]	0.0008 (0.0007) [0.967]	0.0012 (0.0009) [0.372]
	Two	-0.0513 (0.0860) [0.491]	0.0114 (0.0198) [0.084]	-0.0009 (0.0040) [0.186]	-0.0003 (0.0027) [0.701]	0.0049 (0.0035) [0.330]
	Three	-0.1272 (0.2052) [0.378]	0.0708 (0.0477) [0.381]	-0.0212 (0.0097) [0.620]	0.0023 (0.0066) [0.437]	0.0117 (0.0085) [0.581]
Optimal Poly Order (AIC)	1	1	1	1	1	
1000 N=1786	One	-0.0511 (0.0808) [0.122]	0.0280 (0.0190) [0.399]	-0.0073 (0.0039) [0.066]	0.0008 (0.0026) [0.156]	0.0064 (0.0034) [0.085]
	Two	-0.1687 (0.3199) [0.122]	0.0236 (0.0766) [0.355]	-0.0273 (0.0152) [0.055]	0.0063 (0.0104) [0.154]	0.0011 (0.0132) [0.605]
	Three	-0.3036 (0.8119) [0.078]	-0.1795 (0.1892) [0.764]	0.0476 (0.0392) [0.298]	-0.0222 (0.0251) [0.392]	-0.0018 (0.0333) [0.521]
	Optimal Poly Order (AIC)	1	1	1	1	2

For each polynomial-bandwidth specification, the first row reports the RKD estimate, the second row the robust standard error, and the third row the p-value from the Goodness of Fit test. For each bandwidth, the optimal order of polynomial is indicated. "West" indicates that the Job Service Center at which the claimant files a claim is in western Washington.