EVALUATING THE REGRESSION DISCONTINUITY DESIGN USING EXPERIMENTAL DATA

Dan Black
Department of Economics
Syracuse University
danblack@maxwell.syr.edu

Jose Galdo
Department of Economics
Syracuse University
jcgaldo@maxwell.syr.edu

Jeffrey Smith
Department of Economics
University of Michigan
econjeff@umich.edu

ABSTRACT

The regression discontinuity (RD) design has recently become a standard method for identifying causal effects for policy interventions. We use an unusual "tie breaking" experiment, the Kentucky Working Profiling and Reemployment Services, to investigate selection bias in the RD approach. Two features characterize this program. First, the treatment (reemployment services) is assigned as a discontinuous function of a profiling variable (expected benefit receipt duration), which allows the identification of an experimental sample and two alternative non-experimental groups. Second, we deal with a discontinuity frontier rather than a discontinuity point, which allows the identification of marginal average treatment effects over a wide range of the support of the discontinuous variable. Using a variety of multivariate parametric and nonparametric kernel estimators, we estimate the bias with respect to the benchmark experimental estimates. In general, we find that the RD estimates are sensitive to the sample used in the estimations, the outcome of interest, and the econometric models.

JEL Classification: C13, C14.

Version: June 2005

We thank seminar participants at the Econometric Society World Congress at the University of College London and the Workshop Seminar at Syracuse University. We thank in particular Wilbert van der Klaauw and Christopher Taber for helpful comments and suggestions.

Introduction

The regression-discontinuity design (hereafter RD) has recently become a standard evaluation framework for solving causal issues with non-experimental data. The intrinsic feature of this approach is that treatment is given to individuals if and only if an observed covariate crosses a known threshold. Under weak smoothness conditions, the probability of receiving treatment near the cut-off behaves as if random. This unique design allows one to identify the program's causal effect without imposing arbitrary exclusion restrictions, index assumptions on the selection process, functional forms, or distributional assumptions on errors.

The empirical literature applying the RD approach can be traced back to Thistlethwaite and Campbell's (1960) pioneering work that estimated the effect that receipt of a National Merit Award has on a student's later success. As the award is given to students who achieve a minimum score, differences in future academic achievement between those students above and below that cut-off is attributed to the effect of the award. Recent literature has shown a renewed interest in exploiting information about discontinuities in the treatment assignment. Hahn, Todd, and van der Klauuw (2001) is the first to link RD design to the programme evaluation literature and, along with Porter (2003), to formally establish weaker conditions for identification. The statistical properties of the randomness in treatment status near the cut-off is formally shown by Lee (2003), and several empirical applications, including Angrist and Lavy (1999), Black (1999), van der Klaauw (2001), Lee (2003), Lemieux and Milligan (2004), Chen and van der Klauuw (2004), exploit randomized variation near the point of discontinuity to solve selection bias. The common ground in all these empirical studies is their reliance on observational data, which prevents the evaluation of the performance of RD econometric estimators in solving the evaluation problem.

In this paper, we present evidence about the extent of the selection bias arising from parametric and nonparametric RD estimators following LaLonde (1986) literature that evaluates the performance of non-experimental estimators using experimental data as benchmark. We exploit a unique tie-breaking experiment, the Kentucky Working and Profiling Reemployment Services (hereafter KWPRS), that allows us to identify both experimental and non-experimental data without the need for resorting to "external" comparable groups. In this sense, one of the contributions of this paper is its reliance on high-quality data that put all experimental and non-experimental treated and untreated individuals in the same local labor markets and under the same administrative surveys and questionnaires.

The KWPRS program seeks behavioral effects on Unemployment Insurance (UI) claimants with expected high probabilities of benefit exhaustion through mandatory reemployment and training services early in their spell as a condition of continued receipt of benefits. Claimants starting new spells receive a profiling score that represents their expected benefit receipt duration. For each local employment office in each week, those individuals with the highest scores are the first to be selected for reemployment services. This process continues until the number of slots available for a given office in a given week is reached. Within the marginal profiling score – the one at which capacity is reached – random assignment allocates the claimants into treatment. Because of this tie breaking experiment (Campbell 1969), four different sample sets of individuals can be identified from observed discontinuities in the profiling score. They are the experimentally determined treated group (B) and control group (C), each sharing the same marginal profiling score in a given office and in a given week. Black, Smith, Berger, and Noel (2003) call these sets of claimants “profiling tie groups”, or PTGs, constituting 286 different experimental groups between October 1994 to June 1996.

Claimants with profiling scores above or below the marginal ones form the non-experimental treated (D) and comparison (A) samples. Individuals with scores above the marginal ones are automatically assigned to reemployment services, whereas individuals with scores below the marginal scores are by design left out from treatment. Following Rosenbaum’s (1987) suggestion of using alternative comparison groups to better identify program impacts, we can identify two sets of non-experimental samples.[1] Claimants in groups D, C, and A form the first non-experimental sample, named the “RD data from above”, whereas claimants in B, D, and A groups form the “RD data from below”. Within each RD sample, we match treated and untreated individuals conditional on week and local office. We call the resulting sets “regression discontinuity groups”, or RDGs – groups of claimants with at least one treated and one untreated individual in a given office and in a given week, located in each side of the discontinuous point along the continuous profiling score.

Then, the estimation of the selection bias for several RD estimators comes from comparing non-experimental treatment impacts from above ($\mu_D - \mu_{CA}$) and below ($\mu_{BD} - \mu_A$) to corresponding experimental impacts ($\mu_B - \mu_C$) using different window widths along the point of discontinuity. When testing the null hypothesis of no difference

[1] Rosenbaum (1987) suggest that the use of a second control group in non-experimental settings can sometimes help detect the presence of important variables not observed in the data that affects each sample differently. Fraker and Maynard (1987) conclude that the treatment impacts for the NSW program is sensitive to the choice of control groups.

between the non-experimental and experimental impacts, we bootstrap the test statistic to account for the correlation between the two estimates that follows from using either the B or the C group in the construction of both estimates.

Two additional features characterize these unique data. First, unlike traditional RD designs that allows us to learn about treatment impacts for persons near the single point of discontinuity, the KWPRS program embodies multiple discontinuity points, as matched treated and untreated individuals are located on both sides of the profiling score boundary for a given week and a given local office. This characteristic allows us identify treatment effects over a wider range of the support of the discontinuity variable. Second, the KWPR program gives us the opportunity to evaluate alternative discontinuities by looking at neighbors in two other dimensions: weeks, and local offices.

In addition to having neighbors along the profiling score dimension, while holding fixed the week and local office, we also can identify neighbors along the local office dimension, while holding fixed the profiling score and week, and neighbors along the week dimension, while holding fixed the profiling score and the local office. Which of these three counterfactuals better replicates the experimentally determined treatment effects is a question of significant methodological and substantive interest. In this context, this paper provides useful information about which variables matter in terms of conditioning for designing econometric evaluations of UI programs or policies.

The estimation of RD treatment effects is related to problems of boundary point estimation of conditional expectations, as the size of the discontinuity at the cut-off value is itself the object of interest (Porter 2003). We implemented both parametric and nonparametric models to estimate treatment impacts for three outcomes of interest: weeks receiving UI benefits, amount of UI benefits received, and annual earnings. The parametric fixed-effect specification uses a full set of boundary dummies to control for correlations within RDG. Then, in addition to implementing simple local Wald estimates that take mean differences on raw outcome variables for "neighbors" at both sides of the discontinuity frontier, we estimate a smooth version of the Wald estimator through a multivariate kernel approach with mixed categorical and continuous data (Li and Racine 2004). This approach allows us to control within-group specific effects by using "hybrid" kernel functions with profiling scores, weeks, and local offices as conditioning variables. Least-squares cross-validation determines the optimal bandwidths. Finally, we estimate Hahn's et al. (2001) one-side unconditional kernel estimator.

We have four main findings. First, by restricting the non-experimental samples to units increasingly closer to the discontinuity frontier, the RD approach leads to the least bias. This is a systematic result for both parametric and nonparametric empirical approaches. It is important to mention, however, that the estimates are sensitive to the sample used in the estimations, the outcome of interest, and the econometric models.

Second, it becomes clear that the narrower the window width, the less important is the degree of flexibility inherent to each estimator. There is a strong variability between parametric and nonparametric impact estimates, however, when the largest window widths samples are the subject of estimation. Indeed, all nonparametric estimators yield highly biased estimates when applied to the full discontinuity samples. This evidence underscores the role that pre-treatment conditioning variables may play in the estimation of UI benefits receipt outcomes. In particular, past-earnings significantly accounts for the good performance of parametric models in estimating the counterfactuals for amount of UI benefits.

Third, we find evidence that alternative simple discontinuity frontiers along the local office and week dimension cannot do better than the intrinsic profiling score discontinuity in replicating experimental treatment impacts. It is evident, however, that counterfactuals along the geographical location perform slightly better than counterfactuals along the time dimension. This result suggests that in predicting the outcomes of the UI claimants had they do not participated in the program, it may be better to use as counterfactuals marginal claimants living in similar (neighbor) local labor market offices than marginal claimants asking UI benefits in similar (neighbor) weeks.

Finally, we estimate the full non-experimental impacts by merging the full sample of treated and untreated non-experimental individuals. In terms of mean impacts, the full sample impacts show that the KWPRS program does what it is intended to do. It shortens the duration of UI claims, reduces total benefits paid, and raises earnings. By looking at the "finest" +/-0.05 sample set, we observe that the treatment group collects payment for about 1.8 fewer weeks than the untreated group, $60 less in benefits, and earnings about $1,500 more than the untreated group.

The paper proceeds as follows. Section I presents the program and the data. In Section II, we discuss the identification strategy, parameters of interest, and assumptions of the RD design. Section III describes the empirical framework used in the estimation of the parameters of interest. Section IV presents the RD estimates and explores the sensitivity of the estimates to several robustness specifications. In Section V, we describe alternative counterfactuals by looking at simple discontinuities along the profiling score, week, and local office dimension. Section VI explores the consequences of expanding the number and compositions of RDGs by merging the full non-experimental sample. The final section concludes.

I. The Program and the Data

The potentially distortionary incentives the UI system provides for workers are well known. The incentives motivates UI claimants to extend their unemployment spells beyond what they would be in the absence of UI benefits, either by subsidizing additional job search or by subsidizing the consumption of leisure.[2] In November 1993, President Clinton signed into law the Unemployment Compensation Amendments of 1993, which requires states to launch worker profiling and reemployment systems in order to reduce the duration of unemployment spells for those with higher probabilities of exhausting the 26 weeks of UI benefits.

In June 1994, the Commonwealth of Kentucky was selected as a prototype state for implementing the KWPRS program, which identifies potential exhaustees of the UI benefits among new initial claimants, and then offers them mandatory reemployment services such as job-training and job-search workshops early in their spell so they may continue receiving benefits. The services themselves can be view either as a valuable opportunity to learn new employment-related skills, or as an in-kind tax on the leisure of the UI claimants (Black, et al 2003). Thus, the KWPRS program combines aspects of two prototypical UI reforms that intend to reduce the incentives for excess benefit receipt

[2] See Mortensen (1970) and McCall (1970) for earlier works in job search models and UI; Ashenfelter (1978a) and Moffitt and Nicholson (1982) for labor supply models and UI. Meyer (1990) documents spikes in the empirical hazard function as claimants approach the exhaustion of their UI benefits. Card and Levine (2000) show that increasing the length of time that claimants may receive benefits causes the empirical hazard function to fall significantly.

either without punishing workers for whom a longer search is optimal (e.g., Illinois bonus experiment) or by enforcing job search requirements (e.g., Connecticut experiment).[3]

The Center for Business and Economic Research (CBER) at the University of Kentucky took responsibility for developing and predicting the fraction of their 26 weeks of UI benefits that claimants will use up. The model was estimated by employing five years of claimant data obtained from the Kentucky unemployment insurance mainframe computer databases, supplemented with data from other administrative data sources.[4] Two main features distinguish the Kentucky model from prior prototypical profiling models implemented in other states (see Eberts, O'Leary, and Wandner 2002). First, the dependent variable is not represented by a dichotomous variable of whether the claimant exhausted UI benefits, but rather used the fraction of benefits received as a continuous variable. Second, the KWPRS model relies on more than 140 covariates, including past earnings, schooling, past job characteristics, prior unemployment receipt, prior welfare receipt, industry and occupation controls, and local economic and labor market conditions.[5] With these data, a double-limit tobit model was implemented resulting in monotonic increases in the weekly benefit amount, months of job experience, and the previous year's earnings as the fraction of benefits exhausted increases. Most important, the richness of the data yield significant gains in predictive power with respect to profiling models from other states (Berger, Black, Chandra, and Allen 1997).[6]

[3] See Meyer (1995) for a comprehensive survey; Woodbury and Spiegelman (1987) for a detailed analysis of Illinois Bonus experiment; and Ashenfelter, Ashmore, and Deschenes (1999) for evidence about work search enforcement programs.
[4] Enhanced National Data System (ENDS), U.S Department of Labor ES-202 database, U.S 1990 Census.
[5] It is against the law to profiled based on ethnicity, age, sex, and veteran status.
[6] The difference in the exhaustion rate between claimants with predicted benefit receipt duration above the 60th percentile and claimants with predicted benefit durations below this level is 24.8 percent, much greater than the maximum difference -12.5 percent and 10.7 percent- reported by O'Leary, Decker, and Wander (1998) for the Pennsylvania and Washington profiling models.

I.1 The Nature of the KWPRS Design

The Kentucky model produces a single continuous measure of the fraction of UI benefits each claimant will collect. This profiling score is collapsed into a discrete score ranging from 1 to 20. Claimants predicted by the model to exhaust between 95 and 100 percent of their unemployment benefits receive a score of 20, claimants predicted to exhaust between 90 and 95 percent of their unemployment benefits receive 19, and so on. For each local employment office in each week, claimants starting new spells are ranked by their assigned scores. Those individuals with the highest scores are the first to be selected for reemployment services, and this process continues until the number of slots available for each office in each week is reached. Those claimants selected to receive reemployment services are contacted via mail to inform them about their rights and responsibilities under the program. Importantly, the treatment consists of the requirement to receive reemployment services, not actual receipt of these services. As many selected claimants may leave the UI system before receiving services but after being required to receive services, the KWPRS treatment can be considered as the intent-to-receive-treatment.

If the maximum number of claimants to receive reemployment in a given local office and in a given week is reached, and there are two or more claimants receiving the same discrete profiling score, a random number generator assigns the appropriate number of claimants to treatment. Therefore, only claimants with ***marginal*** profiling scores - the one at which the capacity constraint is reached in a given week and in a given local office - are randomly assigned into experimental treated and control groups. Black et al. (2003) call these marginal sets of claimants “profiling tie groups”, or PTGs. This design differs from typical experimental evaluations wherein all program applicants are randomly assigned.

This unique random assignment allows the identification of four different samples following observed discontinuities in the profiling scores. Figure 1 describes the location of these groups. The experimental treated and control samples, B and C samples, are over a region of random overlap in the distribution of profiling scores. D represents the non-experimental treated sample, or claimants that automatically receive treatment because of their high scores. Finally, those claimants with scores below the marginal ones are by design left out from treatment and they represent the non-experimental comparison group A.

From June 1994 to October 1996, the period for which we currently have data, 1,236 and 745 claimants are in the experimental treated and control groups, representing 286 PTGs ranging in size from 2 to 54.[7] For the same period, 47,889 and 9,032 claimants form the non-experimental treated and comparison groups. This means that the experimental design uses only about 2.6 percent of the treated population and 7.6 percent of the untreated population. Table 1 presents descriptive statistics for key pre-treatment covariates for each one of the four samples after discarding individuals with missing information for any covariate of interest.[8] The average discretized profiling scores are 15.2 and 14.7 for the experimental treated and control units, and 16 and 11.9 for non-experimental treatment and comparison units. These results show the ability of the profiling system to select UI claimants into treatment.[9]

[7] The combination of 87 weeks and 32 local offices give 2,742 potential PTGs. Empty cells, however, for many weeks and local offices gives a final number of 286.
[8] Appendix A describes the data attrition process.
[9] The corresponding average continuous profiling scores are 0.83, 0.79, 0.92, and 0.60, respectively.

The large difference in annual and quarterly earnings between the experimental and non-experimental samples is remarkable. In particular, the non-experimental comparison units present much lower annual earnings ($16,489) with respect to the other groups (over $19,000). This result seems counterintuitive, since individuals with lower predicted probability of benefit exhaustion are supposedly individuals with relatively better labor market attachment. A plausible explanation is that poor individuals who work enough to qualify for UI do not stay unemployed very long. In terms of schooling and age, all groups are very homogenous, having on average 12 years of schooling and an average age of 37. In addition, the table depicts some differences in the percentage of females and blacks among the groups, although we do not observe significant differences.

In order to determine if the individuals from the experimental sample were drawn from the same population, we present in column 4 the *p*-values for test of differences in means between the B and C samples. Since there are as many experiments as the number of PTGs that exist, the test is based on a linear regression that conditions on a treatment dummy variable and PTGs. The *p*-values do not reject the null for all covariates. On the other hand, when we applied the test to both D and A non-experimental groups after conditioning on local office and week variables instead of PTGs, the null hypothesis is rejected for almost all covariates.

I.2 <u>The Determinants of the Profiling Tie Groups</u>

In this section, we explore the determinants of a discontinuity having a PTG. Figure 2 shows the asymmetric distribution of PTGs across the 32 local employment offices in the Commonwealth of Kentucky. The average number of PTGs per office is 8.9, ranging in size from 0 to 33. The figure also reveals a high concentration of PTGs in few local offices. For instance, Northern Kentucky (e.g., Covington, Louisville, and Fern Valley) counts for almost 30 percent of the total number of experimental groups, whereas the western region (e.g., Paducah, Mayfield, and Murray) has less than 3 percent of them, and other areas like Maysville and Danville have none.

Two factors explain the rationale for considering these regional differences. First, it is clear that those offices in which most of the experimental groups are concentrated will drive the experimental treatment impact. Therefore, comparing non-experimental treated and untreated individuals from the same local labor markets is crucial in the estimation of the RD bias. In addition, if the goal is to compare comparable samples, then it is important to recognize that the probabilities of having a PTG determines simultaneously the probability of having a corresponding RDG in a given office and in a given week..

In general, the probability of belonging to a PTG depends on both local-level economic conditions and demographic or socioeconomic individual characteristics. For instance, an office-week with a larger number of claimants should be more likely to have a PTG, because they are more likely to have insufficient slots to serve all their claimants and because it is more likely that the marginal slot will not fall at a profiling score boundary. Likewise, individuals living in unpopulated geographic areas, like rural and Appalachian towns, are less likely to participate in the PTGs. In addition, given the relatively low female labor force participation in the Appalachian areas, we expect women will have less probability of participation in the PTGs.

Table 2 presents the estimated coefficients of the probit model *Pr(PTG)*. The regression considers local-level characteristics, including number of claimants, unemployment rate, per capita income, and size of the labor force. To build these variables we map local offices to counties served in order to use county-level official data. For instance, the unemployment rate added to the record of each claimant is a weighted average of the county-level unemployment rate in the month each individual starts claiming UI benefits, where the weights are proportional to the county labor force. In addition, we include individual characteristics such as gender, first quarter earnings before participation, area of residence, age distribution, education, race, and dummy variables to indicate labor force status transitions between four and one quarter before participation. Furthermore, we include dummy variables for each of the 32 local offices and each of the 87 weeks for which we currently have data. These two variables increase the predicted ability of the model, at the cost of having 2,396 out of 56,908 observations that perfectly predict *Pr(PTG=0)*. We drop these units in the final estimation.

The results are consistent with our basic predictions. The number of claimants and the unemployment rate are positively related to the probability of being profiled into the PTGs, as is labor force size. An increase of 10,000 claimants increases that probability by 4.3 percent. Per-capita income is negatively correlated with participation. Individual characteristics also enter the equation in the manner one would generally expect. Women are less likely to participate in the PTGs, although the coefficient suggests the effect of sex on participation is negligible. Claimants living in Appalachian and Western areas have higher probability of being profiled into PTGs than those living in rural areas. The negative coefficient for Metropolitan areas is explained by the large number of claimants residing in those areas, which implies that a relatively small proportion of them fall in the PTGs. Finally, the magnitude of both age and schooling coefficients are statistically insignificant, and the result for past earnings, suggests that individuals with fewer earnings are more likely to participate in the PTGs.

I.3. The Regression Discontinuity Groups (RDGs)

Following Rosenbaum's (1987) suggestion of using alternative comparison groups to better identify program impacts, we can identify two alternative non-experimental samples each composed of treated and untreated individuals located in each side of a boundary along the continuous profiling score. Claimants in groups D, C, and A form the first non-experimental sample, named "RD data from above", whereas claimants in groups B, D, and A groups form the "RD data from below". Within each RD sample, we match treated and untreated individuals conditional on week and local office. We call the resulting sets "regression discontinuity groups", or RDGs – groups of claimants with at least one treated and one untreated individual in a given office and in a given week, located in each side of the discontinuous point along the continuous profiling score.

It is important to notice that the potential number of RDGs, either from above or below, is bounded by the total number of PTGs, which follows from using either the B group or the C group in the construction of the RDGs. As the goal of this paper is to evaluate the selection bias of RD econometric estimators, we consider only those RDGs that have corresponding PTGs in a given office and in a given week.[10] This feature guarantees the availability of high quality data, thereby overcoming one of the main criticisms of matching incomparable non-experimental samples to experimentally determined samples (Smith and Todd 2003). Using data for the period of analysis, 281

[10] For instance, the PTG # 10 includes B and C claimants in week 1, local office 3, and continuous scores in the range [0.50, 0.44]. Its corresponding RDG from above includes D, C, and A claimants in the same week and local office, but with scores in the range [1.06, 0.53] for the treated D group, and [0.50, 0.48] for the untreated C and A groups.

RDGs with 11,235 claimants from above emerge, ranging in size from 2 to 219. The mean size is 60.1, with a 25th percentile of 33, and a 75th percentile of 79. Likewise, 266 RDGs with 11,479 claimants from below arises, ranging in size from 4 to 231. The mean size is 64.5, with a 25th percentile of 35, and a 75th percentile of 83.

II. The RD Approach

The evaluation problem is the identification of the counterfactual as a measure of the effect of some binary treatment variable (T_i) has on an outcome variable (Y_i). When T_i is not randomly distributed in the population of interest, the correlation between the treatment variable and unobservables affecting the outcome of interest prevents the estimation of unbiased and consistent treatment effects. The RD design uses discontinuities in the treatment assignment process to identify and estimate a causal effect. In this model, the participation process is a discontinuous function of an observable variable (S) that is also related to the outcome of interest. Let $T_i = f(S_i)$, where $\overline{S}$, the point at which $f(S_i)$ is discontinuous, determines participation status. If Y depends on S, then this assignment rule will induce a discontinuity (jump) in the observed relationship between Y and S at $S = \overline{S}$, where the size of the discontinuity itself is the object of interest. Two features distinguish this approach from the standard selection on observable models. First, unlike matching models, there is no common support for participants and nonparticipants as for all values of S, $P(T = 1 | S) \in \{0,1\}$. Second, the selection rule depends deterministically on a known and observable variable, constituting what the literature calls a “sharp” design (Trochim 1984).

Figure 3 illustrates the model. The underlying relationship between Y and S is assumed to be nonlinear. The segment below the cutoff level reflects this relationship, which would continue to higher levels of *S* in the absence of the treatment (dashed region). We can observe from the data $E(Y_1 | S)$ when $S \geq \bar{S}$, and $E(Y_0 | S)$ when $S < \bar{S}$. What we cannot identify is either the dashed curves or the counterfactual mean $E(Y_0 | S = \bar{S})$ at the point of discontinuity. Under the assumptions of a common effect model and of linearity in the relationship between *Y* and *S,* the treatment effect can be estimated by simple OLS estimation of *Y* on *T* and *S* (Heckman, LaLonde, and Smith 1999). The crucial issue with this approach is the selection of the functional form for $E(Y_0 | S)$, which is done either by formal tests or visual inspection of the data.

A more flexible approach uses data from individuals arbitrarily close to the cut-off point, $E(Y_0 | S = \bar{S} - \varepsilon)$, to identify causal effects. Because in a small neighborhood around the cutoff point $\bar{S}$ the direct impact of *S* on the potential outcomes is likely to vary only a little with *S*, differences between the mean outcome for individuals marginally below and above the threshold represent the causal effect. This approach has the advantage of avoiding extrapolation bias, but does so at the cost of identifying only a local treatment effect.

In many situations, however, the participation decision is not completely determined by *S.* For instance, case workers may have some discretion about whom they offer the program, or they may base their decision on criteria that are not observed by the econometrician. In this "fuzzy" design world, the assignment rule changes the problem from one of selection on observables to one of selection on unobservables (e.g., van der Klauuw 2002, Chen et al. 2004). The identification and estimation of causal effects is more challenging than in the standard case because the relationship between *T* and *S* may be discontinuous at $\bar{S}$ only on average, resulting in the loss of much of the simplicity of the regression discontinuity design.

II.1 <u>The RD Approach Applied to KWPRS and Identifying Conditions</u>

The KWPRS program constitutes a special case of selection on observables, where T_i is determined by the estimated profiling score (S_i) following a "sharp" assignment rule. Let Y_{i1} and Y_{i0} be the potential outcomes for the individual *ith* in the treated and untreated states, and $Y_i = T_i Y_{1i} + (1 - T_i) Y_{0i}$ be the observed outcome. This is the Neyman-Fisher-Cox-Rubin model of potential outcomes. It is also the "two regime" or "switching regression" model (Quandt 1972) or the Roy model of income distribution. Using a general nonlinear and nonseparable outcome model, we describe the potential outcome equation for the participation state as $Y_{1i} = m_1(S_i, U_{1i})$ and the potential outcome for the non-participation state as $Y_{0i} = m_0(S_i, U_{0i})$, where *S* is the profiling score and (U_1, U_0) are the unobservables affecting the potential outcomes.

The treatment assignment follows the structure of the RD design. For a given office and a given week, treatment outside the PTGs is assigned based on whether the profiling score added to each new claimant record, S_i, crosses a particular threshold,

$$T_i = \begin{Bmatrix} 1 \text{ if } S_i \geq \bar{S}_{\text{office,week}} \\ 0 \text{ if } S_i < \bar{S}_{\text{office,week}} \end{Bmatrix}. \qquad (1)$$

This mere treatment assignment rule itself is insufficient to identify causal impacts because the selection bias comes from the potential relationship between *Y* and S, which causes a correlation between the treatment indicator and unobservables. For instance, those individuals with higher profiling scores are by definition those with higher probabilities of exhausting their UI benefits. Hence, it is plausible to expect that they have less attachment to the labor market than individuals with lower profiling scores, yielding downward bias estimates when the outcomes of participants and non-participants are compared by simple means differences.

The identification of causal effects is possible by matching discontinuities in the relationship between profiling scores and outcomes to discontinuities in the relationship between participation and profiling scores. In the presence of one or very few cutoff values, a practical approach is the combination of formal testing with visual inspection of the data (e.g., Angrist and Lavy 1999, van der Klauuw 2002). In this application, we have as many unordered cutoff points as the number of RDGs that exist, which prevent us from visualizing jumps in the underlying outcome distribution at each discontinuity point along the profiling score space. Under the assumption of a common effect model, however, we still can illustrate discontinuities in the relationship between profiling score and outcomes by re-centering the data to a unique discontinuity point. For any particular RDG *j*, we build a variable defined as the continuous profiling score minus the cut-off score such that the treatment status is defined when this variable, named "profiling margin of treatment", crosses the 0 threshold. After pulling together all RDGs, we have a unique discontinuity point at 0, with all treated units having a profiling margin of treatment equal to or higher than 0 and all non-treated units less than 0.

Figure 4 illustrates the jumps in the outcomes of interest, including weeks receiving UI benefits, amount of UI benefits received, and annual earnings, at the unique threshold for both RD data from above and below. Nowhere else is there a jump, and a smooth relationship between the outcomes and the profiling margin of treatment along the profiling score dimension can be observed, except at the cut-off that determines the treatment status. The dash curves represent conditional means when the outcomes are regressed on a third-order polynomial in the profiling margin of treatment, separately for each side of the threshold. All figures exhibit important nonlinearities, which highlight the importance of determining the functional form of $E(Y_0 | S)$. In addition, there is a negative relationship between weeks receiving UI benefits and profiling margin of treatment on both sides of the threshold. This indicates selection bias, and a simple comparison of means for all units above and below the threshold would yield biased treatment effect estimates of the program. The same result holds for amount of UI benefits received.

The assumption that $Y_1 - Y_0$ is independent of T conditional on S is insufficient for nonparametric identification. The following proposition states the conditions for identifying causal effects, involving only the sizes of discontinuities in conditional expectations:

Proposition (Hahn, et al. 2001): Let $\delta = Y_1 - Y_0$ denote the treatment impact and $e > 0$ denote an arbitrary small number. If $E(Y_{0i} | S_i)$ and $E(Y_{1i} | S_i)$ are continuous in S at $\bar{S}$, and the limit $\lim_{\varepsilon \to 0^+}[\delta_i | S_i = \bar{S} + \varepsilon]$ is well defined, then it can be shown $E[\delta_i | S = \bar{S}] = \lim_{\varepsilon \to 0^+}\{E[Y_i | S = \bar{S} + \varepsilon] - E[Y_i | S = \bar{S} - \varepsilon]\}$.

This result shows that the identification of average treatment effects relies on the assumption that individuals marginally above and below the cut-off score ($\bar{S}$) have the same distribution of observable variables and, therefore, the outcomes of those individuals marginally below represent what would have been the outcomes for those individuals marginally above in the absence of the program. In the context of the KWPRS program, it is equivalent to stating that nonrandom selection in the RDGs mimics a randomized event in the neighborhood of the marginal profiling scores. Thus, when we use the RDGs from above, we are estimating the treatment effect at the C-D discontinuity. Comparing that to the experimental treatment effect impacts constructed using the B and C groups, then, presumes some amount of smoothness in the estimated mean treatment effect along the profiling score dimension at this point. If that smoothness is not there, then the experimental impact is itself a poor estimate of the impact at the discontinuity. One reason why this assumption could be violated is that claimants selectively exploit their profiling scores to avoid being assigned to treatment. Clearly, this possibility is not possible under the rules governing the UI profiling programs, where administrative data is used and it is impossible for each UI claimant to know and manipulate his score such that the density of S_{ij} is zero at the cut-off scores.

It is a limitation of the RD design that the treatment effects are only estimated for individuals around the cutoff point. As a different threshold defines each RDG, the KWPRS non-experimental data is based on a discontinuity frontier rather than in a simple point of discontinuity. The full set of discontinuous points allows us to identify treatment effects over a wider range of the support of *S*, and by then pulling the data together, it is possible to identify a weighted average effect across all discontinuity points.

II.2 The Parameter of Interest

The parameter we are interested in is the marginal average treatment effect (MATE) that estimates the mean impact on persons at some margin of participation (Björklund and Moffitt 1987, Heckman 1997). A small expansion or contraction of the discontinuity point within RDG will bring claimants into the program or push them out of it, driven by a change in the average impact of participants. Let *S* be the profiling score that takes on two values $\bar{S}$ and $\bar{S}'$, with *T=1* if $S=\bar{S}$ and *T=0* otherwise. Denote $S(x)$ the support of the distribution of profiling scores, conditional on *X=x*. For any $(\bar{S},\bar{S}')\in S(x)\times S(x)$, such that $P(\bar{S})>P(\bar{S}')$, the parameter of interest is:

$$\Delta^{MATE}=E(Y_1-Y_0\mid T_{\bar{s}}=1,T_{\bar{s}'}=0). \qquad (2)$$

that gives the mean gain for claimants who would be induced to switch from *T=0* to *T=1* if the profiling score were exogenously manipulated from $\bar{S}'$ to $\bar{S}$. This parameter is similar to the local average treatment effect (LATE) proposed by Imbens and Angrist (1994), although the key difference is that in the MATE case the variable causing the change in participation need not be an instrument.[11]

A second important distinction in our application is that MATE defines a policy relevant parameter that answers well-posed economic questions: What is the gross gain from a marginal expansion of the KWPRS program? What is the average gain for UI claimants who are in the margin between participating or not participating given $S=\bar{S}$?. This is not always the case with LATE, where different instruments defined different parameters. When the instruments are indicator variables that denote either different

[11] In the presence of an instrument, MATE is the limit form of the LATE parameter and it is term local instrumental variable (LIV) by Heckman and Vytlacil (1999). Conditional on *X=x*, LIV is the derivative of the conditional expectation of *Y* with respect to *P(S)=s:* $\Delta^{LIV}=\partial E(Y\mid X,P(S)=s)/\partial s$.

policy regimes or different levels of intensity of a policy within a given regime, LATE has a natural interpretation as the response to policy changes for those who change their participation status. When the instruments refer to personal or neighborhood characteristics used to predict an endogenous variable, LATE has a less unambiguous interpretation (Heckman, et al. 1999).

Another feature that arises from the MATE parameter is that it can be related to the most common parameters of interest in the treatment literature. The average treatment effect (ATE), the mean impact for an individual chosen at random from the population of UI claimants; and the average treatment effect on the treated (ATT), the mean impact for claimants who are selected to receive reemployment services, can be written as weighted averages of MATE (Heckman and Vytlacil 1999). Thus, the average treatment effect, given by $\Delta^{ATE} = E(Y_1 - Y_0 \mid S)$, represents an average of the MATE parameters over the full support of *S,* where *S* is treated uniformly. Likewise, the impact of treatment on the treated, $\Delta^{ATT} = E(Y_1 - Y_0 \mid S, T = 1)$, represents a weighted average of the MATE parameters, where claimants who have characteristics that make them the most inclined to receive reemployment services - those with higher profiling scores *S*- receive on average the most weights.

There is no heterogeneity in gains moving from "0" to "1" as long as $U_1 = U_0$ among people with the same *S*. Therefore, in a homogenous treatment effect world, marginal treatment effects are average treatment effects, so all the above evaluation parameters are the same. If impacts are heterogeneous, then $U_1 \neq U_0$, and the difference between marginal treatment effects and average treatment effects can be important. For instance, Björklund and Moffitt (1987) analyzed a program that had a positive average treatment on the treated but a negative marginal treatment impact, implying that the program was overextended to include individuals who were made worse off by it.

In the heterogeneous case, we can differentiate two cases. If claimants select purposely in the program based on their unobservables affecting the gains, then the average gain experienced by individuals who sort into treatment compared to what the average person would experience is positive, leading to $\Delta^{ATT} > \Delta^{ATE}$. The deterministic selection rule assumed in the RD design, however, precludes claimants choosing to participate in the program based on their expected gains since the assignment is determined exogenously to them. In this case, $E(U_1 - U_0 \mid S, T = 1) = 0$ and, therefore, $\Delta^{ATT} = \Delta^{ATE} = \Delta^{MATE}$. If treatment impacts vary with the profiling score *S*, Δ^{MATE} is nonconstant and the mean impact on the treated may differ from the mean impact on a randomly selected claimant in the population.

III. The Measurement Framework

We use a multivariate framework that controls correlations within RDG, due to the multiple discontinuities that characterized the KWPRS program. Beginning by parametric specifications, we then relax the linear specifications through nonparametric models, including the local Wald estimator, one-side local linear kernel, and a multivariate kernel approach with mixed data.

III. 1 Parametric Estimators

A model for individual outcomes is used to describe the causal relationship to be estimated. For the *ith* individual in discontinuity group *j*, we can write

$$Y_{ij} = \alpha + X_{ij}'\beta + \delta T_{ij} + g(S_{ij}) + \eta_{RDG,j} + \varepsilon_{ij}. \qquad (3)$$

where Y_{ij} is individual *i*'s outcome, X_{ij} is a vector of individual socioeconomic characteristics, including age, sex, and race/ethnicity that were not used in building the profiling score. T_{ij} is an indicator variable such that $T_{ij} = 1\{S_{ij} > \overline{S}_j\}$, $g(S_{ij})$ is an assumed smooth continuous function of the continuous profiling score. The term $\eta_{RDG,j}$ is introduced to parameterize the discontinuity groups fixed effects with the purpose of capturing shifts in the regression line or correlation in outcomes within RDGs, and ε_{ij} is the error component specific to each individual. This methodology is equivalent to calculating differences in mean outcomes on opposite sides of the discontinuity frontier, controlling for individual characteristics and continuous profiling scores and relating this to differences in treatment.

The identification of the causal effect is allowed by matching discontinuities in the relationship between outcomes and profiling scores to discontinuities in the relationship between participation and profiling scores. In this case, the marginal profiling score ($\overline{S}_j$) determines a discontinuous participation in the program; at the same time, any other relationship between profiling scores (S_{ij}) and individual outcomes (Y_{ij}) is controlled by including an assumed smooth function $g(S_{ij})$. Hence, the underlying assumption is that participation in the program is the only source of discontinuity in outcomes at the cut-off profiling scores within each discontinuity group, and there is not reason to expect sudden jumps in the outcomes at those cut-off values in the absence of the program.

To assess the plausibility of this assumption, it is important to highlight the purposes of profiling people in the UI system and see if there is any theory that may back up differential potential outcomes for those that received the same profiling scores. Since more than 140 relevant covariates get involved in the computation of *S*, and after conditioning in *X=x*, it is very difficult to argue that claimants with marginally the same score for the same week and same local office will have different outcomes in the treated and untreated states. Treated and not treated claimants within RDG not only share the same set of standard variables that human capital or job search models consider as first order explanatory variables of labor outcomes, but they are also individuals with the same UI benefits history, UI receipt history, prior employment characteristics, labor market problems, and participation in welfare programs. What it is untestable, however, is the assumption that $g(S_{ij})$ will capture any correlation between the treatment variable and unobservables affecting the outcomes. That is the reason why the empirical literature (e.g., Angrist et al. 1999, Black 1999, Lee 2003, van der Klaauw 2001, Lemieux et al. 2004, Chen, et al. 2004) considers a wide range of alternative specifications for $g(S_{ij})$, such as polynomials and splines. In this paper, the order of the polynomial approximation to the $g(S_{ij})$ function is selected by the data via generalized cross-validation.

When the impact of treatment is the same for each RDG, estimation of equation (3) provides efficient and consistent estimates for ATT, ATE, and MATE. When the impact of treatment varies across RDGs, the ordinary least-squares estimate of δ corresponds to a weighted average of the RDG-specific treatment effects, $\hat{\delta} = \sum_{j=1}^{K} w_j(\overline{Y}_1 - \overline{Y}_0)$, where $w_j = \frac{p_j(1-p_j)N_{1j}}{\sum_{k=1}^{K} p_k N_{1k}}$, K is the number of RDGs, N_{1j} is the number of treated claimants in the *jth* corresponding PTG group, and p_j is the probability that a unit of *jth* corresponding PTG receives treatment. It implies that, for a given probability of receiving treatment within a PTG, those RDGs whose corresponding PTGs have larger number of treated units will receive larger weights. Furthermore, our unique KWPRS design, unlike traditional RD designs, allows us to estimate heterogeneous treatment effects that vary with *S* over the range of values that have treated and untreated individuals. In this case, it is necessary to assume that the functional form of $E(Y_0 \mid S)$ can be determined through extrapolation such that the estimated treatment impacts are equal to the difference between the extrapolated $E(Y_0 \mid S)$ and the observed outcomes of participants at each value of *S*. In the simplest case, where δ_j is a linear function of *S*, the treatment impact for participants, $\delta_i = \delta_0 + \delta_1 S_i$ is estimated by

$$Y_{ij} = \alpha + X_{ij}'\beta + g(S_{ij}) + \delta_0 T_{ij} + \delta_1 S_{ij} \times T_i + \eta_{RDG}, {}_j + \varepsilon_{ij}. \qquad (4)$$

We can interpret the OLS estimate as a weighted average of MATE; where the weights do not necessarily integrate to one and they are not necessarily nonnegative (see Heckman and Vytlacil 2005).

III.2 Nonparametric Estimators

The parametric regression-based estimation approach depends on the assumption made about the functional form of the relationship between the outcome and the selection variable. The problem of specifying a correct parametric control function is relaxed through the adoption of several nonparametric estimators. In addition to implementing simple Wald estimators that takes mean differences on raw outcome variables for neighbors at each discontinuity point and, then, averages those results across all RDGs using as weights the proportion of treated units in each RDG, we employ a multivariate kernel approach that allows us to control within-group specific effects. This nonparametric measurement framework is built up in two steps. First, we pull the data together and estimate the expected conditional means $\hat{m}_1(.)$ and $\hat{m}_0(.)$ for treated and untreated claimants, separately, using the following multivariate kernel regression,

$$Y_{ij} = m(S_{ij}, W_{ij}, L_{ij}) + \varepsilon_{ij}. \qquad (5)$$

where Y_{ij} is *i's* individual outcome, S_{ij} is the key continuous profiling score variable, W_{ij} represents the week when each unit is profiled in the UI system, and L_{ij} is the local labor office where each UI claimant is profiled. The interaction between W_{ij} and L_{ij} defines each discontinuity group *j*. The local office variable is represented by a one-dimensional index equal to the average earnings in the year prior to the start of the UI spell in each office. Taken together, both variables play the same role as η_j in the parametric model (3).[12] One advantage of this flexible approach is the identification of local treatment effects at each discontinuity point by evaluating the estimated conditional means at each particular point (week and local office) of interest, allowing us to test the common effect assumption. In the second step, we take simple mean differences of these smoothed outcomes within each regression discontinuity group. The average treatment effect is estimated by a weighted average of these differences across all RDGs,

$$\Delta^{RDD} = \sum_{j=1} w_j \{\frac{1}{n_1}\sum_{i=1}^{n_1} \hat{m}_{\mathbf{1i}} I(h_j) - \frac{1}{n_0}\sum_{i=1}^{n_0} \hat{m}_{\mathbf{0i}} I(h_j)\}. \qquad (6)$$

where w_j represents the weights equal to the proportion of treated units in each corresponding PTG group, and $I(h_j)$ is an indicator function that takes the value 1 for

[12] Notice that with just one discontinuity point the standard approach would estimate $Y_i = \mu(S_i) + \varepsilon_i$ for *T=1* and *T=0*, using local linear kernel regression and then just differentiate $\hat{m}_1 - \hat{m}_0$ for some fixed bandwidth value of *S*.

those units in the vicinity of each j discontinuity, and takes 0 otherwise. The size of the vicinity is determined by the selected bandwidth (*h*).

This estimator relies on the feasibility and consistency of multivariate kernel regression in the presence of mixed categorical and continuous data. Li and Racine (2004) show the consistency and the asymptotic distribution of multivariate kernel regression functions with mixed data types and least squares cross-validation selection of the bandwidth parameters. In the presence of *q*-continuous variables (x^c) and *r*-categorical variables (x^d), the local linear kernel regression can be computed from the minimization problem:

$$\min_{a,b_1,\ldots b_q} \sum_{i=1}^{n} \{Y_i - a - b_1(x_{i1}^c - x_1^c) - \ldots\ldots b_q(x_{iq}^c - x_q^c)\}^2 \prod_{s=1}^{q} K_c\left(\frac{x_{is}^c - x_s^c}{h_s}\right) \prod_{s=1}^{r} K_d\left(\frac{x_{is}^d - x_s^d}{\lambda_s}\right). \qquad (7)$$

where *K(.)* is a well-behaved kernel function, h_s and λ_s the bandwidths for continuous and discrete variables, both converging to zero as $n \to \infty$. The estimated conditional mean $\hat{\mu}(.)$ is equal to $\hat{a}$, the local linear estimator. The key difference between this approach and the traditional nonparametric approach lies in the nature of the kernel function that is used. Here we use multivariate "hybrid" kernel functions, which allow a nice interaction between continuous and categorical variables without the need for resorting to the conventional nonparametric frequency estimator that in the presence of small samples leads to finite-sample inefficiencies.

A general concern with a multivariate kernel approach is the slow rate of convergence that depends dramatically on the number of independent variables. Li and Racine (2004) show, however, that the presence of categorical variables does not affect the rate of convergence of this "hybrid" kernel estimation framework and it will depend only on the number of continuous variables involved in the estimation. We use the Epanechnikov kernel function as the weight-assigning function for the continuous covariates (profiling score and local office index), which has a bounded support that is shown to be valuable in the evaluation framework (see Smith and Black 2003, and Frölich 2004).[13] The key additional variable in our model is week (W_i), which is a categorical variable. As the time factor is a crucial component of the discontinuity design on its own, we treat this variable as an ordered categorical variable with the univariate weight-assigning function, proposed by Wang and VanRyzin (1981).[14]

All these weight-assigning kernel functions depend crucially on the size of the bandwidth parameters (h_S, h_L, λ_W) Their optimum selection reveals, for instance, the optimum "neighborhood" around the discontinuity profiling scores over which the regression is being evaluated, controlling for weeks and local offices. We follow the Hall, Racine and Li (2004) least-square cross-validation approach to find h_S^{opt}, h_L^{opt}, and λ_W^{opt},

$$CV(h_S, h_L, \lambda_W) = \frac{1}{n}\sum_{i=1}^{n}\{Y_i - \hat{m}_{i-1}(x_i^c, x_i^d, h_S, h_L, \lambda_W)\}^2. \qquad (8)$$

This method applied in the context of multivariate "hybrid" product kernels has the important property of smoothing away categorical independent variables that are irrelevant in the estimation of conditional means.[15]

[13] $K_c(x_i, x, h) = \begin{cases} \frac{1}{h}\left(\frac{3}{4\sqrt{5}}\left(1 - \frac{1}{5}(\frac{x_i - x}{h})^2\right)\right), \left|\frac{x_i - x}{h}\right| < \sqrt{5} \\ 0, \text{otherwise} \end{cases}$, when the range of h is [0, ∞].

[14] $K_d(x_i, x, \lambda) = \begin{cases} 1, & \text{if } x_i^d = x_j^d \\ \lambda^{|x_i^d = x_j^d|}, & \text{if } x_i^d \neq x_j^d \end{cases}$ Here the range of λ is [0,1]. When λ =0, the kernel function becomes an indicator function, and when λ =1 it becomes a constant weight function.

[15] Finding $\lambda_i^{opt} = 1$ implies the irrelevance of the *i* categorical regressor. Likewise, when h_i^{opt} is very large, $\hat{m}(.)$ is linear in continuous variable i.

We also consider one-side kernel estimators proposed by Hahn et al (2001) in one of the most influential papers in the RD literature. The idea is to estimate $\tilde{Y}_1^+ = \lim_{S \to S^+} E(Y_i \mid S)$ and $\tilde{Y}_0^- = \lim_{S \to S^-} E(Y_i \mid S)$ through boundary kernels, given by the expressions,

$$\tilde{Y}_1^+ = \frac{\sum_i Y_i 1(S_i > \bar{S}) K_h (S_i - \bar{S})}{\sum_i 1(S_i > \bar{S}) K_h (S_i - \bar{S})}, \; \tilde{Y}_0^- = \frac{\sum_i Y_i 1(S_i \leq \bar{S}) K_h (S_i - \bar{S})}{\sum_i 1(S_i \leq \bar{S}) K_h (S_i - \bar{S})}$$

where *K(.)* is the Epanechnikov kernel function, 1(.) is an indicator function that equals one if the condition in parenthesis is satisfied and zero otherwise, h is the bandwidth parameter, and $\bar{S}$ the cutoff point. Claimants receive weights that decrease to zero with increasing distance to the point of discontinuity at which the kernel is being estimated. We applied this estimator to each RDG and the estimates, then weighted up using as weights the proportion of treated units in each corresponding PTG. As Hahn et al (2001) demonstrate, under same conditions this procedure is numerically equivalent to an instrumental variable estimator for the regression of Y_i on T_i that uses the indicator function as an instrument, applied to the subsample for which $\bar{S} - h \leq S \leq \bar{S} + h.$

The focus on differences rather than levels makes the regression discontinuity model a particular case of the classic boundary bias problem in the kernel regression literature (Porter 2003). This idea is illustrated in Figure 3, where the treatment effects at the discontinuity point ($\bar{S}_j$) is equal to the jump size $\delta.$ Taking simple local weighted averages of the observations just to the right of $\bar{S}_j$ yield an upward bias estimate of the intercept $m(\bar{S}_j) + \delta$. In the same way, simple local weighted averages of observations just to the left of $\bar{S}_j$ yield downward bias estimate of $m(\bar{S}_j)$. Therefore, the boundary bias would be equal to the difference of both biases, which only cancel each other out when the estimation is carried out at an interior point of the support of S. This is not the case in the RD design where the support condition is absent.. Therefore, in all the estimations we use a polynomial "correction", the local linear kernel regression (Fan 1992), which accounts for the biased behavior of the conditional expectations at the boundary points. In the context of the RD approach, Hahn et al. (2001) and Porter (2003) show important bias-reduction properties for the local linear kernel estimation.[16]

IV. The RD Bias Estimates

In this paper, we are able to test the assumptions and predictions underlying the identification of the regression discontinuity design by comparing treatment effects estimates with experimentally determined treatment impacts. We use two alternative non-experimental samples, from above and below, to better estimate the selection bias of a variety of estimators. In all estimations, we consider four window widths, including 0.05, 0.10, 0.15, and all sample, centered at each discontinuity point within RDG. Then, we estimate the bias from above and below by $(\mu_D - \mu_{CA}) - (\mu_B - \mu_C)$ and $(\mu_{BD} - \mu_A) - (\mu_B - \mu_C)$. We expect that by restricting the samples to the smallest window widths, the RD approach lead to the least bias.

IV.1. The Experimental Estimates

The simple mean differences of outcomes between treated (B) and control units (C) do not estimate the KWPRS experimental treatment effects, as usually happens in the

[16] Porter (2003) shows that polynomial kernel regression not only has the same asymptotic bias as typical kernel regression at an interior point of the support, but also in some cases can exhibit further bias reductions. He finds that the local polynomial estimator achieves the optimal convergence rate in the Stone's (1980) sense for estimation of a conditional mean at a point. Hahn et al. (2001) shows the asymptotically normality and a rate of convergence equal to $\sqrt{nh_n}$ for the local linear estimator.

context of a simple random assignment. In general, the tie-breaking experiment does not directly identify many commonly estimated parameters because the treatment impact may differ between units in and out the PTGs (Black et al 2003). Because the program ensures a random assignment only within each PTG and the random assignment ratio differs by PTG, we estimate the experimental impacts by differencing the mean outcomes of treated and not treated individuals within RDG, and the estimates, then, weighted up using as weights the proportion of treated units in each PTG. This experimental estimate can be thought as a weighted average of the estimates from many small randomized experiments.

In order to compare comparable samples when estimating the RD bias, we only use those PTGs with corresponding valid RDGs and vice versa. As the total number of valid RDGs from above and below is 281 and 266, we obtain two sets of experimental impacts that slightly vary from each other. They are -1.85 and -1.92 for weeks receiving UI benefits, $-7.30 and $-22.9 for amount of UI benefits received, and $1,338 and $1,376 for annual earnings. We call these estimates full experimental impacts.

IV.2 <u>RD Bias from Parametric Estimators</u>

Table 3 shows the magnitude of the bias emerging from the estimation of equation (3). The upper and lower panels show the estimates when the above and below non-experimental RD samples are used. Each row depicts the RD bias for each outcome of interest. The first column shows the full experimental impacts and the remaining columns represent the four window width samples. As the number of RDGs with at least one treated and one untreated individual varies across the window width samples, the number of corresponding PTGs should also vary. Therefore, two alternative measures of bias are considered. The first, on the left, uses the full experimental estimates, whereas the bias in curly brackets uses experimental impacts estimated from PTGs with corresponding valid RDGs in each window width sample. Standard errors appear in parenthesis. We also present, in brackets, the *p-values* from testing the null hypothesis of no difference between the non-experimental and full experimental estimates. Importantly, we bootstrap the test statistic to account for the correlation between non-experimental and experimental estimates that follows from using either the B group or the C group in the construction of both estimates.

Four main results emerge. First, when restricting the sample to those observations that are at the closest distances to the discontinuity frontier, the RD bias is increasingly smaller. It is consistent with the main prediction of the RD approach, which states that in the limit there is a randomized variation in treatment status. This result, however, varies depending on the sample used to estimate the bias and the variable of interest. Only in the case of the RD sample from above, the biases for all outcomes of interest are small with the *p*-values do not reject the null hypothesis in almost all estimations. For the +/-0.05 widow width, the bias are 0.04, $4, $-70 for weeks receiving benefits, amount of UI benefits received and annual earnings. If we turn the analysis to the RD sample from below, however, the results are somewhat different. Here, the biases are relatively bigger, and they do not necessarily behave in the way we expected. A potential candidate to explain this asymmetric result, the degree of smoothness at the discontinuity points, is discarded because the average difference in the profiling score for treated and untreated units at the points of discontinuity is the same for both samples (0.029).

Second, the ability of parametric models to solve the evaluation problem depends on which outcome of interest is on the left hand side of the regression. For week receiving UI benefits and amount of UI benefits received, the estimates show small bias and relative stability across the above and below RD samples. On the other hand, annual earnings, which is the most common outcome of interest in the evaluation literature, present higher bias and it is very sensitive to the sample used to estimate the bias.

Third, there is no significant variation between the two alternative bias estimates. This is because the experimental impacts are very steady across all window widths. In the case of the impacts from above, these estimates range from -1.66 to -1.85 for weeks receiving UI benefits, -$28 to $14 for amount of UI benefits received, and $1,211 to $1,351 for annual earnings. The corresponding numbers for the experimental estimates from below are -1.92 to -2.10 for weeks receiving UI benefits, -$40 to $8 for amount of UI benefits received, and $1,376 to $1,482 for annual earnings. Finally, from a practical standpoint the optimal selection of the polynomial approximation to the $g(S)$ function is irrelevant to determining the magnitude of the parametric treatment impacts. Although we only presented parametric estimates for the optimal polynomial approximation, unreported estimates for alternative $g(S)$ approximations show no significant variation in the magnitude of the counterfactuals within each window width sample. The set of conditioning variables X, are more important to determine the size and sign of treatment impacts for the outcomes of interest.

Figure 5 shows that non-experimental treatment impacts vary with S. We estimate equation (4) using the full sample from above and below in order to have treated and untreated units in the support $S \in [0,1]$. We find that treated claimants with large values of S collect payments for about two fewer weeks than the comparison group, whereas claimants with small values of S collect payments for only about one fewer week than the comparison group. Likewise, high-score claimants from above and below receives about \$180 and \$220 more in benefits than the control group, but these differences drop to almost \$30 and \$130 for low-score claimants. Finally, high-score claimants from above and below earned, on average, \$1,300 and \$1,000 more than the control group in the year following initiation of the UI claim, but these differences reduce to \$200 and \$550 for low-score claimants.

IV.3 Nonparametric RD Estimates

To minimize the potential misspecification of the control function $g(S_{ij})$, we turn to nonparametric estimates. As discussed before, an intuitive and simple estimator is local Wald estimator, which estimates a mean difference of the raw outcome variables within each RDG and the estimates then weighted up using the weights defined in previous sections. Table 4 shows the results for each one of the window width samples organized in the same order as table 3, with the standard errors shown in parentheses, and *p*-values from testing the null hypothesis in brackets. In general, the simple difference in means for individuals above and below the points of discontinuity does not yield a good approximation to the benchmark estimates for all variables of interest and for both RD samples. Even in the case of weeks receiving UI benefits, which present a small bias and steady behavior in the parametric world, the results show a relatively large bias with an unclear pattern regarding the size of the different discontinuity samples.

The most compelling result that emerges is the inability of the local Wald estimator to replicate the experimentally determined impacts for amount of UI benefits received. This is a solid result for both above and below RD data, wherein the *p*-values show the rejection of the null hypothesis for all estimations. This result highlights the importance of using pre-treatment covariates in the estimation of conditional mean counterfactuals for this outcome of interest. In particular, past earnings largely explains the small bias in the +/0.05 window width and parametric specification.[17] When looking at the annual earnings variable, we observe a smaller bias than that for amount of UI benefits variable, with almost no bias present when the estimation includes all sample observations. Some fortuitous cancellation may help to explain the good performance of the simple mean differences estimator when applied to the full sample.

Table 5 presents the results when we estimate equation (6). As described in section III.2, we estimate the smoothed version of the Wald estimator by multivariate kernel regression, where profiling scores, weeks, and local offices variables are the regressors. We use a local linear kernel regression with mixed data and least square cross-validation for the optimal smoothing parameters. After finding the smoothed outcomes, we applied the local Wald estimator within RDG, and the estimates then weighted up using the weights described in previous sections.

Three patterns emerge from Table 5. First, the estimates generally show a better approximation to the experimental impacts than simple Wald estimates. The bias for weeks receiving UI benefits substantially drops for almost all window widths, as do the estimates for amount of benefits received. After comparing these results with parametric estimates from table 3, it is clear that the absence of pre-treatment covariates still causes a large bias ($700) for amount of UI benefits received when looking at the full sample

[17] Dropping the pre-program annual earnings covariate leads to a change in the treatment impact from $-3.40 to $-67.7 for the RD sample from above with window width +/-0.05. The corresponding change for the RD data from below is from $-243 to $-230.

estimates. Second, there is evidence in favor of the estimates emerging from the RD data from above that shows a smaller bias with respect to the estimates from below. For instance, using the above +/-0.05 discontinuity sample, we get -0.25 and $-5 bias for weeks receiving benefits and amount of UI benefits received, whereas these numbers are -0.73 and $372 for the corresponding samples from below. Finally, even though the bias for annual earnings is somewhat small, no pattern emerges from the table either between the above and below data or across the different window widths.

A third set of additional nonparametric estimates, reported in table 6, consists of results from Hahn's, et al (2001) one-side local linear kernel regression applied to each RDG separately and the estimates then weighted up using the weights defined earlier. The structure of the table is similar to the previous ones, with two alternative measures of bias, standard errors in parentheses, and *p*-values from testing the null hypothesis in brackets. The results show a much better approximation to the experimentally determined treatment impacts than the estimates emerging from the simple Wald estimator for weeks receiving UI benefits and amount of UI benefits received. In addition, the bias for these two outcomes is increasingly smaller when the discontinuity sample is restricted to individuals closer and closer to the smallest window widths. Thus, in terms of mean impacts, the implementation of the one-side local linear estimator pays off. The estimates for annual earnings, however, are very sensitive to the window width and they do not behave in the way we expect for the smallest samples.

V. <u>Evaluating Alternative Simple Discontinuities</u>

The aim of program evaluation is to find the best possible counterfactual for those units that received the treatment. In general, this task is achieved by looking at those untreated units that are similar in dimensions that affect both the probability of participation and potential outcomes. The role of both geographic location and time dimension in the construction of counterfactuals has been documented in the applied literature. By looking at comparison units living in the same local labor markets, Heckman, et al (1998) and Heckman and Smith (1999) show that the quality of the counterfactuals increases. The composition and size of the counterfactual groups are also affected by considering different time frameworks, as is vividly illustrated in the National Work Support Demonstration Program (Smith and Todd 2004).

The KWPR program allows us the opportunity to evaluate alternative discontinuities by looking at neighbors in two alternative dimensions: weeks, and local offices. In addition to having neighbors along the profiling score dimension, holding fixed the week and local office, we can also identify neighbors along the local office dimension, holding fixed the profiling score and week, and neighbors along the week dimension, holding fixed the profiling score and the local office. Hence, we can build RDGs composed of treated and untreated individuals marginally different in their geographical location but with identical profiling score and week; and RDGs composed by treated and untreated individuals marginally different in their profiling week but with identical profiling score and local office. Which of these three counterfactuals best replicates the experimentally determined treatment effects is of substantive and methodological interest.

To place claimants on each side of a geographical boundary, we use a one-dimensional index of similarity among offices that was constructed using the average past annual earnings from high school white claimants within each employment office. This index has the ability to separate "rich" local labor markets (e.g., Louisville and Fern Valley) from "poor" local labor markets (e.g., Somerset and Murray). The office index ranges from $12,311 to $22,948, with an average of $17,669. To compare comparable samples, we construct valid RDGs along the one-dimensional index office by imposing a $1,000 window in either side of the PTGs. Within that window, we select all offices that provide a discontinuity. Thus, for instance, claimants from B and C groups, with index $15,000, week 5, and discretized score 10 form the PTG#10. Claimants from D, A, and C groups from the same week and discretized score, but local index ($15,000, $16,000] for treated (D) individuals and ($14,000, $15,000] for untreated individuals (A, C), form its corresponding RDG group. From October 1994 to June 1996, the period for which we currently have data, the numbers of geographical-based RDGs emerging from the RD data from above and below are 178 and 77.

Likewise, considering a 4-week window on either side of the PTGs carries out the construction of counterfactuals along the week dimension. Within that window, we use all weeks that provided a discontinuity. For instance, the corresponding RDG for the PTG# 10 is formed by D, A, and C claimants with the same index and discretized score, but weeks (5, 8] for treated individuals (D), and weeks (1, 4] for untreated individuals (A and C). Because of this matching process, 135 and 65 RDGs emerge from the RD data from above and below. It is important to mention that the RD design that characterizes the KWPRS program comes from intrinsic discontinuities along the score dimension, while the discontinuities along the week and local office dimensions are an "artificial" structure imposed on the data.

Table 7 presents the estimates for both office and week discontinuity dimensions using the same measurement framework of the previous sections. We consider two estimators, including the fixed-effect parametric model and local Wald estimates, and the same three outcomes of interest. The benchmark experimental estimates that papers in the second column are re-estimated to include only the PTGs with corresponding RDGs. Some results of interest are visible in the table. In general, we find plain evidence that neither the week nor the office discontinuity frontier can do better than the intrinsic score discontinuity in replicating experimental treatment impacts. This result holds independently of the estimator used, the RD sample, or the outcome of interest. This is not surprising, since we expect that marginal differences in the profiling score have less impact in the outcomes of interest than differences in local labor market conditions or time structure. Especially interesting is the evidence that counterfactuals along the geographical location perform somewhat better than counterfactuals along the time dimension. This result suggests that in predicting the outcomes of the UI claimants had they not participated in the program, it may be better to use as counterfactuals marginal claimants living in similar (neighboring) local labor market offices than marginal claimants asking UI benefits in similar (neighboring) weeks. The Wald estimates outperform the parametric fixed-effect model for all outcomes of interest.

VI. Expanding the Discontinuity Frontier: Full Non-Experimental RD Estimates

Because of the peculiar nature of the KWPRS program, the number of PTGs bound the number of RDGs. Of the 2,784 potential PTGs for the period for which we currently have data, there are only 286 valid PTGs. This means that for several weeks and several offices, we have empty experimental cells, and therefore, no corresponding RDGs can arise, even though non-experimental claimants exist for those offices and weeks.

So far, the research question has focused on the ability of several RD estimators to replicate experimentally determined treatment impacts. In this section, given the good performance of some RD estimators already documented, particularly when applied to the above RD sample, we estimate the full sample impacts by merging the full sample of treated and untreated non-experimental individuals (D and A groups). Put differently, and more generally, the point here is more substantive than methodological. One advantage of expanding the number of RDGs is that the treatment effects is identified over a wider range of the support of the discontinuity variable because both treated and untreated claimants can be found across the profiling scores' distribution.

By looking at discontinuities along the profiling score, we match non-experimental claimants conditional on weeks and local offices. As a result, 1,049 RDGs arise, with at least one treated and one not treated claimant on both sides of the score discontinuity. This signifies a five-fold increase in the number of available RDGs with respect to the RD sample either from above or below. In addition, efficiency gains can also be obtained as the average number of claimants per RDG increases, which contributes to the calculation of estimates that are more precise. The mean size of the full RDGs is 52.4, ranging in size from 2 to 209, and involving a total number of 24,407 treated and 7,948 untreated individuals.

We consider three different estimators, corresponding to the fixed-effect parametric model, local Wald estimator, and one-side boundary kernel. With the exception of the way we build the weights assigned to each RDG, the implementation of these models follows exactly the same structure of previous sections, including the same window widths, outcome of interest, test for determining the order of the polynomial approximation to $g(S_{ij})$, and the same control covariates. Now, the weight each RDG receives is equal to the proportion of treated units having the marginal discretized profiling score within each RDG.

Two main results emerge from the parametric model in table 8. First, the full treatment impacts show important variation across different window widths, which indicates that if the econometrician does not control the quality of the counterfactuals, measured for the distance from the cutoff values, he or she may severely bias the size of the treatment impacts. By looking at the "finest" +/-0.05 sample set, we observe estimates similar to those arising from the experimental estimates. Thus, the treatment group collects payment for about 1.8 fewer weeks than the not treated group, $60 less in benefits, and earnings about $1,500 more than the untreated group. Thus, en terms of mean impacts, the full non-experimental sample show that the KWPRS program does what it is intended to do. It shortens the duration of UI claims, reduces total benefits paid, and raises earnings. Finally, efficiency gains are also observed when looking at the standard errors, which indicate that expanding the number of RDGs leads to much better inferences.

Table 9 and 10 shows the full sample impacts arising from local Wald and one-side kernel regressions. Four patterns emerge. First, by comparing tables 9 and 10, we observe a robust stability of the treatment impacts for all outcomes of interest when the "finest" +/-0.05 sample is the subject of the estimation. The treated claimants collect on average payments for about 2 weeks fewer than their counterpart comparison individuals, about $100 less in UI benefits, and earning about $750 more than the comparison group in the year following initiation of the UI claim. The results for UI benefits receipt are also similar to the estimates that emerged from the parametric model in table 8.

Second, by comparing individuals on both sides of the discontinuity frontier when the sample sizes are expanded increasingly further away from the discontinuous points, both the local Wald estimates and one-side kernel estimates change dramatically, especially in the case of UI benefits. One more time, it highlights the importance of using pre-treatment covariates in the estimation of conditional mean counterfactuals when the sample sizes are expanded increasingly further away from the discontinuous points. Finally, annual earnings present strong sensibility to the degree of flexibility of each estimator when the treatment impacts are estimated using the smallest window width. The treatment effects range from $1645 (parametric estimation) to $743 (kernel estimator). In general, however, the narrower the window width, the less important is the degree of flexibility inherent to each estimator.

VII. Conclusions

In this paper, we have investigated the selection bias of several RD estimators using a tie-breaking experiment, the KWPRS program, which allows us to identify two non-experimental samples arising from natural discontinuities along the profiling score dimension conditional on local offices and weeks. Our approach follows LaLonde (1986) literature that evaluates the performance of non-experimental estimators by using as a benchmark experimentally determined treatment impacts. A major advantage of the KWPRS data is that both experimental and non-experimental data come from individuals living in the same local labor markets without need for resorting to "external" comparable groups. In addition, unlike traditional RD designs that allows us to learn about treatment impacts for persons near the single point of discontinuity, the KWPRS program embodies multiple discontinuity points, as matched treated and untreated individuals are located on both sides of the profiling score boundary for a given week and a given local office. Furthermore, the KWPR program gives us the opportunity to evaluate alternative discontinuities by looking at neighbors in two other dimensions: weeks, and local offices.

We have five main findings. First, in general the RD approach leads to the least bias by restricting the samples to individuals increasingly closer to the discontinuity points. This is a systematic result for both parametric and nonparametric empirical approaches, which suggest that the RD treatment effects may be the quantity of interest if one is interested in the causal effect of the program on the marginal group of individuals around the discontinuity points, and it does not apply to the entire population of interest. There is a caveat with this general result. The bias estimates are sensitive to changes in the non-experimental samples used to estimate the counterfactuals and the outcomes of interest.

Second, it becomes clear that the narrower the window width, the less important is the degree of flexibility inherent to each estimator. There is a strong variability between parametric and nonparametric impact estimates, however, when the largest window widths samples are the subject of estimation. Indeed, all nonparametric estimators yield highly biased estimates when applied to the full discontinuity samples. This evidence underscores the role that pre-treatment conditioning variables may play in the estimation of UI benefits receipt outcomes. In particular, past-earnings significantly accounts for the good performance of parametric models in estimating the counterfactuals for amount of UI benefits.

Third, we find evidence against the “common effect” assumption. For the KWPRS program the estimated treatment effects appears to vary across deciles of the profiling score variable. In particular, the treatment impacts for all outcomes of interest are larger and closer to the benchmark estimates for the subsample of individuals with higher values of *S*.

Four, we find evidence that alternative simple discontinuity frontiers along the local office and week dimension cannot do better than the intrinsic profiling score discontinuity in replicating experimental treatment impacts. We show, however, that counterfactuals along the geographical location perform somewhat better than counterfactuals along the time dimension. This result suggests that in predicting the outcomes the UI claimants had they do not participated in the program, it may be better to use as counterfactuals marginal claimants living in similar (neighboring) local labor market offices than marginal claimants asking UI benefits in similar (neighboring) weeks.

Finally, we estimate the full non-experimental impacts by merging the full sample of treated and untreated non-experimental individuals. In terms of mean impacts, the full sample impacts show that the KWPRS program does what it is intended to do. It shortens the duration of UI claims, reduces total benefits paid, and raises earnings. By looking at the “finest” +/-0.05 sample set, we observe that the treatment group collects payment for about 1.8 fewer weeks than the untreated group, $60 less in benefits, and earnings about $1,500 more than the untreated group.

References

J. Angrist, V. Lavy (1999): "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement.", *Quarterly Journal of Economics*, 144, 533-576.

B. Barnow, G. Cain, A. Goldberger (1980): "Issues in the Analysis Of Selectivity Bias", Institute for Research on Poverty, Discussion Paper, University of Wisconsin-Madison.

E. Battistin, E. Rettore (2002): "Testing for Programme Effects in a Regression Discontinuity Design with Imperfect Compliance", *Journal of the Royal Statistical Society*, Series A, 165, 39-57.

S. Black (1999): "Do better Schools Matter", Parental Valuation of Elementary Education', *Quarterly Journal of Economics*, 144, 577-599.

D. Black, J. Smith (2003): "How Robust is the Evidence on the Effects of College Quality? Evidence from Matching.", *Journal of Econometrics*. 121(1): 99-124.

D. Black, J. Smith, M. Berger, B. Noel (2003): "Is the Threat of Reemployment Services More Effective than the Services Themselves? Experimental Evidence from the UI System", *American Economic Review*, 93 (4) : 1313-1327.

M. Berger, D. Black, A. Chandra, and S. Allen (1997): "Kentucky's Statistical Model of Working Profiling for Unemployment Insurance.", Kentucky Journal of Economics and Business, 16: 1-18.

A. Björklund, R. Moffitt (1987): "The Estimation of Wage Gains and Welfare Gains in Self-Selection Models" , *The Review of Economics and Statistics*, Vol. 69, 1:42-49.

A. Bowman (1984): "An Alternative Method of Cross-Validation for the Smoothing of Density Estimates", *Biometrika*, 71:353-360

D. Campbell (1969): "Reforms as Experiments", *American Psychologist*, 24: 409-429

D. Card, and P. Levine (2000): "Extended Benefits and the Duration of UI Spells: Evidence from the New Jersey Benefit Program", *Journal of Public Economics*, 78: 1-2, 107-38.

D. Card, D. Lee (2003): "Regression Discontinuity Estimation with Random Specification Error.", manuscript, University of California at Berkeley.

S. Chen, W. van der Klaauw (2004): "The Effect of Disability Insurance on Labor Supply of Older Individuals in the 1990s", unpublished manuscript.

R. Eberts, C. O'Leary, and S. Wandner eds. (2002): "Targeting Employment Services", Kalamazoo, MI: W.E Upjohn Institute for Employment Research.

J. Fan (1992): "Design-Adaptive Nonparametric Regression.", *Journal of the American Statistical Association*, 87: 998-1004.

J. Hahn, P. Todd, W. van der Klauuw (2001): "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design" , *Econometrica*, Vol 69(1), 201-209.

P. Hall, J. Racine, and Q. Li (2003): "Cross-Validation and the Estimation of Conditional Probability Densities.", *Journal of the American Statistical Association*, (forthcoming).

J. Heckman, H. R. LaLonde, J. Smith, (1999): "The Economics and Econometrics of Active Labor Programs", in O. Ashenfelter and D. Card, eds., *Handbook of Labor Economics* Volume 3A, Amsterdam, 1865-2097.

J. Heckman (1997): “Instrumental Variables A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations”, *The Journal of Human Resources,* Vol. 32, 3: 441-462.

J. Heckman, and E. Vytlacil (1999): “Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects.”, Proceedings of the National Academy of Sciences of the United States of America, Vol. 96, 8:4730-4734.

J. Heckman, and E. Vytlacil (2005): “Structural Equations, Treatment Effects and Econometric Policy Evaluations.”, NBER Working Paper Series # 11259.

G. Imbens, and J. Angrist (1994): “Identification and Estimation of Local Average Treatment Effects.”, Econometrica, 62 (2): 467-75.

R. LaLonde, (1986): “Evaluating the Econometric Evaluation of Training Programs with Experimental Data”, *The American Economics Review*, 76(4) , 604-620

D. Lee (2003): “Randomized Experiments from Non-random Selection in U.S House Elections”, Manuscript, University of California at Berkeley.

T. Lemieux, K. Milligan (2004): “Incentive Effects of Social Assistance: A regression Discontinuity Approach”, NBER Working Paper 10541.

Q. Li and J. Racine (2004), “Cross-Validated Local Linear Nonparametric Regression”, *Statistica Sinica,* Volume 14, Number 2, pp 485-512.

Q. Li and J. Racine (2003), “Nonparametric Estimation of Distributions with Both Categorical and Continuous Data”, *Journal of Multivariate Analysis*, Volume 86, pp 266-292

J. McCall (1970): "Economics of Information and Job Search," *Quarterly Journal of Economics,* 84: 113-126.

B. Meyer (1995): “Lessons from the U.S Unemployment Insurance Experiments”, *Journal of Economic Literature*, 33: 91-131.

R. Moffitt and W. Nicholson (1982): “The Effect of Unemployment Insurance on Unemployment: The Case of Federal Supplemental Benefits”, *Review of Economics and Statistics*, 64:1-11.

D. Mortensen (1970): "Job Search, The Duration of Unemployment, and the Phillips Curve," *American Economic Review,* 60: 505-517.

C. O’Leary, P. Decker, and S. Wandner (1998): “Reemployment Bonuses and Profiling”, W.E Upjohn Institute Staff Working Paper # 98-51

J. Porter (2003): “Estimation on the Regression Discontinuity Model”, unpublished manuscript.

J. Racine and Q. Li (2004): “Nonparametric Estimation of Regression Functions with Both Categorical and Continuous Variables, *Journal of Econometrics,* 119(1):99-130

P. Rosenbaum (1987): “The Role of a Second Control Group in a Observational Study.”, *Statistical Science*, Vol. 2. 3:292-316.

J. Smith, P. Todd (2004): “Does Matching Overcome LaLonde’s Critique of Non-Experimental Estimators?, *Journal of Econometrics*, 125 (1-2):305-353.

D.L. Thistlethwaite, D.T. Campbell (1960): “Regression Discontinuity Analysis: An Alternative to ex post facto Experiment.” *Journal of Educational Psychology*, 51:309-17.

W. van der Klaauw (2002): "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach", *International Economic Review*, Vol 43(4).

S. Woodbury, R. Spiegelman (1987): "Bonuses to Workers and Employers to Reduce Unemployment: Randomized Trials in Illinois", *American Economic Review,* 77:513-30.

M. Wang and J. VanRyzin (1981): "A Class of Smooth Estimators for Discrete Distributions.", *Biometrika*, 68, 301-309.

Data Appendix

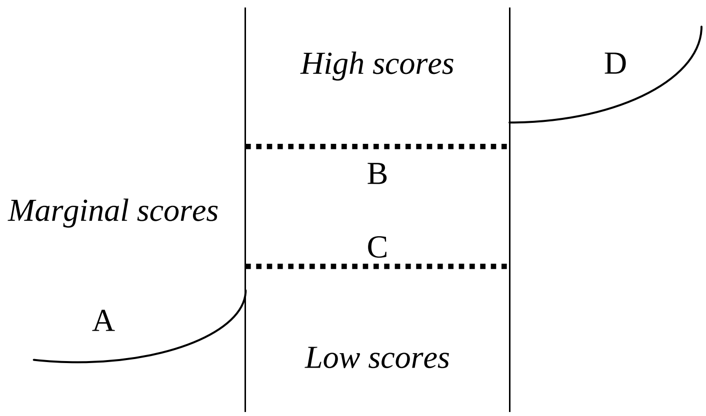
Table A-I: Number of Observations Omitted Due to Sample Restrictions

Kentucky Working and Reemployment Services, October 1994 to June 1996

Restriction	Experimental treated group (B)	Experimental control group (C)	Non-experimental treated group (D)	Non-experimental comparison group (A)
Total number of individuals	1,236	745	46,766	9,032
Number dropped due to missing value for education	14	3	420	398
Number dropped due to missing value for age	0	0	33	3
Number dropped due to corrupted profiling score	0	0	302	19
Final sample size	1,222	742	46,011	8,612

Note: Corrupted profiling score are those continuous scores of size -999.99 or 9999. We consider as valid scores those that are in the range [-2,2].

Figure 1: Discontinuous Treatment Assignment
Kentucky Working and Reemployment Services



Notes:
B=Experimental treated group
C=Experimental control group
D=Non-experimental treated group
A=Non-experimental comparison group

Figure 2: Geographic Distribution of Profiling Tie Groups - PTGs
Kentucky Working and Reemployment Services, October 1994 to June 1996

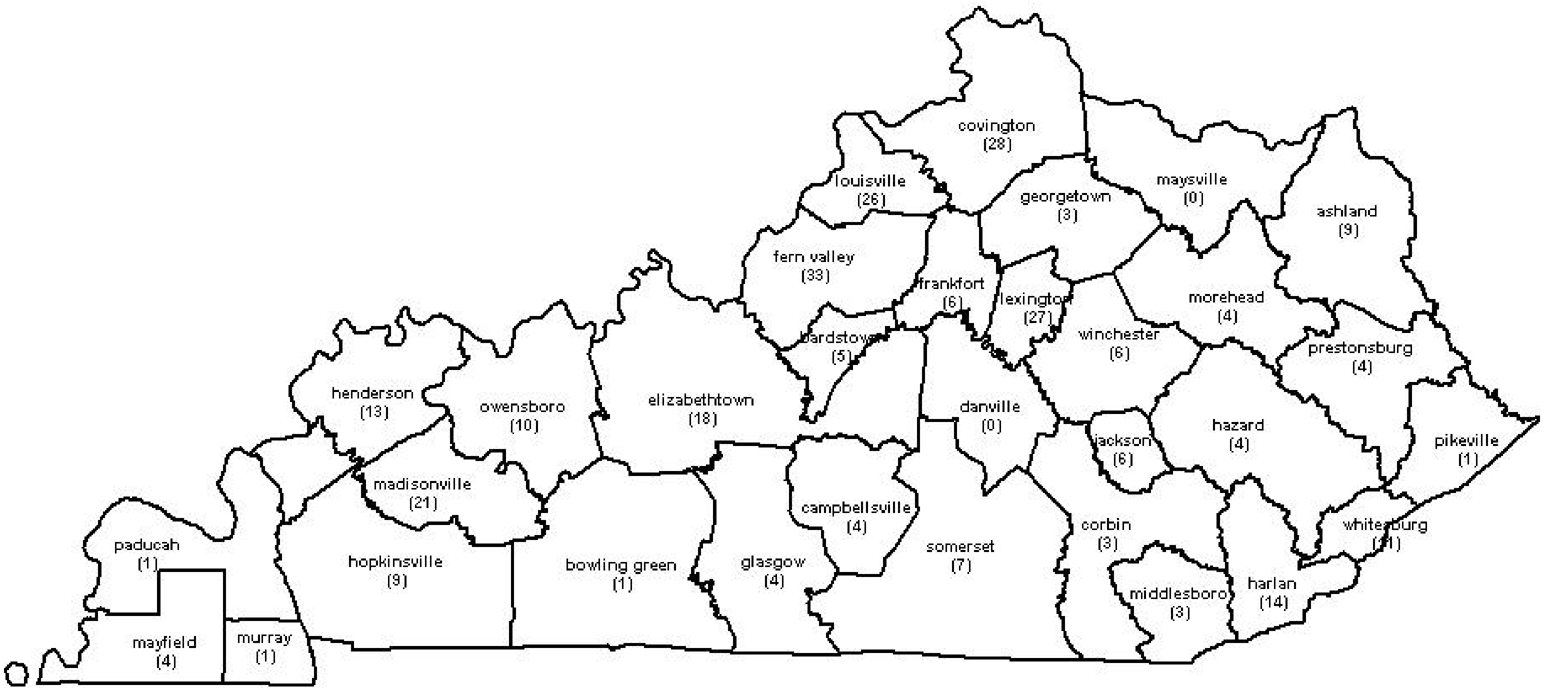


Figure 3
Regression Discontinuity Design

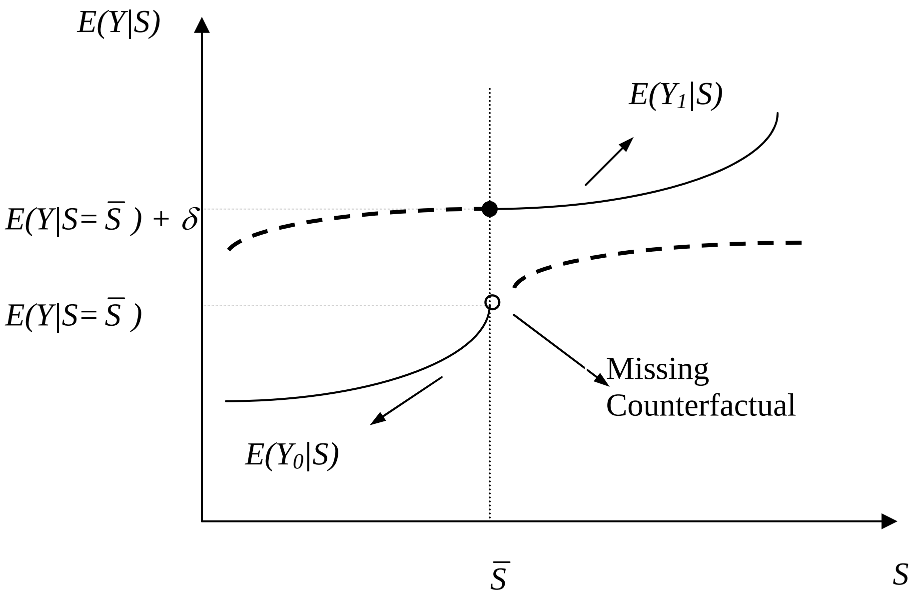
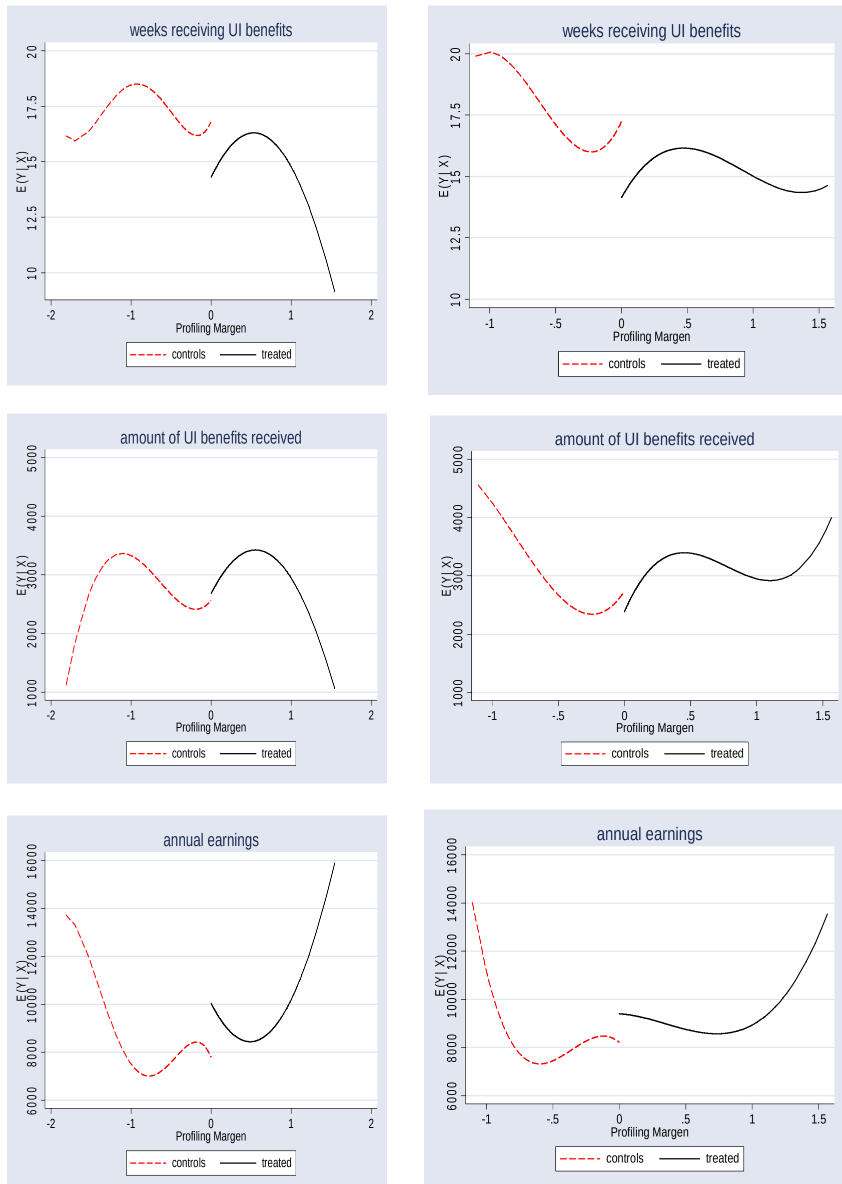


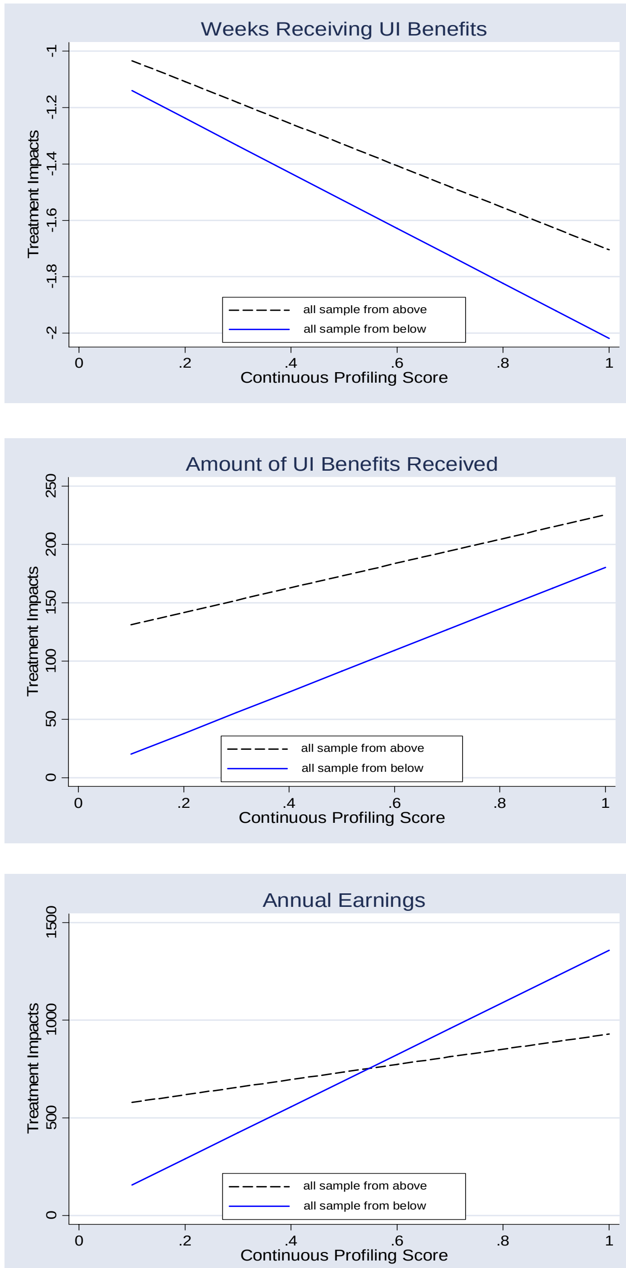
Figure 4: Discontinuities (Jumps) in UI Outcomes by Profiling Margin.
Kentucky Working and Reemployment Services, October 1994 to June 1996

Discontinuities from Above Discontinuities from Below



Note: Conditional means on profiling margin. We use quadratic and cubic specifications for above and below samples. The profiling margin is defined as the difference between the continuous profiling score and the cut-off score.

Figure 5: Heterogeneous Treatment Impacts under a Parametric RD Estimator
Kentucky Working and Reemployment Services, October 1994 to June 1996



Notes: Treatment Impacts are estimated as a linear function of S, $\delta_i = \delta_0 + \delta_1 S_i$, using a parametric fixed-effect model specified by $y_i = \alpha + \delta_0 T_i + g(S_i) + \delta_1 (S_i \times T_i) + \beta_2 age_i + \beta_3 age_i^2 + \beta_4 edu_i + \beta_5 sex_i +$ $+\beta_6 white_i + \beta_7 black_i + \beta_8 S_i * age_i + \beta_9 S_i * edu_i + \beta_{10} S_i * sex_i + \beta_{11} S_i * white + \beta_{12} S_i * black_i + n_{RDG,i} + \varepsilon_i$. *Past annual earnings are also included in the regression for amount of UI benefits received. The order of* $g(S_i)$ *is selected via generalized cross-validation method.*

Table 1: Summary Statistics
Kentucky Working and Reemployment Services, October 1994 to June 1996

	Experimental Sample Set			Non-Experimental Sample Set		
	Treated *(B)*	Control *(C)*	p-values for test of differences in means	Treated *(D)*	Comparison *(A)*	p-values for test of differences in means
Profiling score	15.2	14.7	0.91	16.0	11.9	0.00
	(2.5)	(2.6)		(2.8)	(3.4)	
Annual earnings	$19049	$19775	0.70	$19156	$16489	0.00
	(13639)	(13677)		(14612)	(12722)	
1st quarter earnings	$4560	$5011	0.81	$4566	$3971	0.00
	(3821)	(4074)		(4081)	(3647)	
2nd quarter earnings	$4456	$4683	0.90	$4676	$3846	0.00
	(3829)	(3744)		(4256)	(3578)	
3rd quarter earnings	$4901	$4976	0.84	$4876	$4174	0.00
	(3795)	(3515)		(4029)	(3521)	
4th quarter earnings	$5130	$5103	0.43	$5037	$4496	0.00
	(3735)	(3609)		(4062)	(3396)	
Years of schooling	12.5	12.3	0.22	12.3	12.4	0.66
	(2.1)	(2.0)		(2.0)	(1.9)	
Less than high school (%)	1.9	1.3	0.88	1.2	1.1	0.31
	(0.1)	(0.1)		(0.1)	(0.1)	
Bachelor degree (%)	5.4	5.2	0.31	4.6	4.3	0.04
	(0.2)	(0.2)		(0.2)	(0.2)	
Graduate studies (%)	1.30	0.94	0.44	1.13	0.81	0.00
	(0.11)	(0.0)		(0.10)	(0.08)	
Age	37.0	37.0	0.77	37.3	36.6	0.00
	(11)	(10.8)		(11.2)	(11.4)	
Percent females	43.1	39.6	0.06	44.2	40.8	0.00
	(0.4)	(0.4)		(0.4)	(0.4)	
Percent whites	88.9	91.7	0.76	91.5	90.2	0.04
	(0.3)	(0.2)		(0.2)	(0.2)	
Percent blacks	10.7	7.9	0.84	8.1	9.4	0.04
	(0.3)	(0.2)		(0.2)	(0.2)	
N	1222	742		46,313	8,631	

Notes: Standard deviations are given in parenthesis. Means are unweighted. Test for differences in means for the experimental sample (B versus C) are based on a linear regression that conditions on a treatment dummy variable and on the PTGs. Test for differences in means for the non-experimental sample (D versus A) are based on a linear regression that conditions on a treatment dummy variable and on local office and week.

Table 2: Maximum-Likelihood Probit Model for PTGs
(Coefficient Estimates, Marginal Effects, and *p*-Values)
Dependent Variable: 1 for Claimants in PTGs, 0 for Claimants not in PTGs.

Variables[a]	Coefficient	*dF / dX*	Std Error	*p*-Value[b]
Intercept	-5.018	-----	1.55	0.001
Number of claimants*10000	89.55	4.332	0.000	0.000
Unemployment rate[c]	0.080	0.003	0.015	0.000
Per capita income*10000[d]	-2.240	-0.108	0.000	0.131
Size of labor force *10000[e]	0.881	0.042	0.000	0.000
Appalachia area	2.722	0.475	0.561	0.000
West area	1.707	0.222	0.383	0.000
Metropolitan area	-30.74	-1.00	5.767	0.000
Age 16 to 24	0.088	0.004	0.036	0.016
Age 25 to 35	0.059	0.002	0.025	0.021
Age 56+	-0.037	-0.001	0.046	0.415
Fewer than 10 years schooling	0.030	0.001	0.046	0.518
10-11 years schooling	0.020	0.001	0.040	0.607
13-15 years schooling	0.021	0.001	0.030	0.471
16+ years schooling	0.128	0.006	0.037	0.001
Sex	-0.042	-0.002	0.023	0.077
White	-0.026	-0.001	0.039	0.512
1st quarter earning before profiling	-0.000	-0.00	0.000	0.000
Employed → Employed[f]	0.047	0.002	0.045	0.209
Not employed → Employed	0.191	0.011	0.080	0.018
Not employed → Not employed	0.346	0.023	0.232	0.135
Number of Observations	52782			
R^2	0.135			

[a] The omitted UI area is Rural; the omitted age category is 36 to 55; the omitted schooling category is 12 years; the omitted labor force transition is employed → no employed. Dummy variables for each of the 32 UI local offices and each of the 87 weeks of the KWPRS are also included as regressors. As some weeks and local offices predict perfectly *Pr(PTG=0)* we drop from the estimation 4,126 units for those weeks and local offices, having a final sample of 52,782 UI claimants out of 56,908.

[b] Reported *p*-Values are for two-tailed test of the null hypothesis that the true coefficient equals zero.

[c] The unemployment rate is a county-level monthly variable. As each UI local office serves a group of counties, we assigned to each UI claimant the monthly weighted average unemployment rate of the local office where he or she is profiled in a specific month. Weights are proportional to the labor force size in each county.

[d] The income per-capita is a county-level annual variable. As each UI local office serves a group of counties, we assigned to each UI claimant the annual weighted average income per-capita of the local office where he or she is profiled in a specific year. Weights are proportional to the income per-capita in each county.

[e] The labor force is a county-level annual variable. As each UI local office serves a group of counties, we assigned to each UI claimant the annual weighted average labor force of the local office where he or she is profiled in a specific year. Weights are proportional to the labor force in each county.

[f] Dummy variables indicating labor force transitions between four and one quarter before the program.

Table 3: Selection Bias under Parametric RD Estimator
Kentucky Working and Reemployment Services, October 1994 to June 1996

	Full Experimental Estimates	+/-0.05 from discontinuity	+/-0.10 from discontinuity	+/-0.15 from discontinuity	All units
			Selection Bias Estimates from Above		
week receiving UI benefits	-1.85	0.04 {-0.15} (1.20) [0.46]	-0.02 {-0.01} (0.75) [0.46]	-0.85 {-0.83} (0.60) [0.08]	0.36 {0.36} (0.29) [0.29]
amount of UI benefits received	-7.30	4 {-17} (215) [0.48]	-66 {-46} (136) [0.31]	-161 {-144} (109) [0.13]	95 {95} (59) [0.20]
annual earnings	1338	-70 {57} (1265) [0.48]	-145 {-91} (842) [0.43]	-351 {-364} (679) [0.28]	-518 {-518} (333) [0.11]
			Selection Bias Estimates from Below		
week receiving UI benefits	-1.92	-0.20 {-0.10} (1.05) [0.36]	-0.35 {-0.17} (0.70) [0.32]	0.12 {0.20} (0.56) [0.41]	0.23 {0.23} (0.30) [0.35]
amount of UI benefits received	-22.9	-220 {-235} (187) [0.10]	-135 {-118} (124) [0.17]	-5 {0.70} (100) [0.46]	144 {144} (55) [0.13]
annual earnings	1376	-1216 {-1322} (1183) [0.10]	-524 {-638} (232) [0.18]	-822 {-873} (636) [0.06]	-463 {-463} (334) [0.16]
# RDGs from above	281	169	248	268	281
# N	1949	1132	3104	4848	11235
# RDGs from below	266	168	232	247	266
#N	1915	1324	3107	4848	11479

Estimated bias is equal to the difference between non-experimental and experimental impacts. The first bias is based on the full experimental estimates, whereas the bias in the curly brackets is based only on the number of PTGs with corresponding valid RDGs for each cutoff point.
Experimental impacts are estimated by mean differences applied separately to each PTG and the estimates then weighted up estimates using as weights the proportion of treated units in each PTG.
Non-experimental impacts are estimated by a parametric fixed-effect model specified by $y_i = \alpha + \beta_1 T_i + g(S_i) + \beta_2 age_i + \beta_3 age_i^2 + \beta_4 edu_i + \beta_5 sex_i + \beta_6 white_i + \beta_7 black_i + \beta_8 S_i * age_i + \beta_9 S_i * edu_i +$ $+\beta_{10} S_i * sex_i + \beta_{11} S_i * white + \beta_{12} S_i * black_i + n_{RDG,i} + \varepsilon_i$. Past annual earnings are also included in the regression for amount of UI benefits received. The order of the polynomial approximation to $g(S_i)$ is selected via generalized cross-validation method.
Standard errors in parenthesis. *p-values* for testing the null $Ho: (\mu_1^{non-\exp} - \mu_0^{non-\exp}) - (\mu_1^{\exp} - \mu_0^{\exp}) = 0$ in brackets. We bootstrap the test statistic to account for the correlation between the non-experimental and the full experimental impacts that follows from using either the B group or the C group in the construction of both estimates.

Table 4: Selection Bias under Local Wald Estimator
Kentucky Working and Reemployment Services, October 1994 to June 1996

	Full experimental estimates	+/-0.05 from discontinuity	+/-0.10 from discontinuity	+/-0.15 from discontinuity	All units
			Selection Bias Estimates from Above		
week receiving UI benefits	-1.85	0.88 {0.69} (0.85) [0.10]	0.80 {0.81} (0.50)[0.02]	0.59 {0.61} (0.40)[0.13]	1.01 {1.01} (0.27)[0.03]
amount of UI benefits received	-7.30	422 {401} (172) [0.00]	436 {464} (103) [0.00]	454 {479} (82) [0.00]	778 {778} (57) [0.00]
annual earnings	1338	-287 {-160} (980) [0.39]	-521 {-467} (568) [0.14]	-282 {-295} (458) [0.29]	-58 {58} (311) [0.46]
			Selection Bias Estimates from Above		
week receiving UI benefits	-1.92	0.17 { 0.27} (0.75) [0.38]	0.70 {0.88} (0.51) [0.10]	0.61 {0.69} (0.41) [0.15]	1.17 {1.17} (0.29) [0.01]
amount of UI benefits received	-22.9	223 {209} (152) [0.05]	419 {437} (102) [0.00]	502 {508} (83.7) [0.00]	770 {770} (59) [0.00]
annual earnings	1376	-546 {-652} (863) [0.22]	-504 {-618} (577) [0.15]	-321 {-372} (466) [0.24]	3 {3} (329) [0.48]
# RDGs from above	281	169	248	268	281
# N	1949	1132	3104	4848	11235
# RDGs from below	266	168	232	247	266
# N	1215	1324	3107	4848	11479

Estimated bias is equal to the difference between non-experimental and experimental impacts. The first bias is based on the full experimental estimates, whereas the bias in the curly brackets is based only on the number of PTGs with corresponding valid RDGs for each cutoff point.
Experimental impacts are estimated by mean differences applied separately to each PTG and the estimates then weighted up using as weights the proportion of treated units in each PTG.
Non-experimental impacts are estimated by mean differences applied separately to each RDG and then weighting up the estimates using as weights the proportion of treated units in each corresponding PTG.
Standard errors in parenthesis. *p-values* for testing $Ho:(\mu_1^{non-\exp}-\mu_0^{non-\exp})-(\mu_1^{\exp}-\mu_0^{\exp})=0$ in brackets. We bootstrap the test statistic to account for the correlation between the non-experimental and full experimental impacts that follows from using either the B group or the C group in the construction of both estimates. It is based on 999 replications.

Table 5: Selection Bias under Smooth Local Wald Estimator
Kentucky Working and Reemployment Services, October 1994 to June 1996

	Experimental Estimates	+/-0.05 from discontinuity	+/-0.10 from discontinuity	+/-0.15 from discontinuity	All units
			Selection Bias Estimates from Above		
week receiving UI benefits	-1.85	-0.25 {-0.44} (1.18) [0.35]	-0.58 {-0.57} (1.35) [0.49]	0.02 {0.04} (0.94) [0.68]	0.72 {0.72} (0.20)[0.06]
amount of UI benefits received	-7.30	-5 {-26} (90) [0.45]	195 {216} (91) [0.11]	341 {359} (147) [0.11]	702 {702} (47) [0.00]
annual earnings	1338	-593 {-466} (384) [0.00]	-32 {22} (854) [0.26]	193 {180} (794) [0.57]	-227 {-227} (229) [0.21]
			Selection Bias Estimates from Below		
week receiving UI benefits	-1.92	-0.73 {-0.63} (0.46) [0.14]	0.33 {0.51} (0.35) [0.27]	0.15 {0.23} (0.32) [0.39]	0.92 {0.92} (0.21) [0.07]
amount of UI benefits received	-22.9	72 {58} (95) [0.33]	241 {259} (69) [0.02]	344 {351} (54) [0.00]	728 {728} (41) [0.00]
annual earnings	1376	-433 {-539} (570) [0.22]	-320 {-434} (420) [0.25]	181 {130} (318) [0.33]	-182 {-182} (225) [0.32]
# RDGs from above	281	169	248	268	281
# N	1949	1132	3104	4848	11235
# RDGs from below	266	168	232	247	266
#N	1915	1324	3107	4848	11479

Estimated bias is equal to the difference between non-experimental and experimental impacts. The first bias is based on the full experimental estimates, whereas the bias in the curly brackets is based only on the number of PTGs with corresponding valid RDGs for each cutoff point.
Experimental impacts are estimated by mean differences applied separately to each PTG and the estimates then weighted up the estimates using as weights the proportion of treated units in each PTG.
Non-experimental impacts are estimated in two steps. First, we estimate multivariate local linear kernel regressions with mixed data applied separately to treated and untreated samples. The regressors are profiling score, weeks, and local offices. The local office variable is represented by a one-dimensional index equal to the mean earnings of whites with high school education in the year prior to the start of the UI spell in each office. We use Epanechnikov kernel function for the continuous variables (scores and index) and Wang and VanRyzin's (1981) univariate kernel for the ordered categorical variable weeks. The optimal smoothing parameters are estimated by least square cross-validation. In the second step, we estimated mean differences of the smooth outcomes applied separately to each RDG an the estimates then weighted up using as weights the proportion of treated units in each corresponding PTG.
Bootstrapped standard errors in parenthesis. It is based on 999 replications. *p-values* for testing $Ho:(\mu_1^{non-\exp}-\mu_0^{non-\exp})-(\mu_1^{\exp}-\mu_0^{\exp})=0$ in brackets. We bootstrap the test statistic to account for the correlation between the non-experimental and full experimental impacts that follows from using either the B group or the C group in the construction of both estimates. It is based on 999 replications.

Table 6: Selection Bias under Hahn-Todd-van der Klauuw Kernel Estimator
Kentucky Working and Reemployment Services, October 1994 to June 1996

	Experimental Estimates	+/-0.05 from discontinuity	+/-0.10 from discontinuity	+/-0.15 from discontinuity	All units
			Selection Bias Estimates from Above		
week receiving UI benefits	-1.85	-0.19 {0.38} (1.01) [0.69]	-0.11 {-0.10} (0.68) [0.55]	0.65 {0.67} (0.61) [0.20]	0.77 {0.77} (0.27) [0.06]
amount of UI benefits received	-7.30	198 {177} (189) [0.00]	147 {168} (144) [0.02]	342 {360} (117) [0.00]	597 {597} (53) [0.00]
annual earnings	1338	1321{1450} (941) [0.21]	316 {370} (553) [0.27]	87 {74} (477) [0.47]	33 {33} (338) [0.50]
			Selection Bias Estimates from Below		
week receiving UI benefits	-1.92	-0.51 {-0.41} (0.63) [0.35]	0.44 {0.62} (0.45) [0.27]	0.29 {0.37} (0.37) [0.26]	0.82 {0.82} (0.27) [0.06]
amount of UI benefits received	-22.9	4 {-10} (123) [0.39]	217 {235} (84.9) [0.05]	296 {303} (73) [0.00]	592 {592} (569) [0.00]
annual earnings	1376	-1321 {-1424} (789) [0.06]	-1171{-1285} (585) [0.05]	-734 {-785} (492) [0.10]	-140 {-140} (303) [0.34]
# RDGs from above	281	169	248	268	281
# N	1949	1132	3104	4848	11235
# RDGs from below	266	168	232	247	266
#N	1915	1324	3107	4848	11479

Estimated bias is equal to the difference between non-experimental and experimental impacts. The first bias is based on the full experimental estimates, whereas the bias in the curly brackets is based only on the number of PTGs with corresponding valid RDGs for each cutoff point.

Experimental impacts are estimated by mean differences applied separately to each PTG and the estimates then weighted up using as weights the proportion of treated units in each PTG.

Non-experimental impacts are estimated by local linear kernel regression applied separately to each RDG and the estimates then weighted up using as weights the proportion of treated units in each corresponding PTG. We use Epanechnikov kernel function with smoothing parameters 0.05, 0.10, 0.15, and 1.82 for the RD sample from above; and 0.05, 0.10, 0.15, and 1.56 for the RD sample from below.

Bootstrapped standard errors in parenthesis. It is based on 999 replications. *p-values* for testing $Ho:(\mu_1^{non-\exp}-\mu_0^{non-\exp})-(\mu_1^{\exp}-\mu_0^{\exp})=0$ in brackets. We bootstrap the test statistic to account for the correlation between the non-experimental and full experimental impacts that follows from using either the B group or the C group in the construction of both estimates. It is based on 999 replications.

Table 7: Effect of Geography and Time Dimension on Selection Bias
Evaluating Discontinuities in Local Office and Week Dimensions.
Kentucky Working and Reemployment Services, October 1994 to June 1996

	Local Office Discontinuity Frontier			Week Discontinuity Frontier		
	Experimental estimates	Parametric estimates	Local Wald estimates	Experimental estimates	Parametric estimates	Local Wald estimates
Selection Bias Estimates from Above						
week receiving UI benefits	-2.09	-1.66 (0.65) [0.01]	0.69 (0.54) [0.12]	-1.59	-0.80 (0.60) [0.25]	-0.54 (0.49) [0.31]
amount of UI benefits received	18.8	-358 (131) [0.01]	117 (111) [0.14]	108	-200 (117) [0.19]	-129 (101) [0.30]
annual earnings	1492	-230 (696) [0.33]	1340 (573) [0.00]	1007	-1456 (735) [0.01]	-1440 (566) [0.00]
Selection Bias Estimates from Below						
week receiving UI benefits	-2.72	-1.45 (0.94) [0.05]	0.07 (0.92) [0.48]	-0.88	-1.46 (0.84) [0.19]	-0.27 (0.68) [0.45]
amount of UI benefits received	-28.3	-388 (177) [0.00]	189 (186) [0.11]	-97	-195 (155) [0.28]	50 (142) [0.46]
annual earnings	1863	-715 (999) [0.19]	-414 (999) [0.28]	557	-3167 (980) [0.00]	-1623 (813) [0.10]
# RDGs from above	187	178	178	187	135	135
# N	1532	2794	2794	1428	3276	3276
# RDGs from below	77	77	77	84	65	65
# N	587	1168	1168	536	1811	1811

Estimated bias is equal to the difference between non-experimental and experimental impacts.
Experimental impacts are estimated by mean differences applied separately to each PTG and the estimates then weighted up using as weights the proportion of treated units in each PTG. We consider only those PTGs that are included in the set of valid RDGs.
Specification for office-discontinuity estimation: $y_i = \alpha + \beta_1 T_i + g(I_i) + \beta_2 age_i + \beta_3 age_i^2 + \beta_4 edu_i + \beta_5 sex_i + \beta_6 white_i + \beta_7 black_i + + \beta_8 I_i * age_i + \beta_9 I_i * edu_i + \beta_{10} I_i * sex_i + \beta_{11} I_i * white + \eta_{RDG,i} + \varepsilon_i$. The order of the polynomial approximation to $g(I_i)$, where I is the local office index, is selected via generalized cross-validation method. Valid RDGs in the office dimension are selected by \$1,000 window along the index of similarity in either side of the PTG.
Specification for week-discontinuity estimation: $y_i = \alpha + \beta_1 T_i + g(W_i) + \beta_2 age_i + \beta_3 age_i^2 + \beta_4 edu_i + \beta_5 sex_i + \beta_6 white_i + \beta_7 black_i + + \eta_{RDG,i} + \varepsilon_i$. The order of the polynomial approximation to $g(W_i)$, where W is the week variable, is selected via generalized cross-validation method. Valid RDGs in the week dimension are selected by a four-week window in either side of the PTG. Within that window, we use all weeks that provide a discontinuity.
Standard errors in parenthesis.. *p-values* for testing $Ho: (\mu_1^{non-\exp} - \mu_0^{non-\exp}) - (\mu_1^{\exp} - \mu_0^{\exp}) = 0$ in brackets. We bootstrap the test statistic to account for the correlation between the non-experimental and experimental impacts that follows from using either the B group or the C group in the construction of both estimates. It is based on 999 replications.

Table 8: Full Sample Impact Estimates under Parametric RD Estimator
Kentucky Working and Reemployment Services, October 1994 to June 1996

	+/-0.05 from discontinuity	+/-0.10 from discontinuity	+/-0.15 from discontinuity	All units
week receiving UI benefits	-1.80 (0.88)	-2.45 (0.53)	-1.99 (0.42)	-1.68 (0.18)
amount of UI benefits received	-60.5 (157)	-201 (94)	-103 (73)	13.7 (37.3)
annual earnings	1645 (936)	1182 (600)	701 (462)	548 (218)
# RDGs	296	580	782	1049
# N	1863	5359	9583	32355

Non-experimental impacts (D versus A) estimated by a parametric fixed-effect model specified by $y_i = \alpha + \beta_1 T_i + g(S_i) + \beta_2 age_i + \beta_3 age_i^2 + \beta_4 edu_i + \beta_5 sex_i + \beta_6 white_i + \beta_7 black_i + \beta_8 S_i * age_i + \beta_9 S_i * edu_i + + \beta_{10} S_i * sex_i + \beta_{11} S_i * white + \beta_{12} S_i * black_i + n_{RDG,i} + \varepsilon_i$. Past annual earnings are also included in the regression for amount of UI benefits received. The order of the polynomial approximation to $g(S_i)$ is selected via generalized cross-validation method. Standard errors in parenthesis.

Table 9: Full Sample Impact Estimates under Local Wald Estimator
Kentucky Working and Reemployment Services, October 1994 to June 1996

	+/-0.05 from discontinuity	+/-0.10 from discontinuity	+/-0.15 from discontinuity	All units
week receiving UI benefits	-2.02 (0.64)	-1.75 (0.37)	-1.50 (0.28)	-0.69 (0.18)
amount of UI benefits received	-85.6 (126)	207 (75)	379 (56)	773 (37)
annual earnings	743 (685)	1252 (423)	1200 (313)	1162 (218)
# RDGs	296	580	782	1049
# N	1863	5359	9583	32355

Non-experimental impacts (D versus A) estimated by mean differences applied separately to each RDG and the estimates then weighted up using as weights the proportion of treated units with the marginal discretized score in each RDG. Standard errors in parenthesis.

Table 10: Full Sample Impact Estimates under Hahn-Todd-van der Klauuw Kernel Estimator
Kentucky Working and Reemployment Services, October 1994 to June 1996

	+/-0.05 from discontinuity	+/-0.10 from discontinuity	+/-0.15 from discontinuity	All units
week receiving UI benefits	-2.18 (0.70)	-2.01 (0.32)	-1.68 (0.25)	-0.98 (0.20)
amount of UI benefits received	-145 (150)	37 (100)	282 (70)	606 (45)
annual earnings	764 (700)	1765 (560)	1278 (320)	1101 (240)
# RDGs	296	580	782	1049
# N	1863	5359	9583	32355

Non-experimental impacts (D versus A) estimated by local linear kernel regression applied separately to each RDG and the estimates then weighted up using as weights the proportion of treated units with the marginal discretized score in each RDG. We use Epanechnikov kernel function with smoothing parameters 0.05, 0.10, 0.15, and 2.48. Bootstrapped standard errors in parenthesis. They are based on 999 replications